Nonparametric estimation for Lévy processes with a view towards mathematical finance

Enrique Figueroa-Lopez
Department of Mathematics
Purdue University
150 N. University Street,
West Lafayette, IN 47906
jfiguero@math.purdue.edu

Christian Houdré*
Laboratoire d' Analyse et de
Mathématiques Appliquées
CNRS UMR 8050
Université Paris XII
94010 Créteil Cedex, France
and
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160, USA
houdre@math.gatech.edu

10th November 2004

*Research supported in part by NSF and NSA grants.

Key words and phrases: Lévy processes, Poisson processes, Nonparametric estimation, Minimum contrast estimators, Penalized projection estimator, Model selection, Oracle inequalities, Minimax risk on Besov spaces, Adaptive estimation, Variance Gamma process

AMS Subject Classification (2000): 62G05, 62P05, 60G51, 60E07

Abstract

Nonparametric methods for the estimation of the Lévy density of a Lévy process X are developed. Estimators that can be written in terms of the "jumps" of X are introduced, and so are discrete-data based approximations. A model selection approach made up of two steps is investigated. The first step consists in the selection of a good estimator from a linear model of proposed Lévy densities, while the second is a data-driven selection of a linear model among a given collection of linear models. By providing lower bounds for the minimax risk of estimation over Besov Lévy densities, our estimators are shown to achieve the "best" rate of convergence. A numerical study for the case of histogram estimators and for variance Gamma processes, models of key importance in risky asset price modeling driven by Lévy processes, is presented.

1 Introduction

The class of Lévy processes is central to the theory of stochastic processes (see [34] and [9] for excellent monographs on the topic). Recently, new subclasses of Lévy processes have been introduced and actively investigated mostly because of their relevance to mathematical finance. Among the better known models are the *variance Gamma model* of [16], the *CGMY model* of [14], and the *generalized hyperbolic motion* of [3] and [19] (see also [4] and [18]). This phenomenon is not surprising if one brings to mind the traditional model for risky assets, namely the *Black-Scholes model*. In this model the price $S(t)$ of an asset at time t is assumed to be governed by

$$S(t) = S(0)e^{\sigma B(t)+\mu t},$$

where $B(t)$ is a standard Brownian motion. However, a well-documented empirical evidence against the Black-Scholes model, specially in describing high-frequency data and option prices, have led researchers to consider non-Gaussian based models (see for instance [14], [15], [5], [4], [18], and references therein). The transition to Lévy processes is natural since these preserve the statistical qualities of Brownian Motion's increments, but relax the path continuity by allowing jump-alike discontinuities (a specification that is more consistent with the real evolution of stock prices through time). Another point that naturally led to Lévy processes was the use of economic-relevant random clocks (like volume or number of trades) instead of the historical time. Specifically, a more robust and sensible model is to take

$$S(t) = S(0)e^{\sigma B(T(t))+\mu T(t)}, \tag{1.1}$$

where $T(t)$ is a general increasing stochastic process with $T(0) = 0$. In that case, if $T(t)$ has independent and stationary increments, the log return process is necessarily a Lévy process (see 30.1 in [34]). Such considerations led to the study of *exponential Lévy processes* of the form

$$S(t) = S(0)e^{X(t)}, \tag{1.2}$$

where $X(t)$ is a Lévy process. This paradigm has proved to be successful to account for many of the empirical features of financial data. However, among other drawbacks, the high computational intensity and numerical issues involved in calibrating such models have prevented them from being more widely used in practice. In particular, these difficulties become very serious when dealing with "high-frequency" data.

Lévy processes are determined by three "parameters": a non-negative real σ^2, a real μ, and a measure ν on $\mathbb{R}\backslash\{0\}$. These three parameters characterize the dynamic of a Lévy process $\{X(t)\}_{t\geq 0}$ as the superposition of a Brownian motion with drift, $\sigma B(t) + \mu t$, and a pure-jump Lévy process whose jump behavior is specified by the measure ν as follows:

$$\nu(A) = \frac{1}{t}\mathbb{E}\left[\sum_{s\leq t} \chi_A\left(\Delta X(s)\right)\right],$$

where $\Delta X(t) \equiv X(t) - X(t^-)$ is the jump of X at time t and A is such that the indicator $\chi_A(\cdot)$ vanishes in a neighborhood of the origin (this is a consequence of the so called Lévy-Itô decomposition for the sample paths of processes with independent increments; see Theorem 13.4. of [24] or Section 19 of [34]). We assume throughout that ν is determined by a function $p : \mathbb{R}\backslash\{0\} \to [0, \infty)$, called the *Lévy density*, in the following sense:

$$\nu(A) = \int_A p(x)dx, \quad \forall A \in \mathcal{B}(\mathbb{R}\backslash\{0\}).$$

In that case, the value of p at x_0 provides, roughly speaking, information on the frequency of jumps with sizes "close" to x_0.

Estimating the Lévy density poses a nontrivial problem, even when p takes simple parametric forms. Parsimonious Lévy densities usually produces not only intractable but sometimes not even expressible densities for the marginals $X(t)$, $t > 0$. The current practice of estimation relies on approximations of the density function using *inversion formulas* combined with likelihood methods (see for instance [14]). Such approximations make the estimation particularly susceptible to numerical errors and mis-specification; that is, slight changes in the model can produce quite different results. It is important to notice that these problems become quite critical for "high-frequency" data. Other common calibration methods include simulation based methods and multinomial log likelihoods (see for instance [23] and [10]).

In the present paper, we introduce new estimation methods for the Lévy density. We concentrate on model-free estimation schemes that allow to efficiently retrieve a fairly general Lévy density. Being nonparametric, we relax the dependency on the model and expect that data itself validates the best model. Three theories serve as foundations for our methodology: i) the characterization of the jumps associated with a Lévy process as a spatial Poisson process, ii) some recent methods for the nonparametric estimation of spatial Poisson processes introduced in [30], and iii) the short-term properties of Lévy processes to approximate jump-dependent quantities. To the best of our knowledge, such connection between the Lévy density and the statistical properties of the process in small time spans has not been used for calibration purposes before the present work. It is relevant to point out that our procedures are suitable for high-frequency data, which is widely available nowadays. Furthermore, it is precisely for such data that standard statistical estimation methods are not viable, the traditional *geometric Brownian motion* model is totally inaccurate, and general exponential Lévy models may be more relevant.

Let us describe the outline of the paper. In Section 2, we construct functional estimators which can be written in terms of integrals of deterministic functions with respect to the random measure associated with the jumps of X. The proposed method follows the reasoning of the works on minimum contrast estimation on sieves and model selection developed in the context of density estimation and nonlinear regression in [11] (see [8] and [12]) and recently extended to the estimation of intensity functions for Poisson processes in [30]. Concretely, the procedure addresses two problems: 1) the selection of a good estimator, called the *projection estimator*, from a linear model $\mathcal{S}$

of possible estimators, and 2) the selection of a linear model among a given collection of linear models using a penalization technique that led to a *penalized projection estimator* (p.p.e.). A bound for the *risk* of the p.p.e. is found in Section 3. As a consequence, *Oracle inequalities*, that ensure to approximately reach the best expected error (using projection estimators) up to a constant, are obtained. We also assess the rate of convergence of the p.p.e. on *regular splines*, when the Lévy density belongs to some *Besov spaces*. By analyzing the *minimax risk* of estimation on these Besov spaces, it is actually proved in Section 4 that the p.p.e. attains the best possible rate in the minimax sense, when the estimation is based on jumps bounded away from the origin. In Sections 5 and 6, we examine the problem that the Poisson jump measure cannot be retrieved from discrete observation, and devise an approximation procedure for Poisson integrals based on equally space sampling observations of the process. Finally, in the last part our methods are applied to the estimation of a classical model used in mathematical finance: the *Variance Gamma model* of [16]. The Lévy processes are simulated using time series representations and "discrete skeletons", whereas the considered estimators are mainly regular histograms.

2 A model-free estimation method

Consider a real Lévy process $X = \{X(t)\}_{t\geq 0}$ with Lévy density p. That is, X is a càdlàg process with independent and stationary increments such that the characteristic function of its marginals is given by

$$\mathbb{E}\left[e^{iuX(t)}\right] = \exp\left\{t\left(iub - \frac{u^2\sigma^2}{2} + \int_{\mathbb{R}_0}\left\{e^{iux} - 1 - iux1_{[|x|\leq 1]}\right\}p(x)dx\right)\right\}, \tag{2.1}$$

where $\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ and $p : \mathbb{R}_0 \to \mathbb{R}_+$ satisfies

$$\int_{\mathbb{R}_0}(1 \wedge x^2)p(x)dx < \infty. \tag{2.2}$$

Being a càdlàg process, the set of jump times $\{t > 0 : X(t) - X(t^-) > 0\}$ is countable and, for Borel subsets B of $[0,\infty) \times \mathbb{R}_0$,

$$\mathcal{J}(B) \equiv \#\left\{t > 0 : (t, X(t) - X(t^-)) \in B\right\}, \tag{2.3}$$

is a well-defined random measure on $[0,\infty)\times\mathbb{R}_0$, with # denoting cardinality. The Lévy-Itô decomposition of the sample paths (see Theorem 19.2 of [34]) implies that $\mathcal{J}$ is a Poisson process on the Borel sets of $\mathcal{B}([0,\infty)\times\mathbb{R}_0)$ with mean measure given by

$$\mu(B) = \iint_B p(x)\, dt\, dx. \tag{2.4}$$

We study the problem of estimating the Lévy density p on a Borel set $D \in \mathcal{B}(\mathbb{R}_0)$ using a *projection estimation approach*. According to this paradigm, p is estimated by estimating the best approximating function in a finite-dimensional linear space $\mathcal{S}$. The linear space $\mathcal{S}$ is taken so that it has good approximation qualities in general classes of functions. Typical choices are piecewise polynomials or wavelets. In order for this approach to be general enough but still feassible, it is usually assumed that the function to be estimated is bounded and belongs to an $\mathbb{L}^2$ space on D, simplying the task of specifying the best approximating function. The simplest case is when p is taken bounded and $\int_D p^2(x)dx < \infty$. This condition is quite general if D is away from the origin, since (2.2) entails

$$\int_{|x|>\varepsilon} p^2(x)dx < \infty, \tag{2.5}$$

for any $\varepsilon > 0$, when p is bounded on $\{x : |x| > \varepsilon\}$. However, around the origin the Lévy density is not bounded in most applications. This motivates the use of measures different from the Lebesgue measure. Concretely, it is assumed that the Lévy measure $\nu(dx) \equiv p(x)dx$ is absolutely continuous with respect to a known measure η on $\mathcal{B}(D)$ and that the Radon-Nikodym derivative

$$\frac{d\nu}{d\eta}(x) = s(x), \quad x \in D, \tag{2.6}$$

is positive, bounded, and satisfies

$$\int_D s^2(x)\eta(dx) < \infty. \tag{2.7}$$

Definition 2.1 *If (2.6) and (2.7) are verified, we say that η is a* regularizing measure *for the Lévy density p. In that case, s is referred to as the* regularized (under η) Lévy density *of p on D.*

Notice that under the previous regularization assumption, the measure $\mathcal{J}$ of (2.3), when restricted to $\mathcal{B}([0,\infty) \times D)$, is a Poisson process with mean measure

$$\mu(B) = \iint\limits_{B} s(x)\, dt\, \eta(dx), \quad B \in \mathcal{B}([0,\infty) \times D). \tag{2.8}$$

Our goal will be to estimate the regularized Lévy density s, and using (2.6) to retrieve p on D from s. To illustrate this strategy consider a continuous Lévy density p such that

$$p(x) = O\left(x^{-1}\right), \quad \text{as} \;\; x \to 0.$$

This type of densities admit the regularizing measure $\eta(dx) = x^{-2}dx$ on domains of the form $D = \{x : 0 < |x| < b\}$. Indeed, $s(x) = x^2 p(x)$ will be bounded and fulfills (2.7). Clearly, each estimator $\hat{s}$ for s will induce the natural estimator $x^{-2}\hat{s}(x)$ for p.

The previous methodology is motivated by recent results on the estimation of intensity functions of non-homogeneous Poisson processes (see [30]). In that paper, a type of *projection estimator* is proposed, whereas *penalized projection estimation* is used as a data-driven criterion for selecting the best space among a family of linear spaces. However, these procedures focus on finite Poisson point processes and on classes of intensity functions that are defined with respect to a finite reference measure (see Section 3 for a more detailed description of this hypothesis). Actually, the value of the reference measure plays a key role in the definitions of projection estimators and penalization. Our job in this section is to implement and justify a projection estimation approach that does not rely on the finiteness of the Poisson process.

Let us describe the main ingredients of our procedure. Consider the random functional

$$\gamma_D(f) \equiv -\frac{2}{T} \iint\limits_{[0,T]\times D} f(x)\, \mathcal{J}(dt, dx) + \int\limits_{D} f^2(x)\, \eta(dx), \tag{2.9}$$

which is well defined for any function $f \in L^2\left((D, \eta)\right)$, where $D \in \mathcal{B}\left(\mathbb{R}_0\right)$ and η is as in equations (2.6)-(2.8). Following the terminology of [11] and [30], we call γ_D the *contrast function*. Throughout this section,

$$\|f\|^2 \equiv \int_D f^2(x)\, \eta(dx),$$

for any $f \in L^2((D,\eta))$. Let $\mathcal{S}$ be a finite dimensional subspace of $L^2 \equiv L^2((D,\eta))$. The *projection estimator* of s on S is defined by

$$\hat{s}(x) \equiv \sum_{i=1}^{d} \hat{\beta}_i \varphi_i(x), \tag{2.10}$$

where $\{\varphi_1, \ldots, \varphi_d\}$ is an arbitrary orthonormal basis of $\mathcal{S}$ and

$$\hat{\beta}_i \equiv \frac{1}{T} \iint_{[0,T]\times D} \varphi_i(x) \mathcal{J}(dt, dx). \tag{2.11}$$

Let us give another characterization of the projection estimator.

Remark 2.2 *The projection estimator is the unique minimizer of the contrast function γ_D over S. Indeed, plugging $f = \sum_{i=1}^{d} \beta_i \varphi_i$ in (2.9) gives $\gamma_D(f) = \sum_{i=1}^{d} \left(-2\beta_i \hat{\beta}_i + \beta_i^2\right)$, and thus, $\gamma_D(f) \geq -\sum_{i=1}^{d} \hat{\beta}_i^2$, for all $f \in S$. In particular, this characterization implies that $\hat{s}$ does not depend on the choice of the orthonormal basis, and suggests a mechanism to numerically approximate $\hat{s}$ when we do not have an explicit orthonormal basis for $\mathcal{S}$.*

The remark above helps to make sense of $\hat{s}$ as an estimator of the regularized Lévy density s because the minimizer of $\mathbb{E}[\gamma_D(f)]$ over all $f \in \mathcal{S}$ is precisely the closest function in $\mathcal{S}$ to s. Concretely, the *orthogonal projection* of s on the subspace $\mathcal{S}$, namely

$$s^{\perp} \equiv \sum_{i=1}^{d} \left(\int_D \varphi_i(y) s(y) \eta(dy) \right) \varphi_i(x), \tag{2.12}$$

is such that

$$-\|s^{\perp}\|^2 = \mathbb{E}\left[\gamma_D(s^{\perp})\right] \leq \mathbb{E}\left[\gamma_D(f)\right], \quad \forall f \in \mathcal{S}. \tag{2.13}$$

Moreover, we can readily corroborate that $\hat{s}$ is an unbiased estimator of the orthogonal projection $s^{\perp}$. In order to assess the quality of estimation, we compute the "square error" of $\hat{s}$:

$$\chi^2 \equiv \|s^{\perp} - \hat{s}\|^2 = \sum_{i=1}^{d} \left[\iint_{[0,T]\times D} \varphi_i(x) \frac{\mathcal{J}(dt,dx) - s(x)\, dt\, \eta(dx)}{T} \right]^2. \tag{2.14}$$

Then, by the standard formula for the variance of Poisson integrals, the mean square error takes the form

$$\mathbb{E}\left[\chi^2\right] = \frac{1}{T}\sum_{i=1}^{d}\int_D \varphi_i^2(x)s(x)\,\eta(dx). \tag{2.15}$$

The quantity $\mathbb{E}\left[\chi^2\right]$ is called the *variance term* and the equation above shows that this term will shrink to 0 when the time horizon T goes to infinity. Moreover, the *risk* of $\hat{s}$, $\mathbb{E}\left[\|s-\hat{s}\|^2\right]$, can be decomposed into a nonrandom term plus the previous variance term:

$$\mathbb{E}\left[\|s-\hat{s}\|^2\right] = \|s-s^{\perp}\|^2 + \mathbb{E}\left[\chi^2\right]. \tag{2.16}$$

The first term, called the *bias term*, accounts for the distance of the unknown function s to the model $\mathcal{S}$ and does not depend on the estimation criteria we use within the model.

The next natural problem to tackle is to design a data-driven scheme for selecting a "good" model from a collection of linear models $\{\mathcal{S}_m, m\in\mathcal{M}\}$. Namely, we wish to select a model that approximately realizes the best trade-off between the risk of estimation within the model and the distance of the unknown Lévy density to the model. Let $\hat{s}_m$ and $s_m^{\perp}$ be respectively the projection estimator and the orthogonal projection of s on $\mathcal{S}_m$. For each $m\in\mathcal{M}$, let χ_m^2 be as in (2.14). The following simplifications of (2.16) give insight on a possible solution:

$$\begin{aligned}\mathbb{E}\left[\|s-\hat{s}_m\|^2\right] &= \|s-s_m^{\perp}\|^2 + \mathbb{E}\left[\chi_m^2\right]\\ &= \|s\|^2 - \|s_m^{\perp}\|^2 + \mathbb{E}\left[\chi_m^2\right]\\ &= \|s\|^2 - \mathbb{E}\left[\|\hat{s}_m\|^2\right] + 2\mathbb{E}\left[\chi_m^2\right]\\ &= \|s\|^2 + \mathbb{E}\left[\gamma_D\left(\hat{s}_m\right) + \mathrm{pen}(m)\right],\end{aligned} \tag{2.17}$$

where $\mathrm{pen}(m)$ is defined in terms of an orthonormal basis $\{\varphi_{1,m},\ldots,\varphi_{d_m,m}\}$ for $\mathcal{S}_m$ by the equation:

$$\mathrm{pen}(m) = \frac{2}{T^2}\iint_{[0,T]\times D}\left(\sum_{i=1}^{d_m}\varphi_{i,m}^2(x)\right)\mathcal{J}(dt,dx). \tag{2.18}$$

Equation (2.17) shows that the risk of $\hat{s}_m$ moves "parallel" to the expectation of the *observable statistics* $\gamma_D\left(\hat{s}_m\right)+\mathrm{pen}(m)$. This fact heuristically justifies to choose the model that minimizes such a penalized contrast value. In general, it makes sense to consider **penalized projection estimators** (p.p.e.) of the form

$$\tilde{s} \equiv \hat{s}_{\hat{m}}, \tag{2.19}$$

where pen : $\mathcal{M} \to [0, \infty)$, $\hat{s}_m$ is the projection estimator on $\mathcal{S}_m$ (see (2.10)), and $\hat{m} \equiv \operatorname{argmin}_{m \in \mathcal{M}} \{\gamma_D(\hat{s}_m) + \mathrm{pen}(m)\}$.

Methods of estimation based on the minimization of penalty functions have a long history in the literature on regression and density estimation (for instance, [2], [26], and [35]). The general idea is to choose among a given collection of parametric models the model that minimizes a loss function plus a penalty term that controls the complexity of the model. Such penalized estimation was promoted for nonparametric density estimation in [12], and in the context of non-homogeneous Poisson processes in [30]. There are two main accomplishments obtained in these works both in the context of density estimation and intensity estimation of nonhomogeneous Poisson processes: Oracles inequalities and competitive performance against minimax estimators. The following section shows that the method outlined above preserves Oracle inequalities.

3 Risk bounds, oracle inequalities, and rates of convergence

Consider the problem of model selection among a collection of linear models, $\{\mathcal{S}_m, m \in \mathcal{M}\}$, for the regularized Lévy density s on D as outlined in the previous section. We showed through (2.17) that a sensible criterion to decide for a projection estimator is to penalize its contrast value with a properly chosen penalty function pen : $\mathcal{M} \to [0, \infty)$. Of course, the "best" model, namely

$$\bar{m} \equiv \operatorname{argmin}_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|^2\right], \tag{3.1}$$

is not accessible, but we can aspire to achieve the smallest possible risk up to a constant. In other words, it is desirable that our estimator $\tilde{s}$ comply with an inequality of the form

$$\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|^2\right], \tag{3.2}$$

for a constant C independent of the linear models. The model that achieves the minimal risk of projection estimation is called the *Oracle model* and inequalities of the type (3.2) are called *Oracle inequalities*. Approximate Oracle inequalities were proved in [30] for the intensity function of a nonhomogeneous Poisson process $\{N_A\}_{A\in\mathcal{V}}$ on a measurable space $(\mathrm{V},\mathcal{V})$. Concretely, [30] defines projection estimators $\hat{s}_m$ and penalized projection estimators $\tilde{s}$ satisfying

$$\mathbb{E}\left[\int_V |s(\mathbf{v})-\tilde{s}(\mathbf{v})|^2 \frac{\zeta(d\mathbf{v})}{\zeta(V)}\right] \leq C \inf_{m\in\mathcal{M}} \mathbb{E}\left[\int_V |s(\mathbf{v})-\hat{s}_m(\mathbf{v})|^2 \frac{\zeta(d\mathbf{v})}{\zeta(V)}\right] + \frac{C'}{\zeta(V)}, \tag{3.3}$$

where s and ζ are respectively a bounded measurable function and a finite measure on V such that

$$\mathbb{E}[N_A] = \int_A s(\mathbf{v})d\zeta(\mathbf{v}), \quad A\in\mathcal{V}.$$

The finiteness of ζ plays an important role in the definition of the estimators, and in obtaining the Oracle inequality (3.3). However, such a property is not necessarily satisfied by the mean measure of the Poisson process $\mathcal{J}(\cdot)$ of (2.3) on $\mathcal{B}([0,T]\times D)$ (for instance, if $D=\{|x|>\varepsilon\}$ under $\zeta(d\mathbf{v})=dx\,dt$ as in (2.4), or if $D=\{0<|x|<b\}$ and $\zeta(d\mathbf{v})=x^{-2}dx\,dt$ as in the example described after Definition 2.1). In this section we show that, based on one sample of the Lévy process X on $[0,T]$, the projection estimators $\{\hat{s}_m\}_{m\in\mathcal{M}}$ introduced in Section 2, and certain penalized projection estimators $\tilde{s}$ satisfy the approximate Oracle inequality

$$\mathbb{E}\left[\|s-\tilde{s}\|^2\right] \leq C \inf_{m\in\mathcal{M}} \mathbb{E}\left[\|s-\hat{s}_m\|^2\right] + \frac{C'}{T},$$

where s is a regularized Lévy density, and the constants C, C' depend only on the "complexity" of the family of linear models. Actually, we will be able to estimate the order of the constants C and C' appearing in the Oracle inequality.

The main tool in obtaining Oracle inequalities is an upper bound for the risk of the penalized projection estimator $\tilde{s}$ of (2.19). The proof of this bound is a simple variation of the argument of [30]; however, to overcome the possible lack of finiteness on ζ and to avoid unnecessary use of upper bounds, the dimension of the linear model is explicitly included in the penalization.

Finally, the obtained risk bound is used to assess the rate of convergence of $\tilde{s}$ to s in the long run (as $T \to \infty$) when s is "smooth" and the considered linear spaces are piecewise polynomials.

The following regularity condition was introduced in [30] to make a distinction between not too "large" families of linear models and certain wavelet-type linear models. We will focus here on the simplest case:

Definition 3.1 *A collection of models $\{\mathcal{S}_m, m \in \mathcal{M}\}$ is said to be polynomial if there exist constants $\Gamma > 0$ and $R \geq 0$ such that for every positive integer n*

$$\#\{m \in \mathcal{M} : d_m = n\} \leq \Gamma n^R,$$

where d_m stands for the dimension of the model $\mathcal{S}_m$, while $\#$ denotes cardinality.

Below, we return to the setting of Section 2; that is to say, $X = \{X(t)\}_{0 \leq t \leq T}$ is a Lévy process with Lévy density p and *regularized* Lévy density s on a domain $D \in \mathcal{B}(\mathbb{R}_0)$ under a *regularizing measure* η (see Definition 2.1). Define also

$$D_m = \sup\left\{\|f\|_\infty^2 : f \in S_m, \|f\|^2 \equiv \int_D f^2(x)\eta(dx) = 1\right\}. \tag{3.4}$$

Remark 3.2 *If $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ is an arbitrary orthonormal basis of $\mathcal{S}_m$, then $D_m = \|\sum_{i=1}^{d_m} \varphi_{i,m}^2\|_\infty$ (see Section 9.3 for a verification).*

We now present the main result of this section (see Section 9.1 for the proof):

Theorem 3.3 *Let $\{\mathcal{S}_m, m \in \mathcal{M}\}$ be a polynomial family of finite dimensional linear subspaces of $L^2((D, \eta))$ and let $\mathcal{M}_T \equiv \{m \in \mathcal{M} : D_m \leq T\}$. If $\hat{s}_m$ and $s_m^\perp$ are respectively the projection estimator and the orthogonal projection of the regularized Lévy density s on $\mathcal{S}_m$ then, the penalized projection estimator $\tilde{s}_T$ on $\{\mathcal{S}_m\}_{m \in \mathcal{M}_T}$ defined by (2.19) is such that*

$$\mathbb{E}\left[\|s - \tilde{s}_T\|^2\right] \leq C \inf_{m \in \mathcal{M}_T} \left\{\|s - s_m^\perp\|^2 + \mathbb{E}\left[\mathrm{pen}(m)\right]\right\} + \frac{C'}{T}, \tag{3.5}$$

whenever $\mathrm{pen} : \mathcal{M} \to [0, \infty)$ *takes one of the following forms for some fixed (but arbitrary) constants $c > 1$, $c' > 0$, and $c'' > 0$:*

(a) $\mathrm{pen}(m) \geq c\frac{D_m \mathcal{N}}{T^2} + c' \frac{d_m}{T}$*, where* $\mathcal{N} \equiv \mathcal{J}([0,T] \times D)$ *is the number of jumps prior to* T *with sizes falling in* D *and where it is assumed that* $\rho \equiv \int_D s(x)\eta(dx) < \infty$*;*

(b) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T}$*, where* $\hat{V}_m$ *is defined in terms of an orthonormal basis* $\{\varphi_{i,m}\}_{i=1}^{d_m}$ *of* $\mathcal{S}_m$ *by*

$$\hat{V}_m \equiv \frac{1}{T} \iint\limits_{[0,T]\times D} \left(\sum_{i=1}^{d_m} \varphi_{i,m}^2(x) \right) \mathcal{J}(dt, dx), \tag{3.6}$$

and where it is assumed that $\beta \equiv \inf_{m\in\mathcal{M}} \frac{\mathbb{E}\left[\hat{V}_m\right]}{D_m} > 0$ *and that* $\phi \equiv \inf_{m\in\mathcal{M}} \frac{D_m}{d_m} > 0$*;*

(c) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T} + c' \frac{D_m}{T} + c'' \frac{d_m}{T}$*.*
Moreover, the constant C *depends only on* c*,* c' *and* c''*, while* C' *varies with* c*,* c'*,* c''*,* Γ*,* R*,* $\|s\|$*,* $\|s\|_\infty$*,* ρ*,* β*, and* ϕ*.*

Remark 3.4 *In the Remark 9.5, the order of the constants* C *and* C' *is analyzed. We will show that for* $c \geq 2$ *and for arbitrary* $\varepsilon > 0$*, there is a constant* $C'(\varepsilon)$ *(increasing) so that*

$$\mathbb{E}\|s - \tilde{s}\|^2 \leq (1+\varepsilon) \inf_{m\in\mathcal{M}} \left\{ \|s - s_m^{\perp}\|^2 + \mathbb{E}\left[\mathrm{pen}(m)\right] \right\} + \frac{C'(\varepsilon)}{T}. \tag{3.7}$$

As a first use of the previous risk bound, we obtain Oracle inequalities for our p.p.e. The next corollary immediately follows from the first equality in (2.17), equation (2.15), and part (b) above:

Corollary 3.5 *In the setting of Theorem 3.3, if the penalty function is of the form* $\mathrm{pen}(m) \equiv c\frac{\hat{V}_m}{T}$*, for every* $m \in \mathcal{M}_T$*,* $\beta > 0$*, and* $\phi > 0$*, then*

$$\mathbb{E}\left[\|s - \tilde{s}_{T}\|^2\right] \leq \tilde{C} \inf_{m\in\mathcal{M}_T} \left\{ \mathbb{E}\left[\|s - \hat{s}_m\|^2\right] \right\} + \frac{\tilde{C}'}{T}, \tag{3.8}$$

for a constant C_1 *depending only on* c*, and a constant* C_2 *depending on* c*,* Γ*,* R*,* $\|s\|$*,* $\|s\|_\infty$*,* β*, and* ϕ*.*

As a second application of (3.5), we analyze the "long run" ($T \to \infty$) rate of convergence of penalized projection estimators on regular piecewise polynomials, when the Lévy density is "smooth". More precisely, restricted to the window of estimation $D \equiv [a, b] \subset \mathbb{R}_0$, the Lévy density s is assumed to belong to the *Besov space*[1] $\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p([a, b]))$ with some $p \in [2, \infty]$ and $\alpha > 0$ (see Section 2.9-10 of [17] for the definition). An important reason for the choice of this class of functions is the availability of estimates for the error of approximation by *splines*[2], trigonometric polynomials, and wavelet expansions (see for instance Chapter 12 of [17], and Lemma 13 of [8]). In particular, if $\mathcal{S}^k_m$ denotes the space of piecewise polynomials of degree bounded by k, based on the regular partition of $[a, b]$ with m pieces ($m \geq 1$), Theorem 12.2.4 in [17] implies that for any $s \in \mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p([a, b]))$ with $k > \alpha - 1$, there exists a constant $C(s)$ such that

$$\mathrm{d}_p\left(s, \mathcal{S}^k_m\right) \leq C(s) m^{-\alpha}, \tag{3.9}$$

where d_p is the distance induced by the $\mathbb{L}^p$-norm on $([a, b], dx)$. Actually, $C(s)$ can be taken to be increasing on $|s|_{\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p)}$, the standard seminorm on $\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p([a, b]))$ (see (10.1) Chapter 2 in [17]). Combining (3.9) with (3.5), we obtain the following result (see Section 9.3 for a proof).

Corollary 3.6 *Let $D \equiv [a, b] \subset \mathbb{R}_0$ and let $\mathcal{S}^k_m$ be the space of piecewise polynomials of degree at most k based on the regular partition of $[a, b]$ with m pieces ($m \geq 1$). Following the notation of Theorem 3.3, let $\tilde{s}_{T}$ be the penalized projection estimator on $\{\mathcal{S}^k_m\}_{m \in \mathcal{M}_T}$ with penalization* $\mathrm{pen}(m) \equiv c\frac{\hat{V}_m}{T} + c'\frac{D_m}{T} + c''\frac{d_m}{T}$ *(for some fixed $c > 1$ and $c', c'' > 0$). Then, if the restriction of the Lévy density s to $[a, b]$ is a member of $\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p([a, b]))$ with $2 \leq p \leq \infty$ and $0 < \alpha < k + 1$, then*

$$\limsup_{T \to \infty} T^{2\alpha/(2\alpha+1)} \mathbb{E}\left[\|s - \tilde{s}_{T}\|^2\right] < \infty.$$

Moreover, for any $R > 0$ and $L > 0$,

$$\limsup_{T \to \infty} T^{2\alpha/(2\alpha+1)} \sup_{s \in \Theta(R,L)} \mathbb{E}\left[\|s - \tilde{s}_{T}\|^2\right] < \infty, \tag{3.10}$$

where $\Theta(R, L)$ consists of all Lévy densities s such that $\|s\|_{\mathbb{L}^\infty([a,b])} < R$, and s restricted to $[a, b]$ is a member of $\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p([a, b]))$ with seminorm $|s|_{\mathcal{B}^{\alpha}_{\infty}(\mathbb{L}^p)} < L$.

[1] These Besov spaces are also called *Lipschitz* or *Hölder* spaces.

[2] Piecewise polynomial functions f such that on each compact interval, f is made up of only finitely many polynomial pieces.

The previous result implies that the p.p.e. on regular splines has a rate of convergence of order $T^{-2\alpha/(2\alpha+1)}$ for the class of Besov Lévy densities $\Theta(R, L)$. We will see in the next section that the rate cannot be improved (see Corollary 4.3 and Remark 4.4).

4 On the minimax risk for the estimation of smooth Lévy densities

This section presents some results on the *minimax risk* of estimation for certain families of smooth Lévy densities. Roughly speaking, a *minimax risk* on a given family Θ of "parameters" has the following general form:

$$\inf_{\hat{s}} \sup_{s\in\Theta} \mathbb{E}_s \left[d\left(s, \hat{s}\right)\right],$$

where the inf is taken over all the estimators $\hat{s}$ (based on the available random data, whose law distribution is itself determined by the parameter s), and $d(s, \hat{s})$ is a function that measures how distant s and $\hat{s}$ are from each other. In some sense, $\sup_{s\in\Theta} \mathbb{E}_s \left[d\left(s, \hat{s}\right)\right]$ measures the maximum error that can arise when using the estimator $\hat{s}$. Therefore, an estimator that approximately accomplishes a minimax risk is desirable. Comparisons to the minimax risks is one of the most solicited measures of performance in statistical estimation. In fact, minimax type results have been obtained in very general contexts (see for instance [22] and [8] in the case of density estimation based on i.i.d. random variables, and [25] and [30] in the case of intensity estimation based on finite Poisson point processes).

Since the jumps of a Lévy process can be associated with a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$, many results and techniques for the statistical inference of Poisson processes can be translated into the context of Lévy processes. Following this approach, we adapt below a result of Kutoyants [25] (Theorem 6.5) on the asymptotic minimax risk for the estimation of "smooth" intensity functions of a Poisson point processes on $[0, 1]$, based on n independent copies. The idea of the proof is due to Ibragimov and Has'minskii and is based on the statistical tools for distributions satisfying the *Local Asymptotic Normality* (LAN) property (see Chapters II and Section IV.5 of [22]). Some generalizations and consequences are also deduced.

Let us introduce a *loss function* $\ell : \mathbb{R} \to \mathbb{R}$ with the following properties:

(i) $\ell(\cdot)$ is nonnegative, $\ell(0) = 0$ but not identically 0, and continuous at 0;

(ii) it is symmetric: $\ell(u) = \ell(-u)$ for all u;

(iii) for any $c > 0$, $\{u : \ell(u) < c\}$ is a convex set;

(iv) $\ell(u)\exp\{\varepsilon|u|^2\} \to 0$ as $|u| \to \infty$, for any $\varepsilon > 0$.

Consider the problem of estimating the Lévy density s of a Lévy process $\{X(t)\}_{0\leq t\leq T}$. We are interested in the error of estimation at a fixed point $x_0 \in \mathbb{R}_0$ and in minimax results of the form:

$$\liminf_{T\to\infty}\left\{\inf_{\hat{s}_T}\sup_{s\in\Theta}\mathbb{E}_s\left[\ell\left(T^\gamma\left(\hat{s}_T(x_0) - s(x_0)\right)\right)\right]\right\} > 0, \tag{4.1}$$

where the infimum is over all the "estimators" $\hat{s}_T$ based on the jumps of the Lévy process $\{X(t)\}_{0\leq t\leq T}$, Θ is a collection of Lévy densities, and $\gamma > 0$ is a constant depending on the family Θ. In other words, (4.1) implies the existence of a lower bound $B > 0$ and a time T_0 such that from that time on, all *non-anticipative*[3] estimators $\hat{s}_T$ will not do better than $T^{-\gamma}$ uniformly on Θ, in the sense that there would exist an $s \in \Theta$ for which

$$\mathbb{E}_s\left[l\left(T^\gamma\left(\hat{s}_T(x_0) - s(x_0)\right)\right)\right] > B.$$

Therefore, the inequality (4.1) impose a constraint on the rate of convergence at x_0 that the estimators can attain. By estimators, we mean a "process" $\hat{s} : \mathbb{R}_0 \times \Omega \to \mathbb{R}$ such that for each $x \in \mathbb{R}_0$, the random variable $\hat{s}(x;\cdot)$ is measurable with respect to the σ-field generated by the point process $\mathcal{J}$, while for each $\omega \in \Omega$, $\hat{s}(\cdot;\omega)$ is measurable with respect to the product σ-field.

The considered Lévy densities satisfy a Hölder condition of order β on a given window of estimation. Concretely, fix an interval $[a, b] \subset \mathbb{R}\backslash\{0\}$, and let $k \in \{0, 1, \dots\}$ and $\beta \in (0, 1]$. Define the family $\Theta_{k+\beta}(L; [a, b])$ of functions $f : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that f is k times differentiable on $[a, b]$ and

$$|f^{(k)}(x_1) - f^{(k)}(x_2)| \leq L|x_1 - x_2|^\beta, \;\; \forall \;\; x_1, x_2 \in [a, b]. \tag{4.2}$$

[3]Here, non-anticipative means that the estimator is based on the jumps that occurred up to the present.

Below, $\mathcal{L}$ stands for the class of all Lévy densities; that is, all functions $s : \mathbb{R}_0 \to \mathbb{R}_+$ such that

$$\int_{\mathbb{R}_0} \left(x^2 \wedge 1\right) s(x)dx < \infty.$$

The following result is a minor variation of Theorem 6.5 of [25]. For completeness, we present its proof in Section 9.2.

Theorem 4.1 *If x_0 is an interior point of the interval $[a,b] \subset \mathbb{R}\backslash\{0\}$, then*

$$\liminf_{T\to\infty} \left\{ \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_p \left[\ell \left(T^{\alpha/(2\alpha+1)} \left(\hat{s}_{T}(x_0) - s(x_0) \right) \right) \right] \right\} > 0, \tag{4.3}$$

where $\alpha := k + \beta$, $\Theta := \mathcal{L} \cap \Theta_\alpha (L; [a,b])$, and the infimum is over all the estimators $\hat{s}_T$ based on those jumps of the Lévy process $\{X(t)\}_{0\leq t\leq T}$ whose sizes lie in $[a,b]$.

As already noticed in [22], the previous result can be strengthen to be in a certain sense uniform in $x_0 \in (a,b)$ (see Section 9.2 for a proof)

Corollary 4.2 *With the notation and hypothesis of Theorem 4.1,*

$$\liminf_{T\to\infty} \left\{ \inf_{\hat{s}_T} \inf_{x\in(a,b)} \sup_{s\in\Theta} \mathbb{E}_s \left[\ell \left(T^{\alpha/(2\alpha+1)} \left(\hat{s}_{T}(x) - s(x) \right) \right) \right] \right\} > 0. \tag{4.4}$$

Let us now apply the above assertion to obtain the long run minimax risk of *measurable* estimators, under the $\mathbb{L}^2$-norm. Here, measurable means that for each $\omega \in \Omega$, $\hat{s}(\cdot;\omega)$ is a measurable function on $([a,b], \mathcal{B}([a,b]))$. In Section 9.2 a proof is given.

Corollary 4.3 *Let $[a,b]$ be a closed interval of $\mathbb{R}\backslash\{0\}$, then*

$$\liminf_{T\to\infty} T^{2\alpha/(2\alpha+1)} \left\{ \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_s \left[\int_a^b \left(\hat{s}_{T}(x) - s(x) \right)^2 dx \right] \right\} > 0, \tag{4.5}$$

where $\alpha := k + \beta$, $\Theta := \mathcal{L} \cap \Theta_\alpha (L; [a,b])$, and the infimum is over all the measurable estimators $\hat{s}_T$ based on the jumps of the Lévy process $\{X(t)\}_{0\leq t\leq T}$ whose sizes lie on $[a,b]$.

Remark 4.4 *The proofs of the previous results can be readily modified to cover even smaller classes of Lévy densities* Θ. *For instance,* $\Theta = \mathcal{L} \cap \Theta_\alpha(L;[a,b]) \cap \{s : \|s\|_{\mathbb{L}^\infty([a,b])} < R\}$. *This class has a very close relationship with the family of Besov densities* $\Theta(R,L)$ *introduced in (3.10). Indeed, the class* $\Theta_\alpha(L;[a,b])$ *is contained in* $\mathcal{B}^\alpha_\infty(\mathbb{L}^\infty([a,b]))$ *(see Section 2.9 of [17]). Since* $\mathcal{B}^\alpha_\infty(\mathbb{L}^\infty) \subset \mathcal{B}^\alpha_\infty(\mathbb{L}^p)$, *(4.5) holds true on* $\Theta = \Theta(R,L)$. *Therefore, the p.p.e. on regular splines, described in the previous section, has the best possible rate of convergence and moreover, achieves the minimax rate of convergence on* $\Theta(R,L)$. *This type of property is called* adaptivity *in that, without knowing the smoothness of* s *(controlled by* α*), the p.p.e. reaches asymptotically the minimax risk up to a constant. See for instance Section 4 of [8] for a discussion on adaptivity.*

5 Calibration based on discrete time data: approximation of Poisson integrals

One drawback to the method outlined in Section 2 is that in general we do not observe the jumps of a Lévy process $X = \{X(t)\}_{t\geq 0}$. In practice, we can aspire to sample the process $X(t)$ at discrete times, but we are neither able to measure the size of the jumps $\Delta X(t) \equiv X(t) - X(t^-)$ nor the times of jumps $\{t : \Delta X(t) > 0\}$. Poisson integrals of the type

$$I(f) \equiv \iint_{[0,T]\times\mathbb{R}_0} f(x)\mathcal{J}(dt,dx) = \sum_{t\leq T} f(\Delta X(t)), \tag{5.1}$$

are simply not accessible. In this section, we discuss the approximation of the integral (5.1) based on time series of the form $\{X(t^n_k)\}^n_{k=0}$, where $t^n_k = \frac{kT}{n}$. Let us motivate our approximation scheme. The natural way of interpolating the sample path of a Lévy process from the sampling observations $\{X(t^n_k)\}^n_{k=0}$ is to take a càdlàg piecewise constant approximation of the form

$$X^n(t) \equiv \sum_{k=1}^n X\left(t^n_{k-1}\right) \mathbf{1}\left(t \in [t^n_{k-1}, t^n_k)\right), \quad t \in [0,T), \tag{5.2}$$

where as usual $\mathbf{1}$ is the indicator function of the corresponding set. It is quite simple to prove that X^n converges to X at finitely many points with probability one (a quality shared by any right-continuous process X). Furthermore, the approximated process X^n, having independent increments, converges to $\mathbf{X}$ in $D[0,\infty)$, under the Skorohod metric (see Example VI.18 of [28]). Hence, we might expect that

$$I_n(f) \equiv \sum_{t\le T} f(\Delta X^n(t)) = \sum_{k=1}^{n} f\left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right)\right), \tag{5.3}$$

converges to (5.1) as $n \to \infty$.

The weak convergence of (5.3) to (5.1) can be proved using well-know facts on the transition distributions of X in small time. Namely, if $[a,b] \subset \mathbb{R}_0$ and $h(x) = \mathbf{1}_{(a,b]}(x)g(x)$ for a continuous function g, then

$$\lim_{t\to 0} \frac{1}{t}\mathbb{E}\left[h\left(X(t)\right)\right] = \int_{\mathbb{R}_0} h(x)\nu(dx); \tag{5.4}$$

see for instance pp. 39 of [9] or Corollary 8.9 of [34]. In [33] stronger results are provided. For instance, if h is continuous, bounded, and if $\lim_{|x|\to 0} h(x)|x|^{-2} = 0$, then (5.4) continues to hold.

Limiting results like (5.4) are useful to establish the convergence in distribution of $\mathrm{I}_n(f)$ since

$$\mathbb{E}\left[e^{\mathrm{i}uI_n(f)}\right] = \left(\mathbb{E}\left[e^{\mathrm{i}uf\left(X\left(\frac{T}{n}\right)\right)}\right]\right)^n = \left(1 + \frac{a_n}{n}\right)^n,$$

where $a_n \equiv \mathbb{E}\left[e^{\mathrm{i}uf(X\left(\frac{T}{n}\right))} - 1\right]$. So, if f is such that (5.4) holds for $h(x) \equiv e^{\mathrm{i}uf(x)} - 1$, then

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = \lim_{n\to\infty} e^{n\log\left(1+\frac{a_n}{n}\right)} = e^{T\int_{\mathbb{R}_0} h(x)\nu(dx)},$$

which is the characteristic function of the Poisson integral $I(f)$ (see Lemma 10.2 of [24]). The following result summarizes the previous discussion:

Proposition 5.1 *Let $X = \{X(t)\}_{t\ge 0}$ be a Lévy process with Lévy measure ν. Then,*

$$\lim_{n\to\infty} \mathbb{E}\left[e^{\mathrm{i}uI_n(f)}\right] = \exp\left\{T\int_{\mathbb{R}_0}\left(e^{\mathrm{i}uf(x)} - 1\right)\nu(dx)\right\},$$

if f satisfies either one of the following conditions:

(1) $f(x) = \mathbf{1}_{(a,b]}(x)g(x)$ *for an interval* $[a,b] \subset \mathbb{R}_0$ *and a continuous function* $g : \mathbb{R} \to \mathbb{R}$*;*

(2) $f(x)$ *is continuous on* $\mathbb{R}_0$ *and* $\lim_{|x|\to 0} f(x)|x|^{-2} = 0$.

In particular, $I_n(f)$ *converges in distribution to* $I(f)$ *under any of the two previous conditions.*

Remark 5.2 *Notice also that if* f *and* f^2 *satisfy (5.4), then the mean and variance of* $I_n(f)$ *obey the asymptotics:*

$$\lim_{n\to\infty} \mathbb{E}\left[I_n(f)\right] = T \int_{\mathbb{R}_0} f(x)\nu(dx);$$
$$\lim_{n\to\infty} \mathrm{Var}\left[I_n(f)\right] = T \int_{\mathbb{R}_0} f^2(x)\nu(dx).$$

6 Estimation Method

Let us summarize the previous sections and outline the proposed algorithm of estimation:

Statistician's parameters: The procedure is fed with a Borel *window of estimation* $D \subset \mathbb{R}_0$, a collection $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$ of finite dimensional *linear models* of $L^2\left((D,\eta)\right)$, and a *level of penalization* $c > 1$.

Model and data: It is assumed that a Lévy process $\{X\left(t\right)\}_{t\in[0,T]}$ is monitored at equally spaced times $t_k^n = k\frac{T}{n}$, $k = 1, \ldots, n$, during the time period $[0,T]$. The data consists of the time series $\{X\left(t_k^n\right)\}_{k=1}^n$. The Lévy process admits a regularized Lévy density s under the measure η on D (see Definition 2.1).

Estimators: Inside the linear model $\mathcal{S}_m$, the estimator of s is the *approximated projection estimator*:

$$\hat{s}_m^n(x) \equiv \sum_{i=1}^{d_m} \hat{\beta}_{i,m}^n \varphi_{i,m}(x), \tag{6.1}$$

where $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ is an orthonormal basis for $\mathcal{S}_m$, and

$$\hat{\beta}^n_{i,m} \equiv \frac{1}{T} \sum_{k=1}^{n} \varphi_{i,m}\left(X\left(t^n_k\right) - X\left(t^n_{k-1}\right)\right), \tag{6.2}$$

is the estimator of the inner product $\beta_{i,m} \equiv \int_D \varphi_{i,m}(x) s(x) \eta(dx)$, for $i = 1, \ldots, d_m$. Across the collection of linear models $\{\mathcal{S}_m : m \in \mathcal{M}\}$, the estimator $\hat{s}^n_m$ which minimizes $-\|\hat{s}^n_m\|^2 + c\ \mathrm{pen}^n(m)$, is selected, where

$$\mathrm{pen}^n(m) = \frac{1}{T^2} \sum_{k=1}^{n} \left(\sum_{i=1}^{d_m} \varphi^2_{i,m}\left(X\left(t^n_k\right) - X\left(t^n_{k-1}\right)\right) \right).$$

Remark 6.1 *It is worthwhile to point out the great similarity of the scheme above to some methods of density estimation introduced in [12]. In this paper, the authors estimate the probability density function f of a random sample $X_1, \cdots, X_n$ by projection estimators of the form:*

$$\hat{f}(x) = \sum_{i=1}^{d} \left\{ \frac{1}{n} \sum_{k=1}^{n} \varphi_i\left(X_k\right) \right\} \varphi_i(x), \tag{6.3}$$

where $\{\varphi_i\}_{i=1}^d$ is an orthonormal basis of a linear space $\mathcal{S}$ of $L^2(\mathbb{R}, dx)$. More generally, f can be the density function with respect to a measure μ in the sense that $\mathbb{P}[X_i \in \cdot] = \int_{\cdot} f(x)\mu(dx)$, and the projection estimator above will be well defined provided that $f \in L^2(\mathbb{R}, \mu(dx))$. To solve the problem of model selection, they introduced penalized projection estimators. One considered penalty function there is

$$\mathrm{pen}(\mathcal{S}) = \frac{2}{n(n+1)} \sum_{k=1}^{n} \sum_{i=1}^{d} \varphi^2_i\left(X_k\right).$$

In some sense, the method outlined at the beginning of this section "works" as a byproduct of the small time qualities of Lévy processes and of standard methods of nonparametric estimation for probability densities. Indeed, consider the statistics

$$\hat{\beta}^{n,j}_{i,m} \equiv \frac{n}{Tj} \sum_{k=1}^{j} \varphi_{i,m}\left(X\left(t^n_k\right) - X\left(t^n_{k-1}\right)\right),$$

where T/n is the time span of the increments and j is the number of increments in the sample. From [11], as j progresses,

$$\hat{s}_m^{n,j}(x) \equiv \sum_{i=1}^{d_m} \hat{\beta}_{i,m}^{n,j}\, \varphi_{i,m}(x), \tag{6.4}$$

estimates the orthogonal projection of $\frac{n}{T} f_{T/n}(x)$ on $\mathcal{S}_m$, where f_t stands for the probability density function of $X(t)$ (if it exists). On the other hand, [33] proves that $\frac{n}{T} f_{T/n}(x)$ converges to the Lévy density p, as $n \to \infty$ (under some regularity conditions). Therefore, for large n and j, (6.4) will approximate the projection of p on $\mathcal{S}_m$. Notice that in general, a.s.

$$\lim_{n\to\infty} \lim_{j\to\infty} \frac{n}{Tj} \sum_{k=1}^{j} \varphi\left(X\left(t_k^n\right) - X\left(t_{k-1}^n\right)\right) = \int_{\mathbb{R}_0} \varphi(x)\nu(dx),$$

whenever φ is such that the limit (5.4) holds. Our penalized projection estimators (6.1) are obtained from (6.4) by taking $n = j$. It is not clear from the references just mentioned whether taking $n = j \to \infty$ will produce good results or not. We will see below that this is the case.

Let $\mathcal{R}(X)$ be the linear space of measurable functions h such that (5.4) is satisfied. For instance, $\mathcal{R}(X)$ contains the functions f satisfying conditions (1) or (2) in Proposition 5.1. The following result holds true (see Section 9.3 for a proof).

Proposition 6.2 *Let $s_m^{\perp}$ be the orthogonal projection of s on $\mathcal{S}_m$. If $\varphi_{i,m}$ and $\varphi_{i,m}^2$ belong to $\mathcal{R}(X)$ for every $m \in \mathcal{M}$ and $i = 1, \ldots, d_m$, then the approximated projection estimator $\hat{s}_m^n$ of s on $\mathcal{S}_m$ (based on n equally spaced observations) satisfies:*

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}_m^n - s_m^{\perp}\|^2\right] = \mathbb{E}\left[\|\hat{s}_m - s_m^{\perp}\|^2\right]. \tag{6.5}$$

Moreover,

$$\lim_{n\to\infty} \mathbb{E}\left[\|\hat{s}_m^n - s\|^2\right] = \mathbb{E}\left[\|\hat{s}_m - s\|^2\right].$$

7 Numerical tests of projection estimators

In this section, we try to assess the performance of some penalized projection estimators based on simulated Lévy processes. Piecewise constant functions are considered, and for their intrinsic relevance in mathematical finance, two classes of Lévy processes are studied: Gamma and variance Gamma processes. A method of least-squares errors is also applied to generate parametric Lévy densities that closely fit the nonparametric outputs.

7.1 Specifications of the statistical methods

Let us describe in greater details the considered projection estimators. To simplify notation, $\mathcal{J}(B)$ is used instead of $\mathcal{J}([0,T]\times B)$ of (2.3) when referring to the number of jumps of sizes in $B \in \mathcal{B}(\mathbb{R}_0)$ occurring prior to T. Let $\mathcal{C} : a = x_0 < x_1 < \cdots < x_m = b$ be a partition of the interval $D \equiv [a,b]$ ($0 < a$ or $b < 0$), and let $\mathcal{S}_{\mathcal{C}}$ be the span of the indicator functions $\chi_{[x_0,x_1)}, \ldots, \chi_{[x_{m-1},x_m)}$. In other words, the linear model $\mathcal{S}_{\mathcal{C}}$ consists of "histogram functions" on the window D with cutoff points in $\mathcal{C}$. We assume that the Lévy process has a Lévy density s bounded outside of any neighborhood of the origin. This assumption is very mild, and yet good enough for the integral $\int_D s^2(x)dx$ to be finite. In that case, the orthogonal projection of s onto $\mathcal{S}_{\mathcal{C}}$ exists (under the standard inner product of $L^2(D, dx)$), and thus the projection estimation on $\mathcal{S}_{\mathcal{C}}$ is meaningful. In the terminology of Section 2, the regularizingn measure is simply dx, the regularized Lévy density coincides with the Lévy density, and the orthonormal basis $\{\varphi_1, \ldots, \varphi_m\}$ for $\mathcal{S}_{\mathcal{C}}$ is

$$\varphi_i(x) = \frac{1}{\sqrt{x_i - x_{i-1}}}\chi_{[x_{i-1},x_i)}(x), \quad i = 1, \ldots, m.$$

According to the *basic estimation method* outlined in Section 2, the projection estimator on the linear model $\mathcal{S}_{\mathcal{C}}$ is given by

$$\hat{s}_{\mathcal{C}}(x) = \frac{1}{T}\sum_{i=1}^{m} \frac{\mathcal{J}([x_{i-1}, x_i))}{x_i - x_{i-1}} \, \chi_{[x_{i-1},x_i)}(x). \tag{7.1}$$

Following the heuristics of Section 2 and Theorem 3.3 part (b), an appealing procedure to select a projection estimator of the form (7.1) is to look for the minimization of the following penalized contrast value

$$\frac{1}{T^2}\sum_{i=1}^{m}\frac{1}{x_i - x_{i-1}}\left\{c\,\mathcal{J}([x_{i-1},x_i)) - [\mathcal{J}([x_{i-1},x_i))]^2\right\}. \tag{7.2}$$

Here, $c > 1$ is a constant that controls the level of penalization. In fact, Theorem 3.3 and Corollary 3.8 ensure that, for large enough T, the previous procedure will yield competitive results against the best projection estimator. For that to happen it is necessary to restrict ourselves to models $\mathcal{C}$ satisfying $D_{\mathcal{C}} \le T$, where $D_{\mathcal{C}}$ is defined as in (3.4). In this case, the constant $D_{\mathcal{C}}$ is $1/\min_{1\le i\le m}\{x_i - x_{i-1}\}$ as seen from Remark 3.2.

The simplest case is to take regular partitions $\{x_i = a + i\Delta x\}_{i=0}^{m}$, where $\Delta x = (b-a)/m$ is the mesh of the partition. Then, the projection estimators of (7.1) becomes

$$\hat{s}_m(x) \equiv \frac{m}{T(b-a)}\sum_{i=1}^{m}\,\mathcal{J}([x_{i-1},x_i))\,\chi_{[x_{i-1},x_i)}(x), \tag{7.3}$$

and penalized projection estimation will look to minimize

$$\frac{m}{T^2(b-a)}\left(c\mathcal{J}([a,b)) - \sum_{i=1}^{m}\left(\mathcal{J}([x_{i-1},x_i))\right)^2\right), \tag{7.4}$$

over all m such that $D_m = m/(b-a)$ is smaller than T.

For comparisons against other procedures and to assess the goodness of fit to specific parametric models, it is useful to determine the parametric model of a given type that "best fits" our non-parametric estimators; for instance, suppose the we want to assess whether or not the nonparametric results supports the parametric Gamma model for the Lévy density. The method of least square errors provides an easy solution to this problem. For instance, if $s_\theta(x)$ is the parametric form of the Lévy density, a plausible estimator of θ is

$$\hat{\theta} = \mathrm{argmin}_\theta\; d(s_\theta, \hat{s}),$$

where $\hat{s}$ is the (penalized) projection estimator on a given family of linear models, and d is a function that accounts for the difference between s_θ and $\hat{s}$. For instance, for a fixed set of points $\{x_i\}_{i=1}^{k} \subset D$, $d(\cdot,\cdot)$ can simply be defined for functions f and g as

$$d(f,g) \equiv \sum_{i=1}^{k} \left[f(x_i) - g(x_i)\right]^2 .$$

It is sometimes preferable to use a least-square method that is linear in the parameters, and hence, is robust against numerical errors. In that case, we can look for a functional L so that $L(s_\theta)$ is linear in θ and define

$$d(f,g) \equiv \sum_{i=1}^{k} \left[L(f)(x_i) - L(g)(x_i)\right]^2 .$$

As an example, consider the Lévy density of a Gamma Lévy process with parameters α and β:

$$s(x) = \frac{\alpha}{x} e^{-x/\beta}, \quad x > 0.$$

Given a projection estimator $\hat{s}$ of s, least-square estimates of α and β can be constructed from

$$\mathrm{argmin}_{\alpha,\beta} \sum_{i=1}^{m} \left(\frac{\alpha}{x_i} \exp\left(-\frac{x_i}{\beta}\right) - \hat{s}(x_i) \right)^2 , \tag{7.5}$$

where $\{x_i\}_{i=1}^{k} \subset D$. Notice that the estimation would be very susceptible to the points close to the origin. Instead, a regression method that is linear in the parameters can be devised using a logarithmic transformation as follows

$$\mathrm{argmin}_{\alpha,\beta} \sum_{i=1}^{m} \left(-\frac{1}{\beta} x_i + \log(\alpha) - \log(x_i \hat{s}(x_i)) \right)^2 . \tag{7.6}$$

7.2 Estimation of Gamma Lévy densities.

7.2.1 The model

Lévy Gamma processes are fundamental building blocks in the construction of other Lévy processes like the variance Gamma model [16] and the generalized Gamma convolutions [13]. Moreover, by Bernstein's theorem, any Lévy density of the form $u(x)/|x|$, where u is a completely monotone function, is the limit of superpositions of Gamma Lévy densities.

The Gamma Lévy process $X = \{X(t)\}_{t\geq 0}$ is determined by two positive parameters α and β so that the probability density function of $X(t)$ is

$$f_t(x) = \frac{x^{\alpha t-1}e^{-x/\beta}}{\Gamma(\alpha t)\beta^{\alpha t}}, \tag{7.7}$$

for $x > 0$. The characteristic function of $X(t)$ is

$$E\left[e^{iuX(t)}\right] = (1 - i\beta t)^{\alpha t} = \exp\left[t\left(\alpha\int_0^\infty \left(e^{iux} - 1\right)\nu(dx)\right)\right],$$

where the Lévy measure ν is

$$\nu(dx) = \frac{\alpha}{x}\exp\left(-\frac{x}{\beta}\right)dx, \quad \text{for } \ x > 0; \tag{7.8}$$

see [20] pp. 87 or Example 8.10. of [34]. From the point of view of the marginal densities, β is a *scale parameter* and α is a *shape parameter*. In terms of the jump activity, α controls the overall activity of the jumps, while β takes charge of the heaviness of the Lévy density tail, and hence, of the frequency of big jumps. Notice that changes in the time units is statistically equivalent to changes of the parameter α, while changes in the units at which the values of X are measured are statistically reflected on changes of the parameter β. That is to say, the scaled process $\{cX(ht)\}_{t\geq 0}$ is also a Gamma Lévy process with shape parameter αh and scale parameter βc. This property is consistent with the previous remark on α taking charge of the jump activity and on β taking charge of the frequency of large jumps.

7.2.2 The simulation procedure

Simulation schemes based on series representation are used to generate Gamma Lévy processes. Such procedures allow to retrieve a sample of the jumps of the process. Concretely, following [31], the process

$$X(t) \equiv \beta\sum_{i=1}^{\infty} V_i \exp\left(-\frac{\Gamma_i}{\alpha}\right)\mathbf{1}[U_i \leq t], \tag{7.9}$$

is a Gamma Lévy process on $[0, T]$ with shape parameter α and scale parameter β provided that $\{\Gamma_i\}_{i\geq 1}$ is a homogenous Poisson process with intensity

1, $\{V_i\}_{i\geq 1}$ are independent exponential r.v. with mean 1, $\{U_i\}_{i\geq 1}$ are i.i.d. uniformly distributed on $[0,T]$, and these three series are mutually independent. Below, we shall truncate the series to n terms in order to generate a sample path, and in particular, to approximate the jump process $\mathcal{J}$ of X by

$$\mathcal{J}_n(\cdot) \equiv \sum_{i=1}^{n} \delta_{(U_i,J_i)}(\cdot), \tag{7.10}$$

where $J_i \equiv \beta V_i \exp\left(-\frac{\Gamma_i}{\alpha}\right)$.

7.2.3 The numerical results

We now present a few examples to illustrate the technique of projection estimation on histogram functions based on regular partitions (see Section 7.1 for the specifications of the estimation method). Figure 1 shows the Gamma Lévy density with $\alpha = 1$ and $\beta = 1$, and the penalized projection histogram of the form (7.3). The estimation is based on 2000 jumps of the Gamma Lévy process on $[0, 365]$, and their resulting Poisson integrals obtained by using (7.10) instead of $\mathcal{J}$. The least-square method (7.6), taking the x_i's as the mid points of the partition intervals, yields the estimators $\hat{\alpha} = 0.932$ and $\hat{\beta} = 1.055$. The maximum likelihood estimators based on the increments of the sample path of time length 1 are 1.015 for α and 0.949 for β (we do not observe real improvement if the time length of the increments is reduced).

In the next simulation, we consider a Gamma density with a lighter tail ($\beta = 0.5$) and more jump activity ($\alpha = 2$). The opposite setting was also studied: a heavier tail determined by a $\beta = 2$ and a lower jump activity given by an $\alpha = 0.5$ (see Figures 2 and 3). In the first scenario, the least-square method estimators are $\hat{\alpha} = 1.907$ and $\hat{\beta} = 0.472$, while the maximum likelihood estimators are 1.924 and 0.527, respectively. For this second Gamma density, the least-square method (7.5), taking the x_i's as the midpoints of the partition intervals, produce estimators $\hat{\alpha} = 0.5$ and $\hat{\beta} = 1.72$, while the maximum likelihood estimators are 0.55 and 1.99, respectively.

Approximate histogram estimation on regular partitions is less successful in case of high activity levels. This problem is particularly evident when we have in addition heavy tails in the Lévy density. For instance, if $\alpha = 3$ and $\beta = 3$, the method requires a large sample size to satisfactorily retrieve the behavior around the origin (see Figures 4 and 5). For 2000 jumps, the least square estimates are $\hat{\alpha} = 1.87$ and $\hat{\beta} = 4.45$, while the estimates are $\hat{\alpha} = 2.8893$ and $\hat{\beta} = 2.9268$ for twice as many jumps. The maximum likelihood estimators based on the increments of time length 0.5 are 2.4134054 for α and 3.30971 for β when the approximating process is made out of 2000 jumps, while when the process is approximated using 4000 jumps, these estimates are 2.8281 and 3.1007 for α and β, respectively. We also notice in our experiments that the estimates for the first simulation improve considerably if the window of estimation is taken "far away" from the origin (for example, $\hat{\alpha} = 3.20944$ and $\hat{\beta} = 2.68775$ on $[a, b] = [1.5, 5]$; see Figure 6).

7.2.4 Regularized projection estimation around the origin

We present another way to estimate the Gamma Lévy density even around the origin based on the regularization technique described in Section 2. The key observation is the following: the Gamma Lévy measure (7.8) can be written as

$$\nu(dx) = \alpha x \exp\left(-\frac{x}{\beta}\right) \eta(dx), \tag{7.11}$$

where $\eta(dx) = \frac{1}{x^2} dx$. Then, $s(x) \equiv \alpha x \exp\left(-\frac{x}{\beta}\right)$ is square integrable with respect to η, opening the possibility to use the projection estimation of s on a linear space $\mathcal{S}$ of $L^2\left((0, \infty), \eta\right)$. Once an estimator $\hat{s}$ for s has been obtained, $\hat{p}$ defined by $\hat{p}(x) = \hat{s}(x)/x^2$ can work as an estimator for the Lévy density $p(x) \equiv \alpha \exp\left(-x/\beta\right)/x$. In the terminology introduced in Section 2, η is a regularizing measure for the Gamma Lévy density p, and s is the corresponding regularized Lévy density (see Definition 2.1).

Let us specify this method for the linear model

$$\mathcal{S}_{\mathcal{C}} = \left\{ f(x) \equiv c_1 x \chi_{[x_0, x_1)}(x) + \sum_{i=2}^{m} c_i\ \chi_{[x_i, x_{i+1})}(x) : c_1, \ldots, c_m \in \mathbb{R} \right\},$$

where $\mathcal{C} : 0 = x_0 < x_1 < \cdots < x_m = b$ is a partition of a chosen interval $D = [0, b]$. The projection estimator, say $\hat{s}_{\mathcal{C}}$, onto $\mathcal{S}_{\mathcal{C}}$, under the standard inner product of $L^2\left((0, \infty), \eta\right)$, takes on the value

$$\hat{s}_{\mathcal{C}}(x) = x \frac{1}{T x_1} \sum_{t \leq T} \Delta X(t) \mathrm{I}\left[\Delta X(t) < x_1\right],$$

if $x < x_1$, while if $x_{i-1} \leq x < x_i$, for some $i \in \{2, \ldots, m\}$, then

$$\hat{s}_{\mathcal{C}}(x) = \frac{x_{i-1} x_i}{T(x_i - x_{i-1})} \mathcal{J}([x_{i-1}, x_i)).$$

We shall use the penalty function of Theorem 3.3 part (b) to perform the model selection procedure. That is, among different partitions $\mathcal{C}$ that satisfy

$$D_{\mathcal{C}} = \max \left\{ \frac{1}{x_1}, \frac{x_2 x_1}{x_2 - x_1}, \ldots, \frac{x_m x_{m-1}}{x_m - x_{m-1}} \right\} \leq T,$$

we choose the projection estimator $\hat{s}_{\mathcal{C}}$ that minimize

$$\gamma(\hat{s}_{\mathcal{C}}) + \hat{V}_{\mathcal{C}} = \frac{1}{T^2} \sum_{i=2}^{m} \frac{x_i x_{i-1}}{x_i - x_{i-1}} \left[c\, \mathcal{J}([x_{i-1}, x_i)) - (\mathcal{J}([x_{i-1}, x_i)))^2 \right]$$

$$+ \frac{c}{T^2 x_1} \sum_{\substack{t \leq T: \\ \Delta X(t) < x_1}} (\Delta X(t))^2 - \frac{1}{x_1} \left(\sum_{\substack{t \leq T: \\ \Delta X(t) < x_1}} \Delta X(t) \right)^2 .$$

The previous formulas are found directly from the definitions and results given in Section 2 (see for instance formulas (2.9), (2.10), (3.4), and (3.6)).

Remark 7.1 *Observe that the previous procedure is appropriate to estimate the density function $s(x) = \frac{\alpha}{x} \exp(-\frac{x}{\beta})$ around the origin as far as*

$$\hat{\alpha} \equiv \frac{1}{T x_1} \sum_{t \leq T} \Delta X(t) \mathrm{I} \left[\Delta X(t) < x_1 \right],$$

is a good estimator of α. It is not hard to check that the bias of $\hat{\alpha}$ tend to zero as $x_1 \searrow 0$. However, the variance of $\hat{\alpha}$ converges to $\frac{\alpha}{2T}$, suggesting that the method works better when T is "large" and α is "small".

We apply the above method to the simulated Lévy process used in Figure 1; i.e. a Gamma process with $\alpha = 1$ and $\beta = 1$. Figure 7 shows the estimator $\hat{p}_2(x) = \hat{s}(x)/x^2$ and the actual Lévy density $p(x) = \exp(-x)/x$ for $x \in [0.02, 1]$ (we used regular partitions on $[0, 1]$). From Figure 1, the improvement is notorious, and moreover, we accomplish a good estimation around the origin of $\hat{p}_2(x) = 0.9/x$, for $x \in [0, 0.2)$.

This regularization procedure was also applied to the simulations of the Gamma Lévy processes with ($\alpha = 3, \beta = 3$) and with ($\alpha = 1/2, \beta = 2$). The results are plotted in Figures 8 and 9 below (compare with Figures 3 and 5). We observe an improvement on both sample data. For instance, for $\alpha = \beta = 3$, the nonparametric estimator $\hat{s}(x)/x^2$ combined with a method of least-squares errors estimate α by 2.7296 and β by 3.2439. Similarly, when $\alpha = 0.5$ and $\beta = 2$, least-square errors estimates $\hat{\alpha} = 0.4825$ and $\hat{\beta} = 2.1131$.

7.2.5 Performance of projection estimation based on finitely many observation

In this part, we study the performance of the (approximate) projection estimators introduced in Section 5, and formally stated in Section 6. Namely, the method obtained by approximating the Poisson process of jumps $\mathcal{J}$ by

$$\mathcal{J}^n(\cdot) = \sum_{i=1}^{n} \delta_{(t_i^n J_i)}(\cdot),$$

where J_i is the i^{th} increment of X from t_{i-1}^n to t_i^n and $t_i^n = iT/n$. The time span between increments is denoted by $\Delta t = T/n$. Again, the considered estimators are histograms as defined in Section 7.1 and applied in Section 7.2.3.

Table 1 compares the (approximate) penalized projection estimators with least-square errors (PPE-LSE) to the maximum likelihood estimators (MLE) for the Gamma Lévy process with $\alpha = \beta = 1$ using different time spans Δt. We also consider two types of simulations: jump-based and increment-based. The method based on jumps uses series representation with $n = 36500$ jumps occurring during the time period $[0, 365]$ (notice that if we think of 365 as days, the number of jumps corresponds to a rate of about 1 jump every 5 minute). The increment-based method is a *discrete skeleton* with mesh of 0.001. Notice that maximum likelihood estimation does not do well for small time spans when the approximate sample path is based on jumps. On the other hand, penalized projection estimation does not provide good results for long time spans when the approximate sample path is based on increments. The sampling distributions of the MLE for α and β are shown in Figures 12 and 13 in the case of $\Delta t = 0.1$. On the other hand, the sampling distributions of the estimates for α and β obtained from fitting the PPE are given in Figures 14 and 15. Even though, the MLE are much more superior, the estimates based PPE have good performance considering that they are model-free.

	Jump-based Simulation				Increment-based Simulation			
Δt	PPE-LSE		MLE		PPE-LSE		MLE	
1	1.01	1.46	.997	.995	.73	1.78	1.09	.99
0.5	1.03	1.09	.972	.978	.9	1.49	1.01	1.06
0.1	.944	.995	1.179	.837	.923	1.03	.989	1.09
0.01	.969	.924	6.92	.5	.955	1.019	.9974	1.083

Table 1: Estimation of a Lévy Gamma process with $\alpha = \beta = 1$. Two types of simulation are considered: series-representation based and increments-based. The estimations are based on equally spaced sampling observation at the time span Δt. Results for the approximate penalized projection estimators with least-squares errors, and for the maximum likelihood estimators are given.

7.3 Estimation of variance Gamma processes.

7.3.1 The model

Variance Gamma processes were proposed in [16] as substitutes to the Brownian Motion in the Black-Scholes model. There are two useful representations for this type of processes. In short, a variance Gamma process $X = \{X(t)\}_{t\geq 0}$ is a Brownian motion with drift, time changed by a Gamma Lévy process. Concretely,

$$X(t) = \theta U(t) + \sigma W(U(t)), \tag{7.12}$$

where $\{W(t)\}_{t\geq 0}$ is a standard Brownian motion, $\theta \in \mathbb{R}$, $\sigma > 0$, and $U = \{U(t)\}_{t\geq 0}$ is an independent Gamma Lévy process with density at time t given by

$$f_t(x) = \frac{x^{t/\nu - 1} \exp\left(-\frac{x}{\nu}\right)}{\nu^{t/\nu}\Gamma\left(\frac{t}{\nu}\right)}. \tag{7.13}$$

Notice that $\mathrm{E}\left[U(t)\right] = t$ and $\mathrm{Var}\left[U(t)\right] = \nu t$; therefore, the random clock U has a "mean rate" of one and a "variance rate" of ν. There is no loss of generality in restricting the mean rate of the Gamma process U to one since, as a matter of fact, any process of the form

$$\theta_1 V(t) + \sigma_1 W(V(t)),$$

where $V(t)$ is an arbitrary Gamma Lévy process, $\theta_1 \in \mathbb{R}$, and $\sigma_1 > 0$, has the same law as a process of the form (7.12) with suitably chosen θ, σ, and ν. This a consequence of the *self-similarity*[4] property of Brownian motion and the fact that ν in (7.13) is a scale parameter.

The process X is itself a Lévy process since Gamma processes are *subordinators* (see Theorem 30.1 of [34]). Moreover, it is not hard to check that "statistically" X is the difference of two Gamma Lévy processes (see 2.1 of [14]):

$$\{X(t)\}_{t\geq 0} \stackrel{\mathfrak{D}}{=} \{X_+(t) - X_-(t)\}_{t\geq 0}, \tag{7.14}$$

where $\{X_+(t)\}_{t\geq 0}$ and $\{X_-(t)\}_{t\geq 0}$ are Gamma Lévy processes with respective Lévy measures

$$\nu_\pm(dx) = \alpha \exp\left(-\frac{x}{\beta_\pm}\right) dx, \quad \text{for } x > 0.$$

Here, $\alpha = 1/\nu$ and

$$\beta_\pm = \sqrt{\frac{\theta^2\nu^2}{4} + \frac{\sigma^2\nu}{2}} \pm \frac{\theta\nu}{2}.$$

As a consequence of this decomposition, the Lévy density of X takes the form

$$s(x) = \begin{cases} \frac{\alpha}{|x|} \exp\left(-\frac{|x|}{\beta_-}\right) & \text{if } x < 0, \\ \frac{\alpha}{x} \exp\left(-\frac{x}{\beta_+}\right) & \text{if } x > 0, \end{cases}$$

where $\alpha > 0$, $\beta_- \geq 0$, and $\beta_+ \geq 0$ (of course, $\beta_-^2 + \beta_+^2 > 0$). As in the case of Gamma Lévy processes, α controls the overall jump activity, while β_+ and β_- take respectively charge of the intensity of large positive and negative jumps. In particular, the difference between $1/\beta_+$ and $1/\beta_-$ determines the frequency of drops relative to rises, while their sum measures the frequency of large moves relative to small ones.

[4] namely, $\{W(ct)\}_{t\geq 0} \stackrel{\mathfrak{D}}{=} \{c^{1/2}W(t)\}_{t\geq 0}$, for any $c > 0$.

7.3.2 The simulation procedure

The above two representations provide straightforward methods to simulate a variance Gamma model. One way will be to simulate the Gamma Lévy processes $\{X_+(t)\}_{0\leq t\leq T}$ and $\{X_-(t)\}_{0\leq t\leq T}$ of (7.14) using the series representation method of Section 7.2.2. The other approach is to first generate random time change $\{U(t)\}_{0\leq t\leq T}$ of (7.12), and then construct a discrete skeleton from the increments $X(i\Delta t) - X((i-1)\Delta t)$, $i \geq 1$. The increments of X are simply simulated using normal random variables with mean and variances determined by the increments of U.

7.3.3 The numerical results

Notice that, from an algorithmic point of view, the estimation for the variance Gamma model using penalized projection is not different from the estimation for the Gamma process. We can simply estimate both tails of the variance Gamma process separately. However, from the point of view of maximum likelihood estimation (MLE), the problem is numerically challenging. Even though the marginal density functions have closed form expressions (see [16]), there are well-documented issues with MLE (see for instance [29]). The likelihood function is highly flat for a wide range of parameters and good starting values as well as convergence are critical. Also, the separation of parameters and the identification of the variance Gamma process from other classes of the generalized hyperbolic Lévy processes is difficult. In fact, difference between subclasses in terms of likelihood is small. It is important to mention that these issues worsen when dealing with "high-frequency" data.

Let us consider a numerical example motivated by the empirical findings of [16] based on daily returns on the S&P stock index from January 1992 to September 1994 (see their Table I). Using maximum likelihood methods, the annualized estimates of the parameters for the variance Gamma model were reported to be $\hat{\theta} = -0.00056256$, $\hat{\sigma}^2 = 0.01373584$, and $\hat{\nu} = 0.002$, from where we obtain $\hat{\alpha} = 500$, $\hat{\beta}_+ = 0.0037056$, and $\hat{\beta}_- = 0.0037067$. Figures 10 and 11 show respectively the left- and right- tails of the Lévy density and their penalized projection estimators as well as their corresponding best-fit variance Gamma Lévy densities using a least-square method, and their marginal probability density functions (pdf) scaled by $1/\Delta t$ (the reciprocal of the time span between observations). The estimation was based on 5000 simulated increments with Δt equal to one-eigth of a day. The figures seem quite comforting. To get a better picture, Figures 16 and 17 show the sampling distributions of the estimates for α_- and β_+ obtained from applying the least-square method to the penalized proyection estimators. The histograms are based on 1000 samples of size 5000 with $\Delta t = 1/8$ of a day. This experiment shows clear, though not critical, underestimation of the parameter α and overestimation of the parameters β's. A simple method of moments (based on the first four moments) yields better results (see Figures 18 and 19). Nonparametric methods are not free-lunches and usually the gain in robustness is paid by a lost in precision.

8 Concluding Remarks

- In the present paper we have developed a new methodology for the estimation of the Lévy density of a Lévy process. Our methods are quite flexible in the sense that different type of estimating functions can be used; for instance, histograms, splines, trigonometric polynomials, and wavelets. The estimation is model free, easily implementable, and suitable for "high-frequency" data. We prove that, based on continuous-time data, our procedures enjoy good asymptotic properties. *Oracle inequalities* imply that, up to a constant, the procedure will achieve the best possible risk among the *projection estimators*. Moreover, it is proved that penalized projection estimators on splines achieve the optimal rate of convergence, from the minimax point of view, on some classes of smooth Lévy densities. Simulations show good results in Lévy models with infinite jump activity such as the variance Gamma model.

- Generalization of our procedures and results to some multivariate Lévy models can be readily obtained, since the results behind our construction have multivariate versions. Indeed, the Lévy-Itô decomposition of the sample paths, the concentration inequalities for compensated Poisson integrals, the inference theory for Locally Asymptotically Normal distributions, and the short-term properties of the marginal distributions are valid in the multivariate setting. More precisely, consider a Lévy process $\mathbf{X} = \{\mathbf{X}(t)\}_{t\geq 0}$ on $\mathbb{R}^d$ with Lévy measure ν. Assume that, on a window of estimation $D \in \mathcal{B}(\mathbb{R}^d\backslash\{0\})$, ν is absolutely continuous with respect to a reference measure η and that $s \equiv d\nu/d\eta$ is bounded with also $\int_D s^2(\mathbf{x})\eta(d\mathbf{x}) < \infty$. Then, given a finite-dimensional subspace $\mathcal{S}$ of $L^2\left((D,\eta)\right)$, the projection estimator of s on $\mathcal{S}$ is defined as in Section 2 with $\mathcal{J}$ being the Poisson measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated with the jumps of $\mathbf{X}$. Similarly, penalized projection estimators can be constructed, and the risk bound of Theorem 3.3, along with the Oracle inequality (3.8), are satisfied. The results of Sections 5 and 6 are valid as well. However, let us point out that further specifications of our methods for some semiparametric models are desirable. Important examples of these models include multivariate stable processes, and the tempered stable Lévy processes, recently introduced in [32].

- We have concentrated here on the estimation of the jump part of the Lévy process. It is natural to address the problem of estimating the continuous part too. In the one-dimensional case, this part is of the form $bt + \sigma W(t)$, where $\{W(t)\}_{t\geq 0}$ is a standard Brownian motion. In the multivariate case, it is characterized by a vector $\mathbf{b}$ and a symmetric nonnegative-definite matrix Σ. There are several approaches to deal with the estimation of Σ, from moment based methods to methods based on high-frequency data. A simple approach is to use the following functional limit:

$$\left\{\frac{1}{\sqrt{h}}\mathbf{X}(ht)\right\}_{t\geq 0} \xrightarrow{\mathfrak{D}} \{\mathbf{Y}(t)\}_{t\geq 0}, \quad h \to 0,$$

where $\{\mathbf{Y}(t)\}_{t\geq 0}$ is a centered Gaussian Lévy process with variance-covariance matrix Σ. This result can be deduced from the proof of the uniqueness of the Lévy-Khintchine representation as in pp. 40 of [34]. Another simple method will be to consider empirical versions of the moments:

$$\mathbb{E}\left[(X_i(t) - X_i(t))(X_k(t) - X_k(t))\right] = t\left(\Sigma_{i,k} + \int_{\mathbb{R}_0^d} x_i x_k \nu(d\mathbf{x})\right),$$

provided that $\int_{\|\mathbf{x}\|>1} \|\mathbf{x}\|^2 \nu(d\mathbf{x}) < \infty$ (see Section 25 [34]). Here, $X_i(t)$ and x_i refer to the i^{th} component of the vectors $\mathbf{X}(t)$ and $\mathbf{x}$, respectively, while $\Sigma_{i,j}$ is the (i,j) entry of Σ. The second term on the left hand side of the above expression can be estimated using our estimators for

ν. In the one-dimensional case, another approach is to use "threshold estimators" of the form:

$$\sum_{k=1}^{n} (\Delta_k X)^2 \, \mathbf{1}\left((\Delta_k X)^2 \le r(h)\right),$$

where $\Delta_k X \equiv X(t_k^n) - X(t_{k-1}^n)$ is the i^{th} increment of the process and $r(h)$ is an appropriate cutoff function (see [27] for details). For a class of semimartingales with finite jump activity, [6] provides other methodology based on the *bipower variation* (see also [7]). In the case of Lévy processes with finite jump activity, [1] disentangles the difussion from jumps using maximum likelihood and the *Generalized Method of Moments*. On the other hand, the estimation of the parameter $\mathbf{b}$ can be done by different methods. For instance, using the empirical version for

$$\mathbb{E}\left[\mathbf{X}(t)\right] = t\left(\mathbf{b} + \int_{\|\mathbf{x}\|>1} \mathbf{x}\nu(d\mathbf{x})\right),$$

valid if $\int_{\|\mathbf{x}\|>1} \|\mathbf{x}\|\nu(d\mathbf{x}) < \infty$. Another approach will be to estimate the "drift" $\mathbf{b}_0 \equiv \mathbf{b} - \int_{\|\mathbf{x}\|\le 1} \mathbf{x}\nu(d\mathbf{x})$ (where the integration is component-wise) using the fact that

$$\mathbb{P}\left[\lim_{h\to 0} \frac{1}{h}\mathbf{X}(h) = \mathbf{b}_0\right] = 1.$$

The above result holds true if $\int_{\|\mathbf{x}\|\le 1} \|\mathbf{x}\|\nu(d\mathbf{x}) < \infty$ and $\Sigma = 0$ (see [34]). Even though our methods are valid for non necessarily pure-jump Lévy processes, it is expected that the presence of a diffusion component will reduce the efficiency (in terms of speed of convergence and accuracy) of our estimators. It would be interesting to study in greater detail this phenomenon.

9 Main Proofs

9.1 Proof of the risk Bound

We will break the proof of Theorem 3.3 into several preliminary results.

Lemma 9.1 *For any penalty function* $\mathrm{pen} : \mathcal{M} \to [0, \infty)$ *and any* $m \in \mathcal{M}$*, the penalized projection estimator* $\tilde{s}$ *satisfies*

$$\|s - \tilde{s}\|^2 \leq \|s - s_m^\perp\|^2 + 2\chi_{\hat{m}}^2 + 2\nu_D\left(s_{\hat{m}}^\perp - s_m^\perp\right) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}), \quad (9.1)$$

where $\chi_m^2 \equiv \|s_m^\perp - \hat{s}_m\|^2$ *and where the functional* $\nu_D : L^2((D, \eta)) \to \mathbb{R}$ *is defined by*

$$\nu_D(f) \equiv \iint_{[0,T]\times D} f(x) \, \frac{\mathcal{J}(dt, dx) - s(x)\, dt\, \eta(dx)}{T}. \quad (9.2)$$

The general idea to deduce (3.5) is to bound the unattainable terms of the right hand side of (9.1) (namely $\chi_{\hat{m}}^2$ and $\nu_D\left(s_{\hat{m}}^\perp - s_m^\perp\right)$) by observable statistics. Then, the form of $\mathrm{pen}(\cdot)$ will be determined by this observable statistics so that the right hand side in (9.1) does not involve $\hat{m}$. To carry out this plan, we use concentration inequalities for $\chi_{\hat{m}}^2$ and for the compensated Poisson integrals $\nu_D(f)$. The following result gives a concentration inequality for general compensated Poisson integrals.

Proposition 9.2 *Let* N *be a Poisson process on a measurable space* $(\mathrm{V}, \mathcal{V})$ *with mean measure* μ *and let* $f : \mathrm{V} \to \mathbb{R}$ *be an essentially bounded measurable function satisfying* $0 < \int_{\mathrm{V}} f^2(v)\mu(dv)$ *and* $\int_{\mathrm{V}} |f(v)|\mu(dv) < \infty$*. Then, for any* $u > 0$*,*

$$\mathbb{P}\left[\int_{\mathrm{V}} f(v)(N(dv) - \mu(dv)) \geq \|f\|_{L^2(\mu)}\sqrt{2u} + \frac{1}{3}\|f\|_\infty u\right] \leq e^{-u}, \quad (9.3)$$

where $\|f\|_{L^2(\mu)}^2 = \int_{\mathrm{V}} f^2(v)\mu(dv)$*. In particular, if* $f : \mathrm{V} \to [0, \infty)$ *then, for any* $\epsilon > 0$ *and* $u > 0$*,*

$$\mathbb{P}\left[(1+\varepsilon)\left(\int_{\mathrm{V}} f(v)N(dv) + \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right) \geq \int_{\mathrm{V}} f(v)\mu(dv)\right] \geq 1 - e^{-u}. \quad (9.4)$$

For a proof of the inequality (9.3), see [30] (Proposition 7) or [21] (Corollary 5.1). Inequality (9.4) is a direct consequence of (9.3) (see Section 9.3 for a proof).

The next result allow us to bound the Poisson functional χ_m^2. This results is essentially Proposition 9 of [30].

Lemma 9.3 *Let N be a Poisson process on a measurable space $(\mathrm{V}, \mathcal{V})$ with mean measure $\mu(dv) = p(v)\eta(dv)$ and intensity function $p \in L^2(\mathrm{V}, \mathcal{V}, \eta)$. Let $\mathcal{S}$ be a finite dimensional subspace of $L^2(\mathrm{V}, \mathcal{V}, \eta)$ with orthonormal basis $\{\tilde{\varphi}_1, \ldots, \tilde{\varphi}_d\}$, and define*

$$\hat{p}(v) \equiv \sum_{i=1}^{d} \left(\int_{\mathrm{V}} \tilde{\varphi}_i(w) N(dw) \right) \tilde{\varphi}_i(v) \tag{9.5}$$

$$p^{\perp}(v) \equiv \sum_{i=1}^{d} \left(\int_{\mathrm{V}} p(w) \tilde{\varphi}_i(w) \eta(dw) \right) \tilde{\varphi}_i(v). \tag{9.6}$$

Then, $\chi^2(\mathcal{S}) \equiv \|\hat{p} - p^{\perp}\|^2_{L^2(\eta)}$ is such that for any $u > 0$ and $\varepsilon > 0$

$$\mathbb{P}\left[\chi(\mathcal{S}) \geq (1+\varepsilon)\sqrt{\mathbb{E}\left[\chi^2(\mathcal{S})\right]} + \sqrt{2k\mathrm{M}_{\mathcal{S}} u} + k(\varepsilon)\mathrm{B}_{\mathcal{S}} u \right] \leq e^{-u}, \tag{9.7}$$

where we can take $k = 6$, $k(\varepsilon) = 1.25 + 32/\varepsilon$, and where

$$M_{\mathcal{S}} \equiv \sup \left\{ \int_{\mathrm{V}} f^2(v) p(v) \eta(dv) : f \in \mathcal{S}, \|f\|_{L^2(\eta)} = 1 \right\}, \tag{9.8}$$

$$B_{\mathcal{S}} \equiv \sup \left\{ \|f\|_{\infty} : f \in \mathcal{S}, \|f\|_{L^2(\eta)} = 1 \right\}. \tag{9.9}$$

Following the same strategy as [30], the idea is to deduce a concentration inequality of the form

$$\mathbb{P}\left[\|s - \tilde{s}\|^2 \leq C\left(\|s - s_m^{\perp}\|^2 + \mathrm{pen}(m)\right) + h(\xi) \right] \geq 1 - C'e^{-\xi},$$

for constants C and C', and a function $h(\xi)$ (all independent of m). This will prove to be enough in view of the following result (see Section 9.3 for a proof).

Lemma 9.4 *Let $h : [0, \infty) \to \mathbb{R}_+$ be an strictly increasing function with continuous derivative and such that $h(0) = 0$ and $\lim_{\xi \to \infty} e^{-\xi} h(\xi) = 0$. If Z is random variable satisfying*

$$\mathbb{P}\left[Z \geq h(\xi) \right] \leq K e^{-\xi},$$

for every $\xi > 0$, then

$$\mathbb{E}Z \leq K \int_0^{\infty} e^{-u} h(u) du.$$

We are now in position to prove the main result of this section. Throughout the proof, we shall have to introduce various constants and inequalities that will hold with high probability. In order to clarify the role that the constants play in these inequalities, we shall make some conventions and give to the letters x, y, f, a, b, ξ, $\mathcal{K}$, c, and C, with various sub- or superscripts, special meaning. The letters with x are reserved to denote positive constants that can be chosen arbitrarily. The letters with y denote arbitrary constants greater than 1. $f, f_1, f_2, \ldots$ denote quadratic polynomials of a variable ξ whose coefficients (denoted by $a's$ and $b's$) are determined by the values of the $x's$ and $y's$. The inequalities will be true with probabilities greater that $1 - \mathcal{K}e^{-\xi}$, where $\mathcal{K}$ is determined by the values of the $x's$ and the $y's$. Finally, $c's$ and $C's$ are used for constants constrained by the $x's$ and $y's$. *It is important to remember that the constants in a given inequality are meant only for that inequality.* The pair of equivalent inequalities below will be repeatedly invoked through the proof:

$$\begin{array}{ll} \text{(i)} \;\; 2ab \leq xa^2 + \frac{1}{x}b^2, \quad \text{and} \\ \text{(ii)} \;\; (a+b)^2 \leq (1+x)\,a^2 + \left(1+\frac{1}{x}\right)b^2, \quad (\text{for } x > 0). \end{array} \tag{9.10}$$

Proof of Theorem 3.3: We consider successive improvements of the inequality (9.1):

Inequality 1: For any positive constants x_1, x_2, x_3, and x_4, there is a positive number $\mathcal{K}$ and an increasing quadratic function $f(\xi)$ (both independent of the family of linear models and of T) such that, with probability larger than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{array}{rl} \|s - \tilde{s}\|^2 & \leq \|s - s_m^{\perp}\|^2 + 2\chi^2_{\hat{m}} + 2x_1\|s^{\perp}_{\hat{m}} - s^{\perp}_m\|^2 \\ & \quad + x_2\frac{D_{\hat{m}}}{T} + x_3\frac{D_m}{T} + x_4\frac{d_{\hat{m}}}{T} \\ & \quad + \operatorname{pen}(m) - \operatorname{pen}(\hat{m}) + \frac{f(\xi)}{T}. \end{array} \tag{9.11}$$

Verification: Let us find an upper bound for $\nu_D\left(s^{\perp}_{m'} - s^{\perp}_m\right)$, $m', m \in \mathcal{M}$. Since the operator ν_D defined by (9.2) is just a compensated integral with respect to a Poisson process with mean measure $\mu(dtdx) = dt\eta(dx)$, we can apply Proposition 9.2 to obtain that, for any $x'_{m'} > 0$, and with probability larger than $1 - e^{-x'_{m'}}$

$$\nu_D\left(s^{\perp}_{m'} - s^{\perp}_m\right) \leq \left\|\frac{s^{\perp}_{m'} - s^{\perp}_m}{T}\right\|_{L^2(\mu)} \sqrt{2x'_{m'}} + \frac{\|s^{\perp}_{m'} - s^{\perp}_m\|_\infty x'_{m'}}{3T}. \tag{9.12}$$

In that case, the probability that (9.12) holds for every $m' \in \mathcal{M}$ is larger than $1-\sum_{m'\in\mathcal{M}} e^{-x_{m'}}$ because $P(A\cap B) \geq 1-a-b$, whenever $P(A) \geq 1-a$ and $P(B) \geq 1-b$. Clearly,

$$\left\|\frac{s^{\perp}_{m'} - s^{\perp}_{m}}{T}\right\|^2_{L^2(\mu)} = \iint\limits_{[0,T]\times D} \left(\frac{s^{\perp}_{m'}(x) - s^{\perp}_{m}(x)}{T}\right)^2 s(x)dt\eta(dx)$$
$$\leq \|s\|_\infty \frac{\|s^{\perp}_{m'} - s^{\perp}_{m}\|^2}{T}.$$

Using (9.10-i), the first term on the right hand side of (9.12) is then bounded as follows:

$$\left\|\frac{s^{\perp}_{m'} - s^{\perp}_{m}}{T}\right\|_{L^2(\mu)} \sqrt{2x'_{m'}} \leq x_1\|s^{\perp}_{m'} - s^{\perp}_{m}\|^2 + \frac{\|s\|_\infty x_{m'}}{2Tx_1}, \tag{9.13}$$

for any $x_1 > 0$. Using (3.4) and (9.10-i),

$$\begin{aligned}
\|s^{\perp}_{m'} - s^{\perp}_{m}\|_\infty x'_{m'} &\leq \left(\|s^{\perp}_{m'}\|_\infty + \|s^{\perp}_{m}\|_\infty\right) x'_{m'} \\
&\leq \left(\sqrt{D_{m'}}\|s^{\perp}_{m'}\| + \sqrt{D_m}\|s^{\perp}_{m}\|\right) x_{m'} \\
&\leq \sqrt{D_{m'}}\|s\|x'_{m'} + \sqrt{D_m}\|s\|x'_{m'} \\
&\leq 3x_2 D_{m'} + 3x_3 D_m + \frac{\|s\|^2 x'^2_{m'}}{12}\left(\frac{1}{x_2} + \frac{1}{x_3}\right),
\end{aligned}$$

for all $x_2 > 0$, $x_3 > 0$. It follows that, for any $x_1 > 0$, $x_2 > 0$, and $x_3 > 0$,

$$\begin{aligned}
\nu_D\left(s^{\perp}_{m'} - s^{\perp}_{m}\right) \leq x_1\|s^{\perp}_{m'} - s^{\perp}_{m}\|^2 + x_2\frac{D_{m'}}{T} + x_3\frac{D_m}{T} \\
+ \frac{\|s\|_\infty x'_{m'}}{2Tx_1} + \frac{\|s\|^2 x'^2_{m'}}{36T\bar{x}},
\end{aligned}$$

where we set $\frac{1}{\bar{x}} = \frac{1}{x_2} + \frac{1}{x_3}$. Next, take

$$x'_{m'} \equiv x_4\sqrt{d_{m'}}\left(\frac{1}{\|s\|} \wedge \frac{1}{\|s\|_\infty}\right) + \xi.$$

Then, for any positive x_1, x_2, x_3, and x_4, there is a $\mathcal{K}$ and a function f such that, with probability greater than $1-\mathcal{K}e^{-\xi}$,

$$\begin{aligned}
\nu_D\left(s^{\perp}_{m'} - s^{\perp}_{m}\right) &\leq x_1\|s^{\perp}_{m'} - s^{\perp}_{m}\|^2 + x_2\tfrac{D_{m'}}{T} + x_3\tfrac{D_m}{T} \\
&\quad + \left(\tfrac{x_4^2}{18\bar{x}} + \tfrac{x_4}{2x_1}\right)\tfrac{d_{m'}}{T} + \tfrac{f(\xi)}{T}, \quad \forall m' \in \mathcal{M}.
\end{aligned} \tag{9.14}$$

Concretely,

$$
\begin{aligned}
f(\xi) &= \tfrac{\|s\|}{18\bar{x}}\xi^2 + \tfrac{\|s\|_\infty}{2x_1}\xi, \\
\mathcal{K} &= \Gamma \textstyle\sum_{n=1}^{\infty} n^R \exp\left(-\sqrt{n}x_4\left(\frac{1}{\|s\|} \wedge \frac{1}{\|s\|_\infty}\right)\right).
\end{aligned}
\tag{9.15}
$$

Here, we use the assumption of polynomial models (Definition3.1) to come up with the constant $\mathcal{K}$. Pluging (9.14) in (9.1), and renaming the coefficient of $d_{m'}/T$, we can corroborate inequality 1.

Inequality 2: For any positive constants $y_1 > 1$, x_2, x_3, and x_4, there are positive constants $C_1 < 1$, $C_1' > 1$, and $\mathcal{K}$, and a strictly increasing quadratic polynomial f (all independent of the class of linear models and T) such that with probability larger than $1 - \mathcal{K}e^{-\xi}$,

$$
\begin{aligned}
C_1\|s - \tilde{s}\|^2 \quad &\leq C_1'\|s - s_m^\perp\|^2 + y_1\chi^2_{\hat{m}} \\
&+x_2\tfrac{D_{\hat{m}}}{T} + x_3\tfrac{D_m}{T} + x_4\tfrac{d_{\hat{m}}}{T} \\
&+ \operatorname{pen}(m) - \operatorname{pen}(\hat{m}) + \tfrac{f(\xi)}{T}.
\end{aligned}
\tag{9.16}
$$

Moreover, if $1 < y_1 < 2$, then $C_1' = 3 - y_1$ and $C_1 = y_1 - 1$. If $y_1 \geq 2$, then $C_1' = 1 + 4x_1$ and $C_1 = 1 - 4x_1$, where x_1 is any positive constant related to f according to equation (9.15).

Verification: Let us combine the term on the left hand side of (9.11) with the first three terms on the right hand side. Using the triangle inequality followed by (9.10-ii),

$$
\|s^\perp_{\hat{m}} - s^\perp_m\|^2 \leq 2\|s - s^\perp_m\|^2 + 2\|s^\perp_{\hat{m}} - s\|^2.
$$

Then, since $\chi^2_{\hat{m}} = \|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2$, and $\|s^\perp_{\hat{m}} - s\|^2 = \|s - \hat{s}_{\hat{m}}\|^2 - \|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2$, it follows that

$$
\begin{aligned}
&\|s - s^\perp_m\|^2 + 2\chi^2_{\hat{m}} + 2x_1\|s^\perp_{\hat{m}} - s^\perp_m\|^2 - \|s - \tilde{s}\|^2 \\
&\leq (1 + 4x_1)\,\|s - s^\perp_m\|^2 + (2 - 4x_1)\,\|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2 \\
&\ + (4x_1 - 1)\,\|s - \tilde{s}\|^2,
\end{aligned}
$$

for every $x_1 > 0$. Then, for any $y_1 > 1$, there are positive constant C, $C_1' > 1$, and $C_1 < 1$ such that

$$
\begin{aligned}
&\|s - s^\perp_m\|^2 + 2\chi^2_{\hat{m}} + 2C\|s^\perp_{\hat{m}} - s^\perp_m\|^2 - \|s - \tilde{s}\|^2 \\
&\leq\ C_1'\|s - s^\perp_m\|^2 + y_1\chi^2_{\hat{m}} - C_1\|s - \tilde{s}\|^2.
\end{aligned}
\tag{9.17}
$$

Combining (9.11) and (9.17), we obtain (9.16).

Inequality 3: For any $y_2 > 1$ and positive constants x_i, $i = 2, 3, 4$, there exist positive numbers $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f_2(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K}_2 > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}_2 e^{-\xi}$,

$$\begin{aligned} C_1\|s - \tilde{s}\|^2 \quad & \leq C_1'\|s - s_m^{\perp}\|^2 \\ & \quad + y_2\tfrac{V_{\hat{m}}}{T} + x_2\tfrac{D_{\hat{m}}}{T} + x_3\tfrac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & \quad + x_4\tfrac{D_m}{T} + \mathrm{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{9.18}$$

Verification: We bound $\chi^2_{m'}$ using Lemma 9.3 with $\mathrm{V} = \mathbb{R}_+ \times D$ and $\mu(d\mathbf{x}) = s(x)dt\eta(dx)$. We regard the linear model $\mathcal{S}_m$ as a subspace of $L^2(\mathbb{R}_+ \times D, dt\eta(dx))$ with orthonormal basis $\left\{\frac{\varphi_{1,m}}{\sqrt{T}}, \ldots, \frac{\varphi_{d_m,m}}{\sqrt{T}}\right\}$. Recall that

$$\chi^2_m = \|s_m^{\perp} - \hat{s}_m\|^2 = \sum_{i=1}^{d}\left[\iint_{[0,T]\times D}\varphi_{i,m}(x)\frac{\mathcal{J}(dt,dx) - s(x)dt\eta(dx)}{T}\right]^2.$$

Then, with probability larger than $1 - \sum_{m'\in\mathcal{M}} e^{-x'_{m'}}$,

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + \sqrt{2k\mathrm{M}_{m'}x'_{m'}} + k(x_1)\mathrm{B}_{m'}x'_{m'}, \tag{9.19}$$

for every $m' \in \mathcal{M}$, where $B_{m'} = \sqrt{D_{m'}/T}$,

$$\begin{aligned} V_{m'} &\equiv \int_D\left(\sum_{i=1}^{d_m}\varphi^2_{i,m}(x)\right)s(x)\eta(dx), \quad \text{and} \\ M_{m'} &\equiv \sup\left\{\int_D f^2(x)s(x)\eta(dx) : f \in \mathcal{S}_{m'}, \|f\| = 1\right\}. \end{aligned} \tag{9.20}$$

Since $\int_D f^2(x)s(x)\eta(dx) \leq \|f\|_\infty\|s\|$, $M_{m'}$ is bounded above by $\|s\|\sqrt{D_{m'}}$. In that case, we can use (9.10-i) to obtain

$$\sqrt{2k\mathrm{M}_{m'}x'_{m'}} \leq x_2\sqrt{D_{m'}} + \frac{k\|s\|}{2x_2}x'_{m'},$$

for any $x_2 > 0$. On the other hand, by hypothesis $D_{m'} \leq T$, and (9.19) implies that

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + \left(\frac{k\|s\|}{2x_2} + k(x_1)\right)x'_{m'},$$

where the constants $x'_{m'}$ are chosen as

$$x'_{m'} = \frac{x_3\sqrt{d_{m'}}}{\frac{k\|s\|}{2x_2} + k(x_1)} + \xi.$$

Then, for any $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and $\xi > 0$,

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + x_3\sqrt{d_{m'}} + f_1(\xi), \tag{9.21}$$

with probability larger than $1 - \mathcal{K}_1 e^{-\xi}$, where

$$\begin{aligned} f_1(\xi) &= \left(\tfrac{k\|s\|}{2x_2} + k(x_1)\right)\xi, \\ \mathcal{K}_1 &= \Gamma \textstyle\sum_{n=1}^{\infty} n^R \exp\left(-\sqrt{n}x_3/\left(\tfrac{k\|s\|}{2x_2} + k(x_1)\right)\right). \end{aligned} \tag{9.22}$$

Squaring (9.21) and using (9.10-ii) repeatedly, we conclude that, for any $y > 1$, $x_2 > 0$, and $x_3 > 0$, there are both a constant $\mathcal{K}_1 > 0$ and a quadratic function of the form $f_2(\xi) = a\xi^2$ (independent of T, m', and the family of linear models) such that, with probability greater than $1 - \mathcal{K}_1 e^{-\xi}$,

$$\chi^2_{m'} \leq y\frac{V_{m'}}{T} + x_2\frac{D_{m'}}{T} + x_3\frac{d_{m'}}{T} + \frac{f_2(\xi)}{T}, \quad \forall m' \in \mathcal{M}. \tag{9.23}$$

Then, (9.18) immediately follows from (9.23) and (9.16).

Proof of (3.5) for case (c):
By the inequality (9.4), we can upper bound $V_{m'}$ by $\hat{V}_{m'}$ on an event of large probability. Namely, for every $x'_{m'} > 0$ and $x > 0$, with probability greater than $1 - \sum_{m'\in\mathcal{M}} e^{-x'_{m'}}$

$$(1 + x)\left(\hat{V}_{m'} + \left(\frac{1}{2x} + \frac{5}{6}\right)\frac{D_{m'}}{T}x'_{m'}\right) \geq V_{m'}, \ \ \forall m' \in \mathcal{M}, \tag{9.24}$$

(recall that $D_m = \|\sum_{i=1}^{d_m}\varphi^2_{i,m}\|_\infty$). Since by hypothesis $D_{m'} < T$, and choosing

$$x'_{m'} = x'd_{m'} + \xi, \ \ (x' > 0),$$

it is seen that for any $x > 0$ and $x_4 > 0$, there are a positive constant $\mathcal{K}_2$ and a function $f(\xi) = b\xi$ (independent of T and of the linear models) such that with probability greater than $1 - \mathcal{K}_2 e^{-\xi}$

$$(1 + x)\hat{V}_{m'} + x_4 d_{m'} + f(\xi) \geq V_{m'}, \ \ \forall m' \in \mathcal{M}. \tag{9.25}$$

Here, we get $\mathcal{K}_2$ from the Polynomial assumption on the class of models. Combining (9.25) and (9.18), it is clear that for any $y_2 > 1$, and positive x_i, $i = 1, 2, 3$, we can choose a pair of positive constants $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$

$$\begin{array}{ll} C_1\|s - \tilde{s}\|^2 & \leq C_1'\|s - s_m^\perp\|^2 \\ & +y_2\frac{\hat{V}_{\hat{m}}}{T} + x_1\frac{D_{\hat{m}}}{T} + x_2\frac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & +x_3\frac{D_m}{T} + \mathrm{pen}(m) + \frac{f(\xi)}{T}. \end{array} \tag{9.26}$$

Next, we take $y_2 = c$, $x_1 = c'$, and $x_2 = c''$ to cancel $-pen(\hat{m})$ in (9.26). By Lemma 9.4, it follows that

$$C_1\mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C_1'\|s - s_m^\perp\|^2 + \left(1 + \frac{x_3}{c'}\right)\mathbb{E}\left[\mathrm{pen}(m)\right] + \frac{C_1''}{T}. \tag{9.27}$$

Since m is arbitrary, we obtain the case (c) of (3.5).

Proof of (3.5) for case (a):

By Remark 3.2, we can bound $V_{m'}$, as given in (9.20), by $D_{m'}\rho$ (assuming that $\rho < \infty$). On the other hand, (9.4) implies that

$$(1 + x_1)\left(\frac{\mathcal{N}}{T} + \left(\frac{1}{2x_1} + \frac{5}{6}\right)\frac{\xi}{T}\right) \geq \rho, \tag{9.28}$$

with probability greater than $1 - e^{-\xi}$. Using these bounds for $V_{m'}$ and the assumption that $D_{m'} \leq T$, (9.18) reduces to

$$\begin{array}{ll} C_1\|s - \tilde{s}\|^2 & \leq C_1'\|s - s_m^\perp\|^2 \\ & +y\frac{D_{\hat{m}}\mathcal{N}}{T^2} + x_1\frac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ & +x_2\frac{D_m\mathcal{N}}{T^2} + \mathrm{pen}(m) + \frac{f(\xi)}{T}, \end{array} \tag{9.29}$$

which is valid with probability $1 - \mathcal{K}e^{-\xi}$. In (9.29), $y > 1$, $x_1 > 0$ and $x_2 > 0$ are arbitrary, while C_1, C_1', the increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ are determined by y, x_1, and x_2 independently of the family of linear models and of T. We point out that we divided and multiplied by ρ the terms $D_{\hat{m}}/T$ and D_m/T in (9.18), and then applied (9.28) to get (9.29). It is now clear that $y = c$, and $x_1 = c'$ will produce the desired cancelation.

Proof of (3.5) for case (b):
We first upper bound $D_{\hat{m}}$ by $\beta^{-1}V_{\hat{m}}$ and $d_{\hat{m}}$ by $(\beta\phi)^{-1}V_{\hat{m}}$ in the inequality (9.18):

$$\begin{aligned} C_1\|s-\tilde{s}\|^2 \; &\leq C_1'\|s-s_m^{\perp}\|^2 + (y + x_1\beta^{-1} + x_2(\beta\phi)^{-1})\,\tfrac{V_{\hat{m}}}{T} \\ &\quad -\text{pen}(\hat{m}) + x_3\beta^{-1}\tfrac{V_m}{T} + \text{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{9.30}$$

Then, using $d_{m'} \leq (\beta\phi)^{-1}V_{m'}$ in (9.25) and letting $x_4(\beta\phi)^{-1}$ vary on $(0,1)$, we verify that for any $x' > 0$, a positive constant $\mathcal{K}_4$ and a polynomial f can be found so that with probability greater than $1 - \mathcal{K}_4 e^{-\xi}$,

$$(1+x')\hat{V}_{m'} + f(\xi) \geq V_{m'}, \;\; \forall m' \in \mathcal{M}. \tag{9.31}$$

Putting together (9.31) and (9.30), it is clear that for any $y > 1$ and $x_1 > 0$, we can find a pair of positive constants $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} C_1\|s-\tilde{s}\|^2 \; &\leq C_1'\|s-s_m^{\perp}\|^2 + y\tfrac{\hat{V}_{\hat{m}}}{T} - \text{pen}(\hat{m}) \\ &\quad + x_1\tfrac{V_m}{T} + \text{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{9.32}$$

In particular, by taking $y = c$, the term $-pen(\hat{m})$ cancels out. Lemma 9.4 implies that

$$C_1\mathbb{E}\left[\|s-\tilde{s}\|^2\right] \leq C_1'\|s-s_m^{\perp}\|^2 + (1+x_1)\,\mathbb{E}\left[\text{pen}(m)\right] + \frac{C_1''}{T}. \tag{9.33}$$

Finally, (3.5) (b) follows since m is arbitrary. $\square$

Remark 9.5 *Let us analyze more carefully the values that the constants C and C' can take in the inequality (3.5). For instance, consider the penalty function of part (c). As we saw in (9.27), the constants C and C' are determined by C_1, C_1', C_1'', and x_3. The constant C_1 was proved to be $y_1 - 1$ if $1 < y_1 < 2$, while it can be made arbitrarily close to one otherwise (see the comment immediately after (9.16)). On the other hand, y_1 itself can be made arbitrarily close to the penalization parameter c since $c = y_2 = y_1(1+x)y$, where x is as in (9.24) and y is in (9.23). Then, when $c \geq 2$, C_1 can be made arbitrarily close to one at the cost of increasing C_1'' in (9.27). Similarly, paying the same cost, we are able to select C_1' as close to one as we* *wish and x_3 arbitrarily small. Therefore, it is possible to find for any $\varepsilon > 0$, a constant $C'(\varepsilon)$ (increasing in ε) so that*

$$\mathbb{E}\|s - \tilde{s}\|^2 \leq (1+\varepsilon) \inf_{m \in \mathcal{M}} \left\{ \|s - s_m^{\perp}\|^2 + \mathbb{E}\left[\text{pen}(m)\right] \right\} + \frac{C'(\varepsilon)}{T}. \tag{9.34}$$

A more thorough inspection shows that

$$\lim_{\varepsilon \to 0} C'(\varepsilon)\varepsilon = K,$$

where K depends only c, c', c'', Γ, R, $\|s\|$, and $\|s\|_\infty$. The same reasoning apply to the other two types of penalty functions when $c \geq 2$. In particular, we point out that $\tilde{C}$ can be made arbitrarily close to 2 in the Oracle inequality (3.8) at the price of having a large $\tilde{C}'$ constant.

9.2 Proof of the minimax results

Proof of Theorem 4.1:
Fix a Lévy density $s_0 \in \Theta_\alpha(L/2; [a,b])$ such that $s_0(x) > 0$, for all $x \in [a,b]$, and a constant $\kappa > 0$. Consider a bounded function $g : \mathbb{R} \to \mathbb{R}_+$, with compact support $\mathbb{K} \subset [-1,1]$, that meets (4.2) with $L/2$ (instead of L) for all $x_1, x_2 \in \mathbb{R}$, $g(0) > 0$, increasing for $x < 0$, and decreasing for $x > 0$. Moreover, the support and the maximum value of g are chosen small enough so that

$$s_0(x) - \kappa^{-\alpha} g\left(\kappa(x - x_0)\right) > 0, \quad \forall x \in [a,b],$$

and the support of $g\left(\kappa(x - x_0)\right)$ is contained in (a,b). Let us consider the parametric model

$$s_\theta(x) := s_0(x) + \theta T^{-\frac{\alpha}{2\alpha+1}} g\left(\kappa T^{\frac{1}{2\alpha+1}}(x - x_0)\right), \quad x \in \mathbb{R}_0,$$

parametrized by $\theta \in (-\kappa^{-\alpha}, \kappa^{-\alpha})$. Notice that the function s_θ is a Lévy density for any $T > 1$ and $|\theta| < \kappa^{-\alpha}$. Now, for $x_1, x_2 \in [a,b]$,

$$\begin{aligned}
|s_\theta^{(k)}(x_1) - s_\theta^{(k)}(x_2)| \leq{}& \left| s_0^{(k)}(x_1) - s_0^{(k)}(x_2) \right| + \\
& |\theta| \kappa^k T^{\frac{-\alpha+k}{2\alpha+1}} \left| g^{(k)}\left(\kappa T^{\frac{1}{2\alpha+1}}(x_1 - x_0)\right) - g^{(k)}\left(\kappa T^{\frac{1}{2\alpha+1}}(x_2 - x_0)\right) \right| \\
\leq{}& \frac{L}{2}|x_1 - x_2|^\beta + \frac{L}{2}|\theta|\kappa^{k+\beta} T^{\frac{-\alpha+k+\beta}{2\alpha+1}} |x_1 - x_2|^\beta \\
\leq{}& L|x_1 - x_2|^\beta,
\end{aligned}$$

implying that $s_\theta \in \Theta$ whenever $|\theta| < \kappa^{-\alpha}$.

Let $\mathcal{M}_0$ be the space of atomic measures on $[0,T] \times [a,b]$ and let $\mathbb{P}_\theta^{(T)}$ be the probability measure on $\mathcal{M}_0$ induced by those jumps of the Lévy process $\{X(t)\}_{0 \le t \le T}$ whose sizes lie on $[a,b]$ and where the Lévy density of the process is $s_{_\theta}$. In other words, $\mathbb{P}_\theta^{(T)}$ is the distribution of a Poisson process on $[0,T] \times [a,b]$ with mean measure $dts_{_\theta}(x)dx$. Using Theorem 1.3 of [25],

$$
\begin{aligned}
\frac{d\mathbb{P}_\theta^{(T)}}{d\mathbb{P}_0^{(T)}}(\xi) = \exp\Bigg\{ &\int_0^T\!\!\int_a^b \ln\left\{1 + \theta T^{-\frac{\alpha}{2\alpha+1}} s_0^{-1}(x) g\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\right\} \xi(dt,dx) \\
&- \theta T^{1-\frac{\alpha}{2\alpha+1}} \int_a^b g\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right) dx \Bigg\}.
\end{aligned}
$$

The goal is to prove the LAN (*local asymptotic normality*) property for the parametric model $\left\{\mathbb{P}_\theta^{(T)} : \theta \in (-\kappa^{-\alpha}, \kappa^{-\alpha})\right\}$ at $\theta = 0$ (see definition 2.1 of [25]). Now, define $R(u) = \ln(1+u) - u + \frac{u^2}{2}$. The right hand side of the above equation can be written as follows:

$$
\frac{d\mathbb{P}_\theta^{(T)}}{d\mathbb{P}_0^{(T)}}(\xi) = \exp\left\{\theta \Delta_{_T} - \frac{\theta^2}{2}\sigma_{_T}^2 + r_{_T}(\theta)\right\},
$$

where

$$
\begin{aligned}
\Delta_{_T} &= T^{-\frac{\alpha}{2\alpha+1}} \int_0^T\!\!\int_a^b s_0^{-1}(x) g\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right) \left[\xi(dt,dx) - s_0(x)dtdx\right], \\
\sigma_{_T}^2 &= T^{1-\frac{2\alpha}{2\alpha+1}} \int_a^b s_0^{-1}(x) g^2\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right) dx, \\
r_{_T}(\theta) &= -\frac{\theta^2}{2} T^{-\frac{2\alpha}{2\alpha+1}} \int_0^T\!\!\int_a^b s_0^{-2}(x) g^2\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right) \left[\xi(dt,dx) - s_0(x)dtdx\right] \\
&\quad + \int_0^T\!\!\int_a^b R\left(\theta T^{-\frac{\alpha}{2\alpha+1}} s_0^{-1}(x) g\left(T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\right) \xi(dt,dx).
\end{aligned}
$$

We want to prove that there are nomalizing constants $\varphi_T > 0$ such that

$$
\mathcal{L}_{\mathbb{P}_0^{(T)}}\left(\varphi_T \Delta_{_T}\right) \xrightarrow{\mathfrak{D}} \mathcal{N}(0,1), \;\; \varphi_T^2 \sigma_{_T}^2 \to 1, \;\; \text{and} \;\; r_{_T}(\theta) \xrightarrow{\mathbb{P}_0^{(T)}} 0
$$

as $T \to \infty$. To prove the first limit, we invoke the CLT for Poisson integrals by verifying the Liapunov condition (see Theorem 1.1 and Remark 1.2 of

[25]). For $T>1$, we have that

$$T^{-\frac{\alpha(2+\delta)}{2\alpha+1}}\int_0^T\int_a^b s_0^{-2-\delta}(x)g^{2+\delta}\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)(s_0(x))\,dxdt=$$
$$\kappa^{-1}T^{1-\frac{\alpha(2+\delta)}{2\alpha+1}-\frac{1}{2\alpha+1}}\int_{\mathbb{K}} s_0^{-1-\delta}(\kappa^{-1}T^{-\frac{1}{2\alpha+1}}u+x_0)g^{2+\delta}(u)\,du\overset{T\to\infty}{\longrightarrow}0.$$

Notice also that, for large enough T,

$$\begin{aligned}\mathrm{Var}\left(\Delta_T\right)=T^{-\frac{2\alpha}{2\alpha+1}}&\int_0^T\int_a^b s_0^{-2}(x)g^2\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)(s_0(x))\,dxdt=\\&=\kappa^{-1}T^{1-\frac{2\alpha}{2\alpha+1}-\frac{1}{2\alpha+1}}\int_{\mathbb{K}} s_0^{-1}(\kappa^{-1}T^{-\frac{1}{2\alpha+1}}u+x_0)g^2(u)\,du\\&\overset{T\to\infty}{\longrightarrow}\kappa^{-1}s_0^{-1}(x_0)\int_{\mathbb{K}}g^2(u)du.\end{aligned}$$

Then, $\mathcal{L}_{\mathbb{P}_0^{(T)}}(\Delta_T)\overset{\mathfrak{D}}{\to}\mathcal{N}(0,I_0^2)$ with $I_0^2:=\kappa^{-1}s_0^{-1}(x_0)\int_{\mathbb{K}}g^2(u)du$. In particular, we also prove that $\sigma_T^2\to I_0^2$. We now verify that $r_T(\theta)$ vanishes in probability. Notice that the first term of r_T has mean 0 and variance

$$\begin{aligned}&\frac{\theta^4}{4}T^{-\frac{4\alpha}{2\alpha+1}}\int_0^T\int_a^b s_0^{-4}(x)g^4\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)(s_0(x))dxdt\\&=\frac{\theta^4}{4}\kappa^{-1}T^{1-\frac{4\alpha}{2\alpha+1}-\frac{1}{2\alpha+1}}\int_{\mathbb{K}} s_0^{-4}(\kappa^{-1}T^{-\frac{1}{2\alpha+1}}u+x_0)g^4(u)\,du\overset{T\to\infty}{\longrightarrow}0.\end{aligned}$$

Then, the first term converges in probability to 0. Similarly, the second term of $r_T(\theta)$ converges to 0 in probability because its mean and variance both goes to 0. Indeed, using that $|R(u)|\le|u|^3/3$, the absolute value of its expectation satisfies

$$\begin{aligned}&\left|\int_0^T\int_a^b R\left(\theta T^{-\frac{\alpha}{2\alpha+1}}s_0^{-1}(x)g\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\right)(s_0(x))dxdt\right|\\&\le\frac{|\theta|^3}{3}T^{1-\frac{3\alpha}{2\alpha+1}}\int_a^b s_0^{-2}(x)g^3\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)dx\overset{T\to\infty}{\longrightarrow}0.\end{aligned}$$

A similar reasoning applies to the variance. Therefore, $\{\mathbb{P}_\theta^{(T)}\}_{\theta\in(-\kappa^{-\alpha},\kappa^{-\alpha})}$ is Locally Asymptotically Normal (LAN) at $\theta=0$ (with the normalizing constants $\varphi_T:=I_0^{-1}$). We are now in position of using the theory for LAN families (see [22] for the general theory and [25] for the case of Poisson processes). In particular, by (2.11) of [25], if for each $T > 0$, $\hat{\theta}_{T}$ is an arbitrary estimator of θ, based on the jumps of the Lévy process happening on or before T and with sizes in $[a, b]$, then

$$\liminf_{T\to\infty} \sup_{|\theta|<\kappa^{-\alpha}} \mathbb{E}_{\theta}\left[\ell_0\left(I_0\left(\hat{\theta}_{T}-\theta\right)\right)\right] \geq B, \tag{9.35}$$

where $B := \mathbb{E}\left[\ell_0(Z)\chi_{[|Z|<I_0\kappa^{-\alpha}/2]}\right]$ and $Z \sim \mathcal{N}(0, 1)$.

Now, for each $T > 0$, let $\hat{s}_{T}$ be an arbitrary estimator, based on the jumps of the Lévy process happening on or before T and with sizes in $[a, b]$. Clearly, $\hat{s}_{T}$ induces the estimator $\hat{\theta}_{T} := T^{\frac{\alpha}{2\alpha+1}} g^{-1}(0) \left(\hat{s}_{T}(x_0) - s_0(x_0)\right)$, and since $\theta = T^{\frac{\alpha}{2\alpha+1}} g^{-1}(0) \left(s_{\theta}(x_0) - s_0(x_0)\right)$, we can write

$$g(0)\left(\hat{\theta}_{T}-\theta\right) = T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_{T}(x_0) - s_{\theta}(x_0)\right).$$

If we take $\ell_0(u) := \ell\left(g(0)I_0^{-1}u\right)$, (9.35) becomes:

$$\begin{aligned} B &\leq \liminf_{T\to\infty} \sup_{|\theta|<\kappa^{-\alpha}} \mathbb{E}_{\theta}\left[\ell_0\left(I_0\left(\hat{\theta}_{T}-\theta\right)\right)\right] \\ &= \liminf_{T\to\infty} \sup_{|\theta|<\kappa^{-\alpha}} \mathbb{E}_{\theta}\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_{T}(x_0) - s_{\theta}(x_0)\right)\right)\right] \end{aligned}$$

Since $\{s_{\theta} : \theta \in (-k^{-\alpha}, k^{-\alpha})\} \subset \Theta$,

$$\liminf_{T\to\infty} \sup_{s\in\Theta} \mathbb{E}_{s}\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_{T}(x_0) - s(x_0)\right)\right)\right] \geq B, \tag{9.36}$$

where

$$B := 2^{-3/2}\pi^{-1/2}\int_{|z|<I_0\kappa^{-\alpha}/2} \ell(g(0)I_0^{-1}z)e^{-z^2/2}dz. \tag{9.37}$$

This implies (4.3) because the lower bound B does not depend on the family of estimators $\hat{s}_{T}$. Indeed, for each $\varepsilon > 0$, let $\hat{s}_{T}^{(\varepsilon)}$ be such that

$$\begin{aligned} &\sup_{s\in\Theta} \mathbb{E}_{s}\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_{T}^{(\varepsilon)}(x_0) - s(x_0)\right)\right)\right] \\ &< \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_{s}\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_{T}(x_0) - s(x_0)\right)\right)\right] + \varepsilon. \end{aligned}$$

Taking the $\liminf$ as $T \to \infty$ on both sides, we obtain (4.3) since ε is arbitrary. $\square$

Proof of Corollary 4.2:
We first notice that the proof of Theorem 4.1 can be modified so that (9.36) holds true even if x_0 is not fixed. That is, for any family of estimators $\hat{s}_{T}$ and points $x_{T} \in (a, b)$,

$$\liminf_{T\to\infty} \sup_{s\in\Theta} \mathbb{E}_s \left[\ell \left(T^{\frac{\alpha}{2\alpha+1}} \left(\hat{s}_{T}(x_{T}) - s(x_{T}) \right) \right) \right] \geq C, \tag{9.38}$$

for a constant $C > 0$, which is independent of the family of estimators and of the points. Indeed, we can construct a parametric model of the form

$$s_{\theta,T}(x) := s_0(x) + \theta T^{-\frac{\alpha}{2\alpha+1}} g_{T} \left(\kappa T^{\frac{1}{2\alpha+1}} (x - x_{T}) \right), \quad x \in \mathbb{R}_0,$$

where $|\theta| < \kappa^{-\alpha}$ and where g_{T} is as in the previous proof. Moreover, without loss of generality, $0 < \inf_T g_{T}(0) \leq \sup_T g_{T}(0) < \infty$, since s_0 is continuous and strictly positive on (a, b). Let $\mathbb{P}_\theta^{(T)}$ be the distribution of a Poisson process on $[0, T] \times [a, b]$ with mean measure $dt s_{\theta,T}(x) dx$. Following the same arguments as above, $\{\mathbb{P}_\theta^{(T)} : \theta \in (-\kappa^{-\alpha}, \kappa^{-\alpha})\}$ is Locally Asymptotically Normal (LAN) at $\theta = 0$ with the normalizing constants

$$\varphi_{T} := \kappa^2 \left(\int_{\mathbb{K}_T} s_0^{-1}(\kappa^{-1} T^{-\frac{1}{2\alpha+1}} u + x_{T}) g_{T}^2 (u) \, du \right)^{-2}$$

Observe that there is an $m > 0$ for which $\inf_T \varphi_T \geq m$. By (2.11) of [25], for any $\delta > 0$,

$$\liminf_{T\to\infty} \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta \left[\ell_0 \left(\varphi_{T}^{-1} \left(\hat{\theta}_{T} - \theta \right) \right) \right] \geq C, \tag{9.39}$$

where $C := \mathbb{E}\left[\ell_0(Z)\chi_{[|Z|<\delta/2]}\right]$ and $Z \sim \mathcal{N}(0, 1)$. Since $\varphi_{T} \geq m$ and $\ell_0(|y|)$ is increasing in y,

$$\liminf_{T\to\infty} \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta \left[\ell_0 \left(m^{-1} \left(\hat{\theta}_{T} - \theta \right) \right) \right] \geq C, \tag{9.40}$$

Now, take

$$\hat{\theta}_{T} := T^{\frac{\alpha}{2\alpha+1}} g_{T}^{-1}(0) \left(\hat{s}_{T}(x_{T}) - s_0(x_{T}) \right).$$

Since $\theta = T^{\frac{\alpha}{2\alpha+1}} g_{T}^{-1}(0)\left(s_{\theta,T}(x_{T}) - s_{0}(x_{T})\right)$,

$$\begin{aligned}\sup_{s\in\Theta} \mathbb{E}_s\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_T(x_T) - s(x_T)\right)\right)\right] &\geq \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta\left[\ell\left(g_T(0)\left(\hat{\theta}_T - \theta\right)\right)\right] \\ &\geq \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta\left[\ell\left(\tilde{m}\left(\hat{\theta}_T - \theta\right)\right)\right],\end{aligned}$$

where $\tilde{m} = \inf_T g_T(0)$. Taking $\liminf$ as $T \to \infty$, (9.38) is obtained with

$$C = 2^{-3/2}\pi^{-1/2}\int_{|z|<\delta/2} \ell(\tilde{m}\, m\, z)e^{-z^2/2}dz. \tag{9.41}$$

Finally, (4.4) can be deduced as follows. For each $\varepsilon > 0$, let $\hat{s}_T^{(\varepsilon)} \in \Theta$ and $x_T^{(\varepsilon)} \in (a,b)$ be such that

$$\begin{aligned}&\sup_{s\in\Theta} \mathbb{E}_s\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_T^{(\varepsilon)}\left(x_T^{(\varepsilon)}\right) - s\left(x_T^{(\varepsilon)}\right)\right)\right)\right] \leq \\ &\inf_{x\in(a,b)} \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_s\left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_T(x) - s(x)\right)\right)\right] + \varepsilon.\end{aligned}$$

Next, take the $\liminf$ as $T \to \infty$ on both sides above, and apply (9.38). Finally, let $\varepsilon \to 0$. □

Proof of Corollary 4.3:
Fix a measurable estimator $\hat{s}_T$ and a $s \in \Theta$. By Fubini's Theorem,

$$\mathbb{E}_s\left[\int_a^b \left(\hat{s}_T(x) - s(x)\right)^2 dx\right] = \int_a^b \mathbb{E}_s\left[\left(\hat{s}_T(x) - s(x)\right)^2\right] dx.$$

Now, for each $\varepsilon > 0$, there exists an $x_0^{(\varepsilon)} \in (a,b)$ satisfying

$$\frac{1}{b-a}\int_a^b \mathbb{E}_s\left[\left(\hat{s}_T(x) - s(x)\right)^2\right] dx \geq \mathbb{E}_s\left[\left(\hat{s}_T\left(x_0^{(\varepsilon)}\right) - s\left(x_0^{(\varepsilon)}\right)\right)^2\right] - \varepsilon.$$

Then,

$$\begin{aligned}\frac{1}{b-a}\sup_{s\in\Theta} \mathbb{E}_s\left[\int_a^b \left(\hat{s}_T(x) - s(x)\right)^2 dx\right] &\geq \sup_{s\in\Theta} \mathbb{E}_s\left[\left(\hat{s}_T(x_0^{(\varepsilon)}) - s(x_0^{(\varepsilon)})\right)^2\right] - \varepsilon \\ &\geq \inf_{x\in(a,b)} \sup_{s\in\Theta} \mathbb{E}_s\left[\left(\hat{s}_T\left(x\right) - s\left(x\right)\right)^2\right] - \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, (4.5) becomes a consequence of (4.4) with $\ell(u) = u^2$. □

9.3 Some additional proofs

Proof of Corollary 3.6: The idea is to estimate the bias and the penalized term in (3.5). Clearly, the dimension d_m of $\mathcal{S}_m^k$ is $m(k+1)$. Also, D_m is bounded by $(k+1)^2 m/(b-a)$ (see (7) in [12]), and

$$\mathbb{E}\left[\hat{V}_m\right] = \int_a^b \left(\sum_i \varphi_{i,m}^2(x)\right) s(x)dx \leq (k+1)m\|s\|_\infty,$$

since the $\varphi_{i,m}$'s are orthonormal. On the other hand, by Chapter 2 (10.1) in [17], if $s \in \mathcal{B}_\infty^\alpha(\mathbb{L}^p([a,b]))$, there is a polynomial $q \in \mathcal{S}_m^k$ such that

$$\|s-q\|_{\mathbb{L}^p} \leq c_{[\alpha]}|s|_{\mathcal{B}_\infty^\alpha(\mathbb{L}^p)}(b-a)^\alpha m^{-\alpha}.$$

Thus,

$$\|s-s_m^\perp\| \leq c_{[\alpha]}(b-a)^{\frac{1}{2}-\frac{1}{p}+\alpha}|s|_{\mathcal{B}_\infty^\alpha(\mathbb{L}^p)}m^{-\alpha}.$$

By (3.5)), there is a constant M (depending on C, c, c', c'', α, k, $b-a$, p, $|s|_{\mathcal{B}_\infty^\alpha(\mathbb{L}^p)}$, and $\|s\|_\infty$), for which

$$\mathbb{E}\left[\|s-\tilde{s}_T\|^2\right] \leq M \inf_{m\in\mathcal{M}_T}\left\{m^{-2\alpha}+\frac{m}{T}\right\}+\frac{C'}{T}.$$

It is not hard to see that, for large enough T, the infimum on the above right hand side is $O_\alpha(T^{-2\alpha/(2\alpha+1)})$ (where O_α means that the term depends only on α). Since M is monotone in $|s|_{\mathcal{B}_\infty^\alpha(\mathbb{L}^p)}$ and $\|s\|_\infty$, (3.10) is verified. □

Verification of Remark 3.2: Suppose that D_m is finite, and thus each $f \in S$, with $\|f\| = 1$ is bounded. It follows using Lagrange multipliers that, for each $x \in D$,

$$D(x) \equiv \sup\left\{\Big|\sum_{i=1}^{d_m} c_i\varphi_i(x)\Big|^2 : \sum_{i=1}^{d_m} c_i^2 = 1\right\} = \sum_{i=1}^{d_m}\varphi_i^2(x).$$

Since $D_m \geq D(x)$ for every $x \in D$, we obtain $D_m \geq \|\sum_{i=1}^{d_m}\varphi_i^2\|_\infty$. On the other hand, for every $\varepsilon > 0$, there are $b_1, \ldots, b_n$ satisfying $\sum_{i=1}^{d_m} b_i^2 = 1$ and an $x \in D$ such that

$$D_m - \varepsilon < \Big|\sum_{i=1}^{d_m} b_i\varphi_i(x)\Big|^2 \leq D(x) = \sum_{i=1}^{d_m}\varphi_i^2(x) \leq \Big\|\sum_{i=1}^{d_m}\varphi_i^2\Big\|_\infty.$$

Letting $\varepsilon \to 0$, it follows that $D_m = \left\| \sum_{i=1}^{d_m} \varphi_i^2 \right\|_\infty$. $\square$

Proof of Lemma 9.1: Clearly, γ_D as defined by (2.9) can be written as

$$\mathbb{E}\gamma_D(f) = \|f\|^2 - 2\langle f, s_D \rangle - 2\nu_D(f) = \|f - s_D\|^2 - \|s_D\|^2 - 2\nu_D(f).$$

By the very definition of $\tilde{s}$ as the penalized projection estimator and by Remark 2.2,

$$\gamma_D(\tilde{s}) + \mathrm{pen}(\hat{m}) \leq \gamma_D(\hat{s}_m) + \mathrm{pen}(m) \leq \gamma(s_m^\perp) + \mathrm{pen}(m),$$

for any $m \in \mathcal{M}$. Using the previous two equations:

$$\begin{aligned} \|\tilde{s} - s_D\|^2 &= \gamma_D(\tilde{s}) + \|s_D\|^2 + 2\nu_D(\tilde{s}) \\ &\leq \gamma(s_m^\perp) + \|s_D\|^2 + 2\nu_D(\tilde{s}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \\ &= \|s_m^\perp - s_D\|^2 + 2\nu_D(\tilde{s} - s_m^\perp) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}). \end{aligned}$$

Finally, notice that $\nu_D(\tilde{s} - s_m^\perp) = \nu_D(\tilde{s} - s_{\hat{m}}^\perp) + \nu_D(s_{\hat{m}}^\perp - s_m^\perp)$ and $\nu_D(\hat{s}_m - s_m^\perp) = \chi_m^2$. $\square$

Verification of inequality (9.4): Notice just that for any $a, b, \varepsilon > 0$:

$$a - \sqrt{2ab} - \frac{1}{3}b \geq \frac{a}{1+\varepsilon} - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right) b. \tag{9.42}$$

Evaluating the integral in (9.3) for $-f$, we can write

$$\mathbb{P}\left[\int_{\mathrm{X}} f(x)N(dx) \geq \int_{\mathrm{X}} f(x)\mu(dx) - \|f\|_\mu \sqrt{2u} - \frac{1}{3}\|f\|_\infty u\right] \geq 1 - e^{-u}.$$

Using that $\|f\|_\mu^2 \leq \|f\|_\infty \int_{\mathrm{X}} |f(x)|\mu(dx)$ and (9.42),

$$\mathbb{P}\left[\int_{\mathrm{X}} f(x)N(dx) \geq \frac{1}{1+\varepsilon}\int_{\mathrm{X}} f(x)\mu(dx) - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right] \geq 1 - e^{-u},$$

which is precisely inequality (9.4). $\square$

Proof of Lemma 9.4:
Let Z^+ be the positive part of Z. First,

$$\mathbb{E}[Z] \leq \mathbb{E}\left[Z^+\right] = \int_0^\infty \mathbb{P}[Z > x]dx.$$

Since h is continuous and strictly increasing, $\mathbb{P}[Z > x] \leq K \exp(-h^{-1}(x))$, where h^{-1} is the inverse of h. Then, changing variables to $u = h^{-1}(x)$,

$$\int_0^\infty \mathbb{P}[Z > x]dx \leq K \int_0^\infty e^{-h^{-1}(x)}dx = K \int_0^\infty e^u h'(u)du.$$

Finally, an integration by parts yields $\int_0^\infty e^u h'(u)du = \int_0^\infty h(u)e^{-u}du$. $\square$

Proof of Proposition 6.2:
From the orthonormality property,

$$\begin{aligned}
\mathbb{E}\left[\|\hat{s}_m^n - s_m^\perp\|^2\right] &= \sum_{i=1}^{d_m} \mathbb{E}\left[\left(\hat{\beta}_{i,m}^n - \beta_{i,m}\right)^2\right] \\
&= \sum_{i=1}^{d_m} \left\{\mathrm{Var}\left(\hat{\beta}_{i,m}^n\right) + \left(\mathbb{E}\left[\hat{\beta}_{i,m}^n\right] - \beta_{i,m}\right)^2\right\}.
\end{aligned}$$

By remark 5.2,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[I_n(\varphi_{i,m})\right] &= T \int_{\mathbb{R}_0} \varphi(x)s(x)\eta(dx), \\
\lim_{n\to\infty} \mathrm{Var}\left[I_n(\varphi_{i,m})\right] &= T \int_{\mathbb{R}_0} \varphi_{i.m}^2(x)s(x)\eta(dx).
\end{aligned}$$

Then, (6.5) is true from (2.14) and (2.15). The second statement in the proof is straightforward since

$$\mathbb{E}\left[\|\hat{s}_m^n - s\|^2\right] = \mathbb{E}\left[\|\hat{s}_m^n - s_m^\perp\|^2\right] + \|s_m^\perp - s\|^2.$$

$\square$

10 Figures

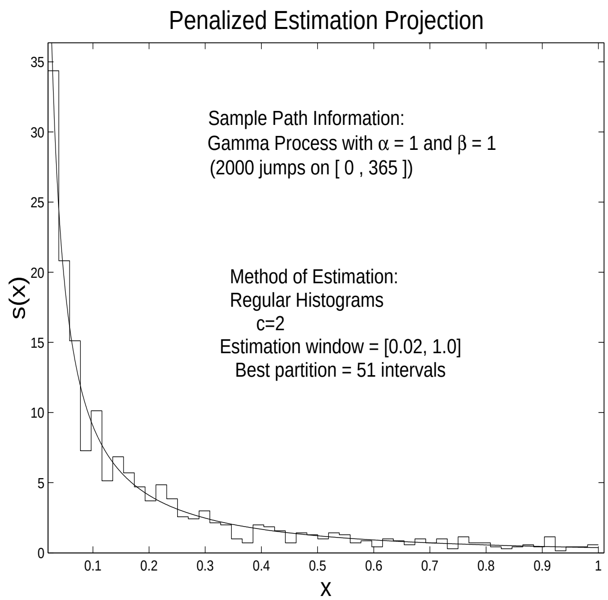

Figure 1: Penalized projection estimation of $\frac{e^{-x}}{x}$.

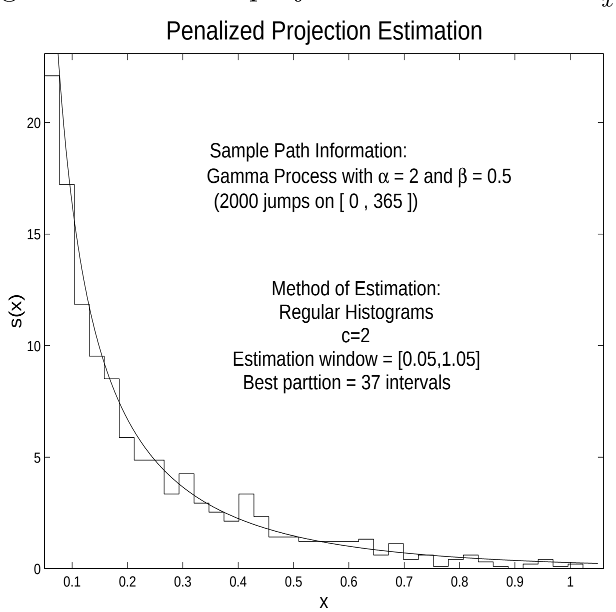

Figure 2: Penalized projection estimation of $\frac{2}{x}\exp(-2x)$.

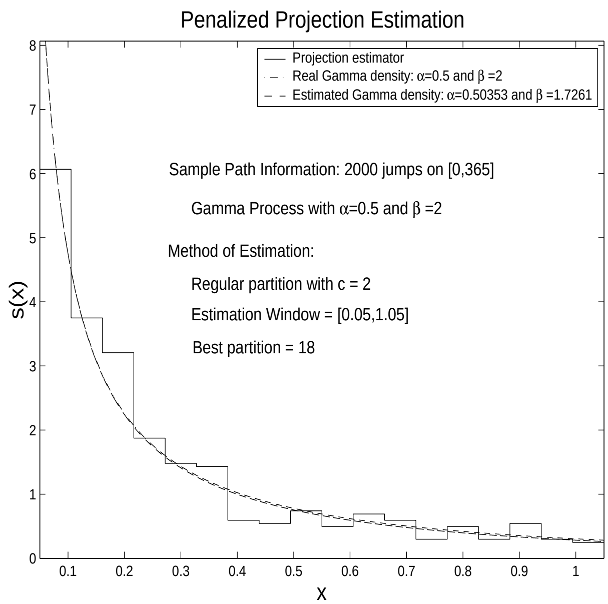

Figure 3: Penalized projection estimation of $\frac{1}{2x}\exp\left(-\frac{x}{2}\right)$.

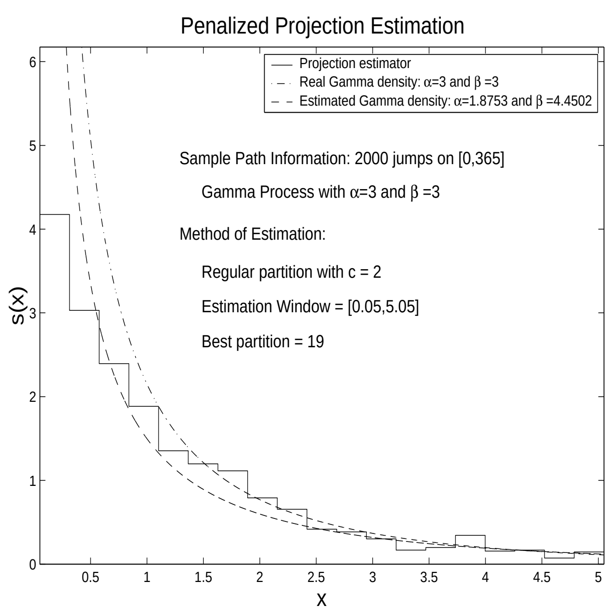

Figure 4: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$.

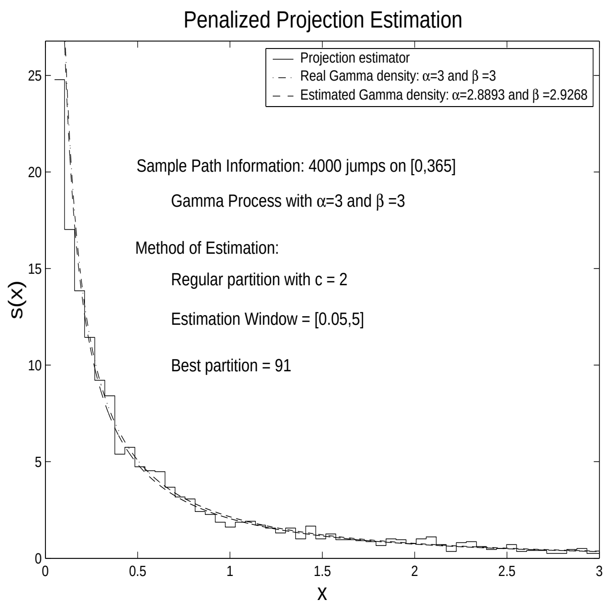

Figure 5: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$.

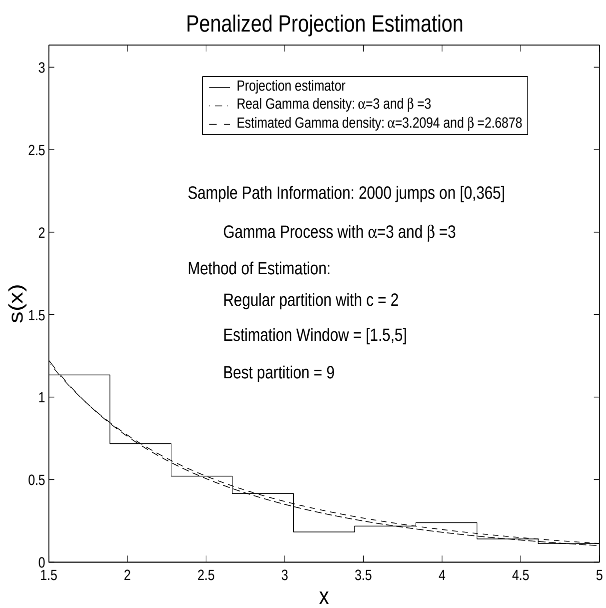

Figure 6: Penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$.

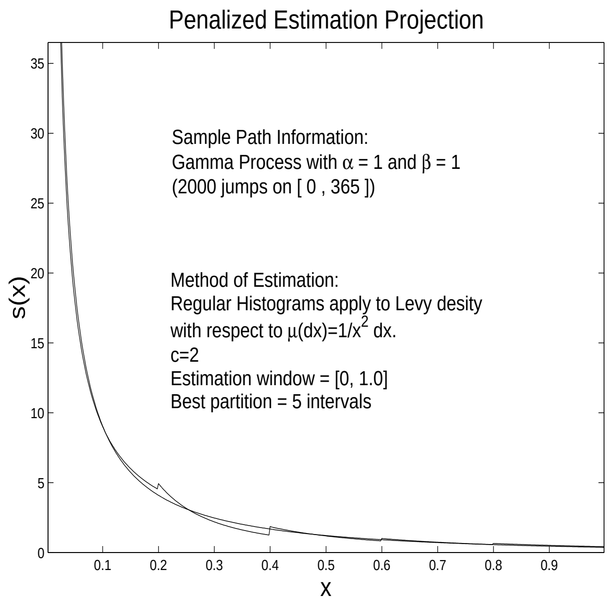

Figure 7: Regularized penalized projection estimation of $\frac{e^{-x}}{x}$.

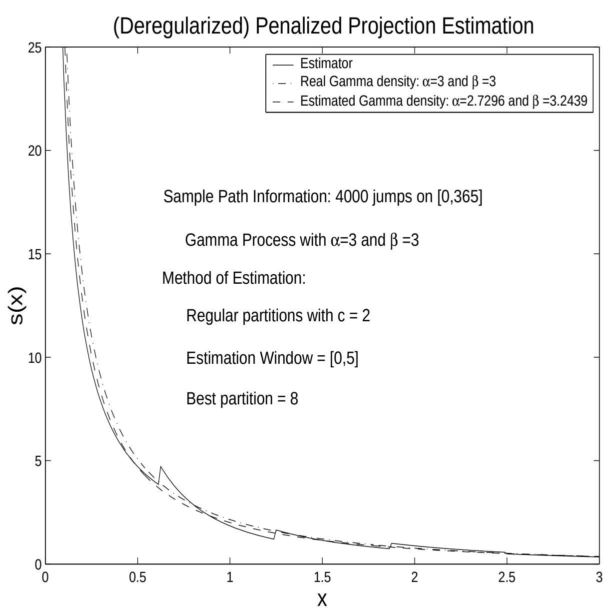

Figure 8: Regularized penalized projection estimation of $\frac{3}{x}\exp\left(-\frac{x}{3}\right)$.

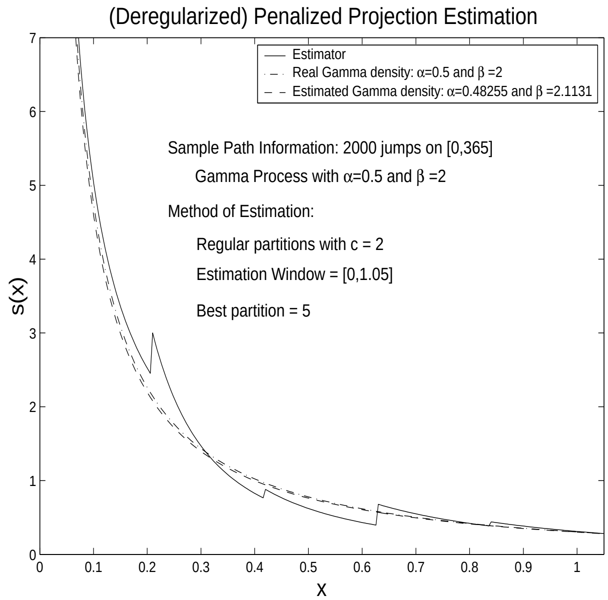

Figure 9: Regularized penalized projection estimation of $\frac{1}{2x}\exp\left(-\frac{x}{2}\right)$.

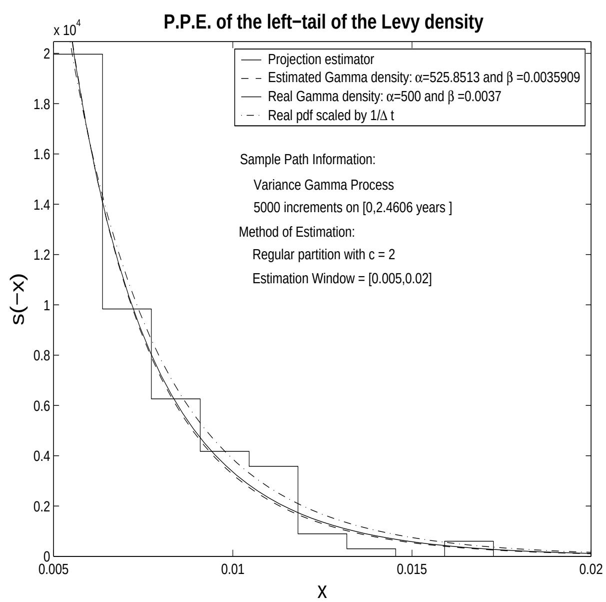

Figure 10: Penalized projection estimation of the left-tail of the variance Gamma Levy density.

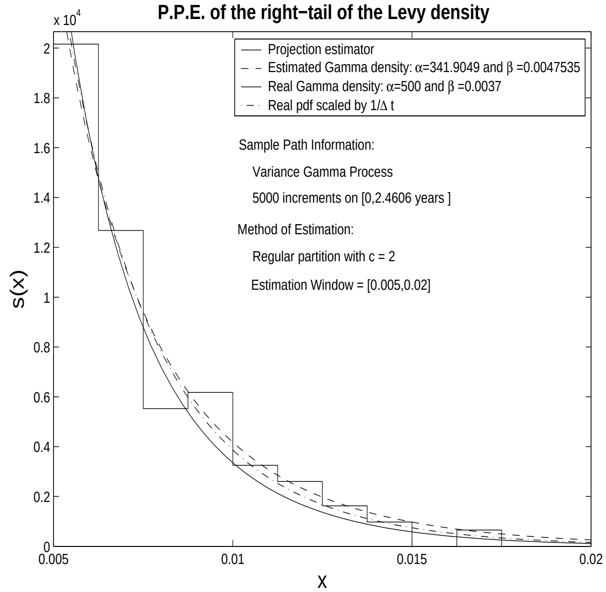

Figure 11: Penalized projection estimation of the right-tail of the variance Gamma Levy density.

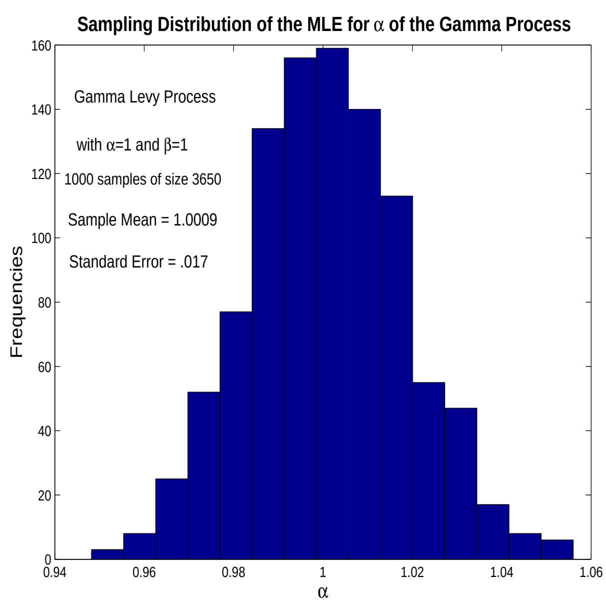

Figure 12: Sampling Distribution for the MLE of the α of the Gamma Lévy Process.

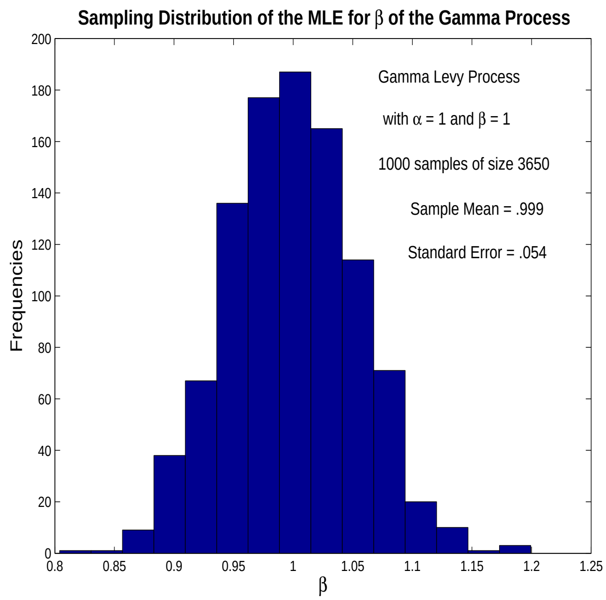

Figure 13: Sampling Distribution for the MLE of the β of the Gamma Lévy Process.

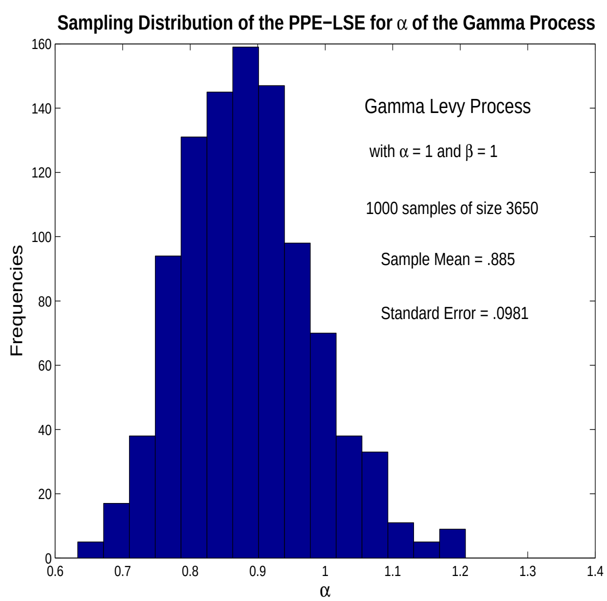

Figure 14: Sampling Distribution for the Estimates of the α of a Gamma Lévy process obtained from the PPE and the LSE method.

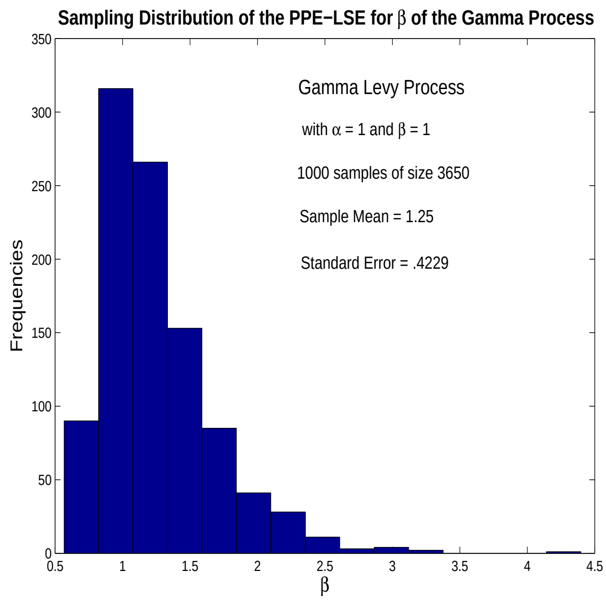

Figure 15: Sampling Distribution for the Estimates of the β of a Gamma Lévy process obtained from the PPE and the LSE method.

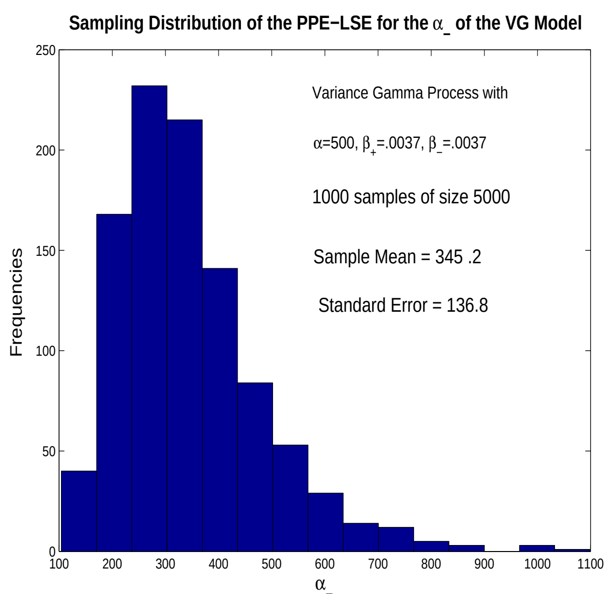

Figure 16: Sampling Distribution for the Estimates of α_- obtained from the PPE and the LSE method.

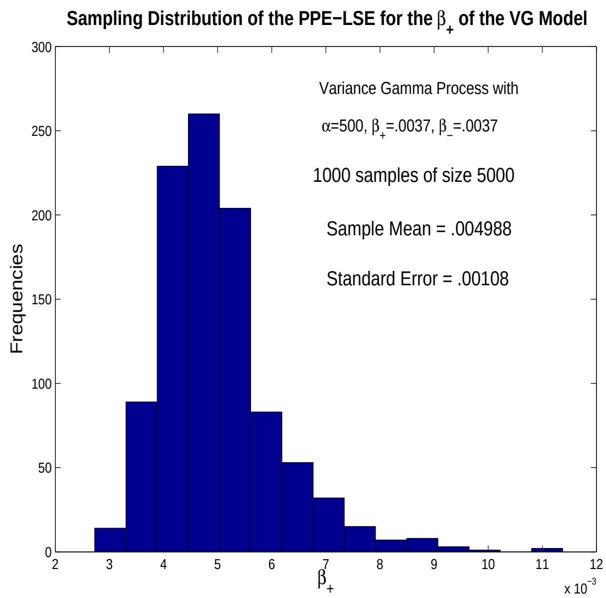

Figure 17: Sampling Distribution for the Estimates of β_+ obtained from the PPE and the LSE method.

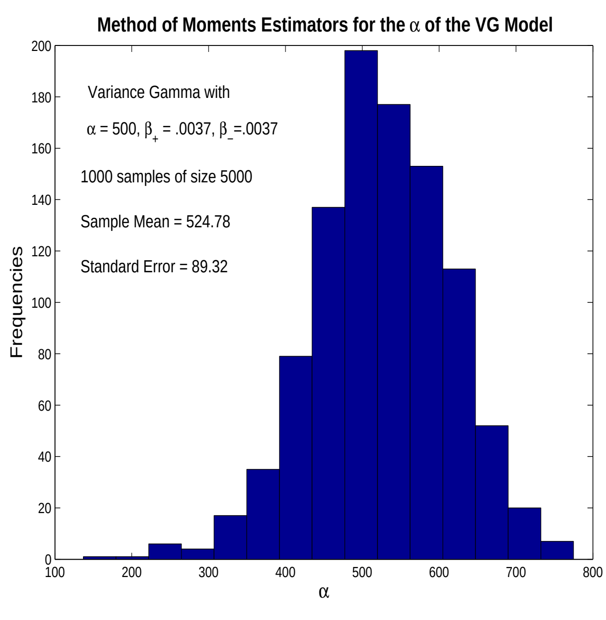

Figure 18: Sampling Distribution for the Estimator of α obtained by the Method of Moments.

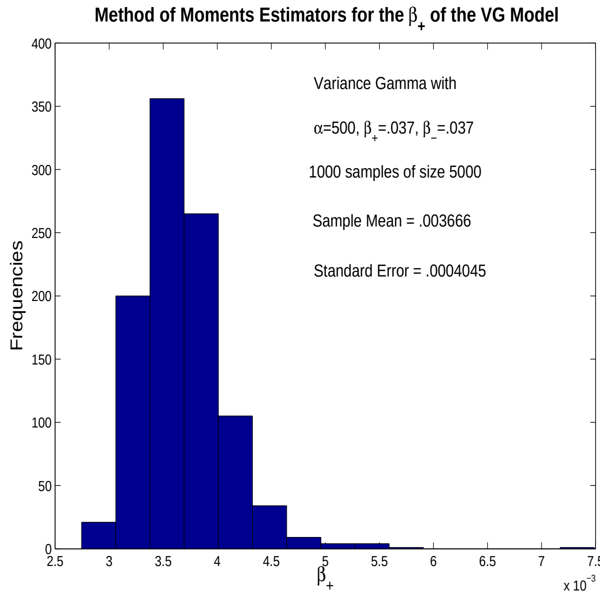

Figure 19: Sampling Distribution for the Estimator of β_+ obtained by the Method of Moments.

Acknowledgments: It is a pleasure to thank P. Reynaud-Bouret for very helpful discussions.

References

[1] Y. Aït-Sahalia. Disentangling difussion from jumps. *Journal of Financial Economics*, July 2004.

[2] H. Akaike. Information theory and an extension of maximum likelihood principle. *In Proceeding 2nd international symposium on information theory by P.N. Petrov & F. Csaki, eds*, pages 267–281, 1973.

[3] O.E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2:41–68, 1998.

[4] O.E. Barndorff-Nielsen and N. Shephard. Modelling by Lévy processess for financial economics. *Lévy Processes. Theory and Applications, by*

O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick, pages 283–318, 2001.

[5] O.E. Barndorff-Nielsen and N. Shephard. Normal modified stable processes. *Theory of Probability and Mathematical Statistics*, 65:1–19, 2001.

[6] O.E. Barndorff-Nielsen and N. Shephard. Econometrics of testing for jumps in financial economics using bipower variation. Technical Report W21, Economics Group, Nuffield College, University of Oxford, 2003.

[7] O.E. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1–48, 2004.

[8] A.R. Barron, L. Birgé, and P. Massart. Model selection via penalization. Technical Report 54, Université Paris-Sud, 1995.

[9] J. Bertoin. *Lévy processes*. Cambridge University Press, 1996.

[10] B.M. Bibby and M. Sorensen. Hyperbolic Processes in Finance. *Handbook of heavy tailed distributions in finance, by S. Rachev (ed.)*, pages 211–248, 2003.

[11] L. Birgé and P. Massart. Minimum contrast estimation on Sieves. Technical Report 34, Université Paris-Sud, 1994.

[12] L. Birgé and P. Massart. From model selection to adaptive estimation. *In Festschrift for Lucien Le Cam*, pages 55–87, 1997.

[13] L. Bondesson. *Generalized Gamma Convolutions and Related Classes of Disributions*. Springer, New York, 1992.

[14] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, April 2002.

[15] P. Carr, H. Geman, D. Madan, and M. Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13:345–382, July 2003.

[16] P. Carr, D. Madan, and E. Chang. The variance Gamma process and option pricing. *European Finance Review*, 2:79–105, 1998.

[17] R.A. DeVore and G.G. Lorentz. *Constructive Approximation*. Springer-Verlag, 1993.

[18] E. Eberlein. Application of Generalized Hyperbolic Lévy Motions to Finance. *Lévy Processes. Theory and Applications, by O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick*, pages 319–336, 2001.

[19] E. Eberlein and U. Keller. Hyperbolic Distribution in Finance. *Bernoulli*, 1:281–299, 1995.

[20] B.V. Gnedenko and A.N. Kolmogorov. *Foundations of Modern Probability*. Springer-Verlag, Berlin, New York, Heidelberg, 1997.

[21] C. Houdré and N. Privault. Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli*, 8(6):697–720, December 2002.

[22] I.A. Ibraginov and R.Z. Has'minskii. *Statistical estimation. Asymptotic Theory*. Springer-Verlag,Berlin, New York, Heidelberg, 1981.

[23] W. Jiang. *Some simulation-based models towards mathematical finance*. PhD thesis, University of Aarhus, April 2000.

[24] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag,Berlin, New York, Heidelberg, 1997.

[25] Y.A. Kutoyants. *Statistical Inference for Spatial Poisson Processes*. Springer, 1998.

[26] C.L. Mallows. Some comments on C_p. *Technometrics*, 15:661–675, 1973.

[27] C. Mancini. Estimating the integrated volatility in stochastic volatility models with Lévy type jumps. Technical report, Dipartimento di Matematica per le Decisioni, Universita di Firenze, July 2004. Presented at the Bachelier Finance Society, 3rd World Congress.

[28] D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

[29] K. Prause. *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. PhD thesis, University of Freiburg, October 1999.

[30] P. Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields*, 126(1):103–153, 2003.

[31] J. Rosiński. Series representations of Lévy processes from the perspective of point processes. *In Lévy processes-Theory and Applications*, pages 401–415, 2001.

[32] J. Rosiński. Tempered stable processes. Technical report, University of Tennessee, 2002.

[33] L. Rüschendorf and J. Woerner. Expansion of transition distributions of Lévy processes in small time. *Bernoulli*, 8:81–96, 2002.

[34] K. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, 1999.

[35] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.