IMS Lecture Notes–Monograph Series
Optimality: The Third Erich L. Lehmann Symposium
Vol. 57 (2009) 117–146

DOI: 10.1214/09-LNMS5709

Nonparametric Estimation for Lévy Models Based on Discrete-Sampling

José E. Figueroa-López[1,*]

Purdue University

Abstract: A Lévy model combines a Brownian motion with drift and a pure-jump homogeneous process such as a compound Poisson process. The estimation of the Lévy density, the infinite-dimensional parameter controlling the jump dynamics of the process, is studied under a discrete-sampling scheme. In that case, the jumps are latent variables whose statistical properties can in principle be assessed when the frequency of observations increase to infinity. We propose nonparametric estimators for the Lévy density following *Grenander's method of sieves*. The associated problem of selecting a suitable approximating sieve is subsequently investigated using regular piece-wise polynomials as sieves and assuming standard smoothness conditions on the Lévy density. By sampling the process at a high enough frequency relative to the time horizon T, we show that it is feasible to choose the dimension of the sieve so that the rate of convergence of the *risk of estimation off the origin* is the best possible from a minimax point of view, and even if the estimation were based on the whole sample path of the process. The sampling frequency necessary to attain the optimal minimax rate is explicitly identified. The proposed method is illustrated by simulation experiments in the case of variance Gamma processes.

Contents

[1]Department of Statistics, Purdue University, West Lafayette, IN 47906, e-mail: figueroa@stat.purdue.edu

[*]It is a pleasure to thank the Department of Statistics and Applied Probability at the University of California at Santa Barbara for their support while part of this research was conducted. The author is indebted to two anonymous referees for their many suggestions that improve it considerably.

AMS 2000 subject classifications: Primary 62G05; secondary 62G20.

Keywords and phrases: Lévy processes, nonparametric estimation, model selection.

1. Introduction

1.1. Motivation and Some Background

In the last decade, Lévy processes have received a great deal of attention, fueled by numerous applications in the area of mathematical finance, to the extend that Lévy processes have become a fundamental building block in the modeling of asset prices with jumps (see e.g. [11] and [30]). The simplest of these models postulates that the price of a commodity (say a stock) at time t is determined by

$$S_t := S_0 e^{X_t}, \tag{1.1}$$

where $X := \{X_t\}_{t\geq 0}$ is a Lévy process. Even this simple extension of the classical Black-Scholes model, in which X is simply a Brownian motion with drift, is able to account for some fundamental empirical features commonly observed in time series of asset returns such as heavy tails, high-kurtosis, and asymmetry. More recently, other Lévy based models have been proposed to account for more stylized features of stock prices. These models include exponential time-changed Lévy processes (cf. [7]-[9]), and stochastic differential equations driven by multivariate Lévy processes (cf. [1], [31]). Lévy processes, as models capturing some of the most important features of returns and as "first-order approximations" to other more accurate models, should be considered first in developing and testing a successful statistical methodology. However, even in such parsimonious models, there are several issues in performing statistical inference by standard likelihood-based methods.

A Lévy process is the "discontinuous sibling" of a Brownian motion. Concretely, $X = \{X_t\}_{t\geq 0}$ is a Lévy process if X has independent and stationary increments, its paths are right-continuous with left limits, and it has no fixed jump times. The later condition means that, for any $t > 0$,

$$\mathbb{P}\left[\Delta X_t \neq 0\right] = 0,$$

where $\Delta X_t := X(t) - \lim_{s\nearrow t} X_s$ is the magnitude of the "jump" of X at time t. It can be proved that the only Lévy process with continuous paths is essentially the Brownian motion $W := \{W_t\}_{t\geq 0}$ up to a drift term bt (hence, the well-known Gaussian distribution of the increments of W is a byproduct of the stationarity and independence of its increments). The only deterministic Lévy process is of the form $X_t := bt$, for a constant b. Another distinguished type of Lévy process is a compound Poisson process defined as

$$Y_t := \sum_{i=1}^{N_t} \xi_i, \tag{1.2}$$

where N is a homogeneous Poisson process and the random variables ξ_i, $i \geq 1$, are mutually independent from one another, independent from N, and with common distribution ρ. The process N dictates the jump times, which can occur "homogeneously" across time with an (average) intensity of λ jumps per unit time, while the sequence $\{\xi_i\}_{i\geq 1}$ determines the sizes of the jumps.

It turns out that the most general Lévy process is the superposition of a Brownian motion with drift, $\sigma W_t + bt$, a compound Poisson process, and the limit process resulting from making the jump intensity of a compensated compound Poisson process, $Y_t - \mathbb{E}\, Y_t$, to go to infinity while simultaneously allowing jumps of smaller sizes. The latter limiting process is governed by a measure ν such that the intensity of jumps is $\lambda_\varepsilon := \nu(\varepsilon \leq |x| < 1)$, the common distribution of the jump sizes is $\rho_\varepsilon(dx) := \mathbf{1}_{\{|x|\geq\varepsilon\}}\nu(dx)/\lambda_\varepsilon$, and the limit is when $\varepsilon \searrow 0$. For such a limit to converge to a "steady" process it must hold that

$$\int_{\{|x|<1\}} x^2 \nu(dx) < \infty.$$

The previous fundamental decomposition of a Lévy process is called the Lévy-Itô decomposition (see Section 19 in [29] for the details).

In summary, Lévy processes are determined by three "parameters": a non-negative real σ^2, a real b, and a measure ν on $\mathbb{R}\backslash\{0\}$ such that $\int (x^2 \wedge 1)\nu(dx) < \infty$. The measure ν controls the jump dynamics of the process X in that for any $A \in \mathcal{B}(\mathbb{R})$ whose indicator χ_A vanishes in a neighborhood of the origin,

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

for any $t > 0$ (see Section 19 of [29]). Thus, $\nu(A)$ gives the average number of jumps (per unit time) whose magnitudes fall in the set A. A common assumption in Lévy-based financial models is that ν is determined by a function $s : \mathbb{R}\backslash\{0\} \to [0, \infty)$, called the *Lévy density*, as follows

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

Intuitively, the value of s at x_0 provides information on the frequency of jumps with sizes "close" to x_0. In the case of the compound Poisson process (1.2), the Lévy measure is $\nu(dx) = \lambda\rho(dx)$.

By allowing a general Lévy process X in (1.1), instead of just a Brownian motion with drift as in the Black-Scholes model, one can incorporate two very appealing features: sudden changes in the price dynamics and some freedom in the distribution for the log return $\log\{S_t/S_s\} = X_t - X_s$. The possible distributions belong to the so-called class of *infinitely-divisible distributions*, a very rich class which include most known parametric families of distributions. We recall that an infinitely divisible distribution μ is characterized by the so-called *Lévy-Khinchin representation* of its characteristic function.

There are two key properties of a Lévy process that are exploited in this work. The first property relates ν with the short-term moments of X_t. Concretely, if φ is *ν-continuous, bounded, and vanishing in a neighborhood of the origin*, then

$$\lim_{\Delta\to 0} \frac{1}{\Delta}\, \mathbb{E}\, \varphi(X_\Delta) = \int \varphi(x)\nu(dx) = \int \varphi(x)s(x)dx. \tag{1.3}$$

The limiting relation (1.3) is straightforward when X is a compound Poisson process. A proof of (1.3) for a general Lévy process can be found in [29] (see his Corollary 8.9). Let us remark that (1.3) is also valid for certain unbounded functions φ, which does not necessarily vanish in a neighborhood of the origin, but rather converge to 0 at a proper rate (see [15] for more details). The second key property is related to the decomposition of X into two *independent* processes: one accounting for the "small" jumps and a compound Poisson process collecting the "big" jumps. Concretely, let

$$\widetilde{X}_t^{\varepsilon} := \sum_{s \le t} \Delta X_s \mathbf{1}_{\{|\Delta X_s| > \varepsilon\}},$$

be the piece-wise constant process associated with those jumps of X with sizes larger than ε. Then, $\widetilde{X}^{\varepsilon}$ is a compound Poisson process independent of $X - \widetilde{X}^{\varepsilon}$.

1.2. The Statistical Problems and Methodology

We are interested in estimating the Lévy density s on a window of estimation $D := [a, b] \subset \mathbb{R}\backslash\{0\}$, based on discrete observations of the process on a finite interval $[0, T]$. *We remark that the domain D is "separated" from the origin; that is to say, the estimation window D lies outside of a neighborhood of the origin.* If the whole path of the process were available (and hence, the jumps of the process would be available), the problem would be identical to the estimation of the intensity of a non-homogeneous Poisson process on a fixed time interval, say $[0, 1]$, based on T independent copies of the process. However, under discrete-sampling, the times and sizes of jumps are latent (unobservable) variables, whose statistical properties can be assessed when the frequency of observations increase to infinity at a certain speed relative to the time horizon. Hence, we will aim at determining the performance of our estimation method as both frequency and time horizon increase.

We adopt the so-called *method of sieves* originally proposed by [18] and implemented by Birgé, Massart, and others (see e.g. [3] & [5]) in several classical nonparametric problems such as density estimation and regression. This approach consists of the following general steps. First, choose a family of finite-dimensional *linear models* of functions, called *sieves*, with good approximation properties. Common sieves are splines, trigonometric polynomials, or wavelets. Second, specify a distance between functions relative to which the best approximation to s, in a given linear model, is going to be defined and characterized. Finally, devise an estimator, called the *projection estimator*, for the best approximation of s in the given linear model. It is important to point out that in principle there is no guarantee that the projection estimator will be nonnegative. In practice, one barely faces this problem when working with a large sample size, which is exactly the situation when nonparametric methods are recommended.

A linear model has the generic form

$$\mathcal{S} := \{\beta_1 \varphi_1 + \cdots + \beta_d \varphi_d : \beta_1, \ldots, \beta_d \in \mathbb{R}\}, \tag{1.4}$$

where $\varphi_1, \ldots, \varphi_d$ are given functions, typically taken to be orthonormal with respect to the inner product $\langle p, q \rangle := \int_D p(x) q(x) dx$. In the sequel, $\|\cdot\|$ stands for the associated norm $\langle \cdot, \cdot \rangle^{1/2}$ on $\mathbb{L}^2(D, dx)$. Relative to the distance induced by $\|\cdot\|$, the element of $\mathcal{S}$ closest to s, i.e. the *orthogonal projection* of s on $\mathcal{S}$, is

$$s^{\perp}(x) := \sum_{j=1}^{d} \nu(\varphi_j) \varphi_j(x), \tag{1.5}$$

where $\nu(\varphi_j) := \langle \varphi_j, s \rangle = \int \varphi_j(x) s(x) dx$. Then, the method of sieves boils down to estimate the *orthogonal projection* (1.5) on an "adequate" sieve $\mathcal{S}$. The core problem in this paper is to determine what a good sieve is. A very large linear model $\mathcal{S}$ will allow to attain a close approximation to s, but will entail necessarily a high estimation variance as the result of the large number of coefficients β_i to be estimated. Therefore, an essential task, called *model selection*, consists of selecting a linear model $\mathcal{S}$ accomplishing a good tradeoff between the error of approximation (or mis-specification error) and the standard error of the estimation. Concretely, one wishes to minimize the risk of the estimator $\hat{s}$, which in turn can be decomposed into two antagonist terms as follows:

$$\mathbb{E}\,\|s-\hat{s}\|^2 = \|s-s^{\perp}\|^2 + \mathbb{E}\,\|s^{\perp}-\hat{s}\|^2. \tag{1.6}$$

The first term, called the *bias term*, accounts for the error of the approximation, while the second, called the *variance term*, accounts for the standard error of the estimation.

1.3. The Sieve Estimators and an Overview of Results

We assume that the Lévy process $\{X_t\}_{t\geq 0}$ is being sampled over a time horizon $[0,T]$ at discrete times $0 < t_1^n < \cdots < t_n^n = T$. In the sequel, $t_0^n := 0$, $\pi^n := \{t_k^n\}_{k=0}^n$, and $\bar{\pi}^n := \max_k\{t_k^n - t_{k-1}^n\}$, the so-called mesh of the partition. We shall sometimes drop the superscript n in π^n and t_i^n. The following statistics are the main building blocks of our estimators:

$$\hat{\beta}^{\pi^n}(\varphi) := \frac{1}{t_n}\sum_{k=1}^{n}\varphi\left(X_{t_k^n} - X_{t_{k-1}^n}\right). \tag{1.7}$$

In the case of a quadratic function $\varphi(x) = x^2$, $\sum_{k=1}^n \varphi(X_{t_k^n} - X_{t_{k-1}^n})$ is called the realized quadratic variation (or variance) of the process. Thus, the statistics (1.7) can be interpreted as the average realized φ-variation of the process per unit time based on the observations $X_{t_1^n},\ldots,X_{t_n^n}$.

To explain the motivation behind the estimator in (1.7), let us assume for now that the sampling observations are equally-spaced in time so that $\Delta_n := t_i^n - t_{i-1}^n = T/n$ for all i, and hence,

$$\mathbb{E}\,\{\hat{\beta}^{\pi^n}(\varphi)\} = \frac{1}{\Delta_n}\,\mathbb{E}\,\varphi\left(X_{\Delta_n}\right),$$
$$\mathrm{Var}\left\{\hat{\beta}^{\pi^n}(\varphi)\right\} = \frac{1}{T}\left(\frac{1}{\Delta_n}\,\mathbb{E}\,\varphi^2\left(X_{\Delta_n}\right)\right) - \frac{1}{n}\left(\frac{1}{\Delta_n}\,\mathbb{E}\,\varphi\left(X_{\Delta_n}\right)\right)^2.$$

In view of (1.3), it is now evident that

$$\lim_{n\to\infty}\,\mathbb{E}\,\{\hat{\beta}^{\pi^n}(\varphi)\} = \int\varphi(x)s(x)dx, \quad \text{and} \quad \lim_{T\to\infty}\sup_n \mathrm{Var}\left(\hat{\beta}^{\pi^n}(\varphi)\right) = 0, \tag{1.8}$$

if φ is *ν-continuous, bounded, and vanishing in a neighborhood of the origin.* In statistical terms, (1.8) means that the statistic $\hat{\beta}^{\pi^n}(\varphi)$ is an asymptotically unbiased estimator of $\int\varphi(x)s(x)dx$ with associated risk vanishing uniformly when time horizon T increases. The previous argument leads us to propose

$$\hat{s}^{\pi^n}(x) := \sum_{j=1}^{d}\hat{\beta}^{\pi^n}(\varphi_j)\varphi_j(x), \tag{1.9}$$

as a natural estimator for the orthogonal projection $s^{\perp}$ defined in (1.5). In view of (1.8), $\hat{s}^{\pi^n}$ is a "consistent" estimator for $s^{\perp}$, in the integrated mean-square sense, as both the time horizon $T = t_n^n$ and the sampling frequency n/t_n^n go to ∞. The general sampling case will be considered in Section 2 as well as other statistical properties. It is worth pointing out that $\hat{s}^{\pi^n}$ is independent of the specific orthonormal basis of $\mathcal{S}$ as it can be proved that $\hat{s}^{\pi^n}$ is the unique solution of the minimization problem

$$\min_{f\in\mathcal{S}} \gamma_{D}^{\pi}(f),$$

where $\gamma_{D}^{\pi^n} : L^2(D, dx) \to \mathbb{R}$ is given by

$$\gamma_{D}^{\pi^n}(f) \equiv -\frac{2}{t_n^n}\sum_{k=1}^{n} f(X_{t_k^n} - X_{t_{k-1}^n}) + \int f^2(x)dx. \tag{1.10}$$

In the literature of model selection (see e.g. [4] and [25]), $\gamma_D^{\pi^n}$ is called the *contrast function.*

Finding the best sieve $\mathcal{S}$ to estimate s, even if we stick with using the class of projection estimators in (1.9), is impossible because s is unknown. However, it is possible to select a reasonably good model under certain qualitative assumptions on the parameter s, typically expressed by requiring s to be a member of a certain class Θ of smooth functions. Concretely, suppose we are interested in selecting a good model out of a family of linear models $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$ (here, $\mathcal{M}$ is a suitable set of labels). Let $m^* := m^*(\pi)$ be the *optimal minimax element of* $\{\hat{s}_m\}_{m\in\mathcal{M}}$ *on* Θ, defined as

$$m^* := \operatorname{arginf}_{m\in\mathcal{M}}\Big\{\sup_{s\in\Theta} \mathbb{E}\,\|s-\hat{s}_m\|^2\Big\}.$$

By requiring certain conditions on Θ and by choosing a suitable family of sieves $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$, we can ensure that

$$\mathbb{E}\,\|s-\hat{s}_{m^*(\pi)}\|^2 \to 0 \tag{1.11}$$

as the mesh of the partition $\pi = \{t_k\}_{k\geq 1}$ vanishes and the time horizon $T := t_n$ goes to infinity. Our goal will be to select a linear model $\hat{m}(\pi) \in \mathcal{M}$ so that the projection estimator on this model, $\hat{s}_{\hat{m}(\pi)}$, "attains" the minimax rate of convergence in (1.11), in the sense that

$$\limsup \frac{\mathbb{E}\,\|s-\hat{s}_{\hat{m}(\pi)}\|^2}{\mathbb{E}\,\|s-\hat{s}_{m^*(\pi)}\|^2} < \infty, \tag{1.12}$$

where the limit is taken as $\bar{\pi} \to 0$ and $T \to \infty$. In order to be able to determine in a "simple" way the rate of convergence of $\hat{s}_{\hat{m}(\pi)}$, we shall control the sampling frequency, measured by $\bar{\pi}$, in terms of the time horizon T. It is intuitive that in general the sampling frequency will depends on how close the window of estimation D is to the origin (see Section 3.4). The limit result (1.12) and the rate of convergence of projection estimators for a certain class of smooth Lévy densities are addressed in Section 3.

In this paper, we will show that the rate of convergence that can be attained using projection estimation on sieves is actually the best possible among all feasible estimators, given the information available on s (namely, that s belongs to a certain class Θ of smooth functions), and even if the estimators were based on continuous-time sampling of the process. Concretely, define $\hat{s}_{T}^{*}$ be the minimax estimator,

$$\hat{s}_{T}^{*} := \arg\inf_{\hat{s}}\sup_{s\in\Theta} \mathbb{E}\,\|s-\hat{s}\|^2 < \infty,$$

where the infimum is over *all the estimators* $\hat{s}$ of s based on $\{X(t)\}_{0\le t\le T}$. Then, by sampling at a high enough frequency (relative to T), we can accomplish that

$$\limsup_{T\to\infty} \frac{\mathbb{E}\,\|s-\hat{s}_{\hat{m}}\|^2}{\mathbb{E}\,\|s-\hat{s}^*_{T}\|^2} < \infty.$$

The rate of convergence of the minimax estimator will be provided in Section 4.

Let us finish by pointing out that the model selection problem was already analyzed in Figueroa-López & Houdré (2006) using the statistics

$$\hat{\beta}^c(\varphi) := \frac{1}{t_n}\sum_{t\le t_n}\varphi\left(\Delta X_t\right), \tag{1.13}$$

which intrinsically required continuous-time sampling of the process to determine the jumps ΔX_t. In the cited paper, the statistics (1.7) were proposed as good proxies of (1.13). Indeed, convergence in distribution is not hard to check, but moreover, recently [20] prove that (1.7) converges in probability to (1.13) when $n\to\infty$ (for fixed T). To the best of our knowledge, an analysis of the model selection problem for Lévy densities, under discrete sampling schemes, has not been considered before the present work.

1.4. Outline

The paper is structured as follows. In Section 2, we introduce the estimators proposed in this paper and study some basic statistical properties. In particular, we prove a CLT for the estimator $\hat{\beta}^\pi(\varphi)$ of (1.7) centered at the inner product $\beta(\varphi) = \int \varphi(x)s(x)dx$. In Section 3, we describe how to control the risk of the projection estimators by imposing three conditions. First, the time horizon T should be large enough (compared to the complexity of the sieves). Second, the time span between consecutive observations should be small enough compared to the time horizon. Finally, the sieves should have good approximating properties in general classes of smooth functions. We show that by ensuring the three previous conditions and by suitably choosing the dimension of the sieve (in terms of the presumed smoothness of the function s), the rate of convergence of the risk is of order $O(T^{-2\alpha/(2\alpha+1)})$ as $T\to\infty$ provided that the parameter s has "degree of smoothness" α.

In Section 4, the minimax risk of estimation, defined by

$$\inf_{\hat{s}_T}\sup_{s\in\Theta} \mathbb{E}_{\,s}\|s-\hat{s}_{T}\|^2,$$

is studied. Here, the infimum is over all estimators $\hat{s}_{T}$ which can be computed from the *whole sample paths* of X on the interval $[0,T]$ and the supremum is over all Lévy densities in a class Θ of functions that are smooth in $D=[a,b]$. We found that the minimax risk converges at an order of $O(T^{-2\alpha/(2\alpha+1)})$, where α is a parameter that measures the smoothness of the functions on Θ. For instance, if s has d continuous derivatives in D, then $\alpha\ge d$. The rate of convergence of the estimation is faster when α increases. Sections 3 and 4 justify the claim of the abstract: "...we show that it is feasible to choose the dimension of the sieve so that the rate of convergence of the *risk of estimation off the origin* is the best possible from a minimax point of view, and even if the estimation were based on the whole sample path of the process".

In Section 5, we propose a data-driven selection method for the sieve. Instead of deciding the dimension of the sieve from a presumed degree of smoothness of s (as it was suggested in Section 3), we propose to choose the sieve that minimizes an unbiased estimator of the risk of the projection estimator corresponding to that sieve. Since the proposed estimator of the risk will require the knowledge of all jumps of X up to time T, we replace it by a natural discrete-based proxy, where the jumps ΔX_t are replaced by the increments $X_{t_k} - X_{t_{k-1}}$. Section 6 illustrates the statistical methods using simulation experiments in the case of a variance gamma Lévy model. We finish with an Appendix where some technical proofs are given.

2. First Properties at the Estimators

In this section, our goal is to survey some statistical properties of the estimators (1.7) and (1.9). We already mentioned a few of these in the case of *regular sampling*[1] and of bounded ν-continuous *test functions* φ vanishing in a neighborhood of the origin. In the framework of this paper, this kind of test functions indeed suffices to recover and estimate the Lévy density off the origin. Our first result is a simple application of the Central Limit Theorem (CLT) for independent random variables (cf. [20] [Theorem 3.2] for the case of regular sampling). In the following results, Z stands for a standard Normal random variable.

Proposition 2.1. *Let φ be ν-continuous, bounded, and such that $\varphi(x) = o(|x|)$, as $x \to 0$. Then,*

$$\sqrt{t_n}\left(\hat{\beta}^{\pi}(\varphi) - \mathbb{E}\,\hat{\beta}^{\pi}(\varphi)\right) \xrightarrow{\mathfrak{D}} \nu(\varphi^2)^{\frac{1}{2}}\, Z, \tag{2.1}$$

as $t_n \to \infty$ and $\bar{\pi} \to 0$.

Proof. Let $\Gamma_t(\varphi) := \mathbb{E}\,\varphi^2(X_t) - \{\mathbb{E}\,\varphi(X_t)\}^2$ and $\Delta_k := t_k - t_{k-1}$. We can write

$$\sqrt{t_n}\left(\hat{\beta}^{\pi}(\varphi) - \mathbb{E}\,\hat{\beta}^{\pi}(\varphi)\right) = \sum_{k=1}^{n} \xi_k^{\pi},$$

where $\xi_k^{\pi} = \frac{1}{\sqrt{t_n}}\{\varphi(X_{t_k} - X_{t_{k-1}}) - \mathbb{E}\,\varphi(X_{t_k - t_{k-1}})\}$. Under the assumption of this Proposition, it turns out that $\lim_{t\to 0} \frac{1}{t}\Gamma_t(\varphi) = \nu(\varphi^2)$ (see Lemma 5.5 in Jacod (2007)), and thus,

$$\bar{\sigma}_{n,\pi}^2 := \mathrm{Var} \sum_{k=1}^{n} \xi_k^{\pi} = \frac{1}{t_n}\sum_{k=1}^{n} \Gamma_{\Delta_k}(\varphi) \longrightarrow \nu(\varphi^2),$$

as the mesh $\bar{\pi} := \max_k\{t_k - t_{k-1}\} \to 0$. Due to the boundedness of φ, we have that, for $\bar{\pi}$ small enough,

$$\frac{|\xi_k^{\pi}|}{\bar{\sigma}_{n,\pi}} \le C\frac{1}{\sqrt{t_n}} \to 0,$$

as $t_n \to \infty$. Then, (2.1) follows from the Central Limit Theorem for independent random variables (see e.g. the Corollary following Theorem 7.1.2 in [10]). ☐

In order to provide an explicit centering in (2.1), we need to estimate the rate of convergence of the bias $\mathbb{E}\,\hat{\beta}^{\pi}(\varphi) - \nu(\varphi)$. Since

$$\mathbb{E}\,\hat{\beta}^{\pi}(\varphi) - \nu(\varphi) = \frac{1}{t_n}\sum_{k=1}^{n} \Delta_k \left\{\frac{1}{\Delta_k}\mathbb{E}\,\varphi(X_{\Delta_k}) - \nu(\varphi)\right\},$$

[1]Sampling equally spaced in time.

the problem is equivalent to analyzing the rate of convergence in (1.3). To achieve this goal, we need to impose some regularity on either the Lévy process or the moment functions φ. Following the second approach, [15] shed light on this problem for functions $\varphi \in C_b^2(\mathbb{R})$; namely, twice-continuously differentiable functions φ such that $\limsup_{|x|\to\infty} |\varphi^{(i)}(x)| < \infty$, for $i = 0, 1, 2$. Below, $*$ denotes the *convolution operator* $\nu_1 * \nu_2(\varphi) := \iint \varphi(x_1+x_2)\nu_1(dx_1)\nu_2(dx_2)$, and L denotes the *infinitesimal generator* of the process X (see e.g. Sato (1999)), which is known to be given by

$$(2.2)\quad (L\varphi)(x) := \frac{\sigma^2}{2}\varphi''(x)+b\varphi'(x)+\int \left(\varphi(y+x)-\varphi(x)-y\varphi'(x)\mathbf{1}_{\{|y|\le 1\}}\right)\nu(dy);$$

see Theorem 31.5 in [29] and also in Proposition 2.3 in [15]. The following result can be found in [15] (see Proposition 3.1), where a proof is provided for a certain class of unbounded functions φ:

Lemma 2.2. *If $\varphi \in C_b^2(\mathbb{R})$ vanishes in a neighborhood of the origin, then*

$$(2.3)\qquad \lim_{t\to 0}\frac{1}{t}\left\{\frac{1}{t}\ \mathbb{E}\,\varphi(X_t)-\nu(\varphi)\right\} = \nu_\varepsilon(L\varphi)-\frac{1}{2}\,\nu_\varepsilon * \nu_\varepsilon(\varphi),$$

where $\nu_\varepsilon(dx) := \mathbf{1}_{\{|x|\ge\varepsilon\}}\nu(dx)$.

The following is an easy consequence of the previous two results.

Theorem 2.3. *Under the assumptions of Lemma 2.2,*

$$(2.4)\qquad \sqrt{t_n}\left(\hat\beta^\pi(\varphi)-\nu(\varphi)\right) \stackrel{\mathfrak{D}}{\longrightarrow} \nu(\varphi^2)^{\frac{1}{2}}\,Z$$

as $t_n \to \infty$ and $\bar\pi \to 0$ so that $\bar\pi\sqrt{t_n} \to 0$.

Proof. It suffices to prove that

$$D_n := \sqrt{t_n}\left(\mathbb{E}\ \hat\beta^\pi(\varphi) - \int \varphi(x)\nu(dx)\right) \longrightarrow 0.$$

Writing $\Delta_k = t_k - t_{k-1}$ and using (2.3), for $\bar\pi = \max_k \Delta_k$ small enough, there exists a constant C such that

$$|D_n| \le \frac{1}{\sqrt{t_n}}\sum_{k=1}^n \Delta_k\left|\frac{1}{\Delta_k}\ \mathbb{E}\,\varphi(X_{\Delta_k})-\nu(\varphi)\right| \le C\frac{1}{\sqrt{t_n}}\sum_{k=1}^n \Delta_k^2 \le C\bar\pi\sqrt{t_n}\to 0,$$

by assumption. □

Remark 2.4. As a direct consequence, it follows that $\hat\beta^\pi(\varphi)$ is a consistent estimator for $\nu(\varphi)$ as $t_n \to \infty$ and $\bar\pi \to 0$ so that $\bar\pi\sqrt{t_n} \to 0$. As a matter of fact, it suffices that $t_n \to \infty$ and $\bar\pi \to 0$, provided that e.g. f is ν-continuous, bounded, and $f(x) = o(|x|^2)$, as $x \to 0$. A proof of this statement is outlined in [32] for regular sampling observations, while the general case is considered in [14].

In view of the linearity of $\hat\beta^\pi(\cdot)$ and $\nu(\cdot)$, we conclude that

Corollary 2.5. *Let Ξ be the class of functions $\varphi \in C_0^2(\mathbb{R})$ that vanish in a neighborhood of the origin. Suppose that the linear model $\mathcal{S}$ in (1.4) is such that $\{\varphi_j\}_{j=1}^d \subseteq \Xi$. Then, the projection estimator $\hat s^\pi(x)$ in (1.9) satisfies the limiting relation*

$$(2.5)\qquad \sqrt{t_n}\left(\hat s^\pi(x)-s^\perp(x)\right) \stackrel{\mathfrak{D}}{\longrightarrow} V^{1/2}(x)\,Z$$

as $t_n \to \infty$ *and* $\bar{\pi} \to 0$ *so that* $\bar{\pi}\sqrt{t_n} \to 0$, *where* $V(x) := \int f^2(y)\nu(dy)$ *with* $f(y) := \sum_{j=1}^{d} \varphi_j(x)\varphi_j(y)$.

Remark 2.6. Notice that we have the following bound for the variance

$$V(x) \leq \|s\|_{\infty,D} \sum_{j=1}^{d} \varphi_j^2(x),$$

where $\|s\|_{\infty,D} := \sup_{y\in D} s(y)$.

We can relax the regularity conditions on the moment functions φ by using a simple integration by parts formula (see Remark 3.3 below). A different approach could be to impose additional regularity conditions on the Lévy process itself. In this direction, [28] studies series expansions for the transition density $p_t(x)$ of X_t as powers of t. For instance, one of their results states that if p_t is monotonically decreasing for $x > b$ and $x < -c$, for some $b, c > 0$, then for any $\eta > 0$, there exists $\varepsilon' > 0$ and $t_0 > 0$, such that

$$\frac{1}{t} p_t(x) = e^{-\int_{\{|y|>\varepsilon\}} s(y)dy} s(x) + O_{\varepsilon,\eta}(t), \tag{2.6}$$

for $|x| > \eta$. Such a result will allow us to estimate the rate of convergence in (1.3) if φ vanishes around the origin, since

$$\frac{1}{\Delta}\ \mathbb{E}\,\varphi(X_\Delta) - \nu(\varphi) = \int \varphi(x) \left\{ \frac{1}{\Delta} p_\Delta(x) - s(x) \right\} dx.$$

However, we should warn that the derivation of (2.6) in [28] is not completely formal[2], and hence, we avoid to use such an approach in the sequel. See [17] for more insight on the small-time polynomial expansions of the transition distributions of the Lévy process.

3. The Model Selection Problem

In this part we describe how to control the risk (1.6) of the projection estimators by imposing two conditions. First, the time horizon T should be large enough (compared to the complexity of the sieves), while the sampling frequency is kept small compared to the time horizon. These conditions will ensure that the variance term of (1.6) is of order $O(T^{-1})$. Second, the sieves should have good approximating properties in general classes of smooth functions so that when the Lévy density is presumed to have "degree of smoothness α", the bias term of (1.6) is of order $O(m^{-\alpha})$, where m is the dimension of the sieve (see Section 3.2 for the details). We prove that under the above conditions, we can tune up the dimension of the sieve to the presumed smoothness of s so that the rate of convergence of the risk is of order $O(T^{-2\alpha/(2\alpha+1)})$.

[2]The main problem arises from the application of Lemma 1 in [28]. The value of t_0 actually depends on δ. Later on in their proof, δ is taken arbitrarily small, which is likely to result in $t_0 \to 0$ (unless otherwise proved).

3.1. Analysis of the Variance Term

Consider the setting and notation of the introduction. For simplicity, we focus on estimation windows D in the positive reals (that is, $D := [a, b]$, for some $0 < a < b \leq \infty$). By making the sampling frequency per unit time high enough relative to the sampling horizon T, we can estimate the rate at which the *variance term* of the risk (1.6) decreases in the time horizon T. In the subsequent sections, we will see that this estimate actually leads to a rate of convergence for the risk which is optimal, even if our estimation were based on the whole sample path $\{X_t\}_{t \leq T}$. We shall need the following technical lemma, which we prove in the appendix for the sake of completeness.

Lemma 3.1. *For any $T > 0$, there exist $\delta_T > 0$ and $k > 0$ (independent of T) such that*

$$\sup_{y \in D} \left| \frac{1}{\Delta} \mathbb{P}\left[X_\Delta \geq y\right] - \nu([y, \infty)) \right| < k \frac{1}{T}$$

for all $\Delta < \delta_T$.

The mesh size δ_T will play a very important role below as the asymptotic results in the sequel will hold true as far as the sampling frequency, measured by $\bar{\pi} := \max\{t_k - t_{k-1}\}$, is such that $\bar{\pi} < \delta_T$. Thus, from a practical point of view, estimating δ_T is crucial. We will discuss this point in more detail in the Section 3.4.

The following easy estimate will be useful in the sequel.

Lemma 3.2. *Suppose that φ has support $[c, d] \subset \mathbb{R}_+ \backslash \{0\}$, where φ is continuous with continuous derivative. Then,*

$$\left| \frac{\mathbb{E}\, \varphi\left(X_\Delta\right)}{\Delta} - \nu(\varphi) \right| \leq \left(|\varphi(c)| + \int_c^d |\varphi'(u)|\, du \right) M_\Delta([c, d]),$$

where $M_\Delta([c, d]) := \sup_{y \in [c,d]} |\frac{1}{\Delta} \mathbb{P}\left[X_\Delta \geq y\right] - \nu([y, \infty))|$.

Proof. The result follows from the following identities

$$\mathbb{E}\, \varphi(X_\Delta) = \varphi(c)\, \mathbb{P}\left[X_\Delta \geq c\right] + \int_c^\infty \varphi'(u)\, \mathbb{P}\left[X_\Delta \geq u\right] du,$$
$$\int \varphi(x) \nu(dx) = \varphi(c) \nu\left([c, \infty)\right) + \int_c^\infty \varphi'(u) \nu\left([u, \infty)\right) du.$$

These are standard consequences of Fubini's Theorem. □

Remark 3.3. We can apply the previous two lemmas to obtain CLTs for $\hat{\beta}^\pi$ and $\hat{s}^\pi$. Indeed, if φ is as in Lemma 3.2 and, for each T, the partition π_T has mesh smaller than δ_T, the critical value in Lemma 3.1, then (2.3) hold true. The projection estimator $\hat{s}^\pi$ will satisfy (2.5) provided that the basis functions φ are as in Lemma 3.2.

We are now ready to estimate the variance term. We shall impose conditions on the approximating linear models so that the estimates of the above lemmas are applicable.

Standing assumption 1. The linear model $\mathcal{S}$ of (1.4) is generated by an orthonormal basis $\mathcal{G} := \{\varphi_j\}_{j=1}^d$ such that each φ_j is bounded with continuous derivative on the interior of its support, which is assumed to be of the form $[x_{j-1}, x_j] \subset D$.

In the sequel, we will need the following notation:

$$D_1(\mathcal{S}) := \inf_{\mathcal{G}} \max \left\{ \|\varphi\|_\infty^2 : \varphi \in \mathcal{G} \right\}, \tag{3.1}$$

$$D_2(\mathcal{S}) := \inf_{\mathcal{G}} \max \left\{ \|\varphi'\|_1^2 : \varphi \in \mathcal{G} \right\}, \tag{3.2}$$

where the infimums are over all orthonormal bases $\mathcal{G}$ of $\mathcal{S}$.

Proposition 3.4. *There exists a constant $K > 0$ such that*

$$\mathbb{E}\,\|s^\perp - \hat{s}^\pi\|^2 \le K\,\frac{\dim(\mathcal{S})}{T}, \tag{3.3}$$

for any linear model $\mathcal{S}$ satisfying the Standing Assumption 1, and for any partition $\pi : 0 = t_0 < \cdots < t_n = T$ such that $T > \max\{D_1(\mathcal{S}), D_2(\mathcal{S})\}$ and $\bar{\pi} < \delta_T$, where δ_T is the "critical" mesh size introduced in Lemma 3.1.

Proof. Fix an orthonormal basis $\mathcal{G} := \{\varphi_j\}_{j=1}^d$ of $\mathcal{S}$. Let $D_t(\varphi) := \frac{1}{t}\,\mathbb{E}\,\varphi(X_t) - \nu(\varphi)$. For any $\varphi_j \in \mathcal{G}$, we have

$$\begin{aligned} \mathbb{E}\left\{\hat{\beta}^\pi(\varphi_j) - \nu(\varphi_j)\right\}^2 &\le \frac{1}{t_n}\int \varphi_j^2(x)\nu(dx) \\ &\quad + \frac{1}{t_n^2}\sum_{k=1}^n \left|D_{\Delta_k}(\varphi_j^2)\right|\Delta_k + \left\{\frac{1}{t_n}\sum_{k=1}^n \left|D_{\Delta_k}(\varphi_j)\right|\Delta_k\right\}^2, \end{aligned}$$

where $\Delta_k := t_k - t_{k-1}$. Then, from the previous two lemmas, when $\bar{\pi} < \delta_T$,

$$\begin{aligned} \mathbb{E}\left\{\hat{\beta}^\pi(\varphi_j) - \nu(\varphi_j)\right\}^2 &\le \frac{1}{T}\int \varphi_j^2(x)\nu(dx) \\ &\quad + \frac{k}{T^2}\left(|\varphi_j^2(x_{j-1})| + \int_{x_{j-1}}^{x_j} \left|2\varphi_j(u)\varphi_j'(u)\right| du\right) \\ &\quad + \frac{k^2}{T^2}\left(|\varphi_j(x_{j-1})| + \int_{x_{j-1}}^{x_j} \left|\varphi_j'(u)\right| du\right)^2, \end{aligned}$$

which can be simplified further as follows

$$\begin{aligned} \mathbb{E}\left\{\hat{\beta}^\pi(\varphi_j) - \nu(\varphi_j)\right\}^2 &\le \frac{1}{T}\int \varphi_j^2(x)\nu(dx) + \frac{2k^2}{T^2}\left(\|\varphi_j\|_\infty + \|\varphi_j'\|_1\right)^2 \\ &\le \frac{\|s\cdot\chi_D\|_\infty}{T} + 8k^2\frac{\max_{j'}\|\varphi_{j'}\|_\infty^2 + \|\varphi_{j'}'\|_1^2}{T^2}. \end{aligned}$$

Then,

$$\mathbb{E}\,\|s^\perp - \hat{s}^\pi\|^2 \le \frac{\dim(\mathcal{S})}{T}\left\{\|s\cdot\chi_D\|_\infty + 8k^2\frac{\max_{j'}\|\varphi_{j'}\|_\infty^2 + \|\varphi_{j'}'\|_1^2}{T}\right\}.$$

Now, it is evident that (3.3) holds whenever $T > \max\{D_1(\mathcal{S}), D_2(\mathcal{S})\}$. □

3.2. The Approximation Error for Besov Type Smooth Functions

As it is customary, the bias term in (1.6) will be estimated by imposing certain degree of smoothness on the function s. Concretely, the restriction of the Lévy density s to $D := [a,b]$ is assumed to belong to the *Besov space* $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ for some $p \in [2,\infty]$ and $\alpha > 0$ (see for instance [12] and references therein for background on these spaces). The space $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ consists of those functions f belonging to $L^p([a,b])$ if $0 < p < \infty$ (or being uniformly continuous if $p = \infty$) such that

$$|f|_{\mathcal{B}^{\alpha}_{\infty}(L^p)} \equiv \sup_{\delta>0} \frac{1}{\delta^{\alpha}} \sup_{0<h\le\delta} \|\Delta^r_h(f,\cdot)\|_p < \infty,$$

with $r := [\alpha]+1$. Here, $\Delta_h(f,x) \equiv f(x+h) - f(x)$ and $\Delta^r_h(f,x)$ is the r^{th}-order difference of f defined recursively by

$$\Delta^r_h(f,x) \equiv \Delta_h(\Delta^{r-1}_h(f,\cdot),x),$$

for x's such that $x + rh \in D$ and $r \in \mathbb{N}$.

The Besov class is closely related to the so-called class of *Lipschitz functions.* For constants $k \in \mathbb{N}$ and $\beta \in (0,1]$, f is said to belong to $\mathrm{Lip}(k+\beta, L^p([a,b]))$ if $f,\dots,f^{(k-1)}$ are absolutely continuous (on $[a,b]$) and $f^{(k)}$ belongs to $L^p((a,b))$ and satisfies

$$\sup_{h>0} \frac{1}{h^{\beta}} \|\Delta_h(f^{(k)},\cdot)\|_p < \infty. \tag{3.4}$$

It is known that if $\beta < 1$ and $1 \le p \le \infty$, then $f \in \mathrm{Lip}(k+\beta, L^p([a,b]))$ if and only if f is a.e. equal to a function in $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ with $\alpha := k+\beta$. In general, $\mathrm{Lip}(k+\beta, L^p([a,b])) \subset \mathcal{B}^{k+\beta}_{\infty}(L^p([a,b]))$, for any $0 < p \le \infty$ (see e.g. [12]). Notice that when $p = \infty$, the condition (3.4) takes the form:

$$|f^{(k)}(x) - f^{(k)}(y)| \le L|x-y|^{\beta}, \tag{3.5}$$

for all $x, y \in (a,b)$ and some $L < \infty$.

An important reason for working with the Besov-type smooth functions is the availability of estimates of the approximation error by splines, trigonometric polynomials, and wavelets (see [12] and [3] for more details). For instance, if $\mathcal{S}_{k,m}$ denotes the space of piecewise polynomials of degree at most k, based on a regular partition of $[a,b]$ with m classes, and $s \in \mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ with $\alpha < k+1$, then there exists a constant $c_{[\alpha]} < \infty$ such that

$$\inf_{f\in\mathcal{S}_{k,m}} \|s - f\|_p \le c_{[\alpha]} |s|_{\mathcal{B}^{\alpha}_{\infty}(L^p)} (b-a)^{\alpha} m^{-\alpha}. \tag{3.6}$$

Thus, when $p \ge 2$, the orthogonal projection of s on $\mathcal{S}_{k,m}$, denoted by $s^{\perp}_m$, is such that

$$\|s - s^{\perp}_m\| \le c_{[\alpha]} (b-a)^{\frac{1}{2}-\frac{1}{p}+\alpha} |s|_{\mathcal{B}^{\alpha}_{\infty}(L^p)} m^{-\alpha}. \tag{3.7}$$

Notice that the elements of $\mathcal{S}_{k,m}$ are not necessarily smooth (not even continuous) and hence, they are not "splines" in the standard sense of the literature, where a spline is understood as a smooth piece-wise polynomial. The upper bound (3.6) is actually true if we restrict to certain splines of $\mathcal{S}_{k,m}$ (say B-splines) (see (10.1) in Chapter 2 of [12]). For the sake of completeness let us describe in detail the space $S_{k,m}$ as well as give estimates for the constants (3.1)-(3.2). Let Q_j be the Legendre polynomials of order j on $\mathbb{L}^2([-1,1],dx)$. The space $\mathcal{S}_{k,m}$ is generated by the orthonormal functions

$$\hat{\varphi}_{i,j}(x) := \sqrt{\frac{2j+1}{x_i - x_{i-1}}}\, Q_j\left(\frac{2x - (x_i + x_{i-1})}{x_i - x_{i-1}}\right) \mathbf{1}_{(x_{i-1},x_i)}(x),$$

for $i = 1, \ldots, m$ and $j = 0, \ldots, k$, and where $a = x_0 < \cdots < x_m = b$ are equally-spaced points. It is well-known that $|Q_j(x)| \leq 1$ and $|Q_j'(x)| \leq Q_j'(1) = \frac{j(j+1)}{2}$. Then, fixing $\Delta_x := x_i - x_{i-1} = \frac{b-a}{m}$, we have that

$$\hat{\varphi}_{i,j}'(x) = 2\sqrt{2j+1}\,\Delta_x^{-3/2}\,Q_j'\left(\frac{2x - (x_i + x_{i-1})}{x_i - x_{i-1}}\right)\mathbf{1}_{(x_{i-1},x_i)}(x),$$
$$\|\hat{\varphi}_{i,j}'\|_1 \leq 2\sqrt{2j+1}\Delta_x^{-3/2}\int_{x_{i-1}}^{x_i}\sup_u |Q_j'(u)|dx \leq \sqrt{2j+1}\Delta_x^{-1/2}(j)(j+1).$$

It is now clear that

$$D_2(\mathcal{S}_{k,m}) \leq \max_{i,j}\left\{\|\varphi_{i,j}'\|_1^2\right\} \leq \frac{(k+1)^2k^2(2k+1)}{b-a}\,m.$$

In a similar manner one can check that

$$D_1(\mathcal{S}_{k,m}) \leq \frac{(k+1)^2(2k+1)}{b-a}\,m.$$

3.3. Rate of Convergence for Smooth Functions Via Splines

As a consequence of the variance and bias term estimates given in the previous two parts, we now estimate the rate of convergence on D of the projection estimators (1.9), using the regular piece-wise polynomials $\{\mathcal{S}_{k,m}\}_{m\geq 1}$ as sieves assuming that the Lévy density s is in the Besov class $\mathcal{B}_\infty^\alpha(L^p([a,b]))$ with $p \geq 2$ and $\alpha < k+1$. It turns out that under the stated conditions, projection estimators converge at a rate at least as good as $T^{-2\alpha/(2\alpha+1)}$. The following result is valid provided that, for each time horizon T, the mesh of the sampling times π_T is smaller than the critical mesh δ_T introduced in Lemma 3.1. In Section 4, we will see that this rate is actually the best possible even under continuous sampling.

Proposition 3.5. *Let $\hat{m}_T := [T^{1/(2\alpha+1)}]$ and let $\Theta(R,L)$ be the class of Lévy densities s such that $\|s \cdot \chi_D\|_\infty < R$, and such that the restriction of s to $D := [a,b]$ is a member of $\mathcal{B}_\infty^\alpha(L^p([a,b]))$ with $|s|_{\mathcal{B}_\infty^\alpha(L^p)} < L$ and $p \geq 2$. Then,*

$$\limsup_{T\to\infty} T^{2\alpha/(2\alpha+1)} \sup_{s\in\Theta(R,L)} \mathbb{E}\left[\|s - \hat{s}_T\|^2\right] < \infty, \tag{3.8}$$

where for each T, the estimator $\hat{s}_T$ is given by (1.7) and (1.9) with $\mathcal{S} = \mathcal{S}_{k,\hat{m}_T}$, $k > \alpha - 1$, and a mesh $\bar{\pi}_T$ smaller than δ_T.

Proof. From the two previous parts, there exists a constant K (depending on k, a, b, α, R, p, L) such that

$$\|s - s_m^\perp\| \leq K\,m^{-\alpha} \quad \text{and} \quad \mathbb{E}\,\|s^\perp - \hat{s}_m^\pi\|^2 \leq K\,\frac{m}{T},$$

for $m \in \mathcal{M}_T := \{m' : T > Km'\}$ and $\bar{\pi} < \delta_T$. Then for a constant M and for large enough T,

$$\sup_{s\in\Theta(R,L)} \mathbb{E}\left[\|s - \hat{s}_T\|^2\right] \leq M\left\{\left[T^{1/(2\alpha+1)}\right]^{-2\alpha} + \left[T^{1/(2\alpha+1)}\right]T^{-1}\right\}.$$

The limit (3.8) is now clear. $\square$

Example 3.6. If s has continuous bounded derivative on $D := [a,b] \subset \mathbb{R}\backslash\{0\}$ (hence, $s \in \mathcal{B}^{\alpha}_{\infty}(L^{\infty}([a,b]))$, for any $\alpha < 1$), then one can construct regular histogram estimators converging to s on D at a rate faster than $T^{-1/2}$ if one selects the number of classes approximately equal to $T^{1/2}$ and the mesh of the partition π smaller than δ_T.

3.4. About the Critical Mesh

The critical mesh, introduced in Lemma 3.1, gives a bound on the mesh of the sampling frequency needed to estimate in a simple way the rate of convergence of the variance term (see Proposition 3.4). Of course, any hope for a feasible implementation of this estimation scheme will require an explicit estimate of this critical mesh. In the compound Poisson case (when $\nu(\mathbb{R}\backslash\{0\}) < \infty$), it turns out that $\delta_{_T} = o(\frac{1}{T})$ suffices. In the general case, we have the following result, which tell us, in particular, that the sampling frequency needs to be higher when one wishes to estimate the Lévy density closer to the origin.

Proposition 3.7. *Let $\rho > 0$ such that $a\rho > 1$. Then, there exists $T_0(\rho) > 0$ and $k > 0$ such that*

$$\sup_{y \in D} \left| \frac{1}{\Delta}\, \mathbb{P}\left[X_\Delta \geq y\right] - \nu([y,\infty)) \right| < k\,\frac{1}{T}$$

for all $T > T_0$ and $\Delta < T^{-\frac{1}{\rho}T}$.

Proof. As in the proof of Lemma 3.1, we can obtain

$$\begin{aligned}\sup_{y \in D} \left| \frac{1}{t}\, \mathbb{P}\left[X_t \geq y\right] - \nu([y,\infty)) \right| &\leq \frac{1}{t}\, \mathbb{P}\left[X^{\varepsilon}_t \geq a\right] + 2c\frac{1}{T} + \nu([a,\infty))\, \mathbb{P}\left[|X^{\varepsilon}_t| \geq \eta\right] \\ &\quad + \lambda_\varepsilon\, \mathbb{P}\left[|X^{\varepsilon}_t| \geq \eta\right] + \nu([a,\infty))\, \lambda_\varepsilon t + \lambda_\varepsilon^2 t,\end{aligned}$$

valid for $T > 1/a$, $\eta = \frac{1}{T}$, and $0 < \varepsilon < a - \eta$ (here $c := \sup_{a-\eta \leq x \leq b+\eta} s(x)$). Fix $\varepsilon > 0$ sufficiently small so that $\rho < \frac{1}{\varepsilon}$. Let us recall that there exists $y_0 := y_0(\rho)$ such that

$$\mathbb{P}\left[|X^{\varepsilon}_t| \geq y\right] \leq \exp\{\rho y_0 \log y_0\} \exp\left\{\rho y - \rho y \log y\right\} t^{y\rho}$$

for all $t < \frac{y}{y_0(\rho)}$ (see e.g. [28]). In particular, when $y = \eta = \frac{1}{T}$ and $t < T^{-\frac{1}{\rho}T}$, for T sufficiently large that $T^{-\frac{1}{\rho}T+1} < \frac{1}{y_0(\rho)}$,

$$\mathbb{P}\left[|X^{\varepsilon}_t| \geq \eta\right] \leq kT^{-1}.$$

Similarly, when $y = a$ and $t < T^{-\frac{1}{\rho}T}$,

$$\frac{1}{t}\, \mathbb{P}\left[X^{\varepsilon}_t \geq a\right] \leq kt^{a\rho-1} < kT^{-\frac{1}{\rho}T(a\rho-1)} < kT^{-1},$$

if $T > \frac{\rho}{a\rho-1}$. This proves the result since ε is fixed. ☐

Remark 3.8. The estimate of the critical mesh given in Proposition 3.7 can be improved substantially. Indeed, in a forthcoming paper, we will show that it suffices that $\Delta = o(T^{-1})$.

4. Minimax Risk of Estimation for Smooth Lévy Densities

In this section, we show that the rate of convergence $O(T^{-2\alpha/(2\alpha+1)})$ attained by projection estimators is the best possible, in the sense that there is no estimator $\hat{s}^*_T$ that can converge to s faster than $T^{-2\alpha/(2\alpha+1)}$, for any $s \in \Theta$, even assuming continuous-time sampling. In order to prove this, we will assess the long-run behavior of the *minimax risk* on Θ, roughly defined as

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

where the infimum is taken over *all possible* estimators $\hat{s}$, and $d(s, \hat{s})$ measures the distance between $\hat{s}$ and s.

Traditionally, the performance of nonparametric estimators is gauged by comparing the rate of convergence of the estimator in question to the rate of convergence of the minimax risk when the available data increases. The rates of convergence of minimax risks are available in most of the traditional nonparametric problems. For instance, Ibragimov and Has'minskii [19] and Barron et al. [2] provided this kind of asymptotics for the problem of density estimation based on i.i.d. random variables, while Kutoyants [22] and Reynaud-Bouret [25] considered the problem of intensity estimation of a *finite* Poisson point processes. This last set-up is relevant for our problem since the jumps of a Lévy process can be associated with a (possibly infinite) Poisson point process on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ (see e.g. Theorem 19.2 in [29]). Using this connection, we adapt below a result from [22] to obtain the long-run asymptotics of the minimax risk of estimation of the Lévy density off the origin. The idea of the proof, due to Ibragimov and Has'minskii [19], is based on the statistical toolbox for distributions satisfying the *Local Asymptotic Normality* (LAN) property (see Chapters II and Section IV.5 of [19]).

Let us introduce some notation. Here, $\ell : \mathbb{R} \to \mathbb{R}$ stands for a *loss function* satisfying the following:

(i) $\ell(\cdot)$ is nonnegative, $\ell(0) = 0$ but not identically 0, and ℓ continuous at 0;
(ii) ℓ is symmetric: $\ell(u) = \ell(-u)$ for all u;
(iii) $\{u : \ell(u) < c\}$ is a convex set for any $c > 0$;
(iv) $\ell(u)\exp\{\varepsilon|u|^2\} \to 0$ as $|u| \to \infty$ for any $\varepsilon > 0$.

We consider Lévy densities $s : \mathbb{R}\backslash\{0\} \to \mathbb{R}_+$ that are k times differentiable on an interval $[a, b] \subset \mathbb{R}\backslash\{0\}$ and satisfy (3.5) for all $x, y \in [a, b]$. For given $k \in \mathbb{N}$ and $\beta \in (0, 1]$, we denote such a class of functions by $\Theta_{k+\beta}(L; [a, b])$. The proof of the result below is presented in the Appendix A.

Theorem 4.1. *If x_0 is an interior point of the interval $[a, b]$, then*

$$\liminf_{T\to\infty} \left\{ \inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_s \left[\ell\left(T^{\alpha/(2\alpha+1)} \left(\hat{s}_T(x_0) - s(x_0)\right)\right)\right]\right\} > 0, \tag{4.1}$$

where $\alpha := k + \beta$, $\Theta := \Theta_\alpha(L; [a, b])$ and the infimum is over all the estimators $\hat{s}_T$ of s based on $\{X(t)\}_{0\le t\le T}$.

The previous result can be strengthen to be uniform in $x_0 \in (a, b)$ and as a consequence, the long-run behavior of the minimax risk under the integrated mean-square distance can be assessed. The proof of the next result is given in Appendix A.

Corollary 4.2. *Under the notation and conditions of Theorem 4.1, the following two limits hold:*

$$\liminf_{T\to\infty}\left\{\inf_{\hat{s}_T}\inf_{x\in(a,b)}\sup_{s\in\Theta}\mathbb{E}_s\left[\ell\left(T^{\alpha/(2\alpha+1)}\left(\hat{s}_T(x)-s(x)\right)\right)\right]\right\}>0, \tag{4.2}$$

$$\liminf_{T\to\infty}T^{2\alpha/(2\alpha+1)}\left\{\inf_{\hat{s}_T}\sup_{s\in\Theta}\mathbb{E}_s\left[\int_a^b\left(\hat{s}_T(x)-s(x)\right)^2dx\right]\right\}>0. \tag{4.3}$$

Remark 4.3. The previous result is also valid for classes slightly smaller than $\Theta_\alpha(L;[a,b])$ such as

$$\Theta=\Theta_\alpha(L;[a,b])\cap\{s:\|s\|_{\mathbb{L}^\infty([a,b])}<R\},$$

which is closely related to the Besov class $\Theta(R,L)$ of (3.8). Indeed, $\Theta_\alpha(L;[a,b])$ is contained in $\mathcal{B}^\alpha_\infty(\mathbb{L}^\infty([a,b]))$ (see Section 2.9 of [12]), and thus,

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

We conclude that there is no *reasonable* estimator $\hat{s}_T$ of s capable of outperforming the rate $T^{-2\alpha/(2\alpha+1)}$ *uniformly on* Θ: there is always an $s\in\Theta$ for which

$$T^{2\alpha/(2\alpha+1)}\,\mathbb{E}_s\left[\|\hat{s}_T-s\|^2\right]>B,$$

for some $B>0$ and for large enough T. Therefore, the estimator described in Proposition 3.5 achieves the optimum rate of convergence on $\Theta(R,L)$ from a minimax point of view.

5. A Data-Driven Selection Method and Adaptability

The model selection criterion described in Section 3.3, where one tunes up the number of classes m to the "smoothness" of s, has the obvious drawback of requiring (or at least presuming) the smoothness parameter α. In the literature of nonparametric statistics, one wishes to devise data-driven selection methods that can *adapt* to arbitrary degree of smoothness (see e.g. Birgé and Massart [5] for an extensive exposition of the topic).

A typical approach for adaptive model selection schemes consists of minimizing an unbiased estimator of the risk of estimation. This approach was developed in [13] in the context of Lévy density estimation. Let us briefly discuss the findings there. The key idea comes from the following refinement of (1.6):

$$\mathbb{E}\left[\|s-\hat{s}^c\|^2\right]=\|s\|^2+\mathbb{E}\left[-\|\hat{s}^c\|^2+\mathrm{pen}^c(\mathcal{S})\right], \tag{5.1}$$

where $\hat{s}^c$ is as in (1.9) substituting $\hat{\beta}^\pi(\varphi)$ by the statistics $\hat{\beta}^c(\varphi)$ of (1.13), $s^\perp$ is the orthogonal projection in (1.5), and $\mathrm{pen}^c(\mathcal{S})$ is defined in terms of an orthonormal basis $\mathcal{G}:=\{\varphi_1,\ldots,\varphi_d\}$ of $\mathcal{S}$ by the formula:

$$\mathrm{pen}^c(\mathcal{S})\equiv\frac{2}{T^2}\sum_{t\le T}\sum_{\varphi\in\mathcal{G}}\varphi^2(\Delta X_t). \tag{5.2}$$

Equation (5.1) shows that the risk of $\hat{s}^c$ moves "parallel" to the expectation of the *observable statistics* $-\|\hat{s}^c\|^2+\mathrm{pen}^c(\mathcal{S})$, suggesting the selection of the model that minimizes such statistics. Concretely, given a collection of sieves $\{\mathcal{S}_m, m \in \mathcal{M}\}$, we should choose the projection estimator $\tilde{s}^c \equiv \hat{s}^c_{\hat{m}}$, where

$$\hat{m} \equiv \operatorname*{argmin}_{m \in \mathcal{M}} \left\{ -\|\hat{s}^c_m\|^2 + \mathrm{pen}^c(\mathcal{S}_m) \right\}.$$

Such an estimator $\tilde{s}^c$ is called a *penalized projection estimator* (p.p.e.) since the role of $\mathrm{pen}^c(\mathcal{S})$ is to penalize large linear models.

In [16], it is shown that the p.p.e. $\tilde{s}^c$ is adaptive in the class of Besov Lévy densities of Section 3.2 in the sense that $\tilde{s}^c$ attains the optimal rate of convergence $O(T^{-2\alpha/(2\alpha+1)})$ without using the knowledge of α. Unfortunately, the previous approach intrinsically requires continuous-time sampling of the process to determine the jumps ΔX_t. However, the analysis could still be useful if one uses the natural discrete-based proxies of $\hat{\beta}^c$ and pen^c, where the jumps ΔX_t are replaced by the increments $X_{t_k} - X_{t_{k-1}}$. This idea leads to the estimators $\hat{s}^\pi$ in (1.9) and to the statistic

$$\mathrm{pen}^\pi(\mathcal{S}) = \frac{2}{T^2} \sum_{k=1}^{n} \sum_{\varphi \in \mathcal{G}} \varphi^2(X_{t_k} - X_{t_{k-1}}) \tag{5.3}$$

as the penalization term. In the light of the previous arguments, we proposed a discrete-based model selection criterion as follows

$$\begin{aligned} \hat{m}^\pi &\equiv \operatorname*{argmin}_{m \in \mathcal{M}} \left\{ -\|\hat{s}^\pi_m\|^2 + \mathrm{pen}^\pi(\mathcal{S}_m) \right\} \\ &= \operatorname*{argmin}_{m \in \mathcal{M}} \left\{ -\sum_{\varphi \in \mathcal{G}_m} \{\hat{\beta}^\pi(\varphi)\}^2 + \mathrm{pen}^\pi(\mathcal{S}_m) \right\}, \end{aligned} \tag{5.4}$$

where $\mathcal{G}_m$ is an orthonormal basis of $\mathcal{S}_m$, $\hat{\beta}^\pi$ is given by (1.7), and pen^π is given by (5.3). The resulting estimator

$$\widetilde{s} := s^\pi_{\hat{m}^\pi} \tag{5.5}$$

will be called (discrete-based) *penalized projection estimator*.

We hope to extend in a future work the adaptability result in [16] for this discrete-based p.p.e. In the sequel, we illustrate the performance of these estimators for an infinite-jump activity Lévy process of relevance in the area of mathematical finance.

6. An Example: Estimation of Variance Gamma Processes.

6.1. The Model

Variance Gamma processes were proposed in [23] (see also [8]) as substitutes to the Brownian Motion in the Black-Scholes model. Since their introduction, this kind of processes have received a great dealt of attention, even in the financial industry. For an introduction to many basic properties of variance Gamma processes and other related processes, the reader is referred to Knotz et al. [21].

There are two useful representations for this type of processes. A variance Gamma process $X = \{X(t)\}_{t \geq 0}$ is a Brownian motion with drift, time changed by a Gamma Lévy process. Concretely,

$$X(t) = \theta U(t) + \sigma W(U(t)), \tag{6.1}$$

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

$$f_t(x) = \frac{x^{t/\nu - 1} \exp\left(-\frac{x}{\nu}\right)}{\nu^{t/\nu}\Gamma\left(\frac{t}{\nu}\right)}. \tag{6.2}$$

Notice that $\mathrm{E}[U(t)] = t$ and $\mathrm{Var}[U(t)] = \nu t$; 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 (6.1) with suitably chosen θ, σ, and ν. This a consequence of the *self-similarity*[3] property of Brownian motion and the fact that ν in (6.2) is a scale parameter.

The process X is itself a Lévy process since Gamma processes are *subordinators* (see Theorem 30.1 of [29]). Moreover, it is not hard to check that "statistically" X is the difference of two Gamma Lévy processes (see e.g. (2.1) of [6]):

$$\{X(t)\}_{t\geq 0} \stackrel{\mathfrak{D}}{=} \{X_+(t) - X_-(t)\}_{t\geq 0}, \tag{6.3}$$

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

$$s(x) = \begin{cases} \frac{\alpha}{|x|} \exp\left(-\frac{|x|}{\beta^-}\right) & \text{if } x < 0, \\ \frac{\alpha}{x} \exp\left(-\frac{x}{\beta^+}\right) & \text{if } x > 0, \end{cases} \tag{6.4}$$

where $\alpha > 0$, $\beta^- \geq 0$, and $\beta^+ \geq 0$ (of course, $|\beta^-| + |\beta^+| > 0$). 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.

6.2. 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 (6.3) using the series representation method introduced in Rosiński [26]. The other approach is to generate the random time change $\{U(t)\}_{0\leq t\leq T}$ of (6.1), 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.

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

6.3. The Numerical Results

In this part we illustrate the performance of the projection estimators (1.9) and the model selection criterion described in Section 5 using simulation experiments. The approximating linear models $\mathcal{S}_m$ considered here are the span of the indicator functions $\chi_{[x_0,x_1]},\dots,\chi_{(x_{m-1},x_m]}$, where $x_0<\dots<x_m$ is a regular partition of an interval $D\equiv[a,b]$, with $0<a$ or $b<0$. We perform the following numerical experiment. First, we simulate the variance gamma Lévy process with specified (known) parameter settings. Then, we apply the penalized projection estimator defined by (5.4)-(5.5). Finally, to assess the accuracy of the nonparametric estimator, the true parametric model of s is subsequently fit to the nonparametric estimator using a least-square errors method. Concretely, if $\widetilde{s}:=\hat{s}^{\pi}_{\hat{m}^{\pi}}$ is the discrete-based p.p.e. and s^{θ} is the function (6.4), where we set $\theta:=(\alpha,\beta^-,\beta^+)$, then we find

$$\hat{\theta}_{NP}:=\operatorname*{argmin}_{\theta}\sum_{i=0}^{\hat{m}^{\pi}-1}\left(\widetilde{s}(\bar{x}_i)-s^{\theta}(\bar{x}_i)\right)^2, \tag{6.5}$$

where $\bar{x}_i$ is the midpoint of the interval $[x_i,x_{i+1}]$. This approach provides a non-parametric based estimators for the parameters of the variance Gamma process.

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[4] (see [8]), there are well-documented issues with MLE (see for instance [24]). 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 [8] 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 $\hat{\theta}_{ML}=-0.00056256$, $\hat{\sigma}^2_{ML}=0.01373584$, and $\hat{\nu}_{ML}=0.002$, from where we obtain $\hat{\alpha}_{ML}=500$, $\hat{\beta}^+_{ML}=0.0037056$, and $\hat{\beta}^-_{ML}=0.0037067$.

Figures 1 and 2 show respectively the left- and right- tails of the true Lévy density and the (discrete-based) penalized projection estimator as well as their corresponding best-fit variance Gamma Lévy densities using (6.5), 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-eight of a day. The figures seem quite comforting. To get a better idea of the performance of the method, Figures 3 and 4 show the sampling distributions of the estimates of α^- and β^+ obtained from applying the least-square method to the penalized projection estimators. The histograms are based on 1000 samples of size 5000 with $\Delta t=1/8$ of a day. This experiment shows clear, though

[4] More concretely, the density is terms of Bessel special functions of third kind. For more information, see also Section 4.1 in Knotz et al. [21].

Table 1

Sampling mean in **bold** *and standard errors in parenthesis of the estimators of α^+, β^+, and υ in the CGMY model with theoretical values $\alpha^- = 0$, $\beta^+ = 1$, $\alpha^+ = 1$, and $\upsilon = .1$. Sample size is 100 paths*

Δt	Penalized Projection Estimators/Least-Squares Fit			Misspecified Gamma MLE	
	$\hat{\alpha}^+_{NP}$	$\hat{\beta}^+_{NP}$	$\hat{\upsilon}_{Zolotarev}$	$\hat{\alpha}^+_{MLE}$	$\hat{\beta}^+_{MLE}$
.01	**1.03** (0.15)	**0.97** (0.14)	0.09 (0.0002)	**1.2** (0.08)	**0.89** (0.079)

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 5 and 6). Nonparametric methods are not free-lunches and usually the gain in robustness is paid by a loss in efficiency.

To illustrate the seriousness of applying an efficient estimation method to a misspecified model let us consider a close relative of the variance Gamma process: the CGMY model in [6]. This is defined as a pure-jump Lévy process with Lévy density of the form

$$s_m(x) = \begin{cases} \frac{\alpha^-}{|x|^{\upsilon+1}} \exp\left(-\frac{|x|}{\beta^-}\right) & \text{if } x < 0, \\ \frac{\alpha^+}{x^{\upsilon+1}} \exp\left(-\frac{x}{\beta^+}\right) & \text{if } x > 0, \end{cases} \tag{6.6}$$

where $\upsilon > 0$. In the case when $\alpha^- = 0$ and $\nu = 0$, we recover a Gamma Lévy process, for which MLE are widely available. Let us take $\alpha^+ = \beta^+ = 1$ and $\upsilon = .1$. We can estimate the parameter υ using a Zolotarev type estimator. This can be done so since the CGMY Lévy process is a *tempered stable Lévy process*, whose short-term increments behave like stable processes (see Rosínski [27] for details).

Table 1 shows the sampling average and standard deviations of the estimators of α^+, β^+, and υ by two methods based on 100 simulation runs. The first method estimates υ using the Zolotarev's estimator $\hat{\upsilon}$, then computes the piece-wise constant p.p.e. $\widetilde{s}$ of (5.5), and finally, estimate α^+ and β^+ via the LSE method (6.5) replacing s^θ by the Lévy density s_m of (6.6) with $\theta = (\alpha^+, \beta^+)$ and fixing $\alpha_- = 0$ and $\upsilon = \hat{\upsilon}$. The second method assumes (erroneously) that the underlying model is a Lévy gamma process and performs maximum likelihood estimation.

The results above shows that sometime a modestly efficient robust nonparametric method is preferably to a very efficient estimation method.

Appendix A: Technical Proofs

Proof of Lemma 3.1. The idea is to exploit the well-known decomposition of the Lévy process as a compound Poisson process $\widetilde{X}^\varepsilon$ plus an independent Lévy process $X^\varepsilon := X - \widetilde{X}^\varepsilon$ with compactly supported Lévy measure $\nu_\varepsilon(dx) := \mathbf{1}_{\{|x| \leq \varepsilon\}}\, \nu(dx)$, for a suitable chosen $\varepsilon > 0$. Concretely, here

$$\widetilde{X}^\varepsilon_t = \sum_{i=1}^{N_t} \xi_i,$$

for a homogeneous Poisson process $\{N_t\}_{t\geq 0}$ with intensity $\lambda_\varepsilon := \nu(\{|x| > \varepsilon\})$ and for independent random variables $\{\xi_i\}$ with distribution $\frac{1}{\lambda_\varepsilon}\mathbf{1}_{\{|x|>\varepsilon\}}\nu(dx)$. Clearly,

$$\frac{1}{t}\,\mathbb{P}\,[X_t \geq y] = \frac{1}{t}\,\mathbb{P}\,[X_t^\varepsilon \geq y]\, e^{-\lambda_\varepsilon t} + \mathbb{P}\,[X_t^\varepsilon + \xi_1 \geq y]\, e^{-\lambda_\varepsilon t}(\lambda_\varepsilon) + \sum_{n=2}^{\infty} \mathbb{P}\left[X_t^\varepsilon + \sum_{i=1}^{n} \xi_i \geq y\right] e^{-\lambda_\varepsilon t}\lambda_\varepsilon^n t^{n-1}.$$

Then, we have

$$\left|\frac{1}{t}\,\mathbb{P}\,[X_t \geq y] - \nu([y,\infty))\right| \leq \frac{1}{t}\,\mathbb{P}\,[X_t^\varepsilon \geq y] + |\lambda_\varepsilon\,\mathbb{P}\,[X_t^\varepsilon + \xi_1 \geq y] - \nu([y,\infty))| + \nu([y,\infty))\,\lambda_\varepsilon t + \lambda_\varepsilon^2 t.$$

The second term on the right hand side of this inequality can itself be decomposed as follows:

$$|\lambda_\varepsilon\,\mathbb{P}\,[X_t^\varepsilon + \xi_1 \geq y] - \nu([y,\infty))| \leq \int_{y-\eta}^{y+\eta} s(x)\,dx + \lambda_\varepsilon\,\mathbb{P}\,[|X_t^\varepsilon| \geq \eta] + \nu([y,\infty))\,\mathbb{P}\,[|X_t^\varepsilon| \geq \eta]$$

for each $\eta > 0$ such that $a - \eta > \varepsilon$. Since s is bounded off the origin, there exists a $k > 0$ such that

$$\int_{y-\eta}^{y+\eta} s(x)dx \leq k\,\eta$$

for all $y \in D$. Fix $0 < \eta < \frac{1}{T} \wedge a$ and $0 < \varepsilon < a - \eta$. Then,

$$\sup_{y\in D}\left|\frac{1}{t}\,\mathbb{P}\,[X_t \geq y] - \nu([y,\infty))\right| \leq \frac{1}{t}\,\mathbb{P}\,[X_t^\varepsilon \geq a] + \frac{k}{T} + \nu([a,\infty))\,\mathbb{P}\,[|X_t^\varepsilon| \geq \eta] + \lambda_\varepsilon\,\mathbb{P}\,[|X_t^\varepsilon| \geq \eta] + \nu([a,\infty))\,\lambda_\varepsilon t + \lambda_\varepsilon^2 t.$$

Finally, since $\lim_{t\to 0} \frac{1}{t}\,\mathbb{P}\,[X_t^\varepsilon \geq a] = 0$ and $\lim_{t\to 0}\,\mathbb{P}\,[|X_t^\varepsilon| \geq \eta] = 0$, we can choose $\delta_T > 0$ sufficiently small to make each of the terms smaller than $1/T$ when $t < \delta_T$. □

Proof of Theorem 4.1.

(i) Fix a Lévy density $s_0 \in \Theta_\alpha(L/2;[a,b])$ such that $s_0(x) > 0$, for all $x \in \mathbb{R}_0 := \mathbb{R}\backslash\{0\}$, and a constant $\kappa > 0$. Also, let $g : \mathbb{R} \to \mathbb{R}_+$ be a symmetric function with compact support $\mathbb{K}_g$, satisfying (3.5) with $L/2$ (instead of L). Moreover, the support of $x \to g(\kappa(x - x_0))$, denoted by $\mathbb{K}$, does not contain the origin and also,

$$s_0(x) - \kappa^{-\alpha} g\left(\kappa(x - x_0)\right) > 0, \quad \forall x \in \mathbb{R}_0.$$

Let

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

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

(ii) Without loss of generality we assume that $\mathbb{K} \cap [-1,1] = \emptyset$. We follow the notation in [29] (Section 33). Let $\mathbb{P}_\theta^{(T)}$ be the distribution (on $\mathbb{D}[0,T]$) of a Lévy process $\{X(t)\}_{0\leq t\leq T}$ with Lévy density s_θ (the other two parameters of the generating triplet remain constant). We proceed to prove that $\{\,\mathbb{P}_\theta^{(T)}\,:$

$\theta \in (-\kappa^{-\alpha}, \kappa^{-\alpha})\}$ is LAN at $\theta = 0$ (see e.g. Definition II.2.1 in [19]). By Theorems 33.1 and 33.2 in [29], $\mathbb{P}_{\theta}^{(T)} \approx \mathbb{P}_{0}^{(T)}$ and the likelihood function, $L_\theta(\omega) := \frac{d\mathbb{P}_{\theta}^{(T)}}{d\mathbb{P}_{0}^{(T)}}(\omega)$ is given by

$$L_\theta(\omega) := \exp\left\{\int_0^T\!\!\int_{\mathbb{K}} \ln\left[1 + \frac{\theta T^{-\frac{\alpha}{2\alpha+1}}}{s_0(x)} g\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\right]\xi(dt,dx;\omega)\right.$$
$$\left. - \theta T^{1-\frac{\alpha}{2\alpha+1}} \int_{\mathbb{K}} g\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)dx\right\},$$

where $\xi(dt, dx; \omega)$ is the random measure on $\mathbb{R}_+ \times \mathbb{R}_0$ associated with the jumps of $\omega \in \mathbb{D}[0,T]$; that is,

$$\xi(A;\omega) := \#\{(t,x) : \Delta w_t := w_t - w_{t^-} = x\}, \quad A \subset \mathbb{R}_+ \times \mathbb{R}_0.$$

Under $\mathbb{P}_0^{(T)}$, ξ is a Poisson random measure with mean measure $s_0(x)dxdt$. We denote $\bar{\xi}(dt, dx; \omega) := \xi(dt, dx; \omega) - s_0(x)dxdt$. The likelihood $L_\theta(\omega)$ can be written as follows:

$$L_\theta(\omega) = \exp\left\{\theta\Delta_{T} - \frac{\theta^2}{2}\sigma_{T}^2 + r_{T}(\theta)\right\},$$

where

$$\Delta_{T} = T^{-\frac{\alpha}{2\alpha+1}} \int_0^T\!\!\int_{\mathbb{K}} s_0^{-1}(x) g\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\bar{\xi}(dt,dx),$$
$$\sigma_{T}^2 = T^{1-\frac{2\alpha}{2\alpha+1}} \int_{\mathbb{K}} s_0^{-1}(x) g^2\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)dx,$$
$$r_{T}(\theta) = -\frac{\theta^2}{2} T^{-\frac{2\alpha}{2\alpha+1}} \int_0^T\!\!\int_{\mathbb{K}} s_0^{-2}(x) g^2\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\bar{\xi}(dt,dx)$$
$$+ \int_0^T\!\!\int_{\mathbb{K}} R\left(\theta T^{-\frac{\alpha}{2\alpha+1}} s_0^{-1}(x) g\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)\right)\xi(dt,dx),$$

and $R(u) := \ln(1+u) - u + \frac{u^2}{2}$. We want to prove that there are nomalizing constants $\varphi_T > 0$ such that

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

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 [22]). Indeed, for $T > 1$, we have that

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

Similarly, for large enough T,

$$\operatorname{Var}(\Delta_{T}) = T^{-\frac{2\alpha}{2\alpha+1}} \int_0^T \int_{\mathbb{K}} s_0^{-2}(x) g^2\left(\kappa T^{\frac{1}{2\alpha+1}}(x-x_0)\right)(s_0(x))\,dxdt$$
$$\stackrel{T\to\infty}{\longrightarrow} \kappa^{-1} s_0^{-1}(x_0) \int_{\mathbb{K}_g} g^2(u)du.$$

Then, $\mathcal{L}_{\mathbb{P}_0^{(T)}}(\Delta_T) \xrightarrow{\mathfrak{D}} \mathcal{N}(0, I_0^2)$ with $I_0^2 := \kappa^{-1} s_0^{-1}(x_0) \int_{\mathbb{K}_g} g^2(u)du$, and $\sigma_T^2 \to I_0^2$. We now verify that $r_T(\theta)$ vanishes in probability. Notice that the first term of r_T converges to 0 since its mean is 0 and its variance vanishes. 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 $|R(u)| \leq |u|^3/3$, the absolute value of its expectation satisfies

$$\left| \int_0^T \int_{\mathbb{K}} R\left(\theta T^{-\frac{\alpha}{2\alpha+1}} s_0^{-1}(x) g\left(\kappa T^{\frac{1}{2\alpha+1}}(x - x_0)\right)\right)(s_0(x))dxdt \right|$$
$$\leq \frac{|\theta|^3}{3} T^{1-\frac{3\alpha}{2\alpha+1}} \int_{\mathbb{K}} s_0^{-2}(x) g^3\left(\kappa T^{\frac{1}{2\alpha+1}}(x - x_0)\right) dx \overset{T\to\infty}{\longrightarrow} 0.$$

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 $\varphi_T := I_0^{-1}$).

(iii) By Theorem II.12.1 and Remark II.12.2 in [19], if $\hat{\theta}_T$ is any estimator of θ based on $\{X(t)\}_{0\leq t\leq T}$, then

$$\liminf_{T\to\infty} \sup_{|\theta|<\kappa^{-\alpha}} \mathbb{E}_\theta \left[\ell_0\left(I_0\left(\hat{\theta}_T - \theta\right)\right)\right] \geq B, \tag{A.1}$$

where $B := \mathbb{E}\left[\ell_0(Z)\chi_{[|Z|<I_0\kappa^{-\alpha}/2]}\right]$ and $Z \sim \mathcal{N}(0,1)$. Now, let $\hat{s}_T(\cdot)$ be an arbitrary estimator based on $\{X(t)\}_{0\leq t\leq T}$ and let

$$\hat{\theta}_T := T^{\frac{\alpha}{2\alpha+1}} g^{-1}(0)\left(\hat{s}_T(x_0) - s_0(x_0)\right).$$

Since $\theta = T^{\frac{\alpha}{2\alpha+1}} g^{-1}(0)(s_\theta(x_0) - s_0(x_0))$, 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 $\ell_0(u) := \ell(g(0)I_0^{-1}u)$, (A.1) becomes:

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

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

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

where

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

This implies (4.1) 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

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

Taking the lim inf as $T \to \infty$ on both sides, we obtain (4.1) since ε is arbitrary. □

Proof of Corollary 4.2.

(i) Following the same reasoning as in Theorem 4.1, we first prove that for any family of estimators $\{\hat{s}_T\}_{T>0}$ and arbitrary points $\{x_T\}_{T>0} \subset (a,b)$,

$$\liminf_{T\to\infty} \sup_{s\in\Theta} \mathbb{E}_s \left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_T(x_T) - s(x_T)\right)\right)\right] \geq C \tag{A.4}$$

for some constant $C > 0$, which is independent of the family of estimators and of the points. Fix a Lévy density $s_0 \in \Theta_\alpha(L/2;[a,b])$ such that $s_0(x) > 0$ for all $x \in \mathbb{R}_0 := \mathbb{R}\backslash\{0\}$, and a constant $\kappa > 0$. Again, let $g : \mathbb{R} \to \mathbb{R}_+$ be a symmetric function with compact support $\mathbb{K}_g$, satisfying (3.5) with $L/2$ (instead of L). Moreover, for any $y \in (a,b)$, the support of $x \to g(\kappa(x-y))$ does not contain the origin and

$$s_0(x) - \kappa^{-\alpha} g\left(\kappa(x-y)\right) > 0, \quad \forall x \in \mathbb{R}_0.$$

Let

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

for $|\theta| < \kappa^{-\alpha}$. Let $\mathbb{P}_\theta^{(T)}$ be the distribution (on $\mathbb{D}[0,T]$) of a Lévy process $\{X(t)\}_{0\leq t\leq T}$ with Lévy density $s_{\theta,T}$. Following the proof of Theorem 4.1, $\{\mathbb{P}_\theta^{(T)} : \theta \in (-\kappa^{-\alpha}, \kappa^{-\alpha})\}$ is LAN at $\theta = 0$ with the normalizing constants

$$\varphi_T := \kappa^2 \left(\int_{\mathbb{K}_g} s_0^{-1}(\kappa^{-1}T^{-\frac{1}{2\alpha+1}}u + x_T) g^2(u)\, du\right)^{-2},$$

where $\mathbb{K}_g$ denotes the support of g and it is being assumed that $[-1,1] \cap \bigcup_{y\in[a,b]}\{y + \kappa^{-1}\mathbb{K}_g\} = \emptyset$. Observe that there is an $m > 0$ for which $\inf_{T\geq 1}\varphi_T \geq m$.

(ii) By Theorem II.12.1 and Remark II.12.2 in Ibragimov & Has'minskii (1981), for any $\delta > 0$,

$$\liminf_{T\to\infty} \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta \left[\ell_0\left(\varphi_T^{-1}(\hat{\theta}_T - \theta)\right)\right] \geq C, \tag{A.5}$$

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

$$\liminf_{T\to\infty} \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta \left[\ell_0\left(m^{-1}(\hat{\theta}_T - \theta)\right)\right] \geq C. \tag{A.6}$$

Now, setting,

$$\hat{\theta}_T := T^{\frac{\alpha}{2\alpha+1}} g^{-1}(0)\left(\hat{s}_T(x_T) - s_0(x_T)\right),$$

it follows that

$$\sup_{s\in\Theta} \mathbb{E}_s \left[\ell\left(T^{\frac{\alpha}{2\alpha+1}}\left(\hat{s}_T(x_T) - s(x_T)\right)\right)\right] \geq \sup_{|\theta|<\delta\varphi_T} \mathbb{E}_\theta \left[\ell\left(g(0)(\hat{\theta}_T - \theta)\right)\right].$$

Taking $\liminf$ as $T \to \infty$, (A.4) is obtained with

$$C = 2^{-3/2}\pi^{-1/2} \int_{|z|<\delta/2} \ell(g(0)\, m\, z) e^{-z^2/2} dz. \tag{A.7}$$

(iii) To obtain (4.2), for each $\varepsilon > 0$, let $\hat{s}_T^{(\varepsilon)} \in \Theta$ and $x_T^{(\varepsilon)} \in (a,b)$ be such that

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

Next, take the lim inf as $T \to \infty$ on both sides above and apply (A.4).

(iv) We now prove (4.4). Fix a measurable estimator $\hat{s}_T$ and a $s \in \Theta$. By Fubini's Theorem,

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

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

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

Then,

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

Letting $\varepsilon \to 0$, (4.4) becomes a consequence of (4.2) with $\ell(u) = u^2$. □

Appendix B: Figures

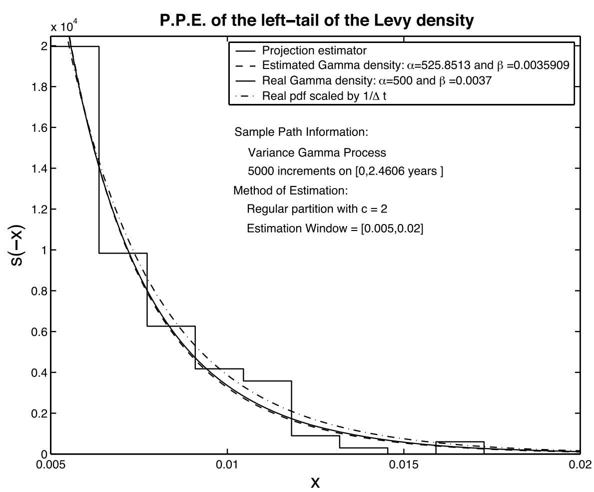

FIG 1. *Penalized projection estimation of the left-tail of the variance gamma Levy density.*

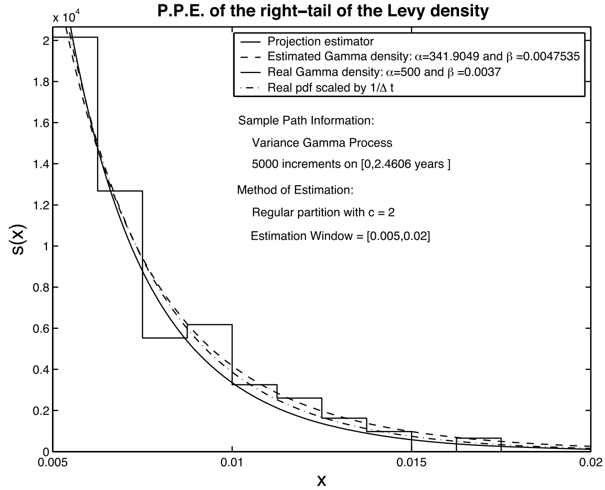

FIG 2. *Penalized projection estimation of the right-tail of the variance Gamma Levy density.*

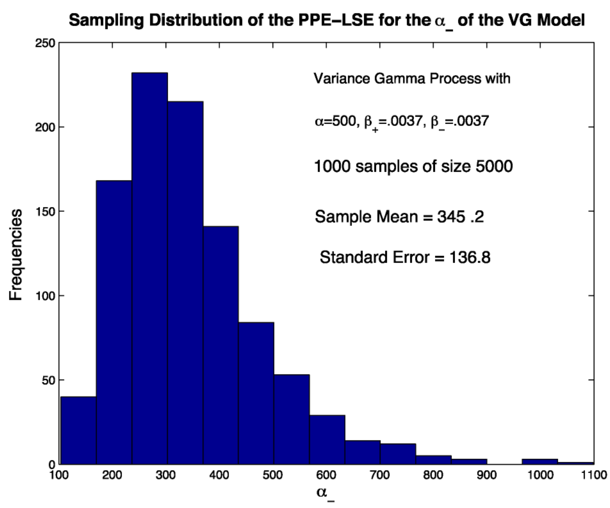

FIG 3. *Sampling Distribution for the Estimates of α^- obtained from the PPE and the LSE method.*

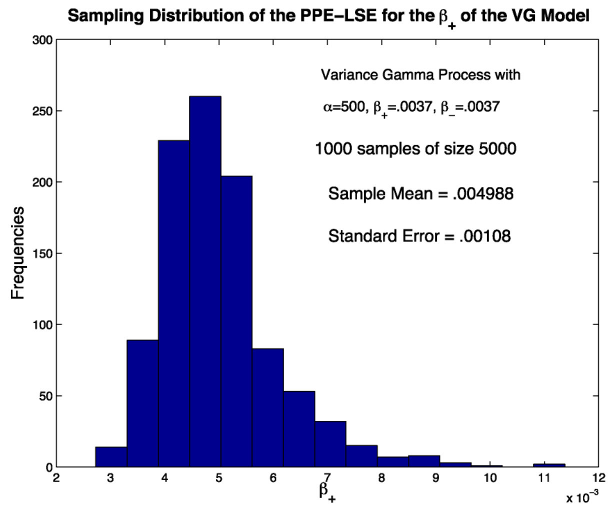

FIG 4. *Sampling Distribution for the Estimates of β^+ obtained from the PPE and the LSE method.*

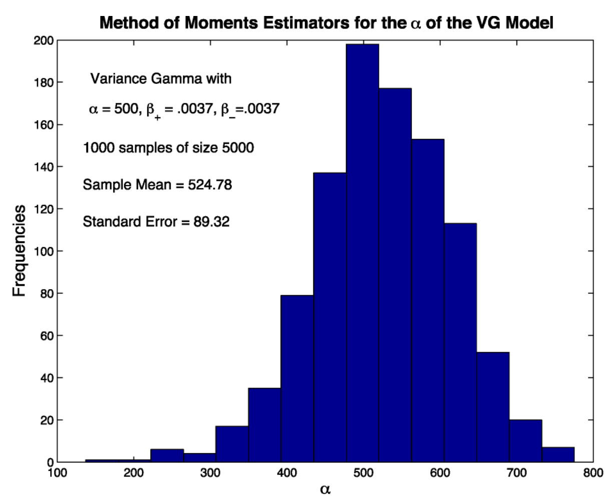

FIG 5. *Sampling Distribution for the Estimator of α obtained by the Method of Moments.*

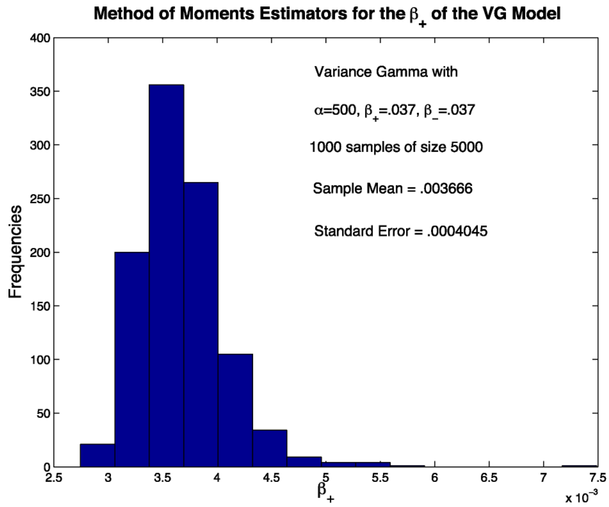

FIG 6. *Sampling Distribution for the Estimator of β^+ obtained by the Method of Moments.*

References

[1] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2001). Modelling by Lévy processess for financial economics. In *Lévy Processes. Theory and Applications* (Barndorff-Nielsen, O. E., Mikosch, T. and Resnick, S. I., eds.) 283–318. Birkhäuser, Boston, MA.

[2] BARRON A., BIRGÉ L., and MASSART P. (1995). Model selection via penalization. Technical Report 54, Université Paris-Sud.

[3] BARRON A., BIRGÉ L., and MASSART P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.

[4] BIRGÉ L., and MASSART P. (1994). Minimum contrast estimation on sieves. Technical Report 34, Université Paris-Sud.

[5] BIRGÉ L., and MASSART P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam* 55–87. Springer, New York.

[6] CARR P., GEMAN, H., MADAN, D. and YOR M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business* 305–332.

[7] CARR P., GEMAN, H., MADAN, D. and YOR M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance* **13** 345–382.

[8] CARR P., MADAN, D., and CHANG E. (1998). The variance Gamma process and option pricing. *European Finance Review* **2** 79–105.

[9] CARR P., and WU, L. (2005). Time-changed Levy processes and option pricing. *Journal of Financial Economics* **71** 113–141.

[10] CHUNG K. L. (2001). *A course in Probability Theory.* Academic Press, San Diego, CA.

[11] CONT R. and TANKOV P. (2003). *Financial modelling with Jump Processes.* Chapman & Hall.

[12] DEVORE R. A., and LORENTZ G. G. (1993). *Constructive Approximation.* Springer, Berlin.

[13] FIGUEROA-LÓPEZ J. E. (2004). Nonparametric estimation of Lévy processes with a view towards mathematical finance. PhD Thesis. Georgia Institute of Technology, Atlanta, GA 30332, April 2004. Available at `http://etd.gatech.`

`edu.No.etd-04072004-122020`.

[14] Figueroa-López J. E. (2007). On the non-parametric estimation of some time-changed Lévy models. In preparation.

[15] Figueroa-López J. E. (2008). Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.* DOI:10.1016/j.spl.2008.07.012.

[16] Figueroa-López J. E. and Houdré C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. *High Dimensional Probability. IMS Lecture Notes - Monograph Series* **51** 96–116.

[17] Figueroa-López J. E. and Houdré C. (2008). Small-time expansions for the transition distributions of Lévy processes. Available at arXiv:0809.0849v1 [math.PR].

[18] Grenander U. (1981). *Abstract Inference*. Wiley, New York.

[19] Ibragimnov I. A. and Has'minskii R. Z. (1981). *Statistical Estimation. Asymptotic Theory*. Springer, New York.

[20] Jacod J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM:P&S* **11** 173–196.

[21] Knotz S., Kozubowski T. J. and Podgórski K. (2001). *The Laplace Distribution and Generalizations: A revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhauser, Boston.

[22] Kutoyants Y. A. (1998). *Statistical Inference for Spatial Poisson Processes*. Springer, New York.

[23] Madan D. B. and Seneta E. (1990). The variance Gamma model for share market returns. *Journal of Business* **63** 511–524.

[24] Prause K. (1999). The generalized hyperbolic model: estimation, financial derivatives, and risk measures. Ph.D. thesis, Univ. Freiburg.

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

[26] Rosiński J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy processes-Theory and Applications* 401–415. Birkhäuser, Boston, MA.

[27] Rosiński J. (2007). Tempering stable processes. *Stochastic Process. Appl.* **117**, 677–707.

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

[29] Sato K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Univ. Press.

[30] Schoutens W. (2003). *Lévy Processes: Pricing Financial Derivatives*. Wiley.

[31] Todorov V. (2005). Econometric analysis of jump-driven stochastic volatility models. Working paper, Duke Univ.

[32] Woerner J. (2003). Variational sums and power variation: A unifying approach to model selection and estimation in semimartingale models. *Statistics and Decisions* **21** 47–68.