Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Optimal calibration of exponential Lévy models

Denis Belomestny [1] and Markus Reiß[2]

submitted: 21th March 2005

[1] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: belomest@wias-berlin.de

[2] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: mreiss@wias-berlin.de

No. 1017

Berlin 2005

2000 *Mathematics Subject Classification.* 60G51; 62G20; 91B28.

Key words and phrases. European option, jump diffusion, minimax rates, severely ill-posed, nonlinear inverse problem, spectral cut-off.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

Based on options data at the market the problem of calibrating an exponential Lévy model for the underlying asset is investigated. It is shown that this statistical inverse problem is in general severely ill-posed and exact minimax rates of convergence are derived. The estimation procedure we propose is based on the explicit inversion of the option price formula in the spectral domain and a cut-off scheme for high frequencies as regularisation. Its performance is illustrated by numerical simulations.

1 Introduction

Already shortly after the introduction of the Black-Scholes model Merton (1976) argued that based on empirical evidence share price models should incorporate a jump component. Nowadays, standard problems of mathematical finance like derivative pricing have been successfully solved for many general Lévy models, as has become manifest in the monograph by Cont and Tankov (2004a). On the other hand, the investigation of calibration methods for Lévy models has mainly focused on certain parametrisations of the underlying Lévy process. Since the characteristic triplet of a Lévy process is a priori an infinite-dimensional object, this approach is always exposed to the problem of misspecification, in particular when there is no inherent economic foundation of the parameters and they are only used to generate different shapes of possible jump distributions.

The goal of this paper is to investigate mathematically the problem of nonparametric inference for the Lévy triplet when the asset price (S_t) follows an exponential Lévy model

$$S_t = Se^{rt+X_t} \text{ with a Lévy process } X_t \text{ for } t \geqslant 0. \tag{1.1}$$

We suppose that at time $t = 0$ we dispose of prices for vanilla European call and put options on this asset with different strike prices and possibly different maturities. By basing our estimation on option data we draw inference on the underlying risk neutral price process, which in general cannot be determined from historical price data due to the incompleteness of the Lévy market.

The observed option prices will be slightly unprecise due to bid-ask spreads or other frictions in the market. It is well known that in the ideal case of precise observations for all possible strike prices the state price density and hence the Lévy triplet can be uniquely identified, see e.g. Aït-Sahalia and Duarte (2003). Under the realistic model of finitely many noisy observations we cannot hope to determine the triplet correctly, we should rather try to provide an estimator which is as good as possible for the given accuracy of the data. This optimality property is usually assessed by the minimax paradigm, which measures the inherent complexity of the statistical problem class. One main result of the present paper is a lower bound, showing that already in the simple exponential Lévy model the estimation problem is in general severely ill-posed, that is, the estimation error for any part of the Lévy triplet as a function of the accuracy of the observations will only converge with a logarithmic rate for any conceivable estimation procedure.

On the other hand, we propose an explicit construction of an estimator that attains this optimal minimax rate. The procedure is based on the inversion of the explicit pricing formula via Fourier transforms by Carr and Madan (1999) and a regularisation in the spectral domain. Using the Fast Fourier Transformation, the procedure is easy to implement and yields good results in simulations in view of the severe ill-posedness. In comparison with standard statistical ill-posed problems, the main challenges are the nonlinearity involved and the complex interplay between the jump measure as nonparametric part and the drift and diffusion coefficient as parametric parts.

The exponential Lévy model reflects the assumption that the log returns of the asset evolve independently and with identical distribution for the same time steps, which is plausible for liquid markets and not too long time horizons. This basic model class has been considered recently for a variety of pricing and optimisation problems in finance. Let us mention here Mordecki (2002) for pricing American-type perpetual options, Cont and Voltchkova (2005) for pricing other path-dependent options and Eberlein and Papapantoleon (2004) for a good survey and generalisations to the time-inhomogeneous case. Kallsen (2000) and Emmer and Klüppelberg (2004) study market models in a multidimensional framework.

When no model for the price process is specified, calibration from option data can be used to estimate the state price density, see Aït-Sahalia and Duarte (2003). This density yields the distribution of the asset price at the times of maturity, but does not provide any information on the evolution of the price in time. A structural assumption on the price process allows to find prices for path-dependent options or to perform a dynamic risk management. In financial engineering information about the time evolution expected at the market is obtained by smoothing implied Black-Scholes volatilities, cf. Fengler, Härdle, and Mammen (2003). For the generalised Black-Scholes model Dupire's formula permits the calibration from option prices, see e.g. Jackson, Süli, and Howison (1999) for a numerical approach and Crépey (2003) for a theoretical study. The calibration of parametric exponential Lévy models has been studied for example by Eberlein, Keller, and Prause (1998) and Carr, Geman, Madan, and Yor (2002).

The study by Cont and Tankov (2004b), also described in Cont and Tankov (2004a), is closest to our nonparametric approach for exponential Lévy models. In order to cope with the involved ill-posedness, these authors employ a least squares method penalized by the relative entropy with respect to an a priori chosen Lévy triplet. This type of penalisation has certain genuine features: the method takes into account prior information and the resulting functional is convex. However, the value of the diffusion coefficient is thus fixed in advance, and the regularising effect does not take place for independent random errors in the observations, essentially because white noise can only be considered as an element in a Sobolev space of negative regularity, cf. the Hilbert scales approach in Engl, Hanke, and Neubauer (1996). In contrast, we strive for a method that has only few tuning parameters, permits the calibration of the diffusion coefficient and is suited for observations with random errors. The method we present below will have all these properties and is in addition provably rate-optimal over standard smoothness classes. Instead of minimizing some data-dependent criterion, for which in each step the option price for the current triplet value has to be evaluated, we prefer using the explicit nonlinear inversion directly. This results in an efficient straight-forward algorithm. Combining this method with a stage-wise aggregation procedure, a robust data-driven method is obtained.

After introducing the financial and statistical model in Section 2, the estimation method for the finite intensity case is developed in Section 3. The main theoretical results are formulated in Section 4. A typical infinite intensity case is treated in Section 5 and we conclude in Section 6. The proofs of the upper and lower bounds are deferred to Sections 7 and 8, respectively, while the Appendix provides some further technical results.

2 The model

2.1 The exponential Lévy model and option prices

We suppose that the price S_t of an asset at time t follows the Lévy model (1.1), 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. An excellent reference for this model in finance is the monograph by Cont and Tankov

(2004a). In this paper we shall only 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

$$\varphi_T(u) := \mathbb{E}[\exp(iuX_T)] = \exp\Big(T\Big(-\frac{\sigma^2}{2}u^2 + i\gamma u + \int_{-\infty}^{\infty}(e^{iux}-1)\nu(x)\,\mathrm{d}x\Big)\Big). \tag{2.1}$$

$\sigma \geqslant 0$ is called volatility, $\gamma \in \mathbb{R}$ drift and the non-negative function ν, satisfying $\int(|x|\wedge 1)\nu(x)\,\mathrm{d}x < \infty$, is the jump density. Its jump intensity is defined as $\lambda := \|\nu\|_{L^1(\mathbb{R})}$. The characteristic triplet $\mathcal{T} := (\sigma^2, \gamma, \nu)$ has for finite λ the intuitive explanation that X is the sum of three independent classical processes, namely a Wiener process of volatility σ, a deterministic linear process with trend γ and a compound Poisson process of intensity λ with jump distribution ν/λ. Processes with infinite activity are obtained by a limiting procedure and their sample paths have infinitely many jumps, but with jump sizes accumulating at zero. By excluding Lévy processes of unbounded variation we ensure an intuitive explanation of the parameters and we are in line with the empirical parametric findings of Carr, Geman, Madan, and Yor (2002), though some useful parametric models like generalized hyperbolic distributions are excluded.

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

$$C(K,T) = e^{-rT}\,\mathbb{E}_{\mathbb{Q}}[(S_T - K)^+], \tag{2.2}$$

where $(A)^+ := \max(A, 0)$ 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 $\mathbb{Q}$ and we assume from now on that S follows an exponential Lévy model (1.1) under $\mathbb{Q}$ and that the discounted price process $e^{-rt}S_t$ is a martingale on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q}, (\mathcal{F}_t))$, fixed throughout the paper. As is standard in the calibration literature, the measure $\mathbb{Q}$ is assumed to be settled by the market and to be identical for all options traded.

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

$$\forall t \geqslant 0 : \mathbb{E}[e^{X_t}] = 1 \iff \frac{\sigma^2}{2} + \gamma + \int_{-\infty}^{\infty}(e^x - 1)\nu(x)\,\mathrm{d}x = 0. \tag{2.3}$$

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

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

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 $\mathcal{P}(x,T) = S\,\mathbb{E}[(e^x - e^{X_T})^+]$. Then the well-known put-call parity is easily established:

$$\mathcal{C}(x,T) - \mathcal{P}(x,T) = S\,\mathbb{E}[e^{X_T} - e^x] = S(1 - e^x). \tag{2.4}$$

2.2 The observations

We focus on the calibration from options with a fixed maturity $T > 0$ and mention the straightforward extension to several maturities in Section 3.1. We observe the prices of N call options (or by the put-call parity (2.4) alternatively put options) at different strikes K_j, $j = 1, \dots, N$, corrupted by noise

$$Y_j = C(K_j, T) + \sigma_j \varepsilon_j, \quad j = 1, \dots, N. \tag{2.5}$$

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 need to employ Fourier techniques, we introduce the function

$$\mathcal{O}(x) := \begin{cases} S^{-1}\mathcal{C}(x,T), & x \geqslant 0, \\ S^{-1}\mathcal{P}(x,T), & x < 0 \end{cases} \tag{2.6}$$

in the spirit of Carr and Madan (1999). $\mathcal{O}$ records normalised call prices for $x \geqslant 0$ and normalised put prices for $x \leqslant 0$. The following important properties of $\mathcal{O}$ are proved in the Appendix.

Proposition 2.1.

(a) *We have* $\mathcal{O}(x) = S^{-1}\mathcal{C}(x,T) - (1-e^x)^+$ *for all* $x \in \mathbb{R}$.

(b) $\mathcal{O}(x) \in [0, 1 \wedge e^x]$ *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* $\mathcal{O}(x) \leqslant C_\alpha e^{(1-\alpha)x}$ *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* $\mathcal{O}$ *is twice differentiable with*

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

The first derivative $\mathcal{O}'$ *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* $\mathcal{O}'$ *at* γT.

(e) *The Fourier transform of* $\mathcal{O}$ *satisfies*

$$\mathcal{F}\mathcal{O}(v) = \frac{1-\varphi_T(v-i)}{v(v-i)}, \quad v \in \mathbb{R}. \tag{2.7}$$

This identity extends to all complex values v *with* $\mathrm{Im}(v) \in [0,1]$. *Note the properties* $\varphi_T(0) = 1$ *and* $\varphi_T(-i) = 1$ *derived from the general property of characteristic functions and the martingale condition* (2.3), *respectively.*

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

$$O_j := Y_j/S - (1 - K_j e^{-rT}/S)^+ = \mathcal{O}(x_j) + \delta_j \varepsilon_j, \tag{2.8}$$
$$x_j := \log(K_j/S) - rT, \tag{2.9}$$

where $\delta_j = S^{-1}\sigma_j$. In practice, the design (x_j) will be rather dense around $x = 0$ and sparse for options further out of the money or in the money, cf. Fengler, Härdle, and Mammen (2003) for a study on the German DAX index.

In order to facilitate the subsequent analysis we make a mild moment assumption on the price process, which guarantees by Proposition 2.1(b,c) the exponential decay of $\mathcal{O}$.

Assumption 1. *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:* $\mathbb{E}[S_T^2] < \infty$.

3 The estimation for bounded jump densities

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. This has the advantage that no numerical minimization technique needs to be employed and the propagation of errors is more transparent.

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

$$\begin{aligned} \psi(v) &:= \frac{1}{T}\log\Big(1 + iv(1+iv)\mathcal{F}\mathcal{O}(v)\Big) = \frac{1}{T}\log(\varphi_T(v-i)) \\ &= -\frac{\sigma^2 v^2}{2} + i(\sigma^2+\gamma)v + (\sigma^2/2 + \gamma - \lambda) + \mathcal{F}\mu(v), \end{aligned} \tag{3.1}$$

which is a simple consequence of the formulae (2.1) and (2.7). 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 (3.1) shows that the Lévy triplet is uniquely identifiable given the observation of the whole option price function $\mathcal{O}$ without noise: $\mathcal{F}\mu(v)$ 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 $\mathcal{F}\mu(v)$. A properly refined application of this approach will equip us with estimators for the whole triplet $\mathcal{T} = (\sigma^2, \gamma, \mu)$ (we parametrize Lévy triplets equivalently with μ or ν).

3.1 The basic procedure

Let us formulate the basic algorithm to be used when a certain smoothness property is imposed on μ, that is under the prior knowledge $\mu \in \mathcal{G}$, where $\mathcal{G}$ is a smoothness class. The procedure consists of four steps: (a) we build an approximation $\tilde{\mathcal{O}}$ of $\mathcal{O}$ from the data; (b) we obtain an approximation $\tilde{\psi}$ of ψ by formula (3.1); (c) we estimate the coefficients of the quadratic polynomial on the right-hand side in (3.1) from $\tilde{\psi}$ under the presence of a noise component and the nonparametric nuisance part $\mathcal{F}\mu$; (d) we obtain an estimator for $\mathcal{F}\mu$ by considering the remainder.

The model (3.1) has a similar structure as the well-known partial linear models, but in fact there is one substantial difference: the function $\mathcal{F}\mu$ 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 regularize 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 $\mathcal{G}$.

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

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

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

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

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

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 $\mathcal{O}$ at zero: $\beta_0'(0+) - \beta_0'(0-) = -1$. Taking into account the decay properties of $\mathcal{O}$, we interpolate the data by specifying

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

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 $\mathcal{FO}$ 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. Theoretically, we need that the results of Proposition 7.1 and estimate (7.7) are satisfied.

(b) For $\kappa(v) \in (0,1)$, specified later in (4.1), we calculate

$$\tilde{\psi}(v) := \frac{1}{T} \log_{\geqslant \kappa(v)}\Big(1 + iv(1+iv)\mathcal{F}\tilde{\mathcal{O}}(v)\Big), \quad v \in \mathbb{R}, \tag{3.2}$$

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

$$\log_{\geqslant\kappa}(z) := \begin{cases} \log(z), & |z| \geq \kappa \\ \log(\kappa\, z/|z|), & |z| < \kappa \end{cases} \tag{3.3}$$

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 (3.2) does not vanish). 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 (3.1). Upon fixing the spectral cut-off value $U = U(\mathcal{G}, (\delta_j), (x_j))$, we set

$$\hat{\sigma}^2 := \int_{-U}^{U} \mathrm{Re}(\tilde{\psi}(u)) w_\sigma^U(u)\, \mathrm{d}u, \tag{3.4}$$

$$\hat{\gamma} := -\hat{\sigma}^2 + \int_{-U}^{U} \mathrm{Im}(\tilde{\psi}(u)) w_\gamma^U(u)\, \mathrm{d}u, \tag{3.5}$$

$$\hat{\lambda} := \frac{\hat{\sigma}^2}{2} + \hat{\gamma} - \int_{-U}^{U} \mathrm{Re}(\tilde{\psi}(u)) w_\lambda^U(u)\, \mathrm{d}u, \tag{3.6}$$

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

$$\int_{-U}^{U} w_\sigma^U(u)\,\mathrm{d}u = 0, \quad \int_{-U}^{U} u^2 w_\sigma^U(u)\,\mathrm{d}u = -2; \quad \int_{-U}^{U} u w_\gamma^U(u)\,\mathrm{d}u = 1; \tag{3.7}$$

$$\int_{-U}^{U} u^2 w_\lambda^U(u)\,\mathrm{d}u = 0, \quad \int_{-U}^{U} w_\lambda^U(u)\,\mathrm{d}u = 1. \tag{3.8}$$

For standard smoothness classes $\mathcal{G}$ asymptotically optimal choices of the cut-off value U and the weight functions are given in (4.9) and (4.2)-(4.4). 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 $\mathcal{F}\mu$.

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

$$\hat{\mu}(u) := \mathcal{F}^{-1}\left[\left(\tilde{\psi}(\bullet) + \tfrac{\hat{\sigma}^2}{2}(\bullet - i)^2 - i\hat{\gamma}(\bullet - i) + \hat{\lambda}\right)\mathbf{1}_{[-U,U]}(\bullet)\right](u), \quad u \in \mathbb{R}. \tag{3.9}$$

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 $\mathcal{F}b_k$ and the function $\mathcal{F}\beta_0$, which with our choice as linear B-splines can be computed explicitly:

$$\mathcal{F}b_k(u) := u^{-2}\Big(\frac{e^{iux_k}-e^{iux_{k-1}}}{x_k-x_{k-1}}-\frac{e^{iux_{k+1}}-e^{iux_k}}{x_{k+1}-x_k}\Big), \quad \mathcal{F}\beta_0(u) = u^{-2}\Big(1+\frac{e^{iux_{j_0}}x_{j_0-1}-e^{iux_{j_0-1}}x_{j_0}}{x_{j_0}-x_{j_0-1}}\Big) \tag{3.10}$$

with $k = 1, \ldots, N$, 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 defined by $x_{j_0-1} < 0 \leqslant x_{j_0}$.

3.2 A data-driven estimator for the jump density

Let us briefly describe the construction of a data-driven procedure which requires no prior smoothness assumptions on μ to adjust the tuning parameters. The idea is to filter out the parametric part and to obtain a standard 'function in noise'-estimation problem in the spectral domain. Instead of choosing one cut-off value U, we take a geometric grid $U_1 \geqslant U_2 \geqslant \cdots \geqslant U_J$ of cut-off values and aggregate the corresponding estimators adaptively.

For filtering quadratic polynomials we introduce the convolution operator

$$A_g f(x) := f(x) - \frac{1}{2\pi}\int_{-\infty}^{\infty} f(y)\mathcal{F}g(x-y)\,dy \tag{3.11}$$

with a sufficiently regular and nicely decaying function $g : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{-\infty}^{\infty} \mathcal{F}g(y)\,\mathrm{d}y = 1, \quad \int_{-\infty}^{\infty} y^k\mathcal{F}g(y)\,\mathrm{d}y = 0, \quad k = 1,2, \text{ that is } g(0) = 1,\ g'(0) = g''(0) = 0. \tag{3.12}$$

The first two steps of the data-driven procedure are identical to the steps (a) and (b) of the basic procedure. The subsequent steps are as follows:

(c) We apply the operator A_g to $\tilde{\psi}$ and obtain $\tilde{\psi}_g(v) := A_g\tilde{\psi}(v)$, which by (3.1) is a reasonable estimate of $A_g\psi(v) = A_g\mathcal{F}\mu(v) = \mathcal{F}(\mu(1-g))(v)$.

(d) We consider the family of basic estimators $\tilde{\mu}_g^{(j)}$ given by

$$\tilde{\mu}_g^{(j)} := \mathcal{F}^{-1}\Big(\tilde{\psi}_g\mathbf{1}_{[-U_j,U_j]}\Big)(y), \quad j = 1, \ldots, J.$$

(e) We construct the aggregated estimator $\hat{\mu}_g$ as a convex combination of $(\tilde{\mu}_g^{(j)})_{j=1,\ldots,J}$ with data-dependent weights. These weights are obtained by the following algorithm:

(i) Initialize $\hat{\mu}_g^{(1)} = \tilde{\mu}_g^{(1)}$.

(ii) For $j = 2, \ldots, J$ sequentially define

$$\hat{\mu}_g^{(j)} := \alpha_j\tilde{\mu}_g^{(j)} + (1-\alpha_j)\hat{\mu}_g^{(j-1)},$$

where $\alpha_j = K(m^{(j)}/\lambda)$ for some $\lambda > 0$, a compactly supported kernel K and

$$m^{(j)} := \frac{\|\tilde{\mu}_g^{(j)} - \hat{\mu}_g^{(j-1)}\|_{L_2}^2}{\|\operatorname{Var}[\tilde{\mu}_g^{(j)}]\|_{L_1}}.$$

(iii) Put $\hat{\mu}_g := \hat{\mu}_g^{(J)}$.

Although $\operatorname{Var}[\tilde{\mu}_g^{(j)}]$ is not known exactly, it can be easily estimated from above. The parameter λ is taken in accordance with the suggestions given in Belomestny and Spokoiny (2004), where the whole aggregation procedure is explained in detail.

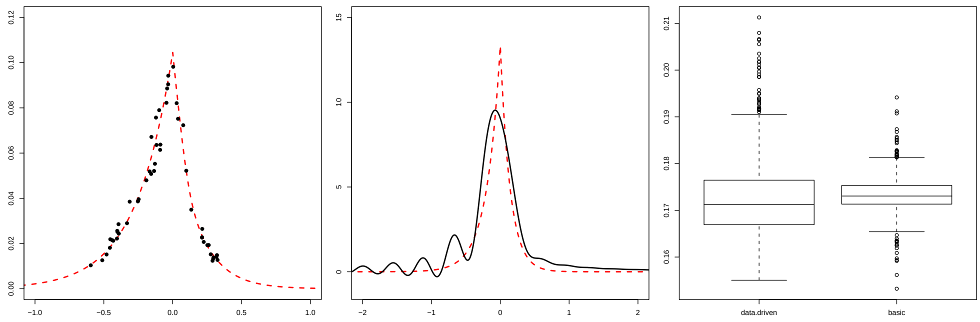

Figure 1: Kou model. Left: Sample (O_j) and true function $\mathcal{O}$ (dashed line). Center: True μ (dashed) and estimated $\hat{\mu}$ (black) modified Lévy densities. Right: Box plot for the data-driven and the basic procedure based on 1000 Monte-Carlo simulations.

(f) The final estimator for $\mu(x)$ is defined as $\hat{\mu}(x) = \hat{\mu}_g(x)/(1-g(x))$ for all $x \in \mathbb{R}$ with $g(x) \neq 1$.

Example 1. *A possible family of functions g satisfying (3.12) is given by $g_\alpha(x) = 1-(1-e^{-x^2/\alpha^2})^2$, $x \in \mathbb{R}$, $\alpha > 0$, which gives rise to the convolution filter $\mathcal{F}g_\alpha(u) = \frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 u^2/4} - \frac{\alpha}{\sqrt{8\pi}} e^{-\alpha^2 u^2/8}$. Observe that $1 - g_\alpha$ only vanishes at zero and that for smaller values of α the weight $1 - g_\alpha$ is closer to one outside the origin, but the filter $\mathcal{F}g_\alpha$ does not decay so rapidly.*

3.3 Numerical Example

Two empirical phenomena in financial data have attracted much 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 (2002). In his model the Lévy triplet is specified by the jump density

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

and the parameters σ, λ, λ_+, $\lambda_- \geqslant 0$ 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 $\delta_j = \mathcal{O}(x_j)/10$.

In Figure 1 (left) the simulated observations (O_j) and the true curve $\mathcal{O}$ 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 1 we compare the performance of the data-driven aggregated estimator 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. In this plot, as provided by the statistical software package R, the box stretches from the 25% percentile to the 75% percentile, crossed by the median, and the position of the remaining 50% of the values is indicated. The fact that the data-driven estimator frequently even outperforms the oracle estimator is to some part due to the hard cut-off at frequency U, which is smoothed out by aggregating the basic estimators. As pointed out by Cavalier and Golubev (2004), standard data-driven estimation procedures often perform badly for inverse problems such that the method of aggregation used here can be considered as comparatively very stable.

4 Risk bounds for bounded jump densities

4.1 Mathematical results

We shall use throughout the notation $A \lesssim B$ if A is bounded by a constant multiple of B, independent 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.

In order to assess the quality of the estimators, we quantify their risks under a Sobolev-type smoothness condition of order s on the transformed jump density μ.

Definition 4.1. *For $s \in \mathbb{N}$ and $R, \sigma_{max} > 0$ let $\mathcal{G}_s(R, \sigma_{max})$ denote the set of all Lévy triplets $\mathcal{T} = (\sigma^2, \gamma, \mu)$, satisfying the martingale condition and Assumption 1 with $C_2 \leqslant R$, such that μ is s-times (weakly) differentiable and*

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

We have enforced $|\tilde{\psi}_T(v)| \geqslant \log(\kappa(v))$ in (3.2) to prevent unboundedness in the case of large stochastic errors. For Lévy triplets in $\mathcal{G}_s(R, \sigma_{max})$ a reasonable choice for $\kappa(v)$ can be obtained from the following calculation using the identity $\frac{\sigma^2}{2} + \gamma + \mathcal{F}\mu(0) = \lambda$ derived from the martingale condition (2.3):

$$\begin{aligned} \tfrac{1}{2}|\varphi_T(v-i)| &= \tfrac{1}{2}\exp\Big(-T\frac{\sigma^2}{2}v^2 - T\mathcal{F}\mu(0) + T\,\mathrm{Re}(\mathcal{F}\mu(v))\Big) \\ &\geqslant \tfrac{1}{2}\exp\Big(-T\frac{\sigma_{max}^2}{2}v^2 - 4TR\Big) =: \kappa(v). \end{aligned} \tag{4.1}$$

The only reason for the factor $1/2$ is the mathematical tractability giving later the bound of Lemma 7.2.

Concerning the choice of the weight functions, we take advantage of the smoothness s of μ by taking functions w such that $\mathcal{F}w$ has s vanishing moments. Equivalently expressed in the spectral domain, the weight functions $w(u)$ grow with frequencies $|u|$ like $|u|^s$ to profit from the decay of $|\mathcal{F}\mu(u)|$. Hence, we define for all $U > 0$ families of weight functions by rescaling those functions satisfying restrictions (3.7) and (3.8) for $U = 1$:

$$w_\sigma^U(u) = U^{-3}w_\sigma^1(U^{-1}u) \text{ with } w_\sigma^1 \text{ satisfying (3.7) and } \|\mathcal{F}(w_\sigma^1(u)/u^s)\|_{L^1} < \infty, \tag{4.2}$$
$$w_\gamma^U(u) = U^{-2}w_\gamma^1(U^{-1}u) \text{ with } w_\gamma^1 \text{ satisfying (3.7) and } \|\mathcal{F}(w_\gamma^1(u)/u^s)\|_{L^1} < \infty, \tag{4.3}$$
$$w_\lambda^U(u) = U^{-1}w_\lambda^1(U^{-1}u) \text{ with } w_\lambda^1 \text{ satisfying (3.8) and } \|\mathcal{F}(w_\lambda^1(u)/u^s)\|_{L^1} < \infty. \tag{4.4}$$

In these definitions it is understood that the support of the weight functions is contained in $[-U, U]$. Note that the property $\mathcal{F}(w(u)/u^s) \in L^1(\mathbb{R})$ means in particular that $w(u)/u^s$ is continuous and bounded such that

$$|w_\sigma^U(u)| \lesssim U^{-(s+3)}|u|^s, \quad |w_\gamma^U(u)| \lesssim U^{-(s+2)}|u|^s \quad \text{and} \quad |w_\lambda^U(u)| \lesssim U^{-(s+1)}|u|^s. \tag{4.5}$$

For the simulations we have used symmetric weight functions that are constant multiples of u^2 except for three (at 0 and $\pm U$ for γ) respectively four (at $\pm U', \pm U$ with some $U' < U$ for σ^2, λ) smoothed out jumps to satisfy the restrictions (3.7), (3.8).

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

$$\Delta := \max_{j=2,\dots,N}(x_j - x_{j-1}) \to 0 \qquad \text{and} \qquad A := \min(x_N, -x_1) \to \infty. \tag{4.6}$$

We use linear B-splines for the basis functions (b_k) and the function β_0. To ease the mathematical treatment of the extrapolation error, we assume that all data is contained in the interval $(-A - \Delta, A + \Delta)$ and add the artificial observations $x_0 = -A - \Delta$, $x_{N+1} = A + \Delta$ with $O_0 = O_{N+1} = 0$.

The reason why we choose a piecewise linear approximation is that this yields rate-optimal interpolation errors for $\tilde{\mathcal{O}}$, knowing that $\mathcal{O}$ is twice differentiable except at finitely many points, cf. Proposition 7.1 below. Of course, when assuming some positive regularity on μ or by some adaptive method, the numerical approximation rate with respect to Δ can be accelerated, but this improvement is only valid for a very small discretisation distance Δ when the stochastic observation error is usually dominant anyway. 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

$$\varepsilon := \Delta^{3/2} + \Delta^{1/2}\|\delta\|_{l^\infty}, \tag{4.7}$$

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

We are now in a position to state the main results about the risk upper bounds of the estimators obtained by the basic procedure and about the risk lower bounds valid for any estimation procedure whatsoever. The proofs are given in Sections 7 and 8 for the upper and lower bounds, respectively.

Theorem 4.2. *Assume $e^{-A} \lesssim \Delta^2$ and $\Delta\|\delta\|_{l_2}^2 \lesssim \|\delta\|_{l^\infty}^2$. Choosing for some $\bar\sigma > \sigma_{max}$ the cut-off $U_{\bar\sigma} := \bar\sigma^{-1}(2\log(\varepsilon^{-1})/T)^{1/2}$, we obtain for the risk of $\hat\sigma^2$ the uniform convergence rate*

$$\sup_{\mathcal{T}=(\sigma^2,\gamma,\mu)\in\mathcal{G}_s(R,\sigma_{max})} \mathbb{E}_{\mathcal{T}}[|\hat\sigma^2 - \sigma^2|^2]^{1/2} \lesssim \bar\sigma^{s+3}(\log(\varepsilon^{-1}))^{-(s+3)/2}. \tag{4.8}$$

The asymptotic risk in the estimation of the other unknown quantities shows a dichotomy. While usually it is larger than the risk for $\hat\sigma^2$, it is much smaller if we know that $\sigma = 0$ holds, that is, for the compound Poisson case.

Theorem 4.3. *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*

$$U_{\bar\sigma} := \bar\sigma^{-1}\big(2\log(\varepsilon^{-1})/T\big)^{1/2}, \quad U_0 := \varepsilon^{-2/(2s+5)}, \tag{4.9}$$

in the cases $\sigma_{max} > 0$ and $\sigma_{max} = 0$, respectively. Then the risk bounds for $\hat\gamma$ and $\hat\lambda$ are

$$\sup_{\mathcal{T}=(\sigma^2,\gamma,\mu)\in\mathcal{G}_s(R,\sigma_{max})} \mathbb{E}_{\mathcal{T}}[|\hat\gamma - \gamma|^2]^{1/2} \lesssim \begin{cases} \bar\sigma^{s+2}(\log(\varepsilon^{-1}))^{-(s+2)/2}, & \sigma \in [0,\sigma_{max}] \text{ unknown},\\ \varepsilon^{(2s+4)/(2s+5)}, & \sigma = \sigma_{max} = 0, \end{cases} \tag{4.10}$$

$$\sup_{\mathcal{T}=(\sigma^2,\gamma,\mu)\in\mathcal{G}_s(R,\sigma_{max})} \mathbb{E}_{\mathcal{T}}[|\hat\lambda - \lambda|^2]^{1/2} \lesssim \begin{cases} \bar\sigma^{s+1}(\log(\varepsilon^{-1}))^{-(s+1)/2}, & \sigma \in [0,\sigma_{max}] \text{ unknown},\\ \varepsilon^{(2s+2)/(2s+5)}, & \sigma = \sigma_{max} = 0. \end{cases} \tag{4.11}$$

Theorem 4.4. *Assume $e^{-A} \lesssim \Delta^2$ and $\Delta\|\delta\|_{l^2}^2 \lesssim \|\delta\|_{l^\infty}^2$. For some $\bar\sigma > \sigma_{max}$ we choose $U_{\bar\sigma}$ and U_0 as in (4.9) to obtain the following risk estimates for $\hat\mu$:*

$$\sup_{\mathcal{T}=(\sigma^2,\gamma,\mu)\in\mathcal{G}_s(R,\sigma_{max})} \mathbb{E}_{\mathcal{T}}\Big[\int_{-\infty}^{\infty} |\hat\mu(x) - \mu(x)|^2\,\mathrm{d}x\Big]^{1/2} \lesssim \begin{cases} \bar\sigma^{s}(\log(\varepsilon^{-1}))^{-s/2}, & \sigma \in [0,\sigma_{max}] \text{ unknown},\\ \varepsilon^{2s/(2s+5)}, & \sigma = \sigma_{max} = 0. \end{cases} \tag{4.12}$$

The two assumptions in each theorem are not very severe: because of the exponential decay of $\mathcal{O}$ 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 appeal to the equivalence between the regression and the Gaussian white noise model, as established by Brown and Low (1996), and consider merely the idealized observation model

$$dZ(x) = \mathcal{O}(x)\,dx + \varepsilon\,dW(x), \quad x \in \mathbb{R}, \tag{4.13}$$

with the noise level asymptotics $\varepsilon \to 0$, a two-sided Brownian motion W and with $\mathcal{O} = \mathcal{O}_{\mathcal{T}}$ denoting the option price function from (2.6) for the given triplet $\mathcal{T}$. This simplification avoids tedious numerical approximations in the proofs.

Theorem 4.5. *Let $s \in \mathbb{N}$, $R > 0$ and $\sigma_{max} \geqslant 0$ be given. For the observation model (4.13) and any quantity $q \in \{\sigma^2, \gamma, \lambda, \mu\}$ the following asymptotic risk lower bounds hold:*

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

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, that is all measurable functions of the observation Z, and the rate $v_{q,\sigma_{max}}$ is given in the following table:

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

4.2 Discussion

We have seen that for $\sigma > 0$ the rate corresponds to a severely ill-posed problem (cf. Engl, Hanke, and Neubauer (1996) and the references there), while for known $\sigma = 0$ the rates are much better, but still ill-posed compared to those obtained in classical nonparametric regression. The reason for the severe ill-posedness for $\sigma > 0$ is that we face an underlying deconvolution problem with a Gaussian distribution: 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. This type of estimation problem has been studied thoroughly by Butucea and Matias (2005) in an idealized density estimation setup. Note the general order in which the (asymptotic) quality of estimation decreases: σ^2, γ, λ and finally μ, which is related to the domination property formulated in Aït-Sahalia and Jacod (2004). In the upper bounds we have kept track of the dependence on σ because for small values of σ and finite samples the performance is not so bad, compare the simulations in Section 3.3; it just needs a lot more observations to improve on that.

At first sight the rates for the parametric estimation part in the case $\sigma = 0$ 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. These rates are, however, easily understood when employing the language of distributions. With δ_0 denoting the Dirac distribution in zero and δ_0' its derivative we have

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

Estimating the density of X_T and similarly its characteristic function from the noisy observations of $\mathcal{O}$ amounts roughly to differentiate the observed function twice, cf. Aït-Sahalia and Duarte (2003) and the remark after equation (7.6) below. 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 $\mathcal{F}^{-1}(\log(\varphi_T))$, 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 $\mathcal{G}_s$ 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. An interesting way to estimate directly γ and λ is suggested by Proposition 2.1(d): a change point detection algorithm for jumps in the derivative of $\mathcal{O}$, as proposed by Goldenshluger, Tsybakov, and Zeevi (2004), can equip us with an estimate of γ and a subsequent estimate of the jump size yields an estimate of λ, which gives the same minimax rates.

As usual, the estimation procedure needs certain tuning parameters. The approximate size of σ_{max} and the noise level is in general known to the practitioner. The stabilisation of the logarithm by the function $\kappa(v)$ was enforced mainly for theoretical reasons to prevent explosions due to large deviations. The usually unknown order s of smoothness of the transformed jump density, however, is needed to determine a good choice of the cut-off frequency U and also appears in the weights w_σ^1, w_γ^1, w_λ^1. Yet, for the latter it suffices to use weight functions satisfying (4.2)-(4.4) for some large s_{max} like in standard nonparametrics where the order of the kernel must only be sufficiently large. We are thus left with only one tuning parameter U, which is the same for all four estimation problems. The data-driven procedure presented in Section 3.2 is one way to cope with this problem for the jump density. Note, however, that a proper mathematical analysis for the general problem seems challenging due to the underlying nonlinear 'change point detection'-structure, for which a data-driven algorithm even in the idealized linear setting of Goldenshluger, Tsybakov, and Zeevi (2004) is not yet available. Finally, 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.

5 Estimation for unbounded jump densities

Let us now discuss the case that ν is a jump density with a singularity at zero. For simplicity we restrict the presentation to the case $\sigma = 0$, which is also in agreement with the empirical parametric findings by Carr, Geman, Madan, and Yor (2002). We then deduce as before for $\mu(x) = e^x \nu(x)$ using (2.3), (2.7) and the definition (3.1) of ψ in terms of $\mathcal{O}$

$$\psi(v) = i\gamma v + \int_{-\infty}^{\infty} (e^{ivx} - 1)\mu(x)\,\mathrm{d}x = i\gamma v + \int_0^v \mathcal{F}(ix\mu(x))(w)\,\mathrm{d}w.$$

Under Assumption 1 $x\mu(x) \in L^1(\mathbb{R})$ holds. By taking derivatives we find

$$\psi'(v) = \frac{(i-2v)\mathcal{F}\mathcal{O}(v) - (v - iv^2)\mathcal{F}(x\mathcal{O}(x))(v)}{T(1 + (iv - v^2)\mathcal{F}\mathcal{O}(v))} = i\gamma + \mathcal{F}(ix\mu(x))(v).$$

We first consider the problem of estimating μ in some weighted L^2-loss with a weight function vanishing in zero. More precisely, we aim at estimating $\mu_g(x) = \mu(x)(1 - g(x))$ in $L^2(\mathbb{R})$-loss for some differentiable nicely decaying function $g : \mathbb{R} \to [0,1]$ with $g(0) = 1$. We obtain $\mu_g \in L^1(\mathbb{R})$ and

$$\begin{aligned}\mathcal{F}\mu_g(v) &= \tfrac{1}{2\pi i}\mathcal{F}(ix\mu(x)) * \mathcal{F}((1-g(x))/x)(-v) \\ &= \gamma g'(0) + \frac{1}{2\pi i}\Big(\frac{(i-2\bullet)\mathcal{F}\mathcal{O} - (\bullet + i\bullet^2)\mathcal{F}(x\mathcal{O}(x))}{T(1 + (i\bullet - \bullet^2)\mathcal{F}\mathcal{O})}\Big) * \Big(\mathcal{F}((1-g(x))/x)\Big)(-v) \qquad (5.1)\end{aligned}$$

The convolution kernel $\mathcal{F}((1-g(x))/x)$ decays rapidly for smooth functions g such that for a good approximation of $\mathcal{F}\mu_g(v)$ it suffices to know the functions $\mathcal{F}\mathcal{O}$ and $\mathcal{F}(x\mathcal{O}(x))$ in a close neighbourhood of $-v$.

Consequently, we can estimate $\mathcal{F}\mu_g(v)$ for $v \in [-U, U]$ by substituting the empirical counterpart $\tilde{\mathcal{O}}$ of $\mathcal{O}$ into formula (5.1) and using some g with $g'(0) = 0$. The noise level for frequencies v in the empirical counterpart of (5.1) will be of order v^2 in the finite intensity case $\lambda = \|\nu\|_{L^1(\mathbb{R})} < \infty$ exactly as in the previous analysis for $\sigma = 0$. For $\lambda = \infty$ the characteristic function tends to zero and the estimation error will deteriorate significantly, see the discussion below.

When drawing inference on the behaviour of μ near zero, we have to specify the kind of singularity and smoothness we expect there. Let us therefore postulate that

$$\mu(x) = \frac{\mu_\alpha(x)}{|x|^\alpha} \text{ with } \alpha \in (0,2) \text{ and } |\mathcal{F}\mu_\alpha(u)| \lesssim (1+|u|)^{-(s+1)}, \quad s \in \mathbb{N}. \tag{5.2}$$

To avoid additional considerations we assume that $\alpha \neq 1$. Note that this model includes for example tempered stable processes with regularity index $s = 1$, when their transformed jump density is given by

$$\mu(x) = e^x \nu(x) = C\Big(\frac{e^{(1+\lambda_-)x}}{|x|^\alpha}\mathbf{1}_{(-\infty,0)}(x) + \frac{e^{(1-\lambda_+)x}}{|x|^\alpha}\mathbf{1}_{(0,\infty)}(x)\Big), \quad C,\ \lambda_- > 0,\ \lambda_+ > 1,\ x \in \mathbb{R},$$

cf. Chapter 4 in Cont and Tankov (2004a) which also gives further examples. Under the model (5.2) an interesting information on the behaviour of μ near zero is given by the value $\mu_\alpha(0)$, for which we now want to derive an estimation procedure. Because of $x\mu(x) = x^{1-\alpha}\mu_\alpha(x)$ (we understand always $x^\rho := |x|^\rho \operatorname{sgn}(x)$) we can draw inference on $\mathcal{F}(x^{1-\alpha}\mu_\alpha(x))$, but this Fourier transform decays slowly due to its non-differentiable argument and will not yield a well performing estimator. Consequently, we have to use more refined fractional differentiation results for the precise structure of this Fourier transform. The following result is derived in the Appendix.

Proposition 5.1. *The following asymptotic estimate holds for $|u| \to \infty$:*

$$\Big|\mathcal{F}(ix^{1-\alpha}\mu_\alpha(x))(u) - 2\Gamma(2-\alpha)\sin(\alpha\pi/2)\sum_{k=0}^{s-1}\binom{\alpha-2}{k}\mu_\alpha^{(k)}(0)i^k u^{\alpha-2-k}\Big| \lesssim |u|^{-s-\min(1,2-\alpha)},$$

where Γ denotes the Euler Gamma function.

Hence, we can expand ψ' in a non-integer power series:

$$\psi'(u) = i\gamma + 2\Gamma(2-\alpha)\sin(\alpha\pi/2)\sum_{k=0}^{s-1}\binom{\alpha-2}{k}\mu_\alpha^{(k)}(0)i^k u^{\alpha-2-k} + \mathcal{R}(u)$$

with the remainder satisfying $|\mathcal{R}(u)| \lesssim |u|^{-s-\max(1,2-\alpha)}$. Exactly as in the expansion (3.1), this permits to estimate $\mu_\alpha(0)$ based on an estimator $\tilde{\psi}$ by $\hat{\mu}_\alpha(0) := \int_{-U}^{U} \tilde{\psi}'(u) w_{\mu_\alpha}^U(u)\, du$ with a weight function $w_{\mu_\alpha}^U$ satisfying

$$\int_{-U}^{U} u^{\alpha-2} w_{\mu_\alpha}^U(u)\, du = \frac{1}{2\Gamma(2-\alpha)\sin(\alpha\pi/2)}; \quad \int_{-U}^{U} u^\rho w_{\mu_\alpha}^U(u)\, du = 0 \text{ for } \rho = 0; \rho = \alpha - 2 - k$$

with $k \in \{1, \dots, s-1\}$. For this estimator a similar analysis as in the case of bounded jump densities can be performed. The main digression is that for $\alpha > 1$, the infinite intensity case, the characteristic function is not bounded away from zero anymore and the risk will be essentially determined by the growth of $|\varphi_T(u-i)|^{-1}$ with $|u| \to \infty$, which is usually exponential ($e^{|u|^{\alpha-1}}$ in the tempered stable case) and thus yields again a severely ill-posed problem.

6 Conclusion

We have developed an estimation procedure for the nonparametric calibration of exponential Lévy models which is mathematically very satisfying because of its minimax properties and which yields a straight-forward algorithm for the implementation. The corresponding lower bound results show that the calibration is in general a hard problem to solve, at least if very high accuracy is desired. Nevertheless the estimation procedure is well suited to gain general insight into the size of the parameters and the structure of the jump density. Even if reasonable parametric models exist that can be better fitted, a goodness-of-fit test based on our nonparametric approach should always be used to check against misspecification.

As already seen in the case of unbounded jump densities, our procedure can be adapted to different models as long as the inverse transformation from the option prices to the characteristic function can be calculated and the unknown quantities can be determined from the structure of the characteristic function. As empirical option data suggests, the risk neutral price process is not homogeneous in time and the exponential Lévy model should be extended in that direction. A suitable model class is for instance given by the affine models of Duffie, Filipovic, and Schachermayer (2003). We believe that the question of calibration for models in financial mathematics should be addressed with the same rigour and intensity as other primary questions like pricing, hedging and risk management.

7 Proof of the upper bounds

All calculations take place in the setting of Section 4. As general reference for Fourier techniques like the Plancherel identity and norm estimates we recommend Rudin (1991). To facilitate the calculations we introduce the exponentially increasing function

$$\mathcal{E}(x) := \frac{e^x - 1}{x}, \quad x > 0, \text{ and set } \mathcal{E}(0) := 1. \tag{7.1}$$

Using linear B-splines (cf. Section 3.1) we encounter the following linear interpolation of $\mathcal{O}$

$$\mathcal{O}_l(x) := \mathbb{E}[\tilde{\mathcal{O}}(x)] = \sum_{j=1}^{N} \mathcal{O}(x_j) b_j(x) + \beta_0(x), \quad x \in \mathbb{R}. \tag{7.2}$$

7.1 A numerical approximation result

Proposition 7.1. *Under the hypothesis $e^{-A} \lesssim \Delta^2$ we obtain uniformly over all Lévy triplets satisfying Assumption 1*

$$\sup_{u\in\mathbb{R}} |\mathbb{E}[\mathcal{F}\tilde{\mathcal{O}}(u) - \mathcal{F}\mathcal{O}(u)]| = \sup_{u\in\mathbb{R}} |\mathcal{F}\mathcal{O}_l(u) - \mathcal{F}\mathcal{O}(u)| \lesssim \Delta^2. \tag{7.3}$$

Proof. By standard Fourier estimates the assertion follows once we have proved $\|\mathcal{O}_l - \mathcal{O}\|_{L^1} \lesssim \Delta^2$.

Note that $\mathcal{O} - \beta_0$ is twice differentiable except at the points $x_{j_0-1}, 0, x_{j_0}$ and possibly γT by Proposition 2.1(d). While the discontinuities of $(\mathcal{O}_l - \beta_0)'$ at the knot points do not do any harm, $\mathcal{O} - \beta_0$ has a derivative near zero which is uniformly bounded by a constant C_0 according to (9.1).

Starting with the case $\sigma > 0$, that is without a jump at γT, we obtain using the mean value theorem with suitable $\xi_j \in (x_{j-1}, x_j)$:

$$\begin{aligned}
&\int_{x_1}^{x_N} |\tilde{\mathcal{O}}_l(x) - \mathcal{O}(x)| \,\mathrm{d}x \\
&\quad = \sum_{j=2}^{N} \int_{x_{j-1}}^{x_j} \left| (\mathcal{O} - \beta_0)(x_j) \frac{x - x_{j-1}}{x_j - x_{j-1}} + (\mathcal{O} - \beta_0)(x_{j-1}) \frac{x_j - x}{x_j - x_{j-1}} + \beta_0(x) - \mathcal{O}(x) \right| \mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{N+1} \int_{x_{j-1}}^{x_j} \Big| \int_{x_{j-1}}^{x} ((\mathcal{O}-\beta_0)'(\xi_j) - (\mathcal{O}-\beta_0)'(y))\,\mathrm{d}y \Big|\,\mathrm{d}x \\
&\leqslant \sum_{j\in\{2,\dots,N\}\setminus\{j_0\}} \int_{x_{j-1}}^{x_j}\int_{x_{j-1}}^{x}\int_{x_{j-1}}^{x_j} |\mathcal{O}''(z)|\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x + 2C_0\Delta^2 \\
&\leqslant \|\mathcal{O}''\|_{L^1}\Delta^2 + 2C_0\Delta^2.
\end{aligned}
$$

By Assumption 1 and Proposition 2.1(b,c) the extrapolation error is bounded by

$$
\int_{[x_0,x_1]\cup[x_N,x_{N+1}]} |\mathbb{E}[\tilde{\mathcal{O}}(x) - \mathcal{O}(x)]|\,\mathrm{d}x \leqslant 4C_2\Delta e^{-(A-\Delta)}.
$$

An application of Proposition 2.1(d) therefore shows for $\sigma > 0$

$$
\int_{-\infty}^{\infty} |\mathbb{E}[\tilde{\mathcal{O}}(x) - \mathcal{O}(x)]|\,\mathrm{d}x \leqslant e^{-A} + 3\Delta^2 + 2C_0\Delta^2 + 4C_2\Delta e^{-(A-\Delta)} \lesssim \Delta^2.
$$

In the case $\sigma = 0$ we consider the index j_* with $x_{j_*-1} \leqslant \gamma T < x_{j_*}$ and face an additional error estimated by

$$
\begin{aligned}
\int_{x_{j_*-1}}^{x_{j_*}} |\mathbb{E}[\tilde{\mathcal{O}}(x) - \mathcal{O}(x)]|\,\mathrm{d}x &\leqslant \int_{x_{j_*-1}}^{x_{j_*}} \|(\mathcal{O}-\beta_0)'\|_{L^\infty} \Big| \frac{2(x - x_{j_*-1})(x_{j_*} - x)}{x_{j_*} - x_{j_*-1}} \Big|\,\mathrm{d}x \\
&\leqslant \|(\mathcal{O}-\beta_0)'\|_{L^\infty}(x_{j_*} - x_{j_*-1})^2
\end{aligned}
$$

With a look at (9.1) we infer that this error term is also of order Δ^2 and thus does not enlarge the convergence rate. □

7.2 Proof of Theorem 4.2

The asserted rate (4.8) follows once the general risk estimate

$$
\mathbb{E}[|\hat{\sigma}^2 - \sigma^2|^2] \lesssim U^{-2(s+3)} + \mathcal{E}(T\sigma^2U^2)U^{-1}\varepsilon^2 + \mathcal{E}(T\sigma_{max}^2U^2)^2U^4\varepsilon^4 \tag{7.4}
$$

has been shown for $U \lesssim \Delta^{-1}$ uniformly over $\mathcal{G}_s(R,\sigma_{max})$, since the explicit choice of U renders the second and third term asymptotically negligible.

Consider in the definition (3.2) of $\tilde{\psi}$ separately the linearisation $\mathcal{L}$, neglecting the stabilisation by κ, and the remainder term $\mathcal{R}$:

$$
\begin{aligned}
\mathcal{L}(u) &:= T^{-1}\varphi_T(u-i)^{-1}(u-i)u\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u), &(7.5)\\
\mathcal{R}(u) &:= \tilde{\psi}(u) - \psi(u) - \mathcal{L}(u). &(7.6)
\end{aligned}
$$

When neglecting the remainder term, we may view $\tilde{\psi}(u)$ as observation of $\psi(u)$ in additive noise, whose intensity grows like $|\varphi_T(u-i)|^{-1}|(u-i)u| \sim u^2e^{T\sigma^2u^2}$ for $|u| \to \infty$. This heteroskedasticity reflects the degree of ill-posedness of the estimation problem.

Lemma 7.2. *For all $u \in \mathbb{R}$ the remainder term satisfies*

$$
|\mathcal{R}(u)| \leqslant T^{-1}\kappa(u)^{-2}(u^4 + u^2)|\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)|^2.
$$

Proof. Let us set $\tilde{\varphi}_T(u-i) := 1 - u(u-i)\mathcal{F}\tilde{\mathcal{O}}(u)$ which equals $e^{T\tilde{\psi}(u)}$ if $|\tilde{\varphi}_T(u-i)| \geqslant \kappa(u)$. Using $|e^{T\tilde{\psi}(u)}| \geqslant \kappa(u)$, $u \in \mathbb{R}$, we obtain by a second-order expansion of the logarithm

$$
|T\tilde{\psi}(u) - \log(\varphi_T(u-i)) - \varphi_T(u-i)^{-1}(e^{T\tilde{\psi}(u)} - \varphi_T(u-i))| \leqslant \tfrac{1}{2}\kappa(u)^{-2}|e^{T\tilde{\psi}(u)} - \varphi_T(u-i)|^2.
$$

This gives the result whenever $|\tilde{\varphi}_T(u-i)| \geqslant \kappa(u)$. For the other values u we use $|\tilde{\varphi}_T(u-i)| < \kappa(u) \leqslant |\varphi_T(u-i)|/2$ to infer

$$\begin{aligned}
|\varphi_T(u-i)^{-1}(e^{T\tilde{\psi}(u)} - \tilde{\varphi}_T(u-i))| &\leqslant \tfrac{1}{2}\kappa(u)^{-1}|e^{T\tilde{\psi}(u)} - \tilde{\varphi}_T(u-i)|\big(|\tilde{\varphi}_T(u-i) - \varphi_T(u-i)|\kappa(u)^{-1}\big)\\
&\leqslant \tfrac{1}{2}\kappa(u)^{-2}|\tilde{\varphi}_T(u-i) - \varphi_T(u-i)|^2\\
&= \tfrac{1}{2}\kappa(u)^{-2}(u^4+u^2)|\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)|^2.
\end{aligned}$$

Together with the previous result this gives for all $u \in \mathbb{R}$ the assertion of the lemma. □

We shall frequently use the following norm bounds for the B-splines (b_k), which follow from $\|b_k\|_\infty = 1$ and $|x_{k+1} - x_{k-1}| \leqslant 2\Delta$:

$$\|\mathcal{F}b_k\|_{L^2} = \sqrt{2\pi}\|b_k\|_{L^2} \leqslant (4\pi\Delta)^{1/2}, \quad \|\mathcal{F}b_k\|_\infty \leqslant \|b_k\|_{L^1} \leqslant 2\Delta. \tag{7.7}$$

We decompose $\hat{\sigma}^2$ in terms of $\mathcal{L}$ and $\mathcal{R}$ from (7.5) and (7.6):

$$\begin{aligned}
\hat{\sigma}^2 &= \int_{-U}^{U}\Big(-\frac{\sigma^2}{2}(u^2-1) + \gamma + \operatorname{Re}(\mathcal{F}\mu(u)) - \lambda + \operatorname{Re}\big(\mathcal{L}(u) + \mathcal{R}(u)\big)\Big) w_\sigma^U(u)\,\mathrm{d}u\\
&= \sigma^2 + \int_{-U}^{U} \operatorname{Re}\big(\mathcal{F}\mu(u) + \mathcal{L}(u) + \mathcal{R}(u)\big) w_\sigma^U(u)\,\mathrm{d}u,
\end{aligned} \tag{7.8}$$

which yields

$$\mathbb{E}[|\hat{\sigma}^2 - \sigma^2|^2] \leqslant 3\Big|\int_{-U}^{U} \mathcal{F}\mu(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2 + 3\,\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{L}(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2\Big] + 3\,\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{R}(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2\Big].$$

Let us consider the three terms in the sum separately. The nuisance of $\mathcal{F}\mu$ causes a deterministic error which can be bounded using $(iu)^s\mathcal{F}\mu(u) = \mathcal{F}\mu^{(s)}(u)$ and the Plancherel isometry:

$$\Big|\int_{-U}^{U} \mathcal{F}\mu(u) w_\sigma^U(u)\,\mathrm{d}u\Big| = 2\pi\Big|\int_{-\infty}^{\infty} \mu^{(s)}(x)\overline{\mathcal{F}^{-1}(w_\sigma^U(u)/(iu)^s)(x)}\,\mathrm{d}x\Big| \leqslant \frac{\|\mu^{(s)}\|_\infty \|\mathcal{F}(w_\sigma^1(u)/u^s)\|_{L^1}}{U^{s+3}}. \tag{7.9}$$

The linear error term can be split into a bias and a variance part ($\operatorname{Var}[Z] := \mathbb{E}[|Z - \mathbb{E}[Z]|^2]$):

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{L}(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2\Big] &= \Big|\int_{-U}^{U} \varphi_T(u-i)^{-1}(u-i)u\,\mathbb{E}[\mathcal{F}(\tilde{\mathcal{O}} - \mathcal{O})(u)] w_\sigma^U(u)\,\mathrm{d}u\Big|^2\\
&\quad + \operatorname{Var}\Big[\int_{-U}^{U} \varphi_T(u-i)^{-1}(u-i)u\mathcal{F}\tilde{\mathcal{O}}(u) w_\sigma^U(u)\,\mathrm{d}u\Big]\\
&=: \mathcal{L}_b^2 + \mathcal{L}_v.
\end{aligned}$$

The bias term is easily bounded by Proposition 7.1, using the uniform bound on $U^{s+3}w_\sigma^U(u)/u^s$:

$$\begin{aligned}
|\mathcal{L}_b| &\leqslant \|\mathcal{F}(\mathcal{O}_l - \mathcal{O})\|_\infty \int_{-U}^{U} |\varphi_T(u-i)|^{-1}(u^4+u^2)^{1/2}|w_\sigma^U(u)|\,\mathrm{d}u\\
&\lesssim \Delta^2 U^{-(s+3)} \int_{-U}^{U} e^{T\frac{\sigma^2}{2}u^2 + 2T\|\mu\|_{L^1}}|u|^{s+2}\,\mathrm{d}u.
\end{aligned}$$

Making use of $\int_0^U 2ue^{cu^2}\,\mathrm{d}u = \frac{e^{cU^2}-1}{c} = \mathcal{E}(cU^2)U^2$ for any $c \geqslant 0$, we estimate the last integral by

$$\int_{-U}^{U} e^{T\frac{\sigma^2}{2}u^2 + 2T\|\mu\|_{L^1}}|u|^{s+2}\,\mathrm{d}u \leqslant e^{2T\|\mu\|_{L^1}}U^{s+3}\mathcal{E}\big(T\tfrac{\sigma^2}{2}U^2\big)$$

and derive from $\|\mu\|_{L^1} = \mathcal{F}\mu(0) \leqslant 2R$ for the bias part in the linear term

$$|\mathcal{L}_b| \lesssim \Delta^2 \mathcal{E}(T\tfrac{\sigma^2}{2}U^2). \tag{7.10}$$

For the variance part of the linear error term we use the support properties $\mathrm{supp}(w_\sigma^U) \in [-U, U]$ and $\mathrm{supp}(b_k) = [x_{k-1}, x_{k+1}]$. Several applications of the Plancherel identity, the Cauchy-Schwarz inequality and estimate (7.7) then yield

$$\begin{aligned}
\mathcal{L}_v &= \int_{-U}^{U}\int_{-U}^{U} \mathrm{Cov}\left(\varphi_T(u-i)^{-1}(u-i)u\mathcal{F}\tilde{\mathcal{O}}(u), \varphi_T(v-i)^{-1}(v-i)v\mathcal{F}\tilde{\mathcal{O}}(v)\right) w_\sigma^U(u) w_\sigma^U(v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \sum_{k=1}^N \delta_k^2 \Big| \int_{-U}^{U} \varphi_T(u-i)^{-1}(u-i)u\mathcal{F}b_k(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2 \\
&= 2\pi \sum_{k=1}^N \delta_k^2 \Big| \int_{-\infty}^{\infty} \mathcal{F}^{-1}\Big(\varphi_T(u-i)^{-1}(u-i)u w_\sigma^U(u)\Big)(x) b_k(-x)\,\mathrm{d}x\Big|^2 \\
&\leqslant 2\pi \sum_{k=1}^N \delta_k^2 \int_{x_{k-1}}^{x_{k+1}} \Big|\mathcal{F}^{-1}\Big(\varphi_T(u-i)^{-1}(u-i)u w_\sigma^U(u)\Big)(-x)\Big|^2\,\mathrm{d}x\,\|b_k\|_{L^2}^2 \\
&\lesssim \Delta\|\delta\|_{l^\infty}^2 \int_{-\infty}^{\infty} \Big|\mathcal{F}^{-1}\Big(\varphi_T(u-i)^{-1}(u-i)u w_\sigma^U(u)\Big)(-x)\Big|^2\,\mathrm{d}x \\
&\sim \Delta\|\delta\|_{l^\infty}^2 \int_{-U}^{U} |\varphi_T(u-i)|^{-2}(u^4+u^2) w_\sigma^U(u)^2\,\mathrm{d}u \\
&\lesssim \Delta U^{-1}\mathcal{E}(T\sigma^2U^2)\|\delta\|_{l^\infty}^2.
\end{aligned}$$

Altogether we obtain for the linear error term

$$\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{L}(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2\Big] \lesssim \mathcal{E}(T\sigma^2U^2)\Big(\Delta^4 + U^{-1}\Delta\|\delta\|_{l^\infty}^2\Big). \tag{7.11}$$

It remains to estimate the quadratic remainder term. We use Lemma 7.2, Proposition 7.1, the independence of (ε_k), the finiteness of their fourth order moments and estimates (4.5), (7.7):

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{R}(u) w_\sigma^U(u)\,\mathrm{d}u\Big|^2\Big] \\
&\lesssim \int_{-U}^{U}\int_{-U}^{U} \mathbb{E}\Big[\Big|\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(u)\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O})(v)\Big|^2\Big] \frac{u^4 w_\sigma^U(u) v^4 w_\sigma^U(v)}{\kappa(u)^2\kappa(v)^2}\,\mathrm{d}u\,\mathrm{d}v \\
&\lesssim \int_{-U}^{U}\int_{-U}^{U} \Big(\|\mathcal{F}(\mathcal{O}_l-\mathcal{O})\|_\infty^4 + \mathbb{E}[|\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O}_l)(u)\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O}_l)(v)|^2]\Big) \frac{u^4 w_\sigma^U(u) v^4 w_\sigma^U(v)}{\kappa(u)^2\kappa(v)^2}\,\mathrm{d}u\,\mathrm{d}v \\
&\lesssim \int_{-U}^{U}\int_{-U}^{U} \Big(\Delta^8 + \mathbb{E}\Big[\Big|\sum_{k,l=1}^N \delta_k\delta_l\varepsilon_k\varepsilon_l \mathcal{F}b_k(u)\mathcal{F}b_l(v)\Big|^2\Big]\Big) \frac{u^4 w_\sigma^U(u) v^4 w_\sigma^U(v)}{\kappa(u)^2\kappa(v)^2}\,\mathrm{d}u\,\mathrm{d}v \\
&\lesssim \int_{-U}^{U}\int_{-U}^{U} \Big(\Delta^8 + \sum_{k,l=1}^N \delta_k^2\delta_l^2 |\mathcal{F}b_k(u)|^2|\mathcal{F}b_l(v)|^2\Big) \frac{u^4 w_\sigma^U(u) v^4 w_\sigma^U(v)}{\kappa(u)^2\kappa(v)^2}\,\mathrm{d}u \\
&= \Big(\Delta^4 \int_{-U}^{U} \frac{u^4 w_\sigma^U(u)}{\kappa(u)^2}\,\mathrm{d}u\Big)^2 + \Big(\int_{-U}^{U} \sum_{k=1}^N \delta_k^2 |\mathcal{F}b_k(u)|^2 \frac{u^4 w_\sigma^U(u)}{\kappa(u)^2}\,\mathrm{d}u\Big)^2 \\
&\lesssim (\Delta^8U^4 + \Delta^4U^4\|\delta\|_{l^2}^2)\mathcal{E}(T\sigma_{max}^2U^2)^2.
\end{aligned}$$

This gives the result that the total risk of $\hat{\sigma}^2$ is of order

$$\mathbb{E}[|\hat{\sigma}^2 - \sigma^2|^2] \lesssim U^{-2(s+3)} + \Big(\Delta^4 + U^{-1}\Delta\|\delta\|_{l^\infty}^2\Big)\mathcal{E}(T\sigma^2U^2) + \Big(\Delta^8U^4 + \Delta^4U^4\|\delta\|_{l^2}^2\Big)\mathcal{E}(T\sigma_{max}^2U^2)^2.$$

Because of $U \lesssim \Delta^{-1}$ and $\Delta\|\delta\|_{l^2}^2 \lesssim \|\delta\|_{l^\infty}^2$ the bound simplifies to (7.4).

7.3 Proof of Theorem 4.3

The rates (4.10) and (4.11) follow from the rate-optimal choice (4.9) of U and the $\mathcal{G}_s(R, \sigma_{max})$-uniform risk estimates

$$\mathbb{E}[|\hat{\gamma} - \gamma|^2] \lesssim U^{-2(s+2)} + \mathcal{E}(T\sigma^2U^2)U\varepsilon^2 + \mathcal{E}(T\sigma_{max}^2U^2)^2U^6\varepsilon^4, \tag{7.12}$$
$$\mathbb{E}[|\hat{\lambda} - \lambda|^2] \lesssim U^{-2(s+1)} + \mathcal{E}(T\sigma^2U^2)U^3\varepsilon^2 + \mathcal{E}(T\sigma_{max}^2U^2)^2U^8\varepsilon^4, \tag{7.13}$$

when inserting $\sigma = 0$ in the case $\sigma_{max} = 0$.

Since the claimed risk bound for $\hat{\gamma}$ is larger than for $\hat{\sigma}^2$, we only need to estimate the risk of $\hat{\gamma} + \frac{\hat{\sigma}^2}{2}$ instead of that for $\hat{\gamma}$. Equally, we can restrict to $\hat{\lambda} - \frac{\hat{\sigma}^2}{2} - \hat{\gamma}$ instead of $\hat{\lambda}$. Then the proof follows exactly the lines of the proof for $\hat{\sigma}^2$, the only difference being the different norming in estimate (4.5) giving rise to a factor U for γ and a factor U^2 for λ. It remains to note that we obtain the bounds in the compound Poisson case by setting $\sigma = \sigma_{max} = 0$ and considering the continuous extension of the bounds for that case: For $\hat{\gamma}$ we obtain as bias

$$\Big|\int_{-U}^{U} \mathcal{F}\mu(u) w_\gamma^U(u)\,\mathrm{d}u\Big| \lesssim U^{-(s+2)}. \tag{7.14}$$

The linear error term is estimated by

$$\mathbb{E}\Big[\Big(\int_{-U}^{U} \mathcal{L}(u) w_\gamma^U(u)\,\mathrm{d}u\Big)^2\Big] \lesssim \begin{cases} \mathcal{E}(T\sigma^2U^2)\Big(U^2\Delta^4 + U\Delta\|\delta\|_{l^\infty}^2\Big), & \sigma \in [0, \sigma_{max}] \text{ unknown}, \\ U^2\Delta^4 + U\Delta\|\delta\|_{l^\infty}^2, & \sigma = \sigma_{max} = 0. \end{cases} \tag{7.15}$$

and the remainder satisfies

$$\mathbb{E}\Big[\Big|\int_{-U}^{U} \mathcal{R}(u) w_\gamma^U(u)\,\mathrm{d}u\Big|^2\Big] \lesssim \begin{cases} \Big(\Delta^8U^6 + \Delta^4U^6\|\delta\|_{l^2}^2\mathcal{E}(T\sigma_{max}^2U^2)^2, & \sigma \in [0, \sigma_{max}] \text{ unknown}, \\ \Delta^8U^6 + \Delta^4U^6\|\delta\|_{l^2}^2, & \sigma = \sigma_{max} = 0. \end{cases} \tag{7.16}$$

Altogether we obtain the risk estimate (7.12).

For $\hat{\lambda}$ we obtain the same asymptotic error bounds as for $\hat{\gamma}$, but multiplied by U when regarding the root mean square error. This gives (7.13) and (4.11).

7.4 Proof of Theorem 4.4

The assertion follows as soon as the following $\mathcal{G}_s(R, \sigma_{max})$-uniform risk bound for general U holds:

$$\mathbb{E}\Big[\int_{-\infty}^{\infty} |\hat{\mu}(x) - \mu(x)|^2\,\mathrm{d}x\Big] \lesssim U^{-2s} + \mathcal{E}(T\sigma^2U^2)U^5\varepsilon^2 + \mathcal{E}(2T\sigma_{max}^2U^2)U^9\varepsilon^4. \tag{7.17}$$

The bias in estimating μ due to the cutoff at U can be estimated by

$$\int_{-\infty}^{\infty} |\mathcal{F}\mu(u)(1 - \mathbf{1}_{[-U,U]})|^2\,\mathrm{d}u \leqslant U^{-2s}\int_{-\infty}^{\infty} |u|^{2s}|\mathcal{F}\mu(u)|^2\,\mathrm{d}u = U^{-2s}\|\mu^{(s)}\|_{L^2}^2. \tag{7.18}$$

The variance term can be split up according to the different risk contributions. For $u \in [-U, U]$ we obtain

$$\begin{aligned}
\mathbb{E}[|\mathcal{F}(\hat{\mu} - \mu)(u)|^2] &\leqslant 4\,\mathbb{E}[|\tilde{\psi}(u) - \psi(u))|^2] + 4(u^2+1)^2\,\mathbb{E}[|\hat{\sigma}^2 - \sigma^2|^2] \\
&\quad + 4(u^2+1)\,\mathbb{E}[|\hat{\gamma} - \gamma|^2] + 4\,\mathbb{E}[|\hat{\lambda} - \lambda|^2] \\
&\lesssim \mathbb{E}[|\mathcal{L}(u)|^2] + \mathbb{E}[|\mathcal{R}(u)|^2] + U^4\,\mathbb{E}[|\hat{\sigma}^2 - \sigma^2|^2] + U^2\,\mathbb{E}[|\hat{\gamma} - \gamma|^2] + \mathbb{E}[|\hat{\lambda} - \lambda|^2] \\
&\lesssim \mathbb{E}[|\mathcal{L}(u)|^2] + \mathbb{E}[|\mathcal{R}(u)|^2] + U^{-2(s+1)} + \mathcal{E}(T\sigma^2U^2)U^3\varepsilon^2 + \mathcal{E}(T\sigma_{max}^2U^2)^2U^8\varepsilon^4.
\end{aligned}$$

In analogy to the previous estimates when proving Theorem 4.2, we find

$$\mathbb{E}[|\mathcal{L}(u)|^2] \leqslant |\varphi_T(u-i)|^{-2}(u^4+u^2)(\|\mathcal{F}(\mathcal{O}-\mathcal{O}_l)\|_\infty^2 + \mathrm{Var}[\mathcal{F}\tilde{\mathcal{O}}(u)]) \lesssim e^{T\sigma^2u^2}u^4\Big(\Delta^4+\Delta^2\|\delta\|_{l^2}^2\Big).$$

With a look at Lemma 7.2 we estimate the remainder by

$$\begin{aligned}\mathbb{E}[|\mathcal{R}(u)|^2] &\leqslant 16\kappa(u)^{-4}(u^4+u^2)^2\,\mathbb{E}[|\mathcal{F}(\mathcal{O}_l-\mathcal{O})(u)|^4 + |\mathcal{F}(\tilde{\mathcal{O}}-\mathcal{O}_l)(u)|^4]\\ &\lesssim e^{2T\sigma_{max}^2u^2}u^8\Big(\Delta^8+\Delta^4\|\delta\|_{l^2}^4\Big).\end{aligned}$$

The Plancherel identity and these estimates yield together (7.17) via

$$\begin{aligned}\int_{-\infty}^{\infty}\mathbb{E}[|\hat{\mu}(x)-\mu(x)|^2]\,dx &\lesssim U^{-2s} + \mathcal{E}(T\sigma^2U^2)U^5\varepsilon^2 + \mathcal{E}(2T\sigma_{max}^2U^2)U^9\varepsilon^4\\ &\quad + \mathcal{E}(T\sigma^2U^2)U^4\varepsilon^2 + \mathcal{E}(T\sigma_{max}^2U^2)^2U^9\varepsilon^4\\ &\sim U^{-2s} + \mathcal{E}(\sigma^2U^2)U^5\varepsilon^2 + \mathcal{E}(2T\sigma_{max}^2U^2)U^9\varepsilon^4.\end{aligned}$$

8 Proof of the lower bounds

We follow the usual Bayes prior technique, see e.g. Korostelev and Tsybakov (1993), and perturb a fixed Lévy triplet $\mathcal{T}_0 = (0,\gamma_0,\nu_0)$ in the interior of $\mathcal{G}_s(R,\sigma_{max})$ such that the perturbations remain in $\mathcal{G}_s(R,\sigma_{max})$.

8.1 Lower bound for μ in the case $\sigma = 0$

Fix a positive integer j. Let $\psi^{(j)} \in C^\infty(\mathbb{R})$ be some function with support in $[0,1]$ satisfying $\|\psi^{(j)}\|_{L^2} = 1$, $\int \psi^{(j)}(x)e^{-2^{-j}x}\,\mathrm{d}x = 0$ and $\int|\mathcal{F}\psi^{(j)}(u)u^{-2}|^2\,\mathrm{d}u < \infty$. Certainly, there are infinitely many functions $\psi^{(j)}$ fulfilling these requirements; the last property follows for instance if ψ is the second derivative of an L^2-function. Introduce the wavelet-like notation

$$\psi_{jk}(x) := 2^{j/2}\psi^{(j)}(2^jx-k), \quad j \geqslant 0,\ k = 0,\dots,2^j-1.$$

Consider for any $r = (r_k) \in \{-1,+1\}^{2^j}$ and some $\beta > 0$ the perturbed Lévy triplets $\mathcal{T}_r = (0,\gamma_0,\mu_r)$ with

$$\mu_r(x) = \mu_0(x) + \beta 2^{-j(s+1/2)}\sum_{k=1}^{2^j} r_k\psi_{jk}(x), \quad x \in \mathbb{R}.$$

We note that due to $\mathcal{F}\psi_{jk}(0) = 0$ and $\int e^{-x}\psi_{jk}(x)\,\mathrm{d}x = 0$ the triplet $\mathcal{T}_r$ satisfies the martingale condition such that $\mathcal{T}_r \in \mathcal{G}_s(R,0)$ holds for a sufficiently small choice of the constant $\beta > 0$.

The Gaussian likelihood ratio of the observations under the probabilities corresponding to $\mathcal{T}_{r'}$ and $\mathcal{T}_r$ under the law of $\mathcal{T}_r$ for some r, r' with $r_k = r'_k$ for all k except one k_0 is given by

$$\Lambda(r',r) = \exp\Big(\int_{-\infty}^{\infty}(\mathcal{O}_{r'}-\mathcal{O}_r)(x)\varepsilon^{-1}\,\mathrm{d}W(x) - \frac{1}{2}\int_{-\infty}^{\infty}|\mathcal{O}_{r'}-\mathcal{O}_r)(x)|^2\varepsilon^{-2}\,\mathrm{d}x\Big).$$

Hence, the Kullback-Leibler divergence (relative entropy) between the two observation models equals

$$KL(\mathcal{T}_{r'}|\mathcal{T}_r) = \frac{1}{2}\int_{-\infty}^{\infty}|(\mathcal{O}_{r'}-\mathcal{O}_r)(x)|^2\varepsilon^{-2}\,\mathrm{d}x.$$

The standard Assouad Lemma (Korostelev and Tsybakov 1993, Thm. 2.6.4) now yields the lower bound for the risk of any estimator $\hat{\mu}$ of μ

$$\inf_{\hat{\mu}}\sup_{\mathcal{T}=(0,\gamma,\mu)\in\mathcal{G}_s(R,0)}\mathbb{E}_{\mathcal{T}}\Big[\int|\hat{\mu}(x)-\mu(x)|^2\,\mathrm{d}x\Big] \gtrsim 2^j\|\mu_r-\mu_{r'}\|_{L^2}^2 \sim 2^{-2js},$$

provided the Kullback-Leibler divergence $KL(\mathcal{T}_{r'}|\mathcal{T}_r)$ stays uniformly bounded by a small constant. It remains to determine a minimal rate for $2^j \to \infty$ such that this holds when the noise level tends to zero.

Arguing in the spectral domain and using the general estimate $|e^z - 1| \leqslant 2|z|$, for $|z| \leqslant \delta$ and some small $\delta > 0$, together with $\|\varphi_{T,r'}/\varphi_{T,r}\|_\infty \to 1$ for $2^j \to \infty$, we obtain for all sufficiently large j

$$\begin{aligned}
KL(\mathcal{T}_{r'}|\mathcal{T}_r) &= \frac{1}{4\pi\varepsilon^2}\int_{-\infty}^{\infty} |\mathcal{F}(\mathcal{O}_{r'} - \mathcal{O}_r)(u)|^2\,\mathrm{d}u \\
&\leqslant \varepsilon^{-2}\int_{-\infty}^{\infty}\left|\frac{\varphi_{T,r}(u-i) - \varphi_{T,r'}(u-i)}{u(u-i)}\right|^2\mathrm{d}u \\
&\leqslant 4\varepsilon^{-2}\int_{-\infty}^{\infty}|\varphi_{T,r}(u-i)|^2T^2|\mathcal{F}(\mu_r - \mu_{r'})(u)|^2(u^4+u^2)^{-1}\,\mathrm{d}u \\
&\lesssim \varepsilon^{-2}2^{-j(2s+1)}\int_{-\infty}^{\infty}|\mathcal{F}\psi_{jk_0}(u)|^2u^{-4}\,\mathrm{d}u \\
&= \varepsilon^{-2}2^{-j(2s+5)}\int_{-\infty}^{\infty}|\mathcal{F}\psi^{(j)}(v)|^2v^{-4}\,\mathrm{d}v.
\end{aligned}$$

Hence, for $2^{j(2s+5)} \sim \varepsilon^2$ with a sufficiently large constant the Kullback-Leibler divergence remains bounded and the asymptotic lower bound for μ follows.

8.2 Lower bound for γ and λ in the case $\sigma = 0$

Let us start with the lower bound for γ. We proceed as before by perturbing a triplet $\mathcal{T}_0 = (0, \gamma_0, \mu_0)$ from the interior of $\mathcal{G}_s(R, 0)$, but this time we only consider one alternative $\mathcal{T}_1 = (0, \gamma_1, \mu_1)$ and choose the perturbation in such a way that the characteristic function $\varphi_T(u-i)$ does not change for small values of $|u|$. For any $\delta > 0$ and $U > 0$ put

$$\gamma_1 := \gamma_0 + \delta, \quad \mathcal{F}\mu_1(u) := \mathcal{F}\mu_0(u) - \delta i(u-i)e^{-u^2/U^2}, \quad u \in \mathbb{R}.$$

Then the function μ_1 is real-valued. Moreover, the martingale condition (2.3) is satisfied:

$$\gamma_1 + \mathcal{F}\mu_1(0) - \mathcal{F}\mu_1(i) = \gamma_0 + \delta + \mathcal{F}\mu_0(0) - \delta - \mathcal{F}\mu_0(i) + 0 = 0.$$

Because of

$$\|\mu_1^{(s)} - \mu_0^{(s)}\|_\infty \leqslant 2\pi\int_{-\infty}^{\infty}|u|^s|\mathcal{F}(\mu_1 - \mu_0)(u)|\,\mathrm{d}u \lesssim \delta\int_{-\infty}^{\infty}|u|^{s+1}e^{-u^2/U^2}\,\mathrm{d}u \sim \delta U^{s+2}$$

and even better bounds for $\|\mu_1^{(k)} - \mu_0^{(k)}\|_{L^2}$, $k = 0, \dots, s$, it suffices to choose $U \sim \delta^{-1/(s+2)}$ small enough to ensure that $\mathcal{T}_1$ still lies in our nonparametric class $\mathcal{G}_s(R, 0)$. The basic lower bound result (Korostelev and Tsybakov 1993, Prop. 2.2.2) then yields

$$\inf_{\hat\gamma}\sup_{(0,\gamma,\mu)\in\mathcal{G}_s(R,0)}\mathbb{E}_{\gamma,\mu}[|\hat\gamma - \gamma|^2] \gtrsim \delta^2,$$

provided the Kullback-Leibler divergence between $\mathcal{T}_1$ and $\mathcal{T}_0$ remains asymptotically bounded. As in the lower bound proof for μ we obtain asymptotically

$$\begin{aligned}
&KL(\mathcal{T}_1|\mathcal{T}_0) \\
&\leqslant 4\varepsilon^{-2}\int_{-\infty}^{\infty}|\varphi_{0,T}(u-i)|^2T^2|i(\gamma_1 - \gamma_0)(u-i) + \mathcal{F}(\mu_1 - \mu_0)(u) - \mathcal{F}(\mu_1 - \mu_0)(i)|^2(u^4+u^2)^{-1}\,\mathrm{d}u \\
&\lesssim \varepsilon^{-2}\delta^2\int_{-\infty}^{\infty}|i(u-i)(1 - e^{-u^2/U^2})|^2(u^4+u^2)^{-1}\,\mathrm{d}u \\
&= \varepsilon^{-2}\delta^2\int_{-\infty}^{\infty}(1 - e^{-v^2})^2U^{-2}v^{-2}U\,\mathrm{d}v \\
&\lesssim \varepsilon^{-2}\delta^2U^{-1} \sim \varepsilon^{-2}\delta^{(2s+5)/(s+2)}.
\end{aligned}$$

The latter remains small for $\delta \sim \varepsilon^{(2s+4)/(2s+5)}$ with a small constant, which gives the asymptotic lower bound for γ.

For the lower bound of λ we perturb the triplet $\mathcal{T}_0$ leaving γ_0 and $\sigma_0 = 0$ fixed and putting

$$\mathcal{F}\mu_1(u) := \mathcal{F}\mu_0(u) + \delta e^{-u(u-i)/U^2}.$$

Then μ_1 is real-valued, $\lambda_1 - \lambda_0 = \mathcal{F}(\mu_1 - \mu_0)(i) = \delta$ and the triplet $\mathcal{T}_1 = (0, \gamma_0, \mu_1)$ satisfies the martingale condition. For $U \sim \delta^{-1/(s+1)}$ with a sufficiently small constant the perturbation μ_1 lies in $\mathcal{G}_s(R, 0)$ due to

$$\|\mu_1^{(s)} - \mu_0^{(s)}\|_\infty \lesssim \delta \int_{-\infty}^{\infty} |u|^s e^{-u^2/U^2}\,du \sim \delta U^{s+1}$$

and even better bounds for $\|\mu_1^{(k)} - \mu_0^{(k)}\|_{L^2}$, $k = 0, \ldots, s$. The Kullback-Leibler divergence is asymptotically bounded by

$$\begin{aligned}
KL(\mathcal{T}_1|\mathcal{T}_0) &\leqslant 4\varepsilon^{-2} \int_{-\infty}^{\infty} |\varphi_{0,T}(u-i)|^{-2} T^2 |\mathcal{F}(\mu_1 - \mu_0)(u) - \mathcal{F}(\mu_1 - \mu_0)(i)|^2 (u^4 + u^2)^{-1}\,du \\
&\lesssim \varepsilon^{-2}\delta^2 \int_{-\infty}^{\infty} |1 - e^{-u(u-i)/U^2}|^2 (u^4 + u^2)^{-1}\,du \\
&= \varepsilon^{-2}\delta^2 \int_{-\infty}^{\infty} |1 - e^{-v^2 + iv/U}|^2 (U^4 v^4 + U^2 v^2)^{-1} U\,dv \\
&\lesssim \varepsilon^{-2}\delta^2 U^{-3} \sim \varepsilon^{-2}\delta^{(2s+5)/(s+1)}
\end{aligned}$$

and we obtain the asymptotic lower bound for λ.

8.3 Lower bound for μ in the case $\sigma > 0$

The interesting deviation from standard proofs of lower bounds (see e.g. Butucea and Matias (2005)) for severely ill-posed problems is that we face the restriction that $\mathcal{F}\mu$ is analytic in a strip parallel to the real line and is uniquely identifiable from its values on any open set. So, let $\mathcal{T}_0 = (\sigma_0^2, \gamma_0, \mu_0)$ with $\sigma_0 > 0$ be a Lévy triplet from the interior of $\mathcal{G}_s(R, \sigma_{max})$. Consider the perturbation $\mathcal{T}_1 = (\sigma_0^2, \gamma_0, \mu_1)$ with

$$\mathcal{F}\mu_1(u) := \mathcal{F}\mu_0(u) + \delta m^{1/4} e^{-(T\sigma_0^2 u^2/m)^m/2} (T\sigma_0^2/m)^m u^m (u-i)^m, \quad u \in \mathbb{R}.$$

for $m \in \mathbb{N}$, $\delta > 0$. Then we have uniformly for $m \to \infty$ and $\delta \to 0$

$$\|\mu_1 - \mu_0\|_{L^2}^2 = 2\pi \|\mathcal{F}(\mu_1 - \mu_0)\|_{L^2}^2 = \frac{2\pi\delta^2}{\sqrt{T\sigma_0^2}} \int_0^\infty e^{-v} v^{(1+2m)/2m} (1 + m^{-1} v^{-1/m})^m\,dv \sim \delta^2.$$

Similarly, for $k = 1, \ldots, s$ we derive uniformly in m and δ

$$\begin{aligned}
\|\mu_1^{(k)} - \mu_0^{(k)}\|_{L^2} &= \sqrt{2\pi} \|u^k \mathcal{F}(\mu_1 - \mu_0)(u)\|_{L^2} \sim \delta m^{k/2}, \\
\|\mu_1^{(s)} - \mu_0^{(s)}\|_\infty &\leqslant \|u^s \mathcal{F}(\mu_1 - \mu_0)(u)\|_{L^1} \leqslant \delta m^{s/2 - 1/4}.
\end{aligned}$$

Therefore choosing $\delta \sim m^{-s/2}$ with a small constant yields $\mathcal{T}_1 \in \mathcal{G}_s(R, \sigma_{max})$ because we then also have that μ_1 is real-valued and $\mathcal{T}_1$ satisfies the martingale condition and Assumption 1.

By the same arguments as before and by Stirling's formula to estimate the Gamma function, the Kullback-Leibler divergence between the observations under $\mathcal{T}_0$ and under $\mathcal{T}_1$ is asymptotically bounded by

$$\begin{aligned}
KL(\mathcal{T}_1|\mathcal{T}_0) &\leqslant 4\varepsilon^{-2}\int_{-\infty}^{\infty}|\varphi_{0,T}(u-i)|^2T^2|\mathcal{F}(\mu_1-\mu_0)(u)|^2(u^4+u^2)^{-1}\,\mathrm{d}u\\
&\lesssim \varepsilon^{-2}\delta^2\int_{-\infty}^{\infty}e^{-T\sigma_0^2u^2}m^{1/2}e^{-(T\sigma_0^2u^2/m)^m}(T\sigma_0^2/m)^{2m}u^{2m-2}|u-i|^{2m-2}\,\mathrm{d}u\\
&=\varepsilon^{-2}\delta^2m^{-7/2}(T\sigma_0^2m)^{-1/2}\int_0^{\infty}e^{-mv^{1/m}}e^{-v}v^{(2m-1)/2m}(1+m^{-1}v^{-1/m})^{m-1}\,\mathrm{d}v\\
&\lesssim \varepsilon^{-2}\delta^2m^{-4}\int_0^{\infty}e^{-mv^{1/m}}\,\mathrm{d}v\\
&=\varepsilon^{-2}\delta^2m^{-4}\int_0^{\infty}e^{-z}z^{m-1}m^{1-m}\,\mathrm{d}z\\
&=\varepsilon^{-2}\delta^2m^{-m-3}\Gamma(m)\lesssim\varepsilon^{-2}\delta^2m^{-m-3}(m-1)^{m-1/2}e^{1-m}\sim\varepsilon^{-2}m^{-3-s}e^{-m}
\end{aligned}$$

Consequently, the Kullback-Leibler divergence remains small when choosing $m\geqslant 2\log(\varepsilon^{-1})$, but $m\lesssim\log(\varepsilon^{-1})$, which gives $\delta\sim\log(\varepsilon^{-1})^{-s/2}$. From the basic general lower bound result we therefore obtain the asymptotic lower bound for μ.

8.4 Lower bound for σ^2, γ and λ in the case $\sigma>0$

Let us start with the lower bound for γ. We proceed as in the case $\sigma=0$ by perturbing the triplet $\mathcal{T}_0=(\sigma_0,\gamma_0,\mu_0)$ with $\sigma_0>0$ in such a way that the characteristic function $\varphi_T(u-i)$ does not change much for small values of $|u|$. For any δ put

$$\gamma_1:=\gamma_0+\delta,\quad \mathcal{F}\mu_1(u):=\mathcal{F}\mu_0(u)-\delta i(u-i)e^{-u^{2m}/U^{2m}}.$$

Then μ_1 is real-valued and the martingale condition (2.3) is satisfied. Because of

$$\|\mu_1^{(s)}-\mu_0^{(s)}\|_\infty\leqslant\int|u|^s|\mathcal{F}(\mu_1-\mu_0)(u)|\,\mathrm{d}u\lesssim\delta\int_{-\infty}^{\infty}|u|^{s+1}e^{-u^{2m}/U^{2m}}\,\mathrm{d}u\sim\delta U^{s+2}$$

and smaller bounds for $\|\mu_1^{(k)}-\mu_0^{(k)}\|_{L^2}$, $k=0,\dots,s$, we choose $U\sim\delta^{-1/(s+2)}$ small enough to ensure that the perturbed triplet $\mathcal{T}_1$ still lies in $\mathcal{G}_s(R,\sigma_{max})$. In the same manner as before and using $|1-e^{-x}|\leqslant|x|$, $x\geqslant0$, as well as Stirling's formula, we obtain

$$\begin{aligned}
KL(\mathcal{T}_1|\mathcal{T}_0) &\leqslant 4\varepsilon^{-2}\int_{-\infty}^{\infty}|\varphi_{0,T}(u-i)|^2T^2|i(\gamma_1-\gamma_0)(u-i)+\\
&\qquad+\mathcal{F}(\mu_1-\mu_0)(u)-\mathcal{F}(\mu_1-\mu_0)(i)|^2(u^4+u^2)^{-1}\,\mathrm{d}u\\
&\lesssim\varepsilon^{-2}\delta^2\int_{-\infty}^{\infty}e^{-T\sigma_0^2u^2}|i(u-i)(1-e^{-u^{2m}/(2U^{2m})})|^2(u^4+u^2)^{-1}\,\mathrm{d}u\\
&\lesssim\varepsilon^{-2}\delta^2\int_{-\infty}^{\infty}e^{-T\sigma_0^2u^2}u^{4m}U^{-4m}u^{-2}\,\mathrm{d}u\\
&\sim\varepsilon^{-2}\delta^2U^{-4m}\Gamma(2m-\tfrac12)\\
&\sim\varepsilon^{-2}\delta^{2+4m/(s+2)}(2m)^{2m}e^{-2m}
\end{aligned}$$

To keep the Kullback-Leibler divergence small, we choose

$$\delta^{(2s+4m+4)/(s+2)}\sim\varepsilon^2(2m)^{-2m}e^{2m}$$

and thus obtain uniformly over m the bound

$$\inf_{\hat\gamma}\sup_{\mathcal{T}=(\sigma^2,\gamma,\mu)\in\mathcal{G}_s(R,\sigma_{max})}\mathbb{E}_{\mathcal{T}}[|\hat\gamma-\gamma|^2]^{1/2}\gtrsim\Big(\varepsilon^2(2m)^{-2m}e^{2m}\Big)^{(s+2)/(2s+4m+4)}.$$

The maximizer of this expression $m^* \sim \log(\varepsilon^{-1})$ then yields the asymptotic lower bound for γ.

For λ we perturb the triplet $\mathcal{T}_0$ leaving σ_0 and γ_0 fixed and putting for an even integer m

$$\mathcal{F}\mu_1(u) := \mathcal{F}\mu_0(u) + \delta e^{-u^m(u-i)^m/U^{2m}}.$$

Then $\lambda_1 - \lambda_0 = \mathcal{F}(\mu_1 - \mu_0)(i) = \delta$ and the triplet $\mathcal{T}_1 = (\sigma_0, \gamma_0, \mu_1)$ satisfies the martingale condition. For $U \sim \delta^{-1/(s+1)}$ with a sufficiently small constant the perturbation μ_1 lies in $\mathcal{G}_s(R, \sigma_{max})$. As before we prove that the Kullback-Leibler divergence remains bounded whenever

$$\delta^{(2s+4m+2)/(s+1)} \sim \varepsilon^2 (2m)^{-2m+1} e^{2m}.$$

Choosing $m^* \sim \log(\varepsilon^{-1})$ as before gives the asymptotic lower bound for λ.

For σ^2 we perturb the triplet $\mathcal{T}_0$ leaving γ_0 invariant and putting

$$\sigma_1^2 := \sigma_0^2 + 2\delta, \quad \mathcal{F}\mu_1(u) := \mathcal{F}\mu_0(u) + \delta(u-i)^2 e^{-u^{2m}/U^{2m}}.$$

Then the martingale condition (2.3) is satisfied and for $U \sim \delta^{-1/(s+3)}$ sufficiently small we remain in $\mathcal{G}_s(R, \sigma_{max})$. It is again routine to prove that the Kullback-Leibler divergence remains bounded whenever

$$\delta^{(2s+4m+6)/(s+3)} \sim \varepsilon^2 (2m)^{-2m-1} e^{2m}.$$

Choosing m^* as before gives the asymptotic lower bound for σ^2.

9 Appendix

9.1 Proof of Proposition 2.1

(a) This follows from the put-call parity (2.4).

(b) $\mathcal{O}(x) \geqslant 0$ follows directly from (2.6) while $\mathcal{O}(x) \leqslant \mathbb{E}[e^{X_T}] - (1-e^x)^+ = 1 \wedge e^x$ follows from (a) and the martingale condition.

(c) We conclude by Hölder's and Markov's inequality for $x \geqslant 0$

$$\mathcal{O}(x) \leqslant \mathbb{E}[e^{X_T} \mathbf{1}_{\{X_T > x\}}] \leqslant C_\alpha^{1/\alpha}\, \mathbb{P}(X_T > x)^{(\alpha-1)/\alpha} \leqslant C_\alpha^{1/\alpha} \Big(\frac{C_\alpha}{e^{\alpha x}}\Big)^{(\alpha-1)/\alpha} = C_\alpha e^{(1-\alpha)x}.$$

(d) Let us denote by f_T the density of the absolutely continuous part of the distribution of X_T. The only atom in the distribution of X_T can occur at γT, namely in the compound Poisson case when no jump until T has taken place. For $x \neq 0$ we have

$$\mathcal{O}'(x) = \Big(\mathbb{E}[(e^{X_T} - e^x)\mathbf{1}_{\{X_T \geqslant x\}}]\Big)' + e^x \mathbf{1}_{\{x<0\}} = e^x\Big(-\mathbb{P}(X_T \geqslant x) + \mathbf{1}_{\{x<0\}}\Big). \tag{9.1}$$

This yields $\mathcal{O}'(0+) - \mathcal{O}'(0-) = -1$ and in the case $\sigma = 0$, $\lambda < \infty$, $\gamma \neq 0$ also

$$\mathcal{O}'(\gamma T+) - \mathcal{O}'(\gamma T-) = e^{\gamma T}\, \mathbb{P}(X_T = \gamma T) = e^{(\gamma-\lambda)T}.$$

At all points $x \neq 0$ where the law of X_T has no atom we obtain

$$\mathcal{O}''(x) = -\Big(e^x\, \mathbb{P}(X_T \geqslant x)\Big)' + e^x \mathbf{1}_{\{x<0\}} = e^x\Big(\mathbb{P}(X_T < x) + f_T(x) - \mathbf{1}_{\{x>0\}}\Big).$$

Consequently, by partial integration and using $\mathbb{E}[e^{X_T}] = 1$ we arrive at

$$\begin{aligned}
\int_{\mathbb{R}\setminus\{0,\gamma T\}} |\mathcal{O}''(x)|\,\mathrm{d}x &= \mathcal{O}'(0-) + \int_0^\infty e^x \Big|\mathbb{P}(X_T < x) - 1 + f_T(x)\Big|\,\mathrm{d}x \\
&\leqslant \mathbb{P}(X_T < 0) + \int_0^\infty e^x(1 - \mathbb{P}(X_T < x))\,\mathrm{d}x + \mathbb{E}\left[\mathbf{1}_{\{X_T \geqslant 0\}} e^{X_T}\right] \\
&= 2\,\mathbb{P}(X_T < 0) - 1 + 2\,\mathbb{E}\left[\mathbf{1}_{\{X_T \geqslant 0\}} e^{X_T}\right] \\
&\leqslant 1 + 2\,\mathbb{E}[e^{X_T}] = 3.
\end{aligned}$$

(e) By definition we have

$$\mathcal{F}\mathcal{O}(v) = S^{-1}\Big(\int_{-\infty}^{0} e^{ivx}\mathcal{P}_T(x)\,\mathrm{d}x + \int_{0}^{\infty} e^{ivx}\mathcal{C}_T(x)\,\mathrm{d}x\Big)$$
$$= \int_{-\infty}^{0} e^{ivx}\,\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant x\}}(e^x - e^{X_T})\right]\mathrm{d}x + \int_{0}^{\infty} e^{ivx}\,\mathbb{E}\left[\mathbf{1}_{\{X_T> x\}}(e^{X_T} - e^{x})\right]\mathrm{d}x.$$

By partial integration we obtain

$$\int_{-\infty}^{0} e^{(iv+1)x}\,\mathbb{P}(X_T\leqslant x)\,\mathrm{d}x = \frac{1}{1+iv}\mathbb{P}(X_T\leqslant 0) - \frac{1}{1+iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{(1+iv)X_T}\right],$$
$$\int_{-\infty}^{0} e^{ivx}\,\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant x\}}e^{X_T}\right]\mathrm{d}x = \frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{X_T}\right] - \frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{(1+iv)X_T}\right]$$

and consequently

$$\int_{-\infty}^{0} e^{ivx}\,\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant x\}}(e^x - e^{X_T})\right]\mathrm{d}x = \frac{1}{1+iv}\mathbb{P}(X_T\leqslant 0) - \frac{1}{1+iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{(1+iv)X_T}\right]$$
$$- \frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{X_T}\right] + \frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T\leqslant 0\}}e^{(1+iv)X_T}\right].$$

In the same way we derive

$$\int_{0}^{\infty} e^{ivx}\,\mathbb{E}\left[\mathbf{1}_{\{X_T> x\}}(e^{X_T} - e^{x})\right]\mathrm{d}x = -\frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T>0\}}e^{X_T}\right] + \frac{1}{iv}\mathbb{E}\left[\mathbf{1}_{\{X_T>0\}}e^{(1+iv)X_T}\right]$$
$$+ \frac{1}{1+iv}\mathbb{P}(X_T>0) - \frac{1}{1+iv}\mathbb{E}\left[\mathbf{1}_{\{X_T>0\}}e^{(1+iv)X_T}\right].$$

Taking into account $\mathbb{E}[e^{X_T}] = 1$, we obtain formula (2.7).

9.2 Proof of Proposition 5.1

We only sketch the main steps in the proof, the reasoning being similar to that for fractional derivatives, cf. Samko, Kilbas, and Marichev (1993). The following formula is easily established and closely related to equation (5.8) in Samko, Kilbas, and Marichev (1993):

$$\mathcal{F}(\mu_\alpha(x)x^{1-\alpha})(u) = \frac{\Gamma(2-\alpha)\sin(\alpha\pi/2)}{\pi i}\int_0^\infty z^{\alpha-2}\Big(\mathcal{F}\mu_\alpha(u+z) - \mathcal{F}\mu_\alpha(u-z)\Big)\,\mathrm{d}z.$$

Let us only consider the case $u>0$ and set

$$G_u(x) := \frac{1}{(s-1)!}\int_0^x (x-\xi)^{s-1}\mathcal{F}\mu_\alpha(u+\xi)\,\mathrm{d}\xi + 2\pi\mathbf{1}_{(-\infty,-u]}(x)\sum_{k=0}^{s-1}\mu_\alpha^{(k)}(0)\frac{(-i)^k(u+x)^{s-k-1}}{(s-k-1)!k!}$$

Then we have in a distributional sense

$$G_u^{(k)}(0) = 0,\quad k=0,\dots,s-1;\quad G_u^{(s)}(x) = \mathcal{F}\mu_\alpha(u+x) + 2\pi\sum_{k=0}^{s-1}\mu_\alpha^{(k)}(0)\frac{(-i)^k\delta_{-u}^{(k)}(x)}{k!}.$$

Hence, by s-fold partial integration we obtain

$$\int_0^\infty z^{\alpha-2}\Big(\mathcal{F}\mu_\alpha(u+z) - \mathcal{F}\mu_\alpha(u-z)\Big)\,\mathrm{d}z - 2\pi\sum_{k=0}^{s-1}\mu_\alpha^{(k)}(0)i^k\binom{\alpha-2}{k}u^{\alpha-2-k}$$
$$= \int_0^\infty z^{\alpha-2}\Big(G_u^{(s)}(z) - G_u^{(s)}(-z)\Big)\,\mathrm{d}z = \Big(\prod_{k=1}^{s}(k+1-\alpha)\Big)\int_0^\infty \frac{G_u(z)-(-1)^sG_u(-z)}{z^{s+2-\alpha}}\,\mathrm{d}z.$$

It therefore suffices to show that the last integral is of order $|u|^{-s-\min(1,2-\alpha)}$, which is accomplished by splitting the integration interval into the parts $[0,1]$, $[1,u]$ and $[u,\infty)$ and making use of $|\mathcal{F}\mu_\alpha(u)| \lesssim (1+|u|)^{-s-1}$ and of the properties of G_u established above. We omit the details.

References

Aït-Sahalia, Y., and J. Duarte (2003): "Nonparametric option pricing under shape restrictions.," *J. Econom.*, 116(1-2), 9–47.

Aït-Sahalia, Y., and J. Jacod (2004): "Fisher's information for discretely sampled Lévy processes," Prépublication 950, Laboratoire de Probabilités et Modèles Aléatoires, Paris.

Belomestny, D., and V. Spokoiny (2004): "Local-likelihood modeling via stage-wise aggregation," Preprint 1000, Weierstraß Institute for Applied Analysis and Stochastics, Berlin.

Brown, L. D., and M. G. Low (1996): "Asymptotic equivalence of nonparametric regression and white noise.," *Ann. Stat.*, 24(6), 2384–2398.

Butucea, C., and C. Matias (2005): "Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model," *Bernoulli*, to appear.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002): "The Fine Structure of Asset Returns: An Empirical Investigation," *J. Business*, 75(2), 305–332.

Carr, P., and D. Madan (1999): "Option valuation using the fast Fourier transform," *J. Comput. Finance*, 2, 61–73.

Cavalier, L., and Y. Golubev (2004): "Risk hull method and regularization by projections of ill-posed inverse problems," Preprint, Université Aix-Marseille 1.

Cont, R., and P. Tankov (2004a): *Financial modelling with jump processes*, Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

——— (2004b): "Nonparametric calibration of jump-diffusion option pricing models," *Journal of Computational Finance*, 7(3), 1–49.

Cont, R., and E. Voltchkova (2005): "Integro-differential equations for option prices in exponential Lévy models," *Finance Stoch.*, to appear.

Crépey, S. (2003): "Calibration of the local volatility in a generalized Black–Scholes model using Tikhonov regularization.," *SIAM J. Math. Anal.*, 34(5), 1183–1206.

Duffie, D., D. Filipovic, and W. Schachermayer (2003): "Affine processes and applications in finance.," *Ann. Appl. Probab.*, 13(3), 984–1053.

Eberlein, E., U. Keller, and K. Prause (1998): "New insights into smile, mispricing, and value at risk: the hyperbolic model," *Journal of Business*, 71(3), 371–405.

Eberlein, E., and A. Papapantoleon (2004): "Symmetries and pricing of exotic options in Lévy models," Preprint, Universität Freiburg.

Emmer, S., and C. Klüppelberg (2004): "Optimal portfolios when stock prices follow an exponential Lévy process," *Finance Stoch.*, 8(1), 17–44.

Engl, H. W., M. Hanke, and A. Neubauer (1996): *Regularization of inverse problems.* Mathematics and its Applications (Dordrecht). 375. Dordrecht: Kluwer Academic Publishers.

Fengler, M., W. Härdle, and E. Mammen (2003): "Implied Volatility String Dynamics," Discussion Paper 54, Humboldt University Berlin, Sonderforschungsbereich 373, `http://sfb.wiwi.hu-berlin.de`.

Goldenshluger, A., A. Tsybakov, and A. Zeevi (2004): "Optimal change-point estimation from indirect observations," Preprint, Columbia University, `http://www-1.gsb.columbia.edu/faculty/azeevi`.

JACKSON, N., E. SÜLI, AND S. HOWISON (1999): "Computation of deterministic volatility surfaces," *Journal Comp. Finance*, 2(2), 5–32.

KALLSEN, J. (2000): "Optimal portfolios for exponential Lévy processes," *Math. Meth. Operations Res.*, 51(3), 357–374.

KOROSTELEV, A., AND A. TSYBAKOV (1993): *Minimax theory of image reconstruction.* Lecture Notes in Statistics (Springer). 82. New York: Springer-Verlag.

KOU, S. (2002): "A jump diffusion model for option pricing," *Management Science*, 48(4), 1086–1101.

MERTON, R. (1976): "Option Pricing When Underlying Stock Returns Are Discontinuous," *J. Financial Economics*, 3(1).

MORDECKI, E. (2002): "Optimal stopping and perpetual options for Lévy processes.," *Finance Stoch.*, 6(4), 473–493.

RUDIN, W. (1991): *Functional analysis. 2nd ed.* International Series in Pure and Applied Mathematics. New York, NY: McGraw-Hill.

SAMKO, S., A. KILBAS, AND O. MARICHEV (1993): *Fractional integrals and derivatives: theory and applications.* New York, NY: Gordon and Breach.