Bernoulli **20**(1), 2014, 109–140
DOI: 10.3150/12-BEJ478

Calibration of self-decomposable Lévy models

MATHIAS TRABS

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
E-mail: trabs@math.hu-berlin.de

We study the nonparametric calibration of exponential Lévy models with infinite jump activity. In particular our analysis applies to self-decomposable processes whose jump density can be characterized by the k-function, which is typically nonsmooth at zero. On the one hand the estimation of the drift, of the activity measure $\alpha := k(0+) + k(0-)$ and of analogous parameters for the derivatives of the k-function are considered and on the other hand we estimate nonparametrically the k-function. Minimax convergence rates are derived. Since the rates depend on α, we construct estimators adapting to this unknown parameter. Our estimation method is based on spectral representations of the observed option prices and on a regularization by cutting off high frequencies. Finally, the procedure is applied to simulations and real data.

Keywords: adaptation; European option; infinite activity jump process; minimax rates; nonlinear inverse problem; self-decomposability

1. Introduction

Since Merton [19] introduced his discontinuous asset price model, stock returns were frequently described by exponentials of Lévy processes. A review of recent pricing and hedging results for these models is given by Tankov [26]. The calibration of the underlying model, that is in the case of Lévy models the estimation of the characteristic triplet (σ, γ, ν), from historical asset prices is mostly studied in parametric models only, consider the survey paper of Eberlein [10] and the references therein. Remarkable exceptions are the nonparametric penalized least squares method by Cont and Tankov [9] and the spectral calibration procedure by Belomestny and Reiß [3]. Both articles concentrate on models of finite jump activity. Our goal is to extend their results to infinite intensity models. A class which attracted much interest in financial modeling is given by self-decomposable Lévy processes, examples are the hyperbolic model (Eberlein, Keller and Prause [11]) or the variance gamma model (Madan and Seneta [18], Madan, Carr and Chang [17]). Moreover, self-decomposable distributions are discussed in the financial investigation using Sato processes (Carr *et al.* [6], Eberlein and Madan [12]). Our results can be applied in this context, too. The nonparametric calibration of Lévy models is not only relevant for stock prices, for instance, it can be used for the Libor market as well (see Belomestny and Schoenmakers [4]). In the context of Ornstein–Uhlenbeck processes, the nonparametric inference of self-decomposable Lévy processes was considered by Jongbloed, van der Meulen and van der Vaart [14].

Owing to the infinite activity, the features of market prices can be reproduced even without a diffusion part (cf. Carr *et al.* [5]) and thus we study pure-jump Lévy processes. More precisely,

1350-7265 © 2014 ISI/BS

we assume that the jump density satisfies

$$\nu(\mathrm{d}x) = \frac{k(x)}{|x|}\,\mathrm{d}x \qquad \text{where } k:\mathbb{R}\to\mathbb{R}_+ \text{ has bounded variation.} \tag{K}$$

When k increases on $(-\infty, 0)$ and decreases on $(0, \infty)$, it is called k-function and the processes is self-decomposable. Further examples which have property (K) are compound Poisson processes and limit distributions of branching processes as considered by Keller-Ressel and Mijatović [16]. Using the bounded variation of k, we show that the estimation problem is only mildly ill-posed. While the Blumenthal–Getoor index, which was estimated by Belomestny [1], is zero in our model, the infinite activity can be described on a finer scale by the parameter

$$\alpha := k(0+) + k(0-).$$

Since k is typically nonsmooth at zero, we face two estimation problems: First, to give a proper description of k at zero, we propose estimators for α and its analogs $k^{(j)}(0+) + k^{(j)}(0-)$, with $j \geq 1$, for the derivatives of k as well as for the drift γ, which can be estimated similarly. We prove convergence rates for their mean squared error which turn out to be optimal in minimax sense up to a logarithmic factor. Second, we construct a nonparametric estimator of k whose mean integrated squared error converges with nearly optimal rates. Owing to bid-ask spreads and other market frictions, we observe only noisy option prices. The definition of the estimators is based on the relation between these prices and the characteristic function of the driving process established by Carr and Madan [7] and on different spectral representations of the characteristic exponent. Smoothing is done by cutting off all frequencies higher than a certain value depending on a maximal permitted parameter α. The whole estimation procedure is computationally efficient and achieves good results in simulations and in real data examples. All estimators converge with a polynomial rate, where the maximal α determines the ill-posedness of the problem. Assuming sub-Gaussian error distributions, we provide an estimator with α-adaptive rates. The main tool for this result is a concentration inequality for our estimator $\hat{\alpha}$ which might be of independent interest.

This work is organized as follows: In Section 2, we describe the setting of our estimation procedure and derive the necessary representations of the characteristic exponent. The estimators are described in Section 3, where we also determine the convergence rates. The construction of the α-adaptive estimator of α is contained in Section 4. In view of simulations and real data, we discuss our theoretical results and the implementation of the procedure in Section 5. All proofs are given in Section 6.

2. The model

2.1. Self-decomposable Lévy processes

A real valued random variable X has a *self-decomposable* law if for any $b \in (0, 1)$ there is an independent random variable Z_b such that $X \overset{d}{=} bX + Z_b$. Since each self-decomposable distribution is infinitely divisible (see Proposition 15.5 in [21]), we can define the corresponding *self-decomposable Lévy process*. Self-decomposable laws can be understood as the class of limit distributions of converging scaled sums of independent random variables (Theorem 15.3 in [21]). This characterization is of economical interest. If we understand the price of an asset as an aggregate of small independent influences and release from the $\sqrt{n}$ scaling, which leads to diffusion models, we automatically end up in a self-decomposable price process.

Sato [21] shows that the jump measure of a self-decomposable distribution is always absolutely continuous with respect to the Lebesgue measure and its density can be characterized through (K) where k needs to be increasing on $\mathbb{R}_-$ and decreasing on $\mathbb{R}_+$. Note that self-decomposability does not affect the volatility σ nor the drift γ of the Lévy process.

Assuming $\sigma = 0$ and property (K), the process X_t has finite variation and the characteristic function of X_T is given by the Lévy–Khintchine representation

$$\varphi_T(u) := \mathbb{E}\big[\mathrm{e}^{\mathrm{i}uX_T}\big] = \exp\left(T\left(\mathrm{i}\gamma u + \int_{-\infty}^{\infty}\big(\mathrm{e}^{\mathrm{i}ux} - 1\big)\frac{k(x)}{|x|}\,\mathrm{d}x\right)\right). \tag{2.1}$$

Motivated by a martingale argument, we will suppose the exponential moment condition $\mathbb{E}[\mathrm{e}^{X_t}] = 1$ for all $t \geq 0$, which yields

$$0 = \gamma + \int_{-\infty}^{\infty}\big(\mathrm{e}^{x} - 1\big)\frac{k(x)}{|x|}\,\mathrm{d}x. \tag{2.2}$$

In particular, we will impose $\int_{-\infty}^{\infty}(\mathrm{e}^x - 1)\frac{k(x)}{|x|}\,\mathrm{d}x < \infty$. In this case, φ_T is defined on the strip $\{z \in \mathbb{C} \,|\, \operatorname{Im} z \in [-1, 0]\}$.

Besides Lévy processes there is another class that is closely related to self-decomposability. Assuming self-similarity, that means $(Y_{at}) \overset{d}{=} (a^H Y_t)$, for all $a > 0$ and some exponent $H > 0$, instead of stationary increments, Y_t is a *Sato processes*. Sato [20] showed that self-decomposable distributions can be characterized as the laws at unit time of these processes. From the self-similarity and self-decomposability follows for $T > 0$

$$\varphi_{Y_T}(u) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}uY_T}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}T^H uY_1}\big] = \exp\left(\mathrm{i}T^H\gamma u + \int_{-\infty}^{\infty}\big(\mathrm{e}^{\mathrm{i}ux} - 1\big)\frac{k(T^{-H}x)}{|x|}\,\mathrm{d}x\right).$$

Since our estimation procedure only depends through equation (2.1) on the distributional structure of the underlying process, we can apply the estimators directly to Sato processes using $T_s = 1, \gamma_s = T^H\gamma$ and $k_s(\cdot) = k(T^{-H}\bullet)$ instead of T, γ and k. However, we concentrate on Lévy processes in the sequel.

For self-decomposable distributions the parameter α captures many of its properties such as the smoothness of the densities of the marginal distributions (Theorem 28.4 in [21]) and the tail behavior of the characteristic function. This holds even for the more general class of Lévy processes that satisfy property (K). Recall that k has bounded variation if and only if

$$\|k\|_{\mathrm{TV}} := \sup\left\{\sum_{i=1}^{n}\big|k(x_i) - k(x_{i-1})\big| : n \in \mathbb{N}, -\infty < x_0 < \cdots < x_n < \infty\right\} < \infty.$$

In particular, $\|k\|_{\mathrm{TV}} < \infty$ implies $\alpha < \infty$. Similarly to deconvolution problems, the stochastic error in our model is driven by $|\varphi_T(u-\mathrm{i})|^{-1}$ and thus we prove the following lemma in the Appendix.

Lemma 2.1. *Let X_t have property* (K) *and $\sigma = 0$ and let the martingale condition* (2.2) *hold.*

(i) *If $\|\mathrm{e}^x k(x)\|_{L^1} < \infty$ and $q_k := \sup_{x\in(0,1]} \frac{k(x)+k(-x)-\alpha}{x} < \infty$ then there exists a constant $C_\varphi = C_\varphi(T, \max\{q_k, \|\mathrm{e}^x k(x)\|_{L^1}, \|k\|_{\mathrm{TV}}\}) > 0$ such that for all $u \in \mathbb{R}$ with $|u| \geq 1$ we obtain the bound*

$$\left|\varphi_T(u-\mathrm{i})\right| \geq C_\varphi |u|^{-T\alpha}.$$

(ii) *Let $\bar{\alpha}, R > 0$ then $|\varphi_T(u-\mathrm{i})| \geq C_\varphi(T,R)|u|^{-T\bar{\alpha}}$ holds uniformly over all $|u| \geq 1$ and all X_T with $\alpha \leq \bar{\alpha}$ and $\max\{q_k, \|\mathrm{e}^x k(x)\|_{L^1}, \|k\|_{\mathrm{TV}}\} \leq R$.*

The value q_k as defined in the lemma can be understood as the largest slop of k near zero. If the process is self-decomposable than $q_k \leq 0$ holds and the bounded variation norm equals α. Otherwise, we can use $q_k \leq \sup_{|x|\leq 1} |k'(x)|$ and $\|k\|_{\mathrm{TV}} \leq \|k'\|_{L^1}$, assuming the derivative k' exists, is bounded on $[-1, 1]$ and integrable on $\mathbb{R}$. If either $\sigma > 0$ or property (K) is violated, φ_T can decay faster than any polynomial order, for example, consider self-decomposable processes with $\alpha = \infty$ (see [21], Lemma 28.5). Hence, the conditions of Lemma 2.1 are sharp.

2.2. Asset prices and Vanilla options

Let $r \geq 0$ be the risk-less interest rate in the market and $S_0 > 0$ denote the initial value of the asset. In an exponential Lévy model the price process is given by

$$S_t = S_0 \mathrm{e}^{rt+X_t},$$

where X_t is a Lévy process described by the characteristic triplet (σ, γ, ν). Throughout these notes, we assume X_t has property (K) and $\sigma = 0$. On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with pricing (or martingale) measure $\mathbb{P}$ the discounted process $(\mathrm{e}^{-rt} S_t)$ is a martingale with respect to its natural filtration $(\mathcal{F}_t)$. This is equivalent to $\mathbb{E}[\mathrm{e}^{X_t}] = 1$ for all $t \geq 0$ and thus, the martingale condition (2.2) holds.

At time $t = 0$ the risk neutral price of an European call option with underlying S, time to maturity T and strike price K is given by $\mathrm{e}^{-rT}\mathbb{E}[(S_T - K)_+]$, where $A_+ := \max\{0, A\}$, and similarly $\mathrm{e}^{-rT}\mathbb{E}[(K - S_T)_+]$ is the price of European put. In terms of the negative log-forward moneyness $x := \log(K/S_0) - rT$ the prices can be expressed as

$$\mathcal{C}(x,T) = S_0 \mathbb{E}\big[\big(\mathrm{e}^{X_T} - \mathrm{e}^x\big)_+\big] \quad \text{and} \quad \mathcal{P}(x,T) = S_0 \mathbb{E}\big[\big(\mathrm{e}^x - \mathrm{e}^{X_T}\big)_+\big].$$

Carr and Madan [7] introduced the option function

$$\mathcal{O}(x) := \begin{cases} S_0^{-1}\mathcal{C}(x,T), & x \geq 0, \\ S_0^{-1}\mathcal{P}(x,T), & x < 0, \end{cases}$$

and set the Fourier transform $\mathcal{F}\mathcal{O}(u) := \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}ux}\mathcal{O}(x)\,\mathrm{d}x$ in relation to the characteristic function φ_T through the pricing formula

$$\mathcal{F}\mathcal{O}(u) = \frac{1-\varphi_T(u-\mathrm{i})}{u(u-\mathrm{i})}, \qquad u \in \mathbb{R} \setminus \{0\}. \tag{2.3}$$

The properties of $\mathcal{O}$ were studied further by Belomestny and Reiß [3]. In particular, they showed that the option function is contained in $C^1(\mathbb{R} \setminus \{0\})$ and decays exponentially under the following assumption.

Assumption 1. *We assume that* $C_2 := \mathbb{E}[\mathrm{e}^{2X_T}]$ *is finite, which is equivalent to the moment condition* $\mathbb{E}[S_t^2] < \infty$.

Our observations are given by

$$O_j = \mathcal{O}(x_j) + \delta_j \varepsilon_j, \qquad j = 1, \ldots, N, \tag{2.4}$$

where the noise (ε_j) consists of independent, centered random variables with $\mathbb{E}[\varepsilon_j^2] = 1$ and $\sup_j \mathbb{E}[\varepsilon_j^4] < \infty$. The noise levels δ_j are assumed to be positive and known. In practice, the uncertainty is due to market frictions such as bid-ask spreads.

2.3. Representation of the characteristic exponent

Using (2.1) and (2.3), the shifted characteristic exponent is given by

$$\psi(u) := \frac{1}{T}\log\big(1 + \mathrm{i}u(1+\mathrm{i}u)\mathcal{F}\mathcal{O}(u)\big) = \frac{1}{T}\log\big(\varphi_T(u-\mathrm{i})\big) \tag{2.5}$$

$$= \mathrm{i}\gamma u + \gamma + \int_{-\infty}^{\infty} \big(\mathrm{e}^{\mathrm{i}(u-\mathrm{i})x} - 1\big)\frac{k(x)}{|x|}\,\mathrm{d}x \tag{2.6}$$

for $u \in \mathbb{R}$. Note that the last line equals zero for $u = 0$ because of the martingale condition (2.2). Throughout, we choose a distinguished logarithm, that is a version of the complex logarithm such that ψ is continuous with $\psi(0) = 0$. Under the assumption that $\int_{-\infty}^{\infty}(1 \vee \mathrm{e}^x)k(x)\,\mathrm{d}x$[1] is finite, we can apply Fubini's theorem to obtain

$$\psi(u) = \mathrm{i}\gamma u + \gamma + \int_0^1 \mathrm{i}(u-\mathrm{i})\mathcal{F}\big(\operatorname{sgn}(x)k(x)\big)\big((u-\mathrm{i})t\big)\,\mathrm{d}t, \tag{2.7}$$

where the Fourier transform $\mathcal{F}(\operatorname{sgn}\cdot k)$ is well defined on $\{z \in \mathbb{C} \,|\, \operatorname{Im} z \in [-1, 0]\}$. Typically, the k and its derivatives are not continuous at zero. Moreover, if $\alpha \neq 0$ the function $x \mapsto \operatorname{sgn}(x)k(x)$ has a jump at zero in every case. Therefore, the Fourier transform decreases very slowly. Let k be smooth on $\mathbb{R} \setminus \{0\}$ and fulfill an integrability condition which will be important later:

[1] We denote $A \wedge B := \min\{A, B\}$ and $A \vee B := \max\{A, B\}$ for $A, B \in \mathbb{R}$.

Assumption 2. *Assume* $k \in C^s(\mathbb{R} \setminus \{0\})$ *with all derivatives having a finite right- and left-hand limit at zero and* $(1 \vee \mathrm{e}^x)k(x), \dots, (1 \vee \mathrm{e}^x)k^{(s)}(x) \in L^1(\mathbb{R})$.

To compensate those discontinuities, we add a linear combination of the functions $h_j(x) := x^j e^{-x}\mathbf{1}_{[0,\infty)}(x)$, $x \in \mathbb{R}$, for $j = \mathbb{N} \cup \{0\}$. Since $h_j \in C^{j-1}(\mathbb{R})$ for $j \geq 1$ and all h_j are smooth on $\mathbb{R} \setminus \{0\}$, we can find α_j, $j = 0, \dots, s-2$, such that $\operatorname{sgn}(x)k(x) - \sum_{j=0}^{s-2} \alpha_j h_j(x)$ is contained in $C^{s-2}(\mathbb{R}) \cap C^s(\mathbb{R} \setminus \{0\})$. This approach yields the following representation. The proof is given in the supplementary article [27].

Proposition 2.2. *Let* $s \geq 2$. *On Assumption* 2, *there exist functions* $D : \{-1, 1\} \to \mathbb{C}$ *and* $\rho : \mathbb{R} \setminus \{0\} \to \mathbb{C}$ *such that* $|u^{s-1}\rho(u)|$ *is bounded in* u *and it holds*

$$\psi(u) = D\big(\operatorname{sgn}(u)\big) + \mathrm{i}\gamma u - \alpha_0 \log\big(|u|\big) + \sum_{j=1}^{s-2} \frac{\mathrm{i}^j (j-1)!\alpha_j}{u^j} + \rho(u), \qquad u \neq 0. \tag{2.8}$$

The coefficients are given by $\alpha_j = \frac{1}{j!}(k^{(j)}(0+) + k^{(j)}(0-)) - \sum_{m=1}^{j} \frac{(-1)^m}{m!}\alpha_{j-m}$, *especially* $\alpha_0 = \alpha$ *holds.*

Representation (2.8) allows us to estimate γ and $\alpha_0, \dots, \alpha_{s-2}$. A plug-in approach yields estimators for $k^{(j)}(0+) + k^{(j)}(0-)$, $j = 0, \dots, s-2$. Since we only apply this representation when ψ is multiplied with weight functions having roots of degree $s-1$ at zero, the poles that appear in (2.8) do no harm.

Proposition 2.2 covers the case $s \geq 2$. For $s = 1$, we conclude from (2.6), the martingale condition (2.2) and Assumption 2

$$\psi(u) = \mathrm{i}\gamma u + \int_{-\infty}^{\infty} \big(\mathrm{e}^{\mathrm{i}ux} - 1\big)\mathrm{e}^x \frac{k(x)}{|x|}\,\mathrm{d}x = \mathrm{i}\gamma u + \mathrm{i}\int_0^u \mathcal{F}\big(\operatorname{sgn}(x)\mathrm{e}^x k(x)\big)(v)\,\mathrm{d}v. \tag{2.9}$$

Hence, ψ is a sum of a constant from the integration, the linear drift $\mathrm{i}\gamma u$ and a remainder of order $\log|u|$, which follows from the decay of the Fourier transform as $|u|^{-1}$. Corollary 8 in [27] even shows, that there exists no L^2-consistent estimator of α for $s = 1$. Therefore, we concentrate on the case $s \geq 2$ in the sequel.

Equation (2.9) allows another useful observation. Defining the exponentially scaled k-function

$$k_e(x) := \operatorname{sgn}(x)\mathrm{e}^x k(x), \qquad x \in \mathbb{R},$$

we obtain by differentiation

$$\psi'(u) = \frac{1}{T}\frac{(\mathrm{i} - 2u)\mathcal{F}\mathcal{O}(u) - (u + \mathrm{i}u^2)\mathcal{F}(x\mathcal{O}(x))(u)}{1 + (\mathrm{i}u - u^2)\mathcal{F}\mathcal{O}(u)} = \mathrm{i}\gamma + \mathrm{i}\mathcal{F}k_e(u). \tag{2.10}$$

Using this relation, we can define an estimator of k_e.

3. Estimation procedure

3.1. Definition of the estimators and weight functions

Given the observations $\{(x_1, O_1), \dots, (x_N, O_N)\}$, we fit a function $\tilde{\mathcal{O}}$ to these data using linear B-splines

$$b_j(x) := \frac{x - x_{j-1}}{x_j - x_{j-1}} \mathbf{1}_{[x_{j-1}, x_j)} + \frac{x_{j+1} - x}{x_{j+1} - x_j} \mathbf{1}_{[x_j, x_{j+1}]}, \qquad j = 1, \dots, N,$$

and a function β_0 with $\beta_0'(0+) - \beta_0'(0-) = -1$ to take care of the jump of $\mathcal{O}'$:

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

We choose β_0 with support $[x_{j_0-1}, x_{j_0}]$ where j_0 satisfies $x_{j_0-1} < 0 \le x_{j_0}$. Replacing $\mathcal{O}$ with $\tilde{\mathcal{O}}$ in the representations (2.5) and (2.10) of ψ and ψ', respectively, allows us to define their empirical versions through

$$\tilde{\psi}(u) := \frac{1}{T} \log\bigl(v_{\kappa(u)}\bigl(1 + \mathrm{i}u(1 + \mathrm{i}u)\mathcal{F}\tilde{\mathcal{O}}(u)\bigr)\bigr),$$

$$\tilde{\psi}'(u) := \frac{1}{T} \frac{(\mathrm{i} - 2u)\mathcal{F}\tilde{\mathcal{O}}(u) - (u + \mathrm{i}u^2)\mathcal{F}(x\tilde{\mathcal{O}}(x))(u)}{v_{\kappa(u)}(1 + \mathrm{i}u(1 + \mathrm{i}u)\mathcal{F}\tilde{\mathcal{O}}(u))}, \qquad u \in \mathbb{R},$$

where κ is a positive function and we apply a trimming function given by

$$v_\kappa(z) : \mathbb{C} \setminus \{0\} \to \mathbb{C}, \qquad z \mapsto \begin{cases} z, & |z| \ge \kappa, \\ \kappa z/|z|, & |z| < \kappa, \end{cases}$$

to stabilize for large stochastic errors. A reasonable choice of κ will be derived below. The function $\tilde{\psi}$ is well defined on the interval $[-U, U]$ on the event

$$A := \bigl\{\omega \in \Omega\colon\ 1 + \mathrm{i}u(1 + \mathrm{i}u)\mathcal{F}\bigl(\tilde{\mathcal{O}}(\omega, \bullet)\bigr)(u) \ne 0\ \forall u \in [-U, U]\bigr\} \in \mathcal{F}.$$

For $\omega \in \Omega \setminus A$, we set $\tilde{\psi}$ arbitrarily, for instance equal to zero. The more $\tilde{\mathcal{O}}$ concentrates around the true function $\mathcal{O}$ the greater is the probability of A. Söhl [23] shows even that in the continuous-time Lévy model with finite jump activity the identity $\mathbb{P}(A) = 1$ holds.

In the spirit of Belomestny and Reiß [3], we estimate the parameters γ and α_j, $j = 0, \dots, s - 2$, as coefficients of the different powers of u in equation (2.8). Using a spectral cut-off value $U > 0$, we define

$$\hat{\gamma} := \int_{-U}^{U} \mathrm{Im}\bigl(\tilde{\psi}(u)\bigr) w_\gamma^U(u)\, \mathrm{d}u$$

and for $0 \le j \le s-2$

$$\hat{\alpha}_j := \begin{cases} \int_{-U}^{U} \mathrm{Re}\big(\tilde{\psi}(u)\big) w_{\alpha_j}^U(u)\,\mathrm{d}u, & \text{if } j \text{ is even,} \\ \int_{-U}^{U} \mathrm{Im}\big(\tilde{\psi}(u)\big) w_{\alpha_j}^U(u)\,\mathrm{d}u, & \text{otherwise.} \end{cases}$$

The weight functions w_γ^U and $w_{\alpha_j}^U$ are chosen such that they filter the coefficients of interest. Owing to (2.10), the nonparametric object k_e can be estimated by

$$\hat{k}_e(x) := \begin{cases} \mathcal{F}^{-1}\big[\big(-\hat{\gamma} - \mathrm{i}\tilde{\psi}'(u)\big)\mathcal{F}W_k(u/U)\big](x), & x > 0, \\ \mathcal{F}^{-1}\big[\big(-\hat{\gamma} - \mathrm{i}\tilde{\psi}'(u)\big)\mathcal{F}W_k(-u/U)\big](x), & x < 0, \end{cases} \tag{3.1}$$

applying a one-sided kernel function W_k with bandwidth U^{-1} since we know that k_e jumps only at zero. The condition on the weights are summarized in the following:

Assumption 3. *We assume*:

- w_γ^U *fulfills for all odd* $j \in \{1, \dots, s-2\}$

$$\int_{-U}^{U} u w_\gamma^U(u)\,\mathrm{d}u = 1, \qquad \int_{-U}^{U} u^{-j} w_\gamma^U(u)\,\mathrm{d}u = 0 \quad \textit{and} \quad \int_0^U w_\gamma^U(\pm u)\,\mathrm{d}u = 0.$$

- $w_{\alpha_0}^U$ *satisfies for all even* $j \in \{1, \dots, s-2\}$

$$\int_{-U}^{U} \log\big(|u|\big) w_{\alpha_0}^U(u)\,\mathrm{d}u = -1, \qquad \int_{-U}^{U} u^{-j} w_{\alpha_0}^U(u)\,\mathrm{d}u = 0 \quad \textit{and} \quad \int_0^U w_{\alpha_0}^U(\pm u)\,\mathrm{d}u = 0.$$

- *For* $j = 1, \dots, s-2$ *the weight functions* $w_{\alpha_j}^U$ *fulfill*[2]

$$\int_{-U}^{U} u^{-j} w_{\alpha_j}^U(u)\,\mathrm{d}u = \frac{(-1)^{\lfloor j/2 \rfloor}}{(j-1)!}, \qquad \int_{-U}^{U} u^{-l} w_{\alpha_j}^U(u)\,\mathrm{d}u = 0 \quad \textit{and}$$
$$\int_0^U w_{\alpha_j}^U(\pm u)\,\mathrm{d}u = 0,$$

where $1 \le l \le s-2$ *and* l *is even for even* j *and odd otherwise. For even* j *we impose additionally*

$$\int_{-U}^{U} \log\big(|u|\big) w_{\alpha_j}^U(u)\,\mathrm{d}u = 0.$$

[2]For $a \in \mathbb{R}$ let $\lfloor a \rfloor$ denote the largest integer which is smaller than a.

- *W_k is of Sobolev smoothness $T\bar{\alpha}+2$, that is, $\int(1+|u|^2)^{T\bar{\alpha}+2}|\mathcal{F}W_k(u/U)|^2\,\mathrm{d}u<\infty$, has support* $\operatorname{supp} W_k\subseteq(-\infty,0]$ *and fulfills for $l=1,\dots,s-1$*

$$\int_{\mathbb{R}} W_k(x)\,\mathrm{d}x=1,\qquad \int_{\mathbb{R}} x^l W_k(x)\,\mathrm{d}x=0\quad\text{and}\quad x^{2s-1}W_k(x)\in L^1(\mathbb{R}).$$

Furthermore, we assume continuity and boundedness of the functions $u\mapsto u^{-s+1}w_q^1(u)$ for $q\in\{\gamma,\alpha_0,\dots,\alpha_{s-2}\}$.

The integral conditions can be provided by rescaling: Let w_q^1 satisfy Assumption 3 for $q\in\{\gamma,\alpha_0,\dots,\alpha_{s-2}\}$ and $U=1$. Since $1=\int_{-1}^1 uw_\gamma^1(u)\,\mathrm{d}u=\int_{-U}^U uU^{-2}w_\gamma^1(u/U)\,\mathrm{d}u$, we can choose $w_\gamma^U(u):=U^{-2}w_\gamma^1(\frac{u}{U})$. Similarly, a rescaling is possible for $w_{\alpha_0}^U$:

$$\begin{aligned}-1&=\int_{-1}^1\log\bigl(|u|\bigr)w_{\alpha_0}^1(u)\,\mathrm{d}u=\int_{-U}^U\log\bigl(|u|\bigr)U^{-1}w_{\alpha_0}^1\left(\frac{u}{U}\right)\mathrm{d}u-\frac{\log(U)}{U}\int_{-U}^U w_{\alpha_0}^1\left(\frac{u}{U}\right)\mathrm{d}u\\&=\int_{-U}^U\log\bigl(|u|\bigr)U^{-1}w_{\alpha_0}^1\left(\frac{u}{U}\right)\mathrm{d}u.\end{aligned}$$

Therefore, we define $w_{\alpha_0}^U(u):=U^{-1}w_{\alpha_0}^1(\frac{u}{U})$ and analogously $w_{\alpha_j}^U(u):=U^{j-1}w_{\alpha_j}^1(\frac{u}{U})$. The continuity condition on w_q^1 in Assumption 3 is set to take advantage of the decay of the remainder ρ. In combination with the rescaling it implies

$$\bigl|w_\gamma^U(u)\bigr|\lesssim U^{-s-1}|u|^{s-1}\quad\text{and}\quad\bigl|w_{\alpha_j}^U(u)\bigr|\lesssim U^{-s+j}|u|^{s-1},\qquad j=0,\dots,s-2.\tag{3.2}$$

Throughout, we write $A\lesssim B$ if there is a constant $C>0$ independent of all parameters involved such that $A\le CB$. In the sequel we assume that the weight functions satisfy Assumption 3 and the property (3.2).

We reduce the loss of $\hat{k}_e$ by truncating positive values on $\mathbb{R}_-$ and negative ones on $\mathbb{R}_+$. In the self-decomposable framework there are additional shape restrictions of the k-function which the proposed estimator $\hat{k}_e$ does not take into account. The monotonicity can be generated by a rearrangement of the function. To this end let $\hat{k}(x):=(\operatorname{sgn}(x)\mathrm{e}^{-x}\hat{k}_e(x)\vee 0)\mathbf{1}_{[-C,C]}(x)$, $x\in\mathbb{R}$, where we bounded the support with an arbitrary large constant $C>0$. The rearranged estimator which is increasing on $\mathbb{R}_-$ and decreasing on $\mathbb{R}_+$ is then given by

$$\hat{k}^*(x):=\begin{cases}\inf\Bigl\{y\in\mathbb{R}_+\Bigm|\int_0^C\mathbf{1}_{\{\hat{k}(z)\ge y\}}\,\mathrm{d}z\le x\Bigr\}, & x>0,\\ \inf\Bigl\{y\in\mathbb{R}_+\Bigm|\int_0^C\mathbf{1}_{\{\hat{k}(-z)\ge y\}}\,\mathrm{d}z\le |x|\Bigr\}, & x<0.\end{cases}\tag{3.3}$$

Chernozhukov, Fernández-Val and Galichon [8] show that the rearrangement reduces weakly the error for increasing target functions on compact subsets. This result carries over to our estimation problem.

3.2. Convergence rates

To ensure a well-defined procedure, an exponential decay of $\mathcal{O}$, the identity (2.9) and to obtain a lower bound of $|\varphi_T(u-\mathrm{i})|$, we consider the class $\mathcal{G}_0(R,\bar{\alpha})$. Uniform convergence results for the parameters will be derived in the smoothness class $\mathcal{G}_s(R,\bar{\alpha})$.

Definition 3.1. *Let $s\in\mathbb{N}$ and $R,\bar{\alpha}>0$. We define*

(i) *$\mathcal{G}_0(R,\bar{\alpha})$ as the set of all pairs $\mathcal{P}=(\gamma,k)$ where k is of bounded variation and the corresponding Lévy process X given by the triplet $(0,\gamma,k(x)/|x|)$ satisfies Assumption 1 with $C_2\le R$, martingale condition (2.2) as well as*

$$\alpha\in[0,\bar{\alpha}]\quad \text{and}\quad \max\left\{\sup_{x\in(0,1]}\left\{\frac{k(x)+k(-x)-\alpha}{x}\right\},\left\|k_e(x)\right\|_{L^1},\|k\|_{\mathrm{TV}}\right\}\le R,$$

(ii) *$\mathcal{G}_s(R,\bar{\alpha})$ as the set of all pairs $\mathcal{P}=(\gamma,k)\in\mathcal{G}_0(R,\bar{\alpha})$ satisfying additionally Assumption 2 with*

$$\left|k^{(l)}(0+)+k^{(l)}(0-)\right|\le R,\qquad \text{for } l=1,\dots,s-1,$$
$$\left\|\left(1\vee \mathrm{e}^x\right)k^{(l)}(x)\right\|_{L^1}\le R,\qquad \text{for } l=0,\dots,s.$$

In the class $\mathcal{G}_0(R,\bar{\alpha})$ Lemma 2.1(ii) provides a common lower bound of $|\varphi_T(u-\mathrm{i})|$ for $|u|\ge 1$. Using $\max_{x\in\mathbb{R}}\frac{1-\cos(x)}{x}\in(0,1]$, we estimate roughly for $u\in(-1,1)\setminus\{0\}$:

$$\left|\varphi_T(u-\mathrm{i})\right|=\exp\left(T\int_{-\infty}^{\infty}\frac{\cos(x)-1}{x}\mathrm{e}^{x/|u|}k\left(x/|u|\right)\mathrm{d}x\right)\ge \mathrm{e}^{-TR}.$$

Hence, the choice

$$\kappa(u):=\kappa_{\bar{\alpha}}(u):=\begin{cases}\frac{1}{3}\mathrm{e}^{-TR}, & |u|<1,\\ \frac{1}{3}C_\varphi(T,R)|u|^{-T\bar{\alpha}}, & |u|\ge 1,\end{cases}$$

satisfies

$$\tfrac{1}{3}\left|\varphi_T(u-\mathrm{i})\right|\ge\kappa(u),\qquad u\in\mathbb{R},\tag{3.4}$$

where the factor $1/3$ is used for technical reasons. As discussed above, we can restrict our investigation to the case $s\ge 2$. Since the Lévy process is only identifiable if $\mathcal{O}$ is known on the whole real line, we consider asymptotics of a growing number of observations with

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

Taking into account the numerical interpolation error and the stochastic error, we analyze the risk of the estimators in terms of the abstract noise level

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

Theorem 3.2. *Let $s \geq 2$, $R, \bar{\alpha} > 0$ and assume $\mathrm{e}^{-A} \lesssim \Delta^2$ and $\Delta\|\delta\|_{l^2}^2 \lesssim \|\delta\|_{l^\infty}^2$. We choose the cut-off value $U_{\bar{\alpha}} := \varepsilon^{-2/(2s+2T\bar{\alpha}+1)}$ to obtain the uniform convergence rates*

$$\sup_{\mathcal{P}=(\gamma,k)\in\mathcal{G}_s(R,\bar{\alpha})} \mathbb{E}_{\mathcal{P}}\big[|\hat{\gamma}-\gamma|^2\big]^{1/2} \lesssim \varepsilon^{2s/(2s+2T\bar{\alpha}+1)} \quad \textit{and}$$

$$\sup_{\mathcal{P}=(\gamma,k)\in\mathcal{G}_s(R,\bar{\alpha})} \mathbb{E}_{\mathcal{P}}\big[|\hat{\alpha}_j-\alpha_j|^2\big]^{1/2} \lesssim \varepsilon^{2(s-1-j)/(2s+2T\bar{\alpha}+1)}, \qquad j=0,\dots,s-2.$$

As one may expect the rates for α_j, $j=0,\dots,s-2$, become slower as j gets closer to its maximal value because the profit from the smoothness of k decreases. Note that the cut-off for all estimators is the same. In contrast to $\mathcal{G}_s(R,\bar{\alpha})$ we assume Sobolev conditions on k_e in the class $\mathcal{H}_s(R,\bar{\alpha})$ in order to apply L^2-Fourier analysis.

Definition 3.3. *Let $s \in \mathbb{N}$ and $R, \bar{\alpha} > 0$. We define $\mathcal{H}_s(R,\bar{\alpha})$ as the set of all pairs $\mathcal{P}=(\gamma,k) \in \mathcal{G}_0(R,\bar{\alpha})$ satisfying additionally $k \in C^s(\mathbb{R}\setminus\{0\})$, $\mathbb{E}_{\mathcal{P}}[|X_T e^{X_T}|] \leq R$ for corresponding Lévy process X as well as*

$$|\gamma| \leq R \quad \textit{and} \quad \big\|k_e^{(l)}\big\|_{L^2} \leq R \qquad \textit{for } l=0,\dots,s.$$

In the next theorem the conditions on A and δ are stronger than for the upper bounds of the parameters which is due to the necessity to estimate also the derivative of ψ. However, the estimation of ψ' does not lead to a loss in the rate. As seen in (3.1), we need $\hat{\gamma}$ to estimate k_e.

Theorem 3.4. *Let $s \geq 1$, $R, \bar{\alpha} > 0$ and assume $Ae^{-A} \lesssim \Delta^2$ as well as $\Delta(\|\delta_j\|_{l^2}^2 + \|(x_j\delta_j)_j\|_{l^2}^2) \lesssim \|\delta\|_{l^\infty}^2$. Using an estimator $\hat{\gamma}$ which satisfies $\sup_{\mathcal{P}} \mathbb{E}_{\mathcal{P}}[|\hat{\gamma}-\gamma|^2] < \infty$ and choosing the cut-off value $U_{\bar{\alpha}} := \varepsilon^{-2/(2s+2T\bar{\alpha}+5)}$, we obtain for the risk of $\hat{k}_e$ the uniform convergence rate*

$$\sup_{\mathcal{P}=(\gamma,k)\in\mathcal{H}_s(R,\bar{\alpha})} \mathbb{E}_{\mathcal{P}}\big[\|\hat{k}_e-k_e\|_{L^2}^2\big]^{1/2} \lesssim \varepsilon^{2s/(2s+2T\bar{\alpha}+5)}.$$

Remark 3.5. The convergence rates in the Theorems 3.2 and 3.4 are minimax optimal up to a logarithmic factor, which is shown in the supplementary article [27].

4. Adaptation

The convergence rate of our estimation procedure depends on the bound $\bar{\alpha}$ of the true but unknown $\alpha \in \mathbb{R}_+$. Therefore, we construct an α-adaptive estimator. For simplicity we concentrate on the estimation of α itself whereas the results can be easily extended to γ, α_j, $j=1,\dots,s-2$, and k_e. In this section, we will require the following assumption.

Assumption 4. *Let $R > 0$, $s \geq 2$ and $\alpha \in [0,\bar{\alpha}]$ for some maximal $\bar{\alpha} > 0$. Furthermore, we suppose $\mathrm{e}^{-A} \lesssim \Delta^2$ and $\Delta\|\delta\|_{l^2}^2 \lesssim \|\delta\|_{l^\infty}^2$.*

These conditions only recall the setting in which the convergence rates of our parameter estimators were proven. Given a consistent preestimator $\hat{\alpha}_{\text{pre}}$ of α, let $\tilde{\alpha}_0$ be the estimator using the data-driven cut-off value and the trimming parameter

$$\tilde{U} := U_{\hat{\alpha}_{\text{pre}}} := \varepsilon^{-2/(2s+2T\hat{\alpha}_{\text{pre}}+1)} \quad \text{and} \tag{4.1}$$

$$\tilde{\kappa}(u) := \kappa_{\bar{\alpha}_{\text{pre}}}(u) := \begin{cases} \frac{1}{2}\mathrm{e}^{-TR}, & |u| < 1, \\ \frac{1}{2}C_{\bar{\alpha}_{\text{pre}}}|u|^{-T\bar{\alpha}_{\text{pre}}}, & |u| \geq 1, \end{cases} \tag{4.2}$$

respectively, with $\bar{\alpha}_{\text{pre}} := \hat{\alpha}_{\text{pre}} + |\log\varepsilon|^{-1}$. If $\hat{\alpha}_{\text{pre}}$ is sufficiently concentrated around the true value, the adaptation does not lead to losses in the rate as the following proposition shows. Note that the condition $\tilde{\alpha}_0 \in [0, \bar{\alpha}]$ is not restrictive since any estimator $\hat{\alpha}$ of $\alpha \in [0, \bar{\alpha}]$ can be improved by using $(0 \vee \hat{\alpha}) \wedge \bar{\alpha}$ instead.

Proposition 4.1. *On Assumption 4 let $\hat{\alpha}_{\text{pre}}$ be a consistent estimator which is independent of the data O_j, $j = 1, \ldots, N$, and fulfills for $\varepsilon \to 0$ the inequality*

$$\mathbb{P}\big(|\hat{\alpha}_{\text{pre}} - \alpha| \geq |\log\varepsilon|^{-1}\big) \leq d\varepsilon^2 \tag{4.3}$$

with a constant $d \in (0, \infty)$. Furthermore, we suppose $\tilde{\alpha}_0 \in [0, \bar{\alpha}]$ almost surely. Then $\tilde{\alpha}_0$ satisfies the asymptotic risk bound

$$\sup_{\mathcal{P}\in\mathcal{G}_s(R,\alpha)} \mathbb{E}_{\mathcal{P},\hat{\alpha}_{\text{pre}}}\big[|\tilde{\alpha}_0 - \alpha|^2\big]^{1/2} \lesssim \varepsilon^{2(s-1)/(2s+2T\alpha+1)},$$

where the expectation is taken with respect to the common distribution $\mathbb{P}_{\mathcal{P},\hat{\alpha}_{\text{pre}}}$ of the observations $O_1, \ldots, O_N$ and the preestimator $\hat{\alpha}_{\text{pre}}$.

To use $\hat{\alpha}_0$ on an independent sample as preestimator, we establish a concentration result for the proposed procedure. We require (ε_j) to be uniformly sub-Gaussian (see, e.g., van de Geer [28]). That means there are constants $C_1, C_2 \in (0, \infty)$ such that the following concentration inequality holds for all $t, N > 0$ and $a_1, \ldots, a_N \in \mathbb{R}$

$$\mathbb{P}\left(\left|\sum_{j=1}^{N} a_j\varepsilon_j\right| \geq t\right) \leq C_1 \exp\left(-C_2 \frac{t^2}{\sum_{j=1}^{N} a_j^2}\right). \tag{4.4}$$

Proposition 4.2. *Additionally to Assumption 4 let (ε_j) be uniformly sub-Gaussian fulfilling (4.4). Then there is a constant $c > 0$ and for all $\kappa > 0$ there is an $\varepsilon_0 \sim \kappa^{(2s+2T\bar{\alpha}+1)/(2s-2)}$, such that for all $\varepsilon < \varepsilon_0 \wedge 1$ the estimator $\hat{\alpha}_0$ satisfies*

$$\mathbb{P}\big(|\hat{\alpha}_0 - \alpha| \geq \kappa\big) \leq \big((7N+1)C_1 + 2\big)\exp\big(-c\big(\kappa^2 \wedge \kappa^{1/2}\big)\varepsilon^{-(s-1)/(2s+2T\bar{\alpha}+1)}\big). \tag{4.5}$$

Concentration (4.5) is stronger than needed in Proposition 4.1. To apply the proposed estimation procedure, let S_{pre} and S be two independent samples with noise levels ε_{pre} and ε as well as sample sizes N_{pre} and N, respectively. Using S_{pre} for the estimator $\hat{\alpha}_{\text{pre}}$, we construct adaptively

$\tilde{\alpha}_0$ on S. We suppose N_{pre} grows at most polynomial in ε_{pre}, that is $N_{\text{pre}} \lesssim \varepsilon_{\text{pre}}^{-p}$ holds for some $p > 0$, cf. [27]. To satisfy (4.3), it is sufficient if there exists a power $q > 0$, which can be arbitrary small, such that $\varepsilon_{\text{pre}} \sim \varepsilon^q$ owing to the exponential inequality (4.5). Using $\varepsilon^2 \gtrsim A_N/N \geq 1/N$, we estimate

$$\frac{N_{\text{pre}}}{N} \lesssim \varepsilon_{\text{pre}}^{-p} \varepsilon^2 \sim \varepsilon^{2-pq} \to 0$$

for $q < 2/p$. Thus, relatively to all available data the necessary number of observations for the preestimator tends to zero.

5. Discussion and application

5.1. Numerical example

We apply the proposed estimation procedure to the variance gamma model. In view of the empirical study [17] we choose the parameters $\nu \in \{0.05, 0.1, 0.2, 0.5\}$, $\sigma = 1.2$ and $\theta = -0.15$. the martingale condition (2.2) yields then $\gamma = \frac{1}{\nu} \log(1 - \theta\nu - \sigma^2\nu/2)$. According to the different choices of ν, we set $\bar{\alpha} = 40$ as maximal value of α.

The deterministic design of the sample $\{x_1, \ldots, x_N\}$ is distributed normally with mean zero and variance $1/3$. The observations O_j are computed from the characteristic function φ_T using the fast Fourier transform method [7]. The additive noise consists of normal centered random variables with variance $|\delta\mathcal{O}(x_j)|^2$ for some $\delta > 0$.

We estimate $q \in \{\gamma, \alpha_0, \alpha_1, k\}$. Hence, we need $s \geq 4$, see Corollary 8 in [27]. By self-decomposablity of the model we apply the rearranged estimator $\hat{k}^*$ given by (3.3). We use maturity $T = 0.25$, interest $r = 0.06$, smoothness $s = 6$, sample size $N = 100$ and noise level $\delta = 0.01$, which generates values of ε on average 0.168. The results of 1000 Monte Carlo simulations are summarized in Tables 1 and 2.

In order to apply the estimation procedure, we need to choose the tuning parameters. Owing to the typically unknown smoothness s, let the weight functions satisfy Assumption 3 for some large value $s_{\max}$. The weights for the parameters can be chosen as polynomial whereas W_k is taken as a polynomial times a smooth function with support $[-1, 0]$. The trimming parameter κ is included mainly for theoretical reasons and is not important to the implementation. The most crucial point is the choice of the cut-off value U. For $\hat{q}$ we implement the oracle method

Table 1. Risk of estimating α with oracle (*middle column*) and adaptive (*right column*) choice of cut-off value U in simulated variance gamma model with $\nu \in \{0.05, 0.1, 0.2, 0.5\}, \sigma = 1.2, \theta = -0.15$

α	$\mathbb{E}[\lvert\hat{\alpha}_0 - \alpha\rvert^2]^{1/2}$	$\mathbb{E}[\lvert\tilde{\alpha}_0 - \alpha\rvert^2]^{1/2}$
40	20.7998	23.3589
20	5.8362	7.7724
10	1.0505	2.4534
4	0.1729	1.1158

Table 2. Risk of estimating the parameters γ, α, α_1 and the k-function with oracle (*middle column*) and adaptive (*right column*) choice of the cut-off value U in simulated variance gamma model ($\nu = 0.2, \sigma = 1.2, \theta = -0.15$)

	q	$\mathbb{E}[\lvert\hat{q}-q\rvert^2]^{1/2}$	$\mathbb{E}[\lvert\tilde{q}-q\rvert^2]^{1/2}$
γ	0.1408	0.0065	0.0126
α_0	10.0000	1.0505	2.4534
α_1	−94.1667	32.1016	77.5311
	$\lVert q\rVert_{L^2}^{1/2}$	$\mathbb{E}[\lVert\hat{q}-q\rVert_{L^2}^2]^{1/2}$	$\mathbb{E}[\lVert\tilde{q}-q\rVert_{L^2}^2]^{1/2}$
k_e	0.9556	0.4075	0.5602

$U = \operatorname{argmin}_{V \geq 0} |\hat{q}(V) - q|$ and an adaptive estimator $\tilde{q}$ based on the construction of Section 4 with sample size $N_{\text{pre}} = 25$ for $\hat{\alpha}_{\text{pre}}$.

5.2. Discussion

Due to the nonparametric setting, our estimators converge more slowly than with $\sqrt{n}$ rate as in parametric models [10,11,17]. Although the studied estimation problem is only mildly ill-posed compared with classical nonparametric regression models and thus the polynomial rates are faster than in nonparametric models with $\sigma > 0$ which achieve logarithmic rates only [3]. In order to understand the convergence rate of the estimators for γ and α_j better, we rewrite equation (2.10) in the distributional sense, denoting the Dirac distribution at zero by δ_0, and differentiate representation (2.8)

$$\psi'(u) = \mathcal{F}(\mathrm{i}\gamma\delta_0 + \mathrm{i}k_e)(u) = \mathrm{i}\gamma - \sum_{j=0}^{s-2} \mathrm{i}^j j!\alpha_j u^{-j-1} + \rho'(u), \qquad u \in \mathbb{R} \setminus \{0\}.$$

Hence, ψ' can be seen as Fourier transform of an s-times weakly differentiable function and estimating γ from noisy observations of ψ' corresponds to a nonparametric regression with regularity s. Since dividing by u on the right-hand side of the above equation corresponds to taking the derivative in the spatial domain, the estimation of α_j is similar to the estimation of the $(j+1)$th derivative in a regression model. The convergence rate of k_e is in line with the results of Belomestny and Reiß [3] for $\sigma = 0$ since their rate equals ours in the compound Poisson case $\alpha = 0$.

For $\hat{k}_e$, the degree of ill-posedness is given by $T\alpha + 2$. This can be seen analytically by observing that the noise is governed by $u^2|\varphi_T(u-\mathrm{i})|^{-1}$, which grows with rate $T\alpha + 2$. From a statistical point of view a higher value of α leads to a more active Lévy process and hence, it is harder to distinguish the small jumps of the process from the additive noise. The influence of the time to maturity T on the convergence rates is an interesting deviation from the analysis of Belomestny and Reiß [3]. The simulation shown in Table 1 demonstrates the improvement of the estimation for small the values of α. The estimators $\hat{\gamma}$ and $\hat{k}_e$ provide a complete calibration of the model. Although, estimating the k-function at zero is most important and thus additional information through $\hat{\alpha}_j$ are crucial. Table 2 contains simulation results for the estimators $\hat{q}$ and $\tilde{q}$, $q \in \{\gamma, \alpha_0, \alpha_1, k_e\}$, corresponding to oracle and α-adaptive cut-off values, respectively. This adaptation to α is a first step to a data-driven procedure and should be developed further.

Since the estimating equation (2.9) holds for all Lévy processes with finite variation, the proposed estimator $\hat{k}_e$ can be more generally understood as estimator of $xe^x\nu(\mathrm{d}x)$. Thus, the estimation procedure can be applied to exponential Lévy models with Blumenthal–Getoor index larger than zero, for example, tempered stable processes. However, the analysis of the convergence rates does not carry over to more general Lévy processes since the polynomial decay of the φ_T, which is guaranteed by property (K), is essential for our proofs. Moreover, if k has no bounded variation the behavior of the Lévy density at zero needs different methods and should be studied further. For instance, Belomestny [1] discusses the estimation of the fractional order for regular Lévy models of exponential type.

Even if the practitioner prefers specific parametric models that might achieve smaller errors and faster rates, the nonparametric method should be used as a goodness-of-fit test against model misspecification. To construct such tests, confidence sets need to be studied which is done by Söhl [24] in the framework of Lévy processes with finite activity. Based on this asymptotic analysis, Söhl and Trabs [25] construct confidence intervals in the self-decomposable model.

5.3. Real data example

We apply our estimation method to a data set from the Deutsche Börse database Eurex.[3] It consists of settlement prices of put and call options on the DAX index with three and six months to maturity from 29 May 2008. The sample sizes are 101 and 106, respectively. The interest rate is chosen according to the put-call parity. The sub-sample for the preestimator consists of every fifth strike while the main estimation is done from the remaining data points. By a rule of thumb, the bid-ask spread is chosen as 1% of the option prices. Therefore, we get noise levels ε with values 0.0138 and 0.069 for the two maturities, respectively. Table 3 shows the result of

Table 3. Adaptive estimation based on DAX options from 29 May 2008 with time to maturity T and $N + N_{\mathrm{pre}}$ observations

T	0.314	0.567
r	0.045	0.044
N_{pre}	20	21
N	81	85
$\tilde{\gamma}$	0.101	0.344
$\tilde{\alpha}_0$	34.848	23.600
$\tilde{\alpha}_1$	239.348	147.699

[3]Provided through the Collaborative Research Center 649 “Economic Risk”.

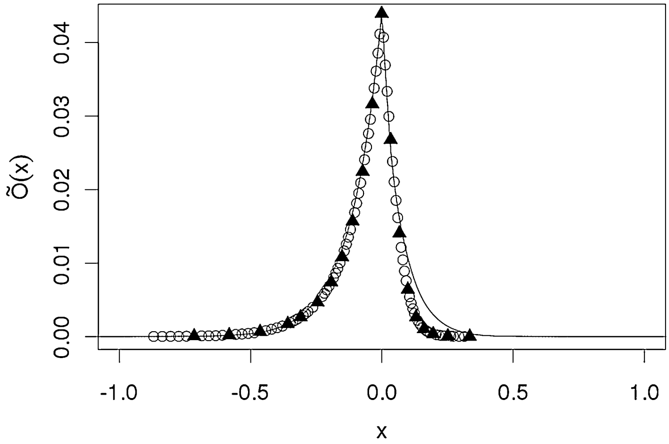

Figure 1. Observed DAX option prices (*points and triangles*) from 29 May 2008 with fixed maturity $T = 0.314$ and different log strike prices x as well as the option function generated from the estimated model (*solid line*).

the proposed method. As one would expect, the jump activity is smaller for a longer time to maturity. The estimator $\hat{k}(x) = \mathrm{e}^{-x}\hat{k}_e(x)$ as well as the rearranged estimator $\hat{k}^*$ are presented in Figure 2. In Figure 1, the calibrated model is used to generate the option function in the case of three months to maturity, where the data points used for the preestimator are marked with triangles in the figure. For a comparison of the outcome of our estimation procedure with the spectral calibration of Belomestny and Reiß [3], we refer to Söhl and Trabs [25].

6. Proofs

6.1. Proof of the upper bounds

Let us recall some results of [3]: Because of the B-spline interpolation we obtain $\mathcal{O}_l(x) := \mathbb{E}[\tilde{\mathcal{O}}(x)] = \sum_{j=1}^N \mathcal{O}(x_j)b_j(x) + \beta_0(x)$, $x \in \mathbb{R}$. Furthermore, the decomposition of the stochastic error $\tilde{\psi} - \psi$ in a linearization $\mathcal{L}$ and a remainder $\mathcal{R}$,

$$\mathcal{L}(u) := T^{-1}\varphi_T(u-\mathrm{i})^{-1}(\mathrm{i}-u)u\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u), \qquad \mathcal{R}(u) := \tilde{\psi}(u) - \psi(u) - \mathcal{L}(u),$$

$u \in \mathbb{R}$, has the following properties.

Proposition 6.1. (i) *Under the hypothesis* $\mathrm{e}^{-A} \lesssim \Delta^2$ *we obtain* $\sup_{u\in\mathbb{R}} |\mathbb{E}[\mathcal{F}\tilde{\mathcal{O}}(u) - \mathcal{F}\mathcal{O}(u)]| = \sup_{u\in\mathbb{R}} |\mathcal{F}\mathcal{O}_l(u) - \mathcal{F}\mathcal{O}(u)| \lesssim \Delta^2$ *uniformly over all Lévy triplets satisfying Assumption* 1.

(ii) *If the function* $\kappa\colon \mathbb{R} \to \mathbb{R}_+$ *satisfies* (3.4), *then for all* $u \in \mathbb{R}$ *the remainder is bounded by* $|\mathcal{R}(u)| \le T^{-1}\kappa(u)^{-2}(u^4+u^2)|\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u)|^2$.

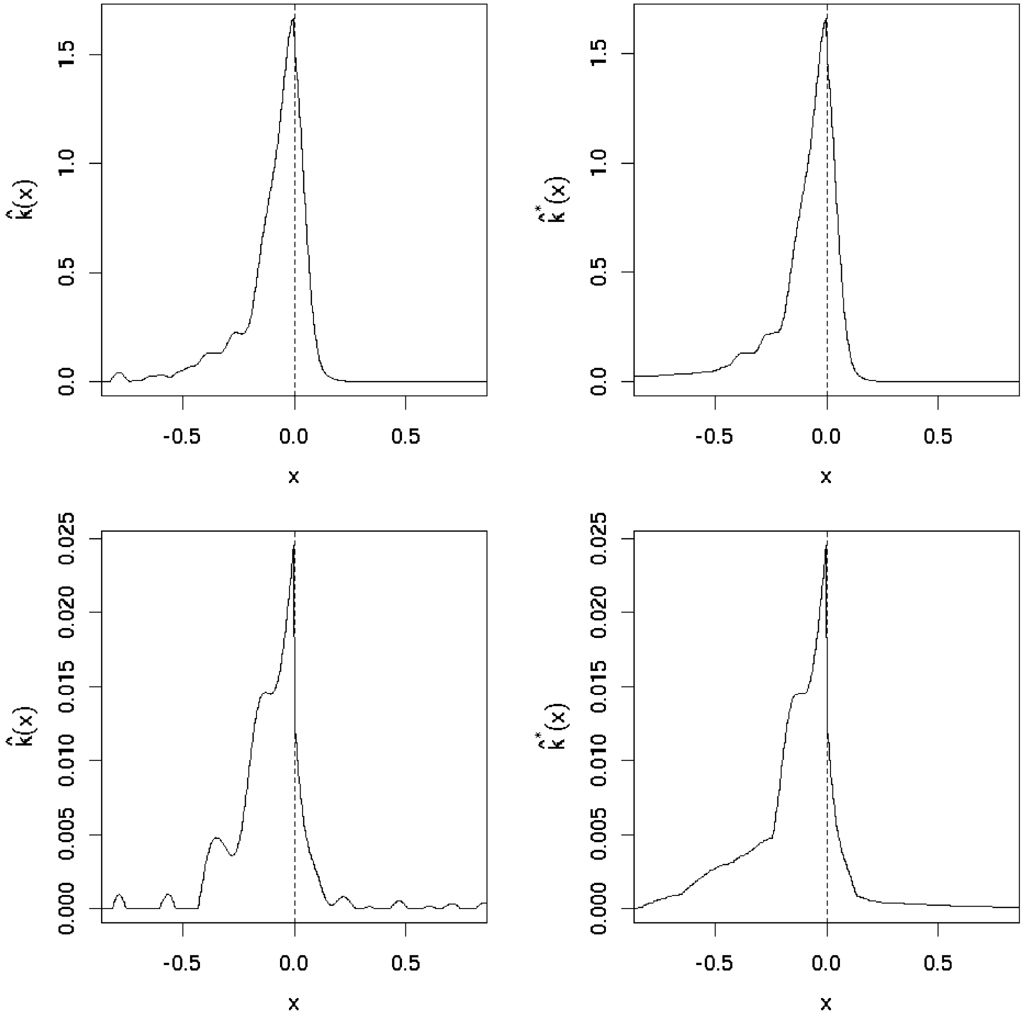

Figure 2. Using ODAX data from 29 May 2008 with three (*top*) and six (*bottom*) months to maturity, estimation of the function k with (*right*) and without (*left*) rearrangement.

Upper bound for γ *and* α_j (*Theorem* 3.2)

Since Theorem 3.2 can be proven analogously to Theorem 4.2 in [3], we only sketch the main steps. Note that in $\mathcal{G}_s(R, \bar{\alpha})$ we can bound uniformly $|u^{s-1}\rho(u)|$ in representation (2.8), cf. Lemma 9 in [27]. Let us consider γ first. The definition of $\hat{\gamma}$ and w_γ^U, the decomposition of $\tilde{\psi}$ and representation (2.8) yield

$$\hat{\gamma} = \int_{-U}^{U} \operatorname{Im}\big(\tilde{\psi}(u)\big) w_\gamma^U(u)\,\mathrm{d}u = \gamma + \int_{-U}^{U} \operatorname{Im}\big(\rho(u) + \mathcal{L}(u) + \mathcal{R}(u)\big) w_\gamma^U(u)\,\mathrm{d}u.$$

Hence, we obtain

$$\mathbb{E}\big[|\hat{\gamma}-\gamma|^2\big]\le 3\bigg|\int_{-U}^{U}\rho(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|^2+3\mathbb{E}\bigg[\bigg|\int_{-U}^{U}\mathcal{L}(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|^2\bigg]$$
$$+3\mathbb{E}\bigg[\bigg|\int_{-U}^{U}\mathcal{R}(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|^2\bigg],$$

where all three summands can be estimated separately. The first one is a deterministic error term. It can be estimated using the decay of $\rho(u)$ and the weight function property (3.2):

$$\bigg|\int_{-U}^{U}\rho(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|\lesssim\int_{-U}^{U}U^{-(s+1)}\big|\rho(u)u^{s-1}\big|\,\mathrm{d}u\lesssim U^{-s}.$$

A bias-variance decomposition, with the definition $\operatorname{Var}(Z):=\mathbb{E}[|Z-\mathbb{E}[Z]|^2]$, of the linear error term yields

$$\mathbb{E}\bigg[\bigg|\int_{-U}^{U}\mathcal{L}(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|^2\bigg]$$
$$=\bigg|\int_{-U}^{U}\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathbb{E}\big[\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u)\big]w_\gamma^U(u)\,\mathrm{d}u\bigg|^2$$
$$+\operatorname{Var}\bigg(\int_{-U}^{U}\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}\tilde{\mathcal{O}}(u)w_\gamma^U(u)\,\mathrm{d}u\bigg)=:\mathcal{L}_b^2+\mathcal{L}_v.$$

Using the approximation result in Proposition 6.1, the bound of $|\varphi_T(u-\mathrm{i})|^{-1}$ given by κ^{-1} and property (3.2), we infer the estimate of the bias term

$$|\mathcal{L}_b|\lesssim\Delta^2U^{-(s+1)}\int_{-U}^{U}\big|\varphi_T(u-\mathrm{i})\big|^{-1}|u|^{s+1}\,\mathrm{d}u\lesssim\Delta^2U^{T\bar{\alpha}+1}.$$

For the variance part, we make use of the properties of the linear spline functions b_k as well as $\operatorname{supp}(w_\gamma^U)\subseteq[-U,U]$ and the independence of (ε_k). We estimate $(\operatorname{Cov}(Y,Z):=\mathbb{E}[(Y-\mathbb{E}[Y])\overline{(Z-\mathbb{E}[Z])}])$

$$\mathcal{L}_v=\int_{-U}^{U}\int_{-U}^{U}\operatorname{Cov}\bigg(\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}\tilde{\mathcal{O}}(u),\frac{(\mathrm{i}-v)v}{T\varphi_T(v-\mathrm{i})}\mathcal{F}\tilde{\mathcal{O}}(v)\bigg)w_\gamma^U(u)w_\gamma^U(v)\,\mathrm{d}u\,\mathrm{d}v$$
$$=\sum_{k=1}^{N}\delta_k^2\bigg|\int_{-U}^{U}\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}b_k(u)w_\gamma^U(u)\,\mathrm{d}u\bigg|^2\lesssim\Delta\|\delta\|_{l^\infty}^2U^{2T\bar{\alpha}+1}.$$

To estimate the remaining term $\mathcal{R}$, we use Proposition 6.1, the property (3.2) of w_γ^U and the choice of κ. In addition the independence of (ε_k) and the uniform bound of their fourth moments comes into play.

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|\int_{-U}^{U}\mathcal{R}(u)w_\gamma^U(u)\,\mathrm{d}u\Big|^2\Big]\\
&\lesssim \int_{-U}^{U}\int_{-U}^{U}\big(\|\mathcal{F}(\mathcal{O}_l-\mathcal{O})\|_\infty^4+\mathbb{E}\big[\big|\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O}_l)(u)\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O}_l)(v)\big|^2\big]\big)\\
&\qquad\qquad\times\frac{u^4w_\gamma^U(u)v^4w_\gamma^U(v)}{\kappa(u)^2\kappa(v)^2}\,\mathrm{d}u\,\mathrm{d}v\\
&\lesssim\Big(\Delta^4\int_{-U}^{U}\frac{u^4w_\gamma^U(u)}{\kappa(u)^2}\,\mathrm{d}u\Big)^2+\Big(\int_{-U}^{U}\sum_{k=1}^{N}\delta_k^2\big|\mathcal{F}b_k(u)\big|^2\frac{u^4w_\gamma^U(u)}{\kappa(u)^2}\,\mathrm{d}u\Big)^2\\
&\lesssim\Big(\Delta^4U^{-(s+1)}\int_{-U}^{U}\kappa(u)^{-2}|u|^{s+3}\,\mathrm{d}u\Big)^2+\Big(\Delta^2\|\delta\|_{l^2}^2U^{-(s+1)}\int_{-U}^{U}\kappa(u)^{-2}|u|^{s+3}\,\mathrm{d}u\Big)^2\\
&\lesssim U^{4T\bar{\alpha}+6}\big(\Delta^8+\Delta^4\|\delta\|_{l^2}^4\big).
\end{aligned}
$$

Therefore, the total risk of $\hat{\gamma}$ is of order

$$\mathbb{E}\big[|\hat{\gamma}-\gamma|^2\big]\lesssim U^{-2s}+U^{2T\bar{\alpha}+1}\big(\Delta^4U+\Delta\|\delta\|_{l^\infty}^2\big)+U^{4T\bar{\alpha}+6}\big(\Delta^8+\Delta^4\|\delta\|_{l^2}^4\big)$$

uniformly over $\mathcal{G}_s(R,\bar{\alpha})$. Since the explicit choice of $U=U_{\bar{\alpha}}=\varepsilon^{-2/(2s+2T\bar{\alpha}+1)}$ fulfills $U\lesssim\Delta^{-1}$ and $\Delta\|\delta\|_{l^2}^2\lesssim\|\delta\|_{l^\infty}^2$ holds by assumption, this bound simplifies to

$$\mathbb{E}\big[|\hat{\gamma}-\gamma|^2\big]\lesssim U^{-2s}+U^{2T\bar{\alpha}+1}\varepsilon^2+U^{4T\bar{\alpha}+6}\varepsilon^4.$$

Here $U_{\bar{\alpha}}$ balances the trade-off between the first and the second term whereby the third term is asymptotically negligible. We obtain the claimed rate.

For α_j, $j=0,\dots,s-2$, the only difference to the analysis for $\hat{\gamma}$ is the rescaling factor of $w_{\alpha_j}^U$ in (3.2). Since its square appears in front of every term, we verify

$$
\begin{aligned}
\mathbb{E}\big[|\hat{\alpha}_j-\alpha_j|^2\big]&\lesssim U^{-2(s-1-j)}+U^{2T\bar{\alpha}+2j+3}\big(\Delta^4U+\Delta\|\delta\|_{l^\infty}^2\big)+U^{4T\bar{\alpha}+2j+8}\big(\Delta^8+\Delta^4\|\delta\|_{l^2}^4\big)\\
&\lesssim U^{-2(s-1-j)}+U^{2T\bar{\alpha}+2j+3}\varepsilon^2+U^{4T\bar{\alpha}+2j+8}\varepsilon^4.
\end{aligned}
$$

Upper bound for k_e (Theorem 3.4)

Similarly to the uniform bound of the bias of $\mathcal{F}\tilde{\mathcal{O}}$ in Proposition 6.1, the following lemma holds true. It can be proved analogously to Proposition 1 in [3] and thus we omit the details.

Lemma 6.2. *Assuming $Ae^{-A}\lesssim\Delta^2$, we obtain $\sup_{u\in\mathbb{R}}|\mathbb{E}[\mathcal{F}(x(\tilde{\mathcal{O}}-\mathcal{O})(x))(u)]|=\sup_{u\in\mathbb{R}}|\mathcal{F}(x(\mathcal{O}_l-\mathcal{O})(x))(u)|\lesssim\Delta^2$ uniformly over all Lévy triplets satisfying Assumption 1 and $\mathbb{E}[|X_Te^{X_T}|]\lesssim 1$.*

For convenience, we write $m := 2s - 1$ and $w_k := \mathcal{F}W_k$ such that $w_k(u/U) = U \times \mathcal{F}(W_k(Ux))(u)$. Using $\|f\|_{L^2}^2 = \int_{\mathbb{R}_+} |f(x)|^2\,\mathrm{d}x + \int_{\mathbb{R}_-} |f(x)|^2\,\mathrm{d}x =: \|f\|_{L^2(\mathbb{R}_+)}^2 + \|f\|_{L^2(\mathbb{R}_-)}^2$ for $f \in L^2(\mathbb{R})$, it is sufficient to consider the loss of $\hat{k}_e$ on $\mathbb{R}_+$. On $\mathbb{R}_-$ one can proceed analogously. We split the risk into a deterministic error, an error caused by $\hat{\gamma}$ and a stochastic error,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}}\big[\|\hat{k}_e - k_e\|_{L^2(\mathbb{R}_+)}^2\big]\\
&= \mathbb{E}_{\mathcal{P}}\Big[\Big\|\mathcal{F}^{-1}\Big(\big(-\hat{\gamma} - \mathrm{i}\tilde{\psi}'(u)\big)w_k\Big(\frac{u}{U}\Big)\Big) - k_e\Big\|_{L^2(\mathbb{R}_+)}^2\Big]\\
&\le \mathbb{E}_{\mathcal{P}}\Big[\int_{\mathbb{R}_+} 3\Big|\mathcal{F}^{-1}\Big(\big(-\gamma - \mathrm{i}\psi'(u)\big)w_k\Big(\frac{u}{U}\Big)\Big)(x) - k_e(x)\Big|^2\\
&\qquad\qquad + 3\Big|\mathcal{F}^{-1}\Big((\gamma - \hat{\gamma})w_k\Big(\frac{u}{U}\Big)\Big)(x)\Big|^2\\
&\qquad\qquad + 3\Big|\mathcal{F}^{-1}\Big(\big(-\mathrm{i}\tilde{\psi}'(u) + \mathrm{i}\psi'(u)\big)w_k\Big(\frac{u}{U}\Big)\Big)(x)\Big|^2\,\mathrm{d}x\Big]\\
&\le 3\int_{\mathbb{R}_+}\Big|\mathcal{F}^{-1}\Big(\mathcal{F}k_e(u)w_k\Big(\frac{u}{U}\Big)\Big)(x) - k_e(x)\Big|^2\,\mathrm{d}x\\
&\quad + 3\mathbb{E}\big[|\hat{\gamma} - \gamma|^2\big]\int_{\mathbb{R}_+}\big|UW_k(Ux)\big|^2\,\mathrm{d}x\\
&\quad + 3\mathbb{E}\Big[\int_{\mathbb{R}_+}\Big|\mathcal{F}^{-1}\Big(\big(\tilde{\psi}'(u) - \psi'(u)\big)w_k\Big(\frac{u}{U}\Big)\Big)(x)\Big|^2\,\mathrm{d}x\Big]\\
&=: D + G + S.
\end{aligned}$$

The support of W_k yields $G = 0$. The deterministic term D can be estimated in the spatial domain, where we use the local smoothness of k_e. For pointwise convergence rates, this was done in [2]. We decompose using $\operatorname{supp} W_k \subset (-\infty, 0]$

$$\begin{aligned}
D &= 3\int_{\mathbb{R}_+}\big|U\big(k_e * W_k(U\bullet)\big)(x) - k_e(x)\big|^2\,\mathrm{d}x\\
&\le 6\int_{\mathbb{R}_+}\Big|\int_U^\infty \big(k_e(x + y/U) - k_e(x)\big)W_k(-y)\,\mathrm{d}y\Big|^2\,\mathrm{d}x\\
&\quad + 6\int_{\mathbb{R}_+}\Big|\int_0^U \big(k_e(x + y/U) - k_e(x)\big)W_k(-y)\,\mathrm{d}y\Big|^2\,\mathrm{d}x\\
&=: 6(D_1 + D_2).
\end{aligned}$$

Cauchy–Schwarz's inequality, the estimate $\int_U^\infty |W_k(-y)|\,\mathrm{d}y \le U^{-m}\int_{\mathbb{R}} |y^m W_k(y)|\,\mathrm{d}y \lesssim U^{-m}$ and Fubini's theorem yield

$$\begin{aligned}
D_1 &\le \int_{\mathbb{R}_+}\int_U^\infty \big|W_k(-y)\big|\,\mathrm{d}y \int_U^\infty \big|k_e(x+y/U)-k_e(x)\big|^2\big|W_k(-y)\big|\,\mathrm{d}y\,\mathrm{d}x \\
&\lesssim U^{-m}\int_{\mathbb{R}_+}\int_U^\infty \big(\big|k_e(x+y/U)\big|^2+\big|k_e(x)\big|^2\big)\big|W_k(-y)\big|\,\mathrm{d}y\,\mathrm{d}x \\
&\lesssim U^{-m}\int_U^\infty \big|W_k(-y)\big|\int_{\mathbb{R}_+}\big|k_e(x+y/U)\big|^2+\big|k_e(x)\big|^2\,\mathrm{d}x\,\mathrm{d}y \lesssim U^{-2m}\|k_e\|_{L^2}^2.
\end{aligned}$$

Using a Taylor expansion, we split D_2 in a polynomial part and a remainder:

$$\begin{aligned}
D_2 \le{}& 2\int_{\mathbb{R}_+}\left|\int_0^U \left(\sum_{j=0}^{s-1}\frac{k_e^{(j)}(x)}{j!U^j}y^j\right)W_k(-y)\,\mathrm{d}y\right|^2\mathrm{d}x \\
&+2\int_{\mathbb{R}_+}\left|\int_0^U\int_x^{x+y/U}\frac{k_e^{(s)}(z)(x+y/U-z)^{s-1}}{(s-1)!}\,\mathrm{d}z\,W_k(-y)\,\mathrm{d}y\right|^2\mathrm{d}x =: 2D_{2P}+2D_{2R}.
\end{aligned}$$

We estimate by $\int_0^U y^j W_k(-y)\,\mathrm{d}y = -\int_U^\infty y^j W_k(-y)\,\mathrm{d}y$ for $j=0,\dots,s-1$

$$D_{2P} \le sU^{-2m}\sum_{j=0}^{s-1}\frac{1}{(j!)^2}\int_{\mathbb{R}_+}\big|k_e^{(j)}(x)\big|^2\,\mathrm{d}x\left(\int_{\mathbb{R}_+}\big|y^m W_k(-y)\big|\,\mathrm{d}y\right)^2 \lesssim U^{-2m}\sum_{j=0}^{s-1}\big\|k_e^{(j)}\big\|_{L^2}^2.$$

With twofold usage of Cauchy–Schwarz and with Fubini's theorem we obtain

$$\begin{aligned}
D_{2R} &= \int_{\mathbb{R}_+}\left|\int_0^U\int_0^{y/U}\frac{k_e^{(s)}(x+z)(y/U-z)^{s-1}}{(s-1)!}\,\mathrm{d}z\,W_k(-y)\,\mathrm{d}y\right|^2\mathrm{d}x \\
&\le \int_{\mathbb{R}_+}\left(\int_0^U\left(\int_0^{y/U}\big|k_e^{(s)}(x+z)\big|^2\,\mathrm{d}z\right)^{1/2}\right. \\
&\qquad\qquad \left.\times\left(\int_0^{y/U}\frac{(y/U-z)^{2s-2}}{((s-1)!)^2}\,\mathrm{d}z\right)^{1/2}\big|W_k(-y)\big|\,\mathrm{d}y\right)^2\mathrm{d}x \\
&\le \int_{\mathbb{R}_+}\int_0^U\int_0^{y/U}\big|k_e^{(s)}(x+z)\big|^2\,\mathrm{d}z\big|W_k(-y)\big|\,\mathrm{d}y\int_0^U\frac{(y/U)^{2s-1}}{(2s-1)((s-1)!)^2}\big|W_k(-y)\big|\,\mathrm{d}y\,\mathrm{d}x \\
&\lesssim U^{-(2s-1)}\int_0^U\int_0^{y/U}\int_{\mathbb{R}_+}\big|k_e^{(s)}(x+z)\big|^2\,\mathrm{d}x\,\mathrm{d}z\big|W_k(-y)\big|\,\mathrm{d}y \\
&\le U^{-(2s-1)}\big\|k_e^{(s)}\big\|_{L^2}^2\int_0^U\frac{y}{U}\big|W_k(-y)\big|\,\mathrm{d}y \lesssim U^{-2s}.
\end{aligned}$$

Therefore, we have $D+G\lesssim U^{-2s}+U\mathbb{E}[|\hat{\gamma}-\gamma|^2]$.

To estimate the stochastic error S, we bound the term $|\tilde{\psi}'(u) - \psi'(u)|$. Let us introduce the notation

$$\tilde{\varphi}_T(u-\mathrm{i}) := v_{\kappa(u)}\big(1 + \big(\mathrm{i}u - u^2\big)\mathcal{F}\tilde{\mathcal{O}}(u)\big),$$
$$\tilde{\varphi}_T'(u-\mathrm{i}) := (\mathrm{i} - 2u)\mathcal{F}\tilde{\mathcal{O}}(u) - \big(u + \mathrm{i}u^2\big)\mathcal{F}\big(x\tilde{\mathcal{O}}(x)\big)(u), \qquad u \in \mathbb{R}.$$

For all $u \in \mathbb{R}$ where $|\tilde{\varphi}_T(u-\mathrm{i})| > \kappa(u)$ we obtain $\tilde{\varphi}_T(u-\mathrm{i}) = 1 + (\mathrm{i}u - u^2)\mathcal{F}\tilde{\mathcal{O}}(u)$. For $|\tilde{\varphi}_T(u-\mathrm{i})| = \kappa(u)$ the estimate $|\tilde{\varphi}_T(u-\mathrm{i}) - \varphi_T(u-\mathrm{i})| \geq 2\kappa(u)$ follows from (3.4). This yields

$$\begin{aligned}\big|\tilde{\varphi}_T(u-\mathrm{i}) - \varphi_T(u-\mathrm{i})\big| &\leq \big|1 + \big(\mathrm{i}u - u^2\big)\mathcal{F}\tilde{\mathcal{O}}(u) - \varphi_T(u-\mathrm{i})\big| + \kappa(u)\\ &\leq \big|1 + \big(\mathrm{i}u - u^2\big)\mathcal{F}\tilde{\mathcal{O}}(u) - \varphi_T(u-\mathrm{i})\big| + \tfrac{1}{2}\big|\tilde{\varphi}_T(u-\mathrm{i}) - \varphi_T(u-\mathrm{i})\big|.\end{aligned}$$

Therefore, $|\tilde{\varphi}_T(u-\mathrm{i}) - \varphi_T(u-\mathrm{i})| \leq 2|1 + (\mathrm{i}u - u^2)\mathcal{F}\tilde{\mathcal{O}}(u) - \varphi_T(u-\mathrm{i})|$ holds for all $u \in \mathbb{R}$. We obtain a similar decomposition as [15],

$$\begin{aligned}\big|\tilde{\psi}'(u) - \psi'(u)\big| &= \frac{1}{T}\left|\frac{\tilde{\varphi}_T'(u-\mathrm{i})}{\tilde{\varphi}_T(u-\mathrm{i})} - \frac{\varphi_T'(u-\mathrm{i})}{\varphi_T(u-\mathrm{i})}\right|\\ &\leq \frac{1}{T|\tilde{\varphi}_T(u-\mathrm{i})|}\big(\big|\tilde{\varphi}_T'(u-\mathrm{i}) - \varphi_T'(u-\mathrm{i})\big| + T\big|\psi'(u)\big|\big|\varphi_T(u-\mathrm{i}) - \tilde{\varphi}_T(u-\mathrm{i})\big|\big)\\ &\leq \frac{1}{2T\kappa(u)}\Big(\Big(\big(1+4u^2\big)^{1/2} + 2T\big|\psi'(u)\big|\big(u^2+u^4\big)^{1/2}\Big)\big|\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)\big|\\ &\qquad + \big(u^2+u^4\big)^{1/2}\big|\mathcal{F}\big(x(\tilde{\mathcal{O}} - \mathcal{O})(x)\big)(u)\big|\Big).\end{aligned}$$

Since $|\psi'(u)| \leq |\gamma| + \|k_e\|_{L^1} \leq 2R$, we have

$$\big|\tilde{\psi}'(u) - \psi'(u)\big| \lesssim \frac{1}{\kappa(u)}\Big(\big(1+u^2\big)\big|\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)\big| + \big(u^2+u^4\big)^{1/2}\big|\mathcal{F}\big(x(\tilde{\mathcal{O}} - \mathcal{O})(x)\big)(u)\big|\Big).$$

It follows with Plancherel's equality

$$\begin{aligned}S &\leq 3\mathbb{E}\big[\big\|\mathcal{F}^{-1}\big(\big(\tilde{\psi}'(u) - \psi'(u)\big)w_k(u/U)\big)\big\|_{L^2}^2\big]\\ &= \frac{3}{2\pi}\int_{\mathbb{R}}\mathbb{E}\big[\big|\tilde{\psi}'(u) - \psi'(u)\big|^2\big]\big|w_k(u/U)\big|^2\,\mathrm{d}u\\ &\lesssim \int_{\mathbb{R}}\frac{u^4}{|\kappa(u)|^2}\Big(\mathbb{E}\big[\big|\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)\big|^2\big]\\ &\qquad + \mathbb{E}\big[\big|\mathcal{F}\big(x(\tilde{\mathcal{O}} - \mathcal{O})(x)\big)(u)\big|^2\big]\Big)\big|w_k(u/U)\big|^2\,\mathrm{d}u\\ &=: S_1 + S_2.\end{aligned}$$

Both terms can be estimated similarly. Thus, we only write it down for S_2, where stronger conditions are needed. Lemma 6.2 and $\|\mathcal{F}(xb_j(x))\|_\infty \le 2\Delta(x_j+\Delta)$, $j=1,\ldots,N$, yield

$$S_2 \le \int_{\mathbb{R}} \frac{u^4}{|\kappa(u)|^2}\Big(\big\|x(O_l-O)(x)\big\|_\infty^2 + \operatorname{Var}\big(\mathcal{F}\big(x\tilde{\mathcal{O}}(x)\big)(u)\big)\Big)\big|w_k(u/U)\big|^2\,\mathrm{d}u$$
$$\lesssim \int_{\mathbb{R}} |u|^{2T\bar{\alpha}+4}\left(\Delta^4 + \sum_{j=1}^N \delta_j^2\big|\mathcal{F}\big(xb_j(x)\big)(u)\big|^2\right)\big|w_k(u/U)\big|^2\,\mathrm{d}u$$
$$\lesssim \big(\Delta^4 + \Delta^2\big\|(x_j\delta_j)\big\|_{l^2}^2 + \Delta^4\|\delta_j\|_{l^2}^2\big)U^{2T\bar{\alpha}+5} \lesssim \varepsilon^2 U^{2T\bar{\alpha}+5}.$$

Therefore, we have shown $\mathbb{E}[\|\hat{k}_e - k_e\|_{L^2,\tau}^2] \lesssim U^{-2s} + \varepsilon^2 U^{2T\bar{\alpha}+5} + U\mathbb{E}[|\hat{\gamma}-\gamma|^2]$. The assertion follows from the asymptotic optimal choice $U = U_{\bar{\alpha}} = \varepsilon^{-2/(2s+2T\bar{\alpha}+5)}$ and the assumption on the risk of $\hat{\gamma}$.

6.2. Proof of Proposition 4.1

Step 1: Let $(a_\varepsilon)_{\varepsilon>0}$ be a deterministic sequence such that there is a constant $C>0$ with $|a_\varepsilon - \alpha| \le C|\log\varepsilon|^{-1}$. Let the estimator $\hat{\alpha}_0$ use the cut-off value $U_\varepsilon := \tilde{U}_{a_\varepsilon}$ and the trimming parameter $\kappa_\varepsilon := \tilde{\kappa}_{\bar{a}_\varepsilon}$, with $\bar{a}_\varepsilon := a_\varepsilon + C|\log\varepsilon|^{-1}$, as defined in (4.1) and (4.2). Then we can show the asymptotic risk bound $\sup_{\mathcal{P}\in\mathcal{G}_s(R,\alpha)} \mathbb{E}_{\mathcal{P}}[|\hat{\alpha}_0-\alpha|^2]^{1/2} \lesssim \varepsilon^{2(s-1)/(2s+2T\alpha+1)}$ as follows: By construction holds $\alpha \le \bar{a}_\varepsilon$. Hence, κ_ε fulfills condition (3.4) for each pair $\mathcal{P}\in\mathcal{G}_s(R,\alpha)$ and thus we deduce from Theorem 3.2

$$\begin{aligned}&\mathbb{E}_{\mathcal{P}}\big[|\hat{\alpha}_0-\alpha|^2\big]\\ &\lesssim U_\varepsilon^{-2(s-1)} + U_\varepsilon^{2T\alpha+3}\varepsilon^2 + U_\varepsilon^{4T\bar{a}_\varepsilon+8}\varepsilon^4 \qquad (6.1)\\ &= \varepsilon^{4(s-1)/(2s+2Ta_\varepsilon+1)}\big(1+\varepsilon^{4T(a_\varepsilon-\alpha)/(2s+2Ta_\varepsilon+1)} + \varepsilon^{(4s-8+8T(a_\varepsilon-\bar{a}_\varepsilon))/(2s+2Ta_\varepsilon+1)}\big).\end{aligned}$$

The first factor has the claimed order, since $\varepsilon^{4(s-1)/(2s+2Ta_\varepsilon+1)} \lesssim \varepsilon^{4(s-1)/(2s+2T\alpha+1)}$ follows with easy calculations from $(\alpha - a_\varepsilon)\log\varepsilon \le C$. Hence, the claim follows once we have bound the sum in the bracket of equation (6.1). For the second term, this is implied by

$$\left|\frac{4T(a_\varepsilon-\alpha)}{2s+2Ta_\varepsilon+1}\log\varepsilon\right| \le \frac{4T|(a_\varepsilon-\alpha)\log\varepsilon|}{2s+1} \le \frac{4TC}{2s+1}.$$

To estimate the third term, we obtain from $s\ge 2$ and $\varepsilon<1$

$$\frac{4s-8+8T(a_\varepsilon-\bar{a}_\varepsilon)}{2s+2Ta_\varepsilon+1}\log\varepsilon \le \frac{-8TC|\log\varepsilon|^{-1}}{2s+1}\log\varepsilon \le \frac{8TC}{2s+1}.$$

Step 2: Let $\mathcal{P}\in\mathcal{G}_s(R,\alpha)$. Note that κ_ε satisfies the condition (3.4) on the set $\{|\hat{\alpha}_{\text{pre}}-\alpha| < |\log\varepsilon|^{-1}\}$. Using the independence of $\hat{\alpha}_{\text{pre}}$ and O_j, the almost sure bound $\tilde{\alpha}_0 \le \bar{\alpha}$ and the concentration of $\hat{\alpha}_{\text{pre}}$, we deduce from step 1:

$$\begin{aligned}\mathbb{E}_{\mathcal{P},\hat{\alpha}_{\text{pre}}}\Big[\big|\tilde{\alpha}_0-\alpha\big|^2\Big] &\le \mathbb{E}_{\mathcal{P},\hat{\alpha}_{\text{pre}}}\Big[\mathbb{E}_{\mathcal{P},\hat{\alpha}_{\text{pre}}}\big[|\tilde{\alpha}_0-\alpha|^2|\hat{\alpha}_{\text{pre}}\big]\mathbf{1}_{\{|\hat{\alpha}_{\text{pre}}-\alpha|<|\log\varepsilon|^{-1}\}}\Big]\\ &\quad+4\bar{\alpha}^2\mathbb{P}_{\hat{\alpha}_{\text{pre}}}\big(|\hat{\alpha}_{\text{pre}}-\alpha|\ge|\log\varepsilon|^{-1}\big)\\ &\lesssim \varepsilon^{4(s-1)/(2s+2T\alpha+1)}+4\bar{\alpha}^2 d\varepsilon^2.\end{aligned}$$

Since the second term decreases faster then the first one for $\varepsilon\to 0$, we obtain the claimed rate.

6.3. Proof of Proposition 4.2

Recall that the cut-off value of $\hat{\alpha}_0$ is given by $U=\varepsilon^{-2/(2s+2T\bar{\alpha}+1)}$. For $\kappa>0$, we obtain from the definition of the estimator and the decomposition of the stochastic error into linear part and remainder:

$$\begin{aligned}\mathbb{P}\big(|\hat{\alpha}_0-\alpha|\ge\kappa\big) &= \mathbb{P}\bigg(\bigg|\int_{-U}^{U}\operatorname{Re}(\rho+\tilde{\psi}-\psi)(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\ge\kappa\bigg)\\ &\le \mathbb{P}\bigg(\bigg|\int_{-U}^{U}\rho(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\ge\frac{\kappa}{3}\bigg)+\mathbb{P}\bigg(\bigg|\int_{-U}^{U}\operatorname{Re}\big(\mathcal{L}(u)\big)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\ge\frac{\kappa}{3}\bigg)\\ &\quad+\mathbb{P}\bigg(\bigg|\int_{-U}^{U}\mathcal{R}(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\ge\frac{\kappa}{3}\bigg)=:P_1+P_2+P_3.\end{aligned}$$

We will bound all three probabilities separately. To that end, let c_j, $j\in\mathbb{N}$, be suitable non-negative constants not depending on κ, ε and N.

The event in P_1 is deterministic. Hence, the same estimate on the deterministic error as in Theorem 3.2

$$\bigg|\int_{-U}^{U}\rho(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\le c_1U^{-(s-1)}=c_1\varepsilon^{2(s-1)/(2s+2T\bar{\alpha}+1)}$$

yields $P_1=0$ for all $\varepsilon<\varepsilon^{(1)}:=(\kappa/(3c_1))^{(2s+2T\bar{\alpha}+1)/(2s-2)}$.

To bound P_2 we infer from the definition of $\mathcal{L}$, the linearity of the errors in $\tilde{\mathcal{O}}=\mathcal{O}_l+\sum_{j=1}^N\delta_j\varepsilon_j bj$ and from the estimate of the term $|\mathcal{L}_b|$ in Theorem 3.2

$$\begin{aligned}&\bigg|\int_{-U}^{U}\operatorname{Re}\big(\mathcal{L}(u)\big)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\\ &\quad=\bigg|\int_{-U}^{U}\operatorname{Re}\bigg(\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u)\bigg)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\\ &\quad\le\int_{-U}^{U}\frac{(u^4+u^2)^{1/2}}{T|\varphi_T(u-\mathrm{i})|}\big|\mathcal{F}(\mathcal{O}_l-\mathcal{O})(u)w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\end{aligned}$$

$$
\begin{aligned}
&+\left|\int_{-U}^{U}\operatorname{Re}\left(\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\sum_{j=1}^{N}\delta_j\varepsilon_j\mathcal{F}b_j(u)\right)w_{\alpha_0}^U(u)\,\mathrm{d}u\right|\\
&\le c_2\Delta^2U^{T\bar\alpha+2}+\left|\sum_{j=1}^{N}\delta_j\varepsilon_j\int_{-U}^{U}\operatorname{Re}\left(\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}b_j(u)\right)w_{\alpha_0}^U(u)\,\mathrm{d}u\right|\\
&\le c_2\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)}+\left|\sum_{j=1}^{N}a_j\varepsilon_j\right|,
\end{aligned}
$$

where the coefficients are given by $a_j:=\delta_j\int_{-U}^{U}\operatorname{Re}(\frac{(\mathrm{i}-u)u}{T\varphi_T(u-\mathrm{i})}\mathcal{F}b_j(u))w_{\alpha_0}^U(u)\,\mathrm{d}u$ for $j=1,\dots,N$. To apply (4.4), we deduce from $\|\mathcal{F}b_j\|_\infty\le 2\Delta$, the weight function property (3.2) and the assumption $\Delta\|\delta\|_{l^2}^2\lesssim\|\delta\|_{l^\infty}^2$

$$
\begin{aligned}
\sum_{j=1}^{N}a_j^2&\le\sum_{j=1}^{N}\delta_j^2\left(\int_{-U}^{U}\frac{(u^4+u^2)^{1/2}}{T|\varphi_T(u-\mathrm{i})|}\big|\mathcal{F}b_j(u)\big|\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\right)^2\le c_3\Delta^2U^{2T\bar\alpha+4}\|\delta\|_{l^2}^2\\
&\le c_4\varepsilon^2U^{2T\bar\alpha+4}=c_4\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)}.
\end{aligned}
$$

This implies through the concentration inequality of (ε_j)

$$
\begin{aligned}
P_2&\le\mathbb{P}\left(\left|\sum_{j=1}^{N}a_j\varepsilon_j\right|\ge\frac{\kappa}{6}\right)+\mathbb{P}\left(c_2\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)}\ge\frac{\kappa}{6}\right)\\
&\le C_1\exp\left(-\frac{C_2}{36c_4}\kappa^2\varepsilon^{-2(s-1)/(2s+2T\bar\alpha+1)}\right)
\end{aligned}
$$

for all $\varepsilon<\varepsilon^{(2)}:=(\kappa/(6c_2))^{(2s+2T\bar\alpha+1)/(2s-2)}$.

It remains to estimate probability P_3. The bound of $\mathcal{R}$ in Proposition 6.1 ii) yields

$$
\begin{aligned}
&\left|\int_{-U}^{U}\mathcal{R}(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\right|\\
&\quad\le\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\big|\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u)\big|^2\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\\
&\quad\le 2\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\big|\mathcal{F}(\mathcal{O}_l-\mathcal{O})(u)\big|^2\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\\
&\qquad+2\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\left|\sum_{j=1}^{N}\delta_j\varepsilon_j\mathcal{F}b_j(u)\right|^2\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u.
\end{aligned}
$$

The first addend gets small owing to Proposition 6.1(i):

$$\begin{aligned}\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\big|\mathcal{F}(\mathcal{O}_l-\mathcal{O})(u)\big|^2\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\\ \le \big\|\mathcal{F}(\mathcal{O}_l-\mathcal{O})\big\|_\infty^2\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\\ \le c_5\Delta^4U^{2T\bar\alpha+4}\le c_5\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)}.\end{aligned}$$

For the second one, we obtain

$$\bigg|\sum_{j=1}^N\delta_j\varepsilon_j\mathcal{F}b_j(u)\bigg|^2=\sum_{j=1}^N\delta_j^2\varepsilon_j^2\big|\mathcal{F}b_j(u)\big|^2+2\sum_{j=2}^N\sum_{k=1}^{j-1}\delta_j\delta_k\varepsilon_j\varepsilon_k\operatorname{Re}\big(\mathcal{F}b_j(u)\mathcal{F}b_k(-u)\big).$$

Thus,

$$\bigg|\int_{-U}^{U}\mathcal{R}(u)w_{\alpha_0}^U(u)\,\mathrm{d}u\bigg|\le 2c_5\varepsilon^{(4s-6)/(2s+2T\bar\alpha+1)}+2\sum_{j=1}^N\delta_j^2\varepsilon_j^2\xi_{j,j}(U)+4\sum_{j=2}^N\sum_{k=1}^{j-1}\delta_j\delta_k\varepsilon_j\varepsilon_k\xi_{j,k}(U)$$

with $\xi_{j,k}(U):=\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\operatorname{Re}(\mathcal{F}b_j(u)\mathcal{F}b_k(-u))|w_{\alpha_0}^U(u)|\,\mathrm{d}u$. Denoting the diagonal term and the cross term as

$$D_N:=\sum_{j=1}^N\delta_j^2\varepsilon_j^2\xi_{j,j}(U)\quad\text{and}\quad U_N:=\sum_{j=2}^N\sum_{k=1}^{j-1}\delta_j\delta_k\varepsilon_j\varepsilon_k\xi_{j,k}(U),$$

respectively, we obtain

$$P_3\le\mathbb{P}\bigg(2c_5\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)}\ge\frac{\kappa}{9}\bigg)+\mathbb{P}\bigg(2D_N\ge\frac{\kappa}{9}\bigg)+\mathbb{P}\bigg(4U_N\ge\frac{\kappa}{9}\bigg).$$

The first summand vanishes for $\varepsilon<\varepsilon^{(3)}:=(\kappa/(18c_5))^{(2s+2T\bar\alpha+1)/(2s-2)}$. To estimate the probabilities on D_N and U_N, we establish the bound

$$\big|\xi_{j,k}(U)\big|\le\|\mathcal{F}b_j\|_\infty\|\mathcal{F}b_k\|_\infty\int_{-U}^{U}\frac{u^4+u^2}{T\kappa(u)^2}\big|w_{\alpha_0}^U(u)\big|\,\mathrm{d}u\le c_6\Delta^2U^{2T\bar\alpha+4}\tag{6.2}$$

for $j,k=1,\ldots,N$. Hence,

$$\bigg|\sum_{j=1}^N\delta_j^2\xi_{j,j}(U)\bigg|\le c_6\Delta^2\|\delta\|_{l^2}^2U^{2T\bar\alpha+4}\le c_7\varepsilon^2U^{2T\bar\alpha+4}\le c_7\varepsilon^{2(s-1)/(2s+2T\bar\alpha+1)},$$

which yields together with (4.4)

$$\begin{aligned}\mathbb{P}\Big(D_N \geq \frac{\kappa}{18}\Big) &\leq \mathbb{P}\bigg(\sup_{k=1,\dots,N} |\varepsilon_k|^2 \bigg|\sum_{j=1}^{N} \delta_j^2 \xi_{j,j}(U)\bigg| \geq \frac{\kappa}{18}\bigg) \\ &\leq \mathbb{P}\bigg(\sup_{k=1,\dots,N} |\varepsilon_k|^2 \geq \frac{\kappa}{18c_7}\varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)}\bigg) \\ &\leq C_1 N \exp\bigg(-\frac{C_2}{18c_7}\kappa\varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)}\bigg).\end{aligned}$$

To derive an exponential inequality for the U-statistic U_N, we apply the martingale idea in [13]. Because of the independence and the centering of the (ε_j), the process $(U_N)_{N\geq 1}$ is a martingale with respect to its natural filtration $(\mathcal{F}_N^U)$ (setting $U_1 = 0$):

$$\mathbb{E}\big[U_N - U_{N-1}|\mathcal{F}_{N-1}^U\big] = \mathbb{E}\Bigg[\sum_{k=1}^{N-1} \delta_N \delta_k \varepsilon_N \varepsilon_k \xi_{N,k}(U)\Big|\mathcal{F}_{N-1}^U\Bigg] = 0.$$

We apply the martingale version of the Bernstein inequality, see Theorem VII.3.6 in [22], which yields for arbitrary $t, Q, S > 0$

$$\begin{aligned}\mathbb{P}\big(|U_N| \geq t\big) \leq 2\mathbb{P}\big(\langle U\rangle_N > Q\big) + 2\mathbb{P}\Big(\max_{k=1,\dots,N} |U_k - U_{k-1}| > S\Big) \\ + 2\exp\bigg(-\frac{t^2}{4(Q+tS)}\bigg).\end{aligned} \tag{6.3}$$

Hence, we consider the increment $|U_N - U_{N-1}| = |\varepsilon_N||\sum_{k=1}^{N-1} \delta_N \delta_k \xi_{N,k}(U)\varepsilon_k|$, for $N \geq 2$. Denoting $a_{N,k} := \delta_N \delta_k \xi_{N,k}(U)$, we estimate using (6.2)

$$\begin{aligned}\sum_{k=1}^{N-1} a_{N,k}^2 &= \delta_N^2 \sum_{k=1}^{N-1} \delta_k^2 \xi_{N,k}(U)^2 \leq c_6^2 \Delta^4 U^{4T\bar{\alpha}+8} \delta_N^2 \|\delta\|_{l^2}^2 \\ &\leq c_6^2 \Delta^4 \|\delta\|_{l^2}^4 U^{4T\bar{\alpha}+8} \leq c_7^2 \varepsilon^4 U^{4T\bar{\alpha}+8} \leq c_7^2 \varepsilon^{4(s-1)/(2s+2T\bar{\alpha}+1)}.\end{aligned} \tag{6.4}$$

Thus, by Assumption (4.4) we obtain for all $S > 0$

$$\begin{aligned}&\mathbb{P}\big(|U_N - U_{N-1}| > S\big) \\ &= \mathbb{P}\Bigg(|\varepsilon_N|\bigg|\sum_{k=1}^{N-1} a_{N,k}\varepsilon_k\bigg| > S\Bigg) \\ &\leq \mathbb{P}\big(|\varepsilon_N| > \sqrt{S}\varepsilon^{-(s-1)/(2s+2T\bar{\alpha}+1)}\big) + \mathbb{P}\Bigg(\bigg|\sum_{k=1}^{N-1} a_{N,k}\varepsilon_k\bigg| > \sqrt{S}\varepsilon^{(s-1)/(2s+2T\bar{\alpha}+1)}\Bigg)\end{aligned}$$

$$\leq C_1 \exp\bigl(-C_2 S \varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)}\bigr) \\ + C_1 \exp\biggl(-\frac{C_2}{c_7^2} S \varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)}\biggr).$$

The quadratic variation of U_N is given by

$$\langle U\rangle_N - \langle U\rangle_{N-1} = \mathbb{E}\bigl[(U_N - U_{N-1})^2 | \mathcal{F}_{N-1}^U\bigr] = \delta_N^2 \Biggl(\sum_{k=1}^{N-1} \delta_k \varepsilon_k \xi_{N,k}(U)\Biggr)^2.$$

W.l.o.g. we can assume $\sum_{j=2}^N \delta_j^2 > 0$. Otherwise follows $\sum_{j=2}^N \delta_j^2 = 0$ which implies $\delta_j = 0$ for all $j = 2, \dots, N$ and thus $\langle U\rangle_N = \sum_{j=2}^N (\langle U\rangle_j - \langle U\rangle_{j-1}) = 0$. Then $\mathbb{P}(\langle U\rangle_N > Q) = 0$ would hold for $Q > 0$. Hence, we obtain:

$$\mathbb{P}\bigl(\langle U\rangle_N > Q\bigr) = \mathbb{P}\Biggl(\sum_{j=2}^N \bigl(\langle U\rangle_j - \langle U\rangle_{j-1}\bigr) > Q\Biggr) \leq \sum_{j=2}^N \mathbb{P}\biggl(\langle U\rangle_j - \langle U\rangle_{j-1} > \frac{\delta_j^2}{\sum_{k=2}^N \delta_k^2} Q\biggr) \\ \leq \sum_{j=2}^N \mathbb{P}\Biggl(\|\delta\|_{l^2} \sum_{k=1}^{j-1} \delta_k \varepsilon_k \xi_{j,k}(U) > \sqrt{Q}\Biggr).$$

To apply inequality (4.4), we estimate $\|\delta\|_{l^2}^2 \sum_{k=1}^{j-1} \delta_k^2 \xi_{j,k}(U)^2 \leq c_6^2 \Delta^4 \|\delta\|_{l^2}^4 U^{4T\bar{\alpha}+8} \leq c_7^2 \varepsilon^{4(s-1)/(2s+2T\bar{\alpha}+1)}$ analogous to (6.4) and obtain

$$\mathbb{P}\bigl(\langle U\rangle_N > Q\bigr) \leq C_1 N \exp\biggl(-\frac{C_2}{c_7^2} Q \varepsilon^{-4(s-1)/(2s+2T\bar{\alpha}+1)}\biggr).$$

We deduce from Bernstein's inequality (6.3)

$$\mathbb{P}\biggl(U_N \geq \frac{\kappa}{36}\biggr) \\ \leq 2\mathbb{P}\bigl(\langle U\rangle_N > Q\bigr) + 2\mathbb{P}\Bigl(\max_{k=2,\dots,N} |U_k - U_{k-1}| > S\Bigr) + 2\exp\biggl(-\frac{\kappa^2}{144(36Q + \kappa S)}\biggr) \\ \leq 2C_1 N \exp\biggl(-\frac{C_2}{c_7^2} Q \varepsilon^{-4(s-1)/(2s+2T\bar{\alpha}+1)}\biggr) \\ + 4C_1 N \exp\biggl(-\frac{C_2}{c_7^2 \vee 1} S \varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)}\biggr) + 2\exp\biggl(-\frac{\kappa^2}{144(36Q + \kappa S)}\biggr).$$

By choosing $Q = \kappa S$ and $S = \sqrt{\kappa}\varepsilon^{(s-1)/(2s+2T\bar{\alpha}+1)}$, we get

$$\mathbb{P}\biggl(U_N \geq \frac{\kappa}{36}\biggr) \leq (6C_1 N + 2) \exp\Bigl(-c_8 \min_{q=1,3} \bigl(\kappa^{1/2} \varepsilon^{-(s-1)/(2s+2T\bar{\alpha}+1)}\bigr)^q\Bigr).$$

For all $\varepsilon < \varepsilon^{(3)}$, we have $\kappa\varepsilon^{-2(s-1)/(2s+2T\bar{\alpha}+1)} > \kappa(\varepsilon^{(3)})^{-2(s-1)/(2s+2T\bar{\alpha}+1)} \sim 1$ and hence,

$$P_3 \le \mathbb{P}\Big(D_N \ge \frac{\kappa}{18}\Big) + \mathbb{P}\Big(U_N \ge \frac{\kappa}{36}\Big) \le (7C_1N+2)\exp\big(-c_8\kappa^{1/2}\varepsilon^{-(s-1)/(2s+2T\bar{\alpha}+1)}\big).$$

Putting the bounds of P_1, P_2 and P_3 together yields for a constant $c \in (0,\infty)$ and all $\varepsilon < \varepsilon_0 \wedge 1$ with $\varepsilon_0 := \min\{\varepsilon^{(1)}, \varepsilon^{(2)}, \varepsilon^{(3)}\}$

$$\mathbb{P}\big(|\hat{\alpha}_0 - \alpha| \ge \kappa\big) \le (7C_1N + C_1 + 2)\exp\big(-c\big(\kappa^2 \wedge \kappa^{1/2}\big)\varepsilon^{-(s-1)/(2s+2T\bar{\alpha}+1)}\big).$$

Appendix: Proof of Lemma 2.1

Part (i) The martingale condition yields

$$\big|\varphi_T(u-\mathrm{i})\big| = \exp\Big(T\int_{-\infty}^{\infty}\big(\cos(ux)-1\big)\frac{\mathrm{e}^x k(x)}{|x|}\,\mathrm{d}x\Big).$$

W.l.o.g. we assume $T = 1$, $\alpha > 0$ and $u \ge 1$ because of the symmetry of the cosine.

We split the integral domain into three parts:

$$\big|\varphi_1(u-\mathrm{i})\big| = \exp\Big(\Big(\Big(\int_0^1 + \int_1^u + \int_u^\infty\Big)\frac{\cos x - 1}{x}\Big(\mathrm{e}^{x/u}k\Big(\frac{x}{u}\Big) + \mathrm{e}^{-x/u}k\Big(-\frac{x}{u}\Big)\Big)\,\mathrm{d}x\Big). \quad (A.1)$$

Using $\|k\|_\infty \le \|k\|_{\mathrm{TV}} < \infty$ by assumption and the constant $C_1 := \int_0^1 \frac{1-\cos x}{x}\,\mathrm{d}x \in (0,\infty)$, we estimate

$$\int_0^1 \frac{\cos x - 1}{x}\Big(\mathrm{e}^{x/u}k\Big(\frac{x}{u}\Big) + \mathrm{e}^{-x/u}k\Big(-\frac{x}{u}\Big)\Big)\,\mathrm{d}x \ge 2\mathrm{e}^{1/u}\|k\|_\infty\int_0^1 \frac{\cos x - 1}{x}\,\mathrm{d}x \ge -2C_1 e\|k\|_\infty.$$

In the second part the dependence on u comes into play. Writing $\tilde{k}(x) := k(x) + k(-x)$, the Taylor series of the exponential function together with dominated convergence yield

$$\begin{aligned}
&\int_1^u \frac{\cos x - 1}{x}\Big(\mathrm{e}^{x/u}k\Big(\frac{x}{u}\Big) + \mathrm{e}^{-x/u}k\Big(-\frac{x}{u}\Big)\Big)\,\mathrm{d}x \\
&\quad\ge \int_1^u \frac{\cos x - 1}{x}\tilde{k}\Big(\frac{x}{u}\Big)\,\mathrm{d}x + \|k\|_\infty\sum_{k=1}^\infty\int_1^u(\cos x - 1)\frac{x^{k-1}}{u^k k!}\,\mathrm{d}x \\
&\quad\ge -\alpha\log(u) + \int_{1/u}^1\big(\alpha - \tilde{k}(x)\big)\frac{\mathrm{d}x}{x} + \int_{1/u}^1\frac{\cos(ux)}{x}\tilde{k}(x)\,\mathrm{d}x - 2\|k\|_\infty\sum_{k=1}^\infty\frac{1}{k!k}\big(1-u^{-k}\big) \\
&\quad\ge -\alpha\log(u) - \sup_{x\in(0,1]}\frac{\tilde{k}(x)-\alpha}{x} - 2e\|k\|_\infty + \int_{1/u}^1\frac{\cos(ux)}{x}\tilde{k}(x)\,\mathrm{d}x.
\end{aligned}$$

To bound the last term in the above display, we proceed as Lemma 53.9 in [21]. By the bounded variation of k, we can define a bounded signed measure ρ via $\rho((a,b]) = \tilde{k}(b+) - \tilde{k}(a+), 0 \le$

$a < b$. Noting that $\int_y^\infty \frac{\cos x}{x}\,\mathrm{d}x$ can be bounded uniformly $y \in [1, \infty)$ with a constant $C_2 > 0$, Fubini's theorem yields

$$\begin{aligned}
\int_{1/u}^{1} \frac{\cos(ux)}{x}\tilde{k}(x)\,\mathrm{d}x &= \tilde{k}\left(\frac{1}{u}+\right)\int_1^u \frac{\cos x}{x}\,\mathrm{d}x + \int_1^u \frac{\cos x}{x}\int_{1/u}^{x/u}\rho(\mathrm{d}y)\,\mathrm{d}x \\
&= \tilde{k}\left(\frac{1}{u}+\right)\int_1^u \frac{\cos x}{x}\,\mathrm{d}x + \int_{1/u}^{1}\int_{uy}^{u}\frac{\cos x}{x}\,\mathrm{d}x\,\rho(\mathrm{d}y) \\
&\geq \tilde{k}\left(\frac{1}{u}+\right)\underbrace{\min_{v\geq 1}\int_1^v \frac{\cos x}{x}\,\mathrm{d}x}_{=:-C_3\leq 0} - 2C_2\int_0^1 |\rho|(\mathrm{d}y) \\
&\geq -2C_3\|k\|_\infty - 4C_2\|k\|_{\mathrm{TV}}.
\end{aligned}$$

Obtaining for the third part in (A.1) $\int_1^\infty \frac{\cos(ux)-1}{x}(\mathrm{e}^x k(x) + \mathrm{e}^{-x}k(x))\,\mathrm{d}x \geq -2\|\mathrm{e}^x k(x)\|_{L^1}$, we have with q_k as defined in Lemma 2.1

$$\left|\varphi_1(u-\mathrm{i})\right| \geq \exp\bigl(-q_k - (2e + 4C_2 + 2C_3)\|k\|_{\mathrm{TV}} - 2\left\|\mathrm{e}^x k(x)\right\|_{L^1}\bigr)u^{-\alpha}.$$

We deduce the estimate $|\varphi_T(u-\mathrm{i})| \geq C_\varphi(T, q_k, \|\mathrm{e}^x k(x)\|_{L^1}, \|k\|_{\mathrm{TV}}\})|u|^{-T\alpha}$ for $|u| \geq 1$ with $C_\varphi(T, R) := \exp(-TR(3 + 2e + 4C_2 + 2C_3))$.

Part (ii) follows immediately from the explicit choice of C_φ.

Acknowledgements

The author thanks Markus Reiß and Jakob Söhl for providing many helpful ideas and comments. The research was supported by the Collaborative Research Center 649 "Economic Risk" of the German Research Foundation (Deutsche Forschungsgemeinschaft).

Supplementary Material

Characteristic exponent and lower risk bounds (DOI: 10.3150/12-BEJ478SUPP; .pdf). First, we derive the representation of the characteristic exponent given in Proposition 2.2. Furthermore, we discuss Le Cam's asymptotic equivalence of our nonparametric regression model to the continuous-time white noise model and show lower bounds in the latter one.

References

[1] Belomestny, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann. Statist.* **38** 317–351. MR2589324

[2] Belomestny, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39** 2205–2242. MR2893866

[3] Belomestny, D. and Reiß, M. (2006). Spectral calibration of exponential Lévy models. *Finance Stoch*. **10** 449–474. MR2276314
[4] Belomestny, D. and Schoenmakers, J. (2011). A jump-diffusion Libor model and its robust calibration. *Quant*. *Finance* **11** 529–546. MR2784473
[5] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus*. **75** 305–332.
[6] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2007). Self-decomposability and option pricing. *Math*. *Finance* **17** 31–57. MR2281791
[7] Carr, P. and Madan, D.B. (1999). Option valuation using the fast Fourier transform. *J. Comput*. *Finance* **2** 61–73.
[8] Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559–575. MR2538757
[9] Cont, R. and Tankov, P. (2004). Non-parametric calibration of jump-diffusion option pricing models. *J. Comput*. *Finance* **7** 1–49.
[10] Eberlein, E. (2012). Fourier based valuation methods in mathematical finance. Preprint. Univ. Freiburg.
[11] Eberlein, E., Keller, U. and Prause, K. (1998). New insights into smile, mispricing and value at risk: The hyperbolic model. *J. Bus*. **71** 371–406.
[12] Eberlein, E. and Madan, D.B. (2009). Sato processes and the valuation of structured products. *Quant*. *Finance* **9** 27–42. MR2504500
[13] Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inqualities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications* (E. Giné, C. Houdré and D. Nualart, eds.). *Progress in Probability* **56** 55–69. Basel: Birkhäuser.
[14] Jongbloed, G., van der Meulen, F.H. and van der Vaart, A.W. (2005). Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes. *Bernoulli* **11** 759–791. MR2172840
[15] Kappus, J. and Reiß, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl*. **64** 314–328. MR2683463
[16] Keller-Ressel, M. and Mijatović, A. (2012). On the limit distributions of continuous-state branching processes with immigration. *Stochastic Process*. *Appl*. **122** 2329–2345. MR2922631
[17] Madan, D.B., Carr, P.P. and Chang, E.C. (1998). The variance gamma process and option pricing. *Europ*. *Finance Rev*. **2** 79–105.
[18] Madan, D.B. and Seneta, E. (1990). The Variance Gamma (VG) model for share market returns. *J. Bus*. **63** 511–524.
[19] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Finan*. *Econ*. **3** 125–144.
[20] Sato, K.i. (1991). Self-similar processes with independent increments. *Probab. Theory Related Fields* **89** 285–300. MR1113220
[21] Sato, K.i. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR1739520
[22] Shiryaev, A.N. (1996). *Probability*, 2nd ed. New York: Springer. MR1368405
[23] Söhl, J. (2010). Polar sets for anisotropic Gaussian random fields. *Statist*. *Probab. Lett*. **80** 840–847. MR2608824
[24] Söhl, J. (2012). Confidence sets in nonparametric calibration of exponential Lévy models. Available at arXiv:1202.6611.
[25] Söhl, J. and Trabs, M. (2014). Option calibration of exponential Lévy models: Confidence intervals and empirical results. *J. Comput. Finance*. To appear.
[26] Tankov, P. (2011). Pricing and hedging in exponential Lévy models: Review of recent results. In *Paris-Princeton Lectures on Mathematical Finance* 2010. *Lecture Notes in Math*. **2003** 319–359. Berlin: Springer. MR2762364

[27] Trabs, M. (2014). Supplement to “Calibration of self-decomposable Lévy models”. DOI:10.3150/12-BEJ478.
[28] van de Geer, S. (2000). *Empirical Processes in M-estimation*. Cambridge: Cambridge Univ. Press.

Received November 2011 and revised July 2012