Sieve-based confidence intervals and bands for Lévy densities

José E. Figueroa-López*

Purdue University
Department of Statistics

West Lafayette, IN 47906
E-mail: figueroa@stat.purdue.edu

Abstract: A Lévy process combines a Brownian motion 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 considered here under a discrete-sampling scheme. In that case, the jumps are latent variables which statistical properties can be assessed when the frequency and time horizon of observations increase to infinity at suitable rates. Nonparametric estimators for the Lévy density based on *Grenander's method of sieves* had been proposed in [11]. In this paper, central limit theorems for these sieve estimators, both point-wise and uniform on an interval away from the origin, are obtained, leading to point-wise confidence intervals and bands for the Lévy density. In the point-wise case, we find feasible estimators which converge to s at a rate that is arbitrarily close to the rate of the minimax risk of estimation for smooth Lévy densities. We determine how frequently one needs to sample to attain the desired rate. In the case of uniform bands and discrete regular sampling, our results are consistent with the case of density estimation, achieving a rate of order arbitrarily close to $\log^{-1/2}(n) \cdot n^{-1/3}$, where n is the number of observations. The rate is valid provided that s is smooth enough, and that the time horizon T_n and the dimension of the sieve are appropriately chosen in terms of n.

Keywords and phrases: confidence bands, confidence intervals, Lévy processes, nonparametric estimation, model selection, sieve estimators.

1. Introduction

1.1. Motivation and some background

In the past 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., [9] and [13] for further information

*Research partially supported by the NSF grant # DMS 0906919. The author is indebted to the referee and editor for their many suggestions that improve the paper considerably. It is also a great pleasure to thank Professor David Mason for pointing out the KMT inequality to me and for other important remarks. The author would also like to thank Professor Jayanta Ghosh and participants of the Workshop on Infinitely Divisible Processes (CIMAT A.C. March 2009) for their helpful feedback.

about the field). The simplest of these models postulates that the price of a commodity (say a stock) at time t is given as an exponential function of a Lévy process $X := \{X_t\}_{t \geq 0}$. 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 several fundamental empirical features commonly observed in time series of asset returns such as heavy tails, high-kurtosis, and asymmetry. Lévy processes, as models capturing some of the most important features of returns and as "first-order approximations" to other more accurate models, are fundamental in developing and testing successful statistical methodologies. 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. Any Lévy process can be constructed from 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, $Y_t - \mathbb{E}\, Y_t$, goes to infinity while simultaneously allowing jumps of smaller sizes. Formally, X admits a decomposition of the form

$$X_t = bt + \sigma B_t + \lim_{\varepsilon \searrow 0} \int_0^t \int_{\varepsilon \leq |x| \leq 1} x\, (\mu - \bar{\mu})(dx, ds) + \int_0^t \int_{|x|>1} x\, \mu(dx, ds), \quad (1.1)$$

where B is a standard Brownian motion and μ is an independent Poisson measure on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ with mean measure $\bar{\mu}(dx, dt) := \nu(dx)dt$. Thus, 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 $\nu(A)$ gives the average number of jumps (per unit time) whose magnitudes fall in a given set $A \in \mathcal{B}(\mathbb{R})$. 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.

1.2. The statistical problem and methodology

We are interested in estimating, in a non-parametric fashion, the Lévy density s over a window of estimation $D := [a, b] \subset \mathbb{R}\backslash\{0\}$, based on discrete observations of the process on a finite interval $[0, T]$. In general, s can blow up around the origin, and hence, we consider only domains D that are "separated" from the origin, in the sense that $D \cap (-\varepsilon, \varepsilon) = \emptyset$ for some $\varepsilon > 0$. If the whole path of the process were available (and hence, the jumps of the process would be observable), 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. Lamentably, under discrete-sampling, the times and magnitudes of jumps are latent (unobservable) variables. Nevertheless, it is expected that the statistical property of the jumps can be inferred when the frequency and time horizon of observations increase to infinity, which is precisely the sampling scheme we adopt in this paper.

Non-parametric estimators for the Lévy density were proposed in [14], under continuous-sampling of the process, and in [11], under discrete-sampling, using the *method of sieves*. The method of sieves was originally proposed by Grenander [17] and applied more recently by Birgé, Massart, and others (see e.g. [1] & [4]) to 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" metric d between functions, relative to which the best approximation of s in a given linear model $\mathcal{S}$ will be characterized. That is, the best approximation $s^{\perp}$ of s on $\mathcal{S}$ is given by $d(s, s^{\perp}) = \inf_{p \in \mathcal{S}} d(s, p)$. Finally, devise an estimator $\hat{s}$, called the *projection estimator*, for the best approximation $s^{\perp}$ of s in $\mathcal{S}$.

The sieves considered here are of the general form

$$\mathcal{S} := \{\beta_1 \varphi_1 + \cdots + \beta_d \varphi_d : \beta_1, \ldots, \beta_d \in \mathbb{R}\}, \tag{1.2}$$

where $\varphi_1, \ldots, \varphi_d$ are orthonormal functions with respect to the inner product $\langle p, q \rangle_D := \int_D p(x) q(x) dx$. In the sequel, $\| \cdot \| := \| \cdot \|_D$ stands for the associated norm $\langle \cdot, \cdot \rangle_D^{1/2}$ on $\mathbb{L}^2(D, dx)$. We recall that, 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 given by

$$s^{\perp}(x) := \sum_{j=1}^{d} \beta(\varphi_j) \varphi_j(x), \tag{1.3}$$

where $\beta(\varphi_j) := \langle \varphi_j, s \rangle_D = \int_D \varphi_j(x) s(x) dx$. Thus, under this setting, the method of sieves boils down to estimate the functional

$$\beta(\varphi) = \int_D \varphi(x) s(x) dx,$$

for certain functions φ. In Section 3, we propose estimators for $\beta(\varphi)$ and as a byproduct, we develop projection estimators $\hat{s}$ on $\mathcal{S}$.

Following [11], we specialize further our approach and take *regular piece-wise polynomials* as sieves, though similar results will hold true if we take other typical classes of sieves such as smooth splines, trigonometric polynomials, or wavelets. For future reference, let us formally define the sieves.

Definition 1.1. *$\mathcal{S}_{k,m}$ stands for the class of functions φ such that for each $i = 0, \dots, m-1$, there exists a polynomial $q_{i,k}$ of degree at most k satisfying that $\varphi(x) = q_{i,m}(x)$ for all x in $(x_{i-1}, x_i]$, where $x_i = a + i(b-a)/m$.*

It is easy to built an orthonormal basis for $\mathcal{S}_{k,m}$ using the orthonormal Legendre polynomials $\{Q_j\}_{j\geq 0}$ on $\mathbb{L}^2([-1,1], dx)$. Indeed, the functions

$$\hat{\varphi}_{i,j}(x) := \sqrt{\frac{2j+1}{x_i - x_{i-1}}}\, Q_j\left(\frac{2x - (x_i + x_{i-1})}{x_i - x_{i-1}}\right) \mathbf{1}_{[x_{i-1}, x_i)}(x), \tag{1.4}$$

with $i = 1, \dots, m$ and $j = 0, \dots, k$, form an orthonormal basis for $\mathcal{S}_{k,m}$. For future reference, let us recall that

$$|Q_j(x)| \leq 1, \text{ and } |Q_j'(x)| \leq Q_j'(1) = \frac{j(j+1)}{2}. \tag{1.5}$$

We now review a few points of [11] in order to motivate the results in this paper. It is proved in [11] that by appropriately choosing the number of classes m and the sampling frequency high enough (both choices determined in function of the time horizon T), the resulting projection estimator on $\mathcal{S}_{m,k}$ attains the same rate of convergence in T as the minimax risk on a certain class Θ of smooth functions. Specifically, the referred minimax risk, defined by

$$\inf_{\hat{s}_T} \sup_{s\in\Theta} \mathbb{E}_s\left[\int_a^b (\hat{s}_T(x) - s(x))^2\, dx\right], \tag{1.6}$$

where infimum is over all estimators $\hat{s}_T$ based on $\{X_t\}_{t\leq T}$, converges to 0 at a rate $O(T^{-2\alpha/(2\alpha+1)})$ as $T \to \infty$ (see [11, Theorem 4.2]). The parameter α characterizes the smoothness of the Lévy densities $s \in \Theta$ on the interval $[a,b]$, in that if s is r-times differentiable on (a,b) $(r = 0, \dots)$ and

$$|s^{(r)}(x) - s^{(r)}(y)| \leq L|x-y|^\kappa, \tag{1.7}$$

for all $x, y \in (a,b)$ and some $L < \infty$ and $\kappa \in (0,1]$, then the smoothness parameter of s is $\alpha := r + \kappa$. We show in [11, Proposition 3.5] that there exists a critical mesh $\delta_T > 0$ such that if the time span between consecutive sampling observations is at most δ_T and $m_T := \left[T^{1/(2\alpha+1)}\right]$, then the resulting projection estimator, denoted by $\widetilde{s}_T$, is such that

$$\limsup_{T\to\infty} T^{\frac{2\alpha}{2\alpha+1}} \sup_{s\in\Theta} \mathbb{E}\,\|s - \widetilde{s}_T\|^2 < \infty. \tag{1.8}$$

Of course, an "explicit" estimate of δ_T is necessary for practical reasons. In Section 2, we show that it suffices $\delta_T = O(T^{-1})$, improving a former result in [11] (see Proposition 3.7 in there).

Notice that the convergence in (1.8) is in the integrated mean-square sense. A natural question, that we consider in this paper, is whether or not one can devise projection estimators $\hat{s}_T$ on $\mathcal{S}_{k,m}$ such that

$$T^{\alpha/(2\alpha+1)}\left(\hat{s}_T(x) - s(x)\right) \xrightarrow{\mathfrak{D}} \bar{\sigma}(x) Z, \tag{1.9}$$

holds for a standard Normal random variable Z, for each fixed $x \in D$. We were unable to attain (1.9) due to the fact that the bias of the estimator $\hat{s}_{T}$, namely $\mathbb{E}\,\hat{s}_{T}(x) - s(x)$, is just $O(T^{-\alpha/(2\alpha+1)})$. However, for any $\beta < \frac{\alpha}{2\alpha+1}$, we can devise a projection estimator $\hat{s}_{T}^{\beta}$ such that

$$T^{\beta}\left(\hat{s}_{T}^{\beta}(x) - s(x)\right) \xrightarrow{\mathfrak{D}} \bar{\sigma}(x) Z. \tag{1.10}$$

The idea is to use "undersmoothing" to make the effect of bias negligible. Our results are in alignment with those obtained in other standard nonparametric problems, such as density estimation or functional regression, using local nonparametric methods such as kernel estimation (see for instance Hall [18]). We were unable to find a reference where undersmoothing is used in a global nonparametric method such as the sieves method, and hence, this could be an additional contribution of the results presented here.

An important extension of the point-wise central limit theorems is the development of global measures of deviation or asymptotic confidence bands for the Lévy density. In this paper, we establish these methods for piece-wise constant and piece-wise linear regular polynomials (though we believe the result holds true for a general degree), following ideas of the seminal work of Bickel and Rosenblatt [3]. There are some important differences, though, starting from the fact that Bickel and Rosenblatt considered kernel estimators for probability densities, while here we consider a global nonparametric method. In spite of these differences, our results are consistent with the case of density estimation, achieving a convergence rate of order arbitrarily close to $\log^{-1/2}(n) \cdot n^{-1/3}$, where n is the number of observations. Again, the rate is valid provided that the time horizon T_n and the dimension of the sieves is appropriately chosen.

The paper is structured as follows. In Section 2, we derive a short-term ergodic property of a Lévy process, which plays a fundamental role in our results. In Section 3, we introduce the projection estimators for the Lévy densities and show point-wise central limit theorems for them. The uniform case and the resulting confidence bands are developed in Section 4. Section 5 illustrates the performance of the projection estimators and confidence bands using a simulation experiment in the case of a variance gamma Lévy model. Finally, two appendices collect the technical details of our results.

2. The critical mesh and an useful small-time asymptotic result

The critical time span δ_{T}, required for the validity of (1.8), was characterized in [11] by the property that

$$\sup_{y \in D} \left| \frac{1}{\Delta}\, \mathbb{P}\left[X_{\Delta} \geq y\right] - \nu([y,\infty)) \right| < k\, \frac{1}{T}, \tag{2.1}$$

for all $0 < \Delta < \delta_T$, where k is a constant (independent of T and Δ). For practical reasons, an "explicit" estimate of this critical mesh is necessary. The following proposition shows that $\delta_{T} = T^{-1}$ suffices and serves as the fundamental property of Lévy processes used for the asymptotic theory developed in this paper. The proof of the proposition is provided in the Appendix A. See also [15] for related higher order polynomial expansions for $\mathbb{P}\left(X_t \geq y\right)$.

Proposition 2.1. *Suppose that the Lévy density s of X is Lipschitz in an open set D_0 containing $D = [a, b] \subset \mathbb{R}\backslash\{0\}$, and that $s(x)$ is uniformly bounded on $|x| > \delta$, for any $\delta > 0$. Then, there exists a $k > 0$ and $t_0 > 0$ such that for all $0 < t < t_0$*

$$\sup_{y\in D}\left|\frac{1}{t}\,\mathbb{P}\left[X_t \geq y\right] - \nu([y,\infty))\right| < k\,t. \tag{2.2}$$

3. Projection estimators and point-wise central limit theorem

Throughout this paper, we assume that the Lévy process $\{X_t\}_{t\geq 0}$ is being sampled over a time horizon $[0, T]$ at discrete times $0 = t^0_{_T} < \cdots < t^{n_T}_{_T} = T$. We also use the notation $\pi_{_T} := \{t^k_{_T}\}_{k=0}^{n_T}$ and $\bar{\pi}_{_T} := \max_k\{t^k_{_T} - t^{k-1}_{_T}\}$, where we will sometimes drop the subscript T. The following statistics are the main building blocks for our estimation:

$$\hat{\beta}^{\pi_T}(\varphi) := \frac{1}{T}\sum_{k=1}^{n_T}\varphi\left(X_{t^k_{_T}} - X_{t^{k-1}_{_T}}\right). \tag{3.1}$$

In the case of a quadratic function $\varphi(x) = x^2$, $\sum_{k=1}^{n_T}\varphi\left(X_{t^k_{_T}} - X_{t^{k-1}_{_T}}\right)$ is the so-called realized quadratic variation of the process. Thus, the statistics (3.1) can be interpreted as the realized $\varphi-$variation of the process per unit time based on the observations $X_{t^0_{_T}}, \ldots, X_{t^{n_T}_{_T}}$. The estimators (3.1) were proposed independently by Woerner [25] and Figueroa-López [10].

The main virtue of the statistics (3.1) lies on its application to recover $\beta(\varphi) := \int\varphi(x)s(x)dx$ as $T \to \infty$ and $\bar{\pi}_{_T} \to 0$ for bounded ν-continuous functions φ such that $\varphi(x) \to 0$ fast enough as $x \to 0$. This result was obtained in [25] (Theorem 5.1 there) for regular sampling schemes and in [12] (Proposition 2.2 there) for general sampling schemes and a more general class of functions φ (see also [11, Theorem 2.3] for related central limit theorems). The consistency of $\hat{\beta}^{\pi}(\varphi)$ for $\beta(\varphi)$ lead us to propose

$$\hat{s}^{\pi}(x) := \sum_{j=1}^{d}\hat{\beta}^{\pi}(\varphi_j)\varphi_j(x), \tag{3.2}$$

as a natural estimator for the orthogonal projection $s^{\perp}$ defined in (1.3). The nonparametric estimator (3.2) was proposed in [10], where the problem of model selection was also considered under continuous-time sampling.

As it was discussed in the introduction, one can construct a projection estimators $\widetilde{s}_{_T}$ on the regular piece-wise polynomials $\mathcal{S} = \mathcal{S}_{k,m}$ of Definition 1.1 that converges to s, under the integrated mean-square distance, at a rate at least as good as $T^{-2\alpha/(2\alpha+1)}$. Such a rate can be ensured by "tuning" the number of classes m in the sieve, as well as the sampling frequency $\bar{\pi}$, to both the degree of smoothness α of s and the time horizon T. It is natural to wonder if it is possible to construct a projection estimator $\hat{s}_{_T}$ such that

$$T^{\alpha/(2\alpha+1)}\left(\hat{s}_{_T}(x)-s(x)\right)\stackrel{\mathfrak{D}}{\longrightarrow}\bar{\sigma}Z,$$

as $T\to\infty$, for $Z\sim\mathcal{N}(0,1)$ and a constant $\bar{\sigma}$. We are unable to conclude this result due to the fact that the bias $\mathbb{E}\,\hat{s}_{_T}(x)-s(x)$ of any projection estimator $\hat{s}_{_T}$ is at the best $O(T^{-\alpha/(2\alpha+1)})$. However, in this section we show that for any $0<\beta<\frac{\alpha}{2\alpha+1}$, there exists a projection estimator $\hat{s}^{\beta}_{_T}$ such that

$$c'_{_T}\left(\hat{s}^{\beta}_{_T}(x)-s(x)\right)\stackrel{\mathfrak{D}}{\longrightarrow}\bar{\sigma}Z,$$

for a normalizing constant $c'_{_T}\asymp T^{\beta}$ (i.e. $\underline{k}\,T^{\beta}\leq c'_{_T}\leq\bar{k}\,T^{\beta}$, for some constants $\underline{k},\bar{k}\in(0,\infty)$ independent of T). As often, our approach consists of obtaining first a CLT for $\hat{s}(x)$ centered at $\mathbb{E}\,\hat{s}(x)$ with normalizing constants $c'_{_T}\asymp T^{\beta}$ and subsequently, making the bias $\mathbb{E}\,\hat{s}(x)-s(x)$ to be $o(c_{_T}^{-1})$. The CLT for $\hat{s}(x)$ follows from a classical CLT for row-wise independent arrays.

Below, Legrendre polynomials $\{Q_j\}_{j\geq 0}$ on $\mathbb{L}^2([-1,1],dx)$ are used to devise an orthonormal basis for the sieve $\mathcal{S}_{k,m}$ of Definition 1.1. Also, we consider Lévy densities s which restrictions to $D:=[a,b]$ belong to the Besov class $\mathcal{B}^{\alpha}_{\infty}(L^{\infty}([a,b]))$ (i.e. functions satisfying (1.7) with $r\in\mathbb{N}$ and $\kappa\in(0,1]$ such that $\alpha=r+\kappa$). The following is our main theorem of this section. Its proof is presented in the Appendix B.

Theorem 3.1. *Suppose that the Lévy density s of X satisfies the conditions of Proposition 2.1 and belongs to $\mathcal{B}^{\alpha}_{\infty}\left(L^{\infty}([a,b])\right)$ for some $\alpha\geq 1$. Let $c_{_T}$ be a normalizing constant and let $\hat{s}_{_T}$ be the projection estimator on $\mathcal{S}_{k,m_T}$ based on sampling times $\pi_{_T}$ such that the below conditions are satisfied:*

$$(i)\ c_{_T}\stackrel{T\to\infty}{\longrightarrow}\infty,\quad (ii)\ \frac{c_{_T}^2 m_{_T}}{T}\stackrel{T\to\infty}{\longrightarrow}1,\quad (iii)\ c_{_T}m_{_T}\bar{\pi}_{_T}\stackrel{T\to\infty}{\longrightarrow}0$$
$$(iv)\ c_{_T}m_{_T}^{-\alpha}\stackrel{T\to\infty}{\longrightarrow}0,\qquad (v)\ k\geq\alpha-1.$$

Then, for any fixed $x\in(a,b)$ for which $s(x)>0$,

$$\frac{c_{_T}}{b_{k,m_T}(x)}\left(\hat{s}_{_T}(x)-s(x)\right)\stackrel{\mathfrak{D}}{\longrightarrow}\bar{\sigma}(x)Z, \tag{3.3}$$

where

$$Z\sim\mathcal{N}(0,1),\quad \bar{\sigma}^2(x):=(b-a)^{-1}s(x),$$
$$b^2_{k,m}(x):=\sum_{j=0}^{k}(2j+1)\sum_{i=1}^{m}Q_j^2\left(\frac{2x-(x_i+x_{i-1})}{x_i-x_{i-1}}\right)\mathbf{1}_{[x_{i-1},x_i)}(x).$$

Also, for any fixed $0<\beta<\frac{\alpha}{2\alpha+1}$, the resulting projection estimator $\hat{s}_{_T}$ with $m_{_T}=\left[T^{1-2\beta}\right]$ is such that

$$\frac{T^{\beta}}{b_{k,m_T}(x)}\left(\hat{s}_{_T}(x)-s(x)\right)\stackrel{\mathfrak{D}}{\longrightarrow}\bar{\sigma}(x)Z,$$

provided that $\bar{\pi}_{T} = T^{-\gamma}$ *with* $\gamma > 1 - \beta$*.*

Remark 3.2.

1. *In view of (1.5),* $1 \leq b_{k,m} \leq \sum_{j=0}^{k}(2j+1)$ *and hence, the normalizing constant* $c'_{T} := c_{T}/b_{k,m_T} \asymp c_{T}$*. Also, note that* $b_{k,m} \equiv 1$ *in the piece-wise constant case* ($k = 0$)*.*
2. *Theorem 3.1 will allow us to construct approximate confidence intervals for* $s(x)$*. Concretely, the* $100(1-\alpha)\%-$*interval for* $s(x)$ *is approximately given by*

$$\hat{s}_{T}(x) \pm \frac{b_{k,m_T}(x)}{c_{T}(b-a)^{1/2}} \hat{s}_{T}^{1/2}(x) z_{\alpha/2},$$

where $z_{\alpha/2}$ *is the* $\alpha/2$ *normal quantile.*

4. Confidence bands for Lévy densities

In this part we address the problem of constructing confidence bands for the Lévy density s of a Lévy process using projection estimators $\hat{s}_{T}^{n}$ on $\mathcal{S}_{k,m}$ based on n evenly-spaced observations of the process at $t_0 = 0 < \cdots < t_n = T$ on $[0, T]$. Confidence bands entail the limit in distribution of the uniform norm

$$\|\hat{s}_{T}^{n} - s\|_{[a,b]} := \sup_{x \in [a,b]} |\hat{s}_{T}^{n}(x) - s(x)|,$$

but as before, we will first work with the uniform norm of

$$Y_{T}^{n}(x) := \hat{s}_{T}^{n}(x) - \mathbb{E}\,\hat{s}_{T}^{n}(x), \quad x \in [a,b], \tag{4.1}$$

and then, estimate the uniform norm of the bias $\mathbb{E}\,\hat{s}_{T}^{n}(x) - s(x)$. We follow ideas from the seminal paper of Bickel and Rosenblatt [3], which constructs confidence bands for probability densities based on kernel estimators. There are two fundamental general directrices in Bickel and Rosenblatt's approach:

(1) The statistics of interest are expressed in terms of the so-called uniform standardized empirical process

$$Z_{n}^{0}(x) := n^{1/2}\left\{F_{n}^{*}(x) - x\right\}, \quad x \in [0,1], \tag{4.2}$$

where, denoting F_t the distribution of X_t and $\delta^n := t_i - t_{i-1}$ the time span between observations, $F_{n}^{*}(\cdot)$ is the empirical distribution of $\{F_{\delta^n}(X_{t_i} - X_{t_{i-1}})\}_{i \leq n}$.

(2) The empirical process Z_{n}^{0} is approximated by a Brownian bridge Z^0 and the error is estimated using Brillinger's result [5] or the Komlós, Major, and Tusnády construction [19].

Once the statistic of interest is related with the Brownian bridge Z^0, we will carry several successive approximations (see Appendix C for the details), which will allows to connect the distribution of $\|Y_T^n\|_{[a,b]}$ with the limiting distribution of the extreme value

$$\bar{M}_m := \max_{1\leq j\leq m}\{\zeta_j^{(k)}\}$$

of independent copies $\{\zeta_j^{(k)}\}_j$ of the random variable

$$\zeta^{(k)} := \sup_{x\in[-1,1]}\left|\sum_{j=0}^{k}\sqrt{2j+1}\,Q_j(x)Z_j\right|, \tag{4.3}$$

where Z_j are i.i.d. standard normal random variables. The problem is then reduced to finding the extreme value distribution of a random sample from (4.3). For instance, in the case $k=0$, $\zeta_j^{(0)} \overset{i.i.d.}{\sim} |Z_0|$, which is known to satisfy that

$$\lim_{n\to\infty}\mathbb{P}\left(\max_{1\leq j\leq m}|\zeta_j^{(0)}| \leq \frac{y}{a_m}+b_m\right) = e^{-2e^{-y}}, \tag{4.4}$$

for any $y>0$, where

$$a_m = (2\log m)^{1/2} \tag{4.5}$$

$$b_m = (2\log m)^{1/2} - \frac{1}{2}(2\log m)^{-1/2}\left(\log\log m + \log 4\pi\right). \tag{4.6}$$

We are also able to tackle the case $k=1$, where $\zeta^{(1)} = |Z_0| + \sqrt{3}|Z_1|$, but the general case is still under research. The following are our assumptions:

Standing assumption 1.

1. *s is positive and continuous on $[a,b]$;*
2. *s is differentiable in (a,b) and moreover, the derivative of $s^{1/2}$ is bounded in absolute value on (a,b).*

We are ready to present our main result in this Section. We defer its proof to the Appendix C.

Theorem 4.1. *Suppose that $\nu(\mathbb{R})=\infty$ or $\sigma\neq 0$. Also, suppose that the Lévy density s satisfies the conditions of Proposition 2.1 and the standing assumptions 1. Let $T_n\to\infty$ and $m_n\to\infty$ be such that*

$$(i)\ \delta^n\log\delta^n\cdot m_n\log m_n \overset{n\to\infty}{\longrightarrow} 0, \quad (ii)\ \frac{\log^2 n}{T_n}\cdot m_n\log m_n \overset{n\to\infty}{\longrightarrow} 0.$$

where $\delta_n := T_n/n$. Then, for $k\in\{0,1\}$, the deviation process $Y_{T_n}^n$ of (4.1) satisfies that

$$\lim_{n\to\infty}\mathbb{P}\left(a_{m_n}\left\{\kappa\bar{T}_n^{1/2}\sup_{x\in[a,b]}|s^{-1/2}(x)Y_{T_n}^n(x)| - b_{m_n}\right\}\leq y\right) = e^{-\kappa' e^{-y}}, \tag{4.7}$$

where $\bar{T}_n := T_n/m_n$, a_m and b_m are defined as in (4.5-4.6), and $(\kappa,\kappa') = ((b-a)^{1/2},2)$ if $k=0$ or $(\kappa,\kappa') = ((b-a)^{1/2}2^{-1},4)$ if $k=1$.

The previous result shows that

$$a_{m_n}\left\{\kappa\,\bar{T}_n^{1/2}\sup_{x\in[a,b]}s^{-1/2}(x)|\hat{s}^n_{_{T_n}}(x)-\mathbb{E}\,\hat{s}^n_{_T}(x)|-b_{m_n}\right\}$$

converges to a Gumbel distribution. The final step to our construct confidence bands consists of finding conditions to replace $\mathbb{E}\,\hat{s}^n_{_T}$ with s. The following result shows this step. Its proof is presented in Appendix C.

Corollary 4.2. *Suppose that the conditions of Theorem 4.1 hold true, that the restriction of s to $[a,b]$ is a member of $\mathcal{B}^{\alpha}_{\infty}(L^{\infty}([a,b]))$, and also that*

$$(iii)\; T_n m_n^{1-2\alpha}\log^2 m_n \stackrel{n\to\infty}{\longrightarrow} 0. \tag{4.8}$$

Then,

$$\lim_{n\to\infty}\mathbb{P}\left(a_{m_n}\left\{\kappa\,\bar{T}_n^{1/2}\sup_{x\in[a,b]}\frac{1}{s^{1/2}(x)}\left|\hat{s}^n_{_{T_n}}(x)-s(x)\right|-b_{m_n}\right\}\leq y\right)=e^{-\kappa' e^{-y}}, \tag{4.9}$$

where we used the same notation for κ and κ' as in Theorem 4.1.

The previous corollary allows us to construct confidence bands for s on $[a,b]$ based on the projection estimators $\hat{s}$ on regular piece-wise linear (or constant) polynomials. Indeed, suppose that y^*_α is such that $\exp\{-k' e^{-y^*_\alpha}\}=1-\alpha$ and let

$$d_n := \frac{1}{\sqrt{2}\kappa}\left(\frac{y^*_\alpha}{a_{m_n}}+b_{m_n}\right)\bar{T}_n^{-1/2}.$$

Then, as $n\to\infty$,

$$s(x)\in\left(\hat{s}^n_{_{T_n}}(x)+\left\{d_n^2\pm\sqrt{(\hat{s}^n_{_{T_n}}(x)+d_n^2)^2-(\hat{s}^n_{_{T_n}}(x))^2}\right\}\right), \tag{4.10}$$

with $100(1-\alpha)\%$-confidence. The above interval is asymptotically equivalent to the following simpler interval:

$$s(x)\in\left(\hat{s}^n_{_{T_n}}(x)\pm\frac{1}{\kappa}\left(\frac{y^*_\alpha}{a_{m_n}}+b_{m_n}\right)\bar{T}_n^{-1/2}\left(\hat{s}^n_{_{T_n}}(x)\right)^{1/2}\right). \tag{4.11}$$

We finish this section with some final remarks.

Remark 4.3. *In the case that, $T_n := c_n\cdot n^{\alpha_1}$ and $m_n=[d_n\cdot n^{\alpha_2}]$, for some $\alpha_1,\alpha_2>0$, $c_n\asymp 1$, and $d_n\asymp 1$, the conditions (i)-(ii) of Theorem 4.1 are satisfied if $0<\alpha_1<1$ and $0<\alpha_2<(1-\alpha_1)\wedge\alpha_1$. Also, one can check that the condition (iii) of Corollary 4.2 is met if*

$$0<\alpha_1<\frac{2\alpha+1}{3\alpha+2},\qquad \text{and}\qquad \frac{\alpha_1}{1+2\alpha}<\alpha_2<(2-3\alpha_1)\wedge\alpha_1. \tag{4.12}$$

Note that $(\alpha_2-\alpha_1)/2$ can be made arbitrarily close to $-\alpha/(3\alpha+1)$ on the range of values (4.12) and thus, $a_{m_n}\bar{T}_n^{-1/2}$ can be made to vanish at a rate arbitrarily *close to* $(\log n)^{-1/2}\, n^{-\frac{\alpha}{3\alpha+1}}$, *provided that* α *is large enough. In particular, if* $0<\varepsilon<<1$ *and* s *is smooth enough, one can choose* m_n *and* T_n *such that*

$$\|\hat{s}^{n}_{T_n}-s\|_{[a,b]}=O\left(\log^{-1/2}(n)\, n^{-1/3+\varepsilon}\right).$$

5. A numerical example: estimation of variance Gamma processes.

Variance Gamma processes (VG) were proposed in Madan and Seneta [20] and Carr et. al. [7] as substitutes to the Brownian Motion in the Black-Scholes model. Since their introduction, VG processes have received a great dealt of attention, even in the financial industry. A variance Gamma process $X=\{X(t)\}_{t\geq 0}$ is a time-changed Brownian motion with drift of the form:

$$X(t)=\theta U(t)+\sigma W(U(t)), \tag{5.1}$$

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 such that $\mathrm{E}\,[U(t)]=t$ and $\mathrm{Var}\,[U(t)]=\nu t$. Since Gamma processes are *subordinators*, the process X is itself a Lévy process (see Theorem 30.1 of [23]) and its Lévy density takes the form

$$s(x)=\begin{cases}\frac{\alpha}{|x|}\exp\left(-\frac{|x|}{\beta^-}\right) & \text{if } x<0,\\ \frac{\alpha}{x}\exp\left(-\frac{x}{\beta^+}\right) & \text{if } x>0,\end{cases} \tag{5.2}$$

where $\alpha>0$, $\beta^-\geq 0$, and $\beta^+\geq 0$ with $|\beta^-|+|\beta^+|>0$ (see, e.g., [9] for expressions of $\beta_{\pm},\alpha$ in terms of θ, σ, and ν). In that case, 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.

The performance of projection estimation for the variance gamma Lévy process was illustrated in [11] via simulation experiments. In this part we want to extend further this analysis to show the performance of confidence bands. As in [11] , we take as sieve the class $\mathcal{S}_{0,m}$; namely, the span of the indicator functions $\chi_{[x_0,x_1]},\dots,\chi_{(x_{m-1},x_m]}$, where $x_0<\cdots<x_m$ is a regular partition of an interval $D\equiv[a,b]$, with $0<a$ or $b<0$. We take parameter values which are partially motivated by the empirical findings of [7] based on daily returns of the S&P500 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 one can easily find that

$$\hat{\alpha}=500,\quad \hat{\beta}^+=0.0037056,\quad\text{and}\quad \hat{\beta}^-=0.0037067. \tag{5.3}$$

These parameter values seem to be consistent with other empirical studies (see, e.g., Seneta [24]), though we admit that parameter values fitted to intraday high-frequency data would have been preferable.

We simulate 100 samples of the VG process with a maximal time horizon of $T = 10$ years and a sampling span between observations of $\delta = 1/(252 * 6.5 * 60 * 12)$. Assuming a business calendar year of 252 days and a trading day of 6.5 hours, the time span between observations corresponds to 5 seconds. Intraday data of such characteristics is available through financial database such as NASDAQ TAQ.

We estimate the sample coverage probabilities

$$c_\alpha := \mathbb{P}\left(s(\cdot) \in \text{the } 100(1-\alpha)\% \text{ confidence band on } [a,b]\right),$$

based on the 100 simulations for two sampling frequencies $\delta = 1/(252 * 6.5 * 60 * 12)$ (5 seconds) and $\delta = 1/(252 * 6.5 * 60)$ (1 minute), and maturities of $T = 1, 3, 5, 10$ years. We use two possible number of classes: $m = 40$ and the data-driven selected m proposed in [11]. Concretely, the selection criterion is given by

$$\hat{m} := \mathrm{argmin}_m \left\{ -\|\hat{s}_m^\pi\|^2 + \mathrm{pen}^\pi(\mathcal{S}_{k,m}) \right\}, \tag{5.4}$$

where $\hat{s}_m^\pi$ is given according to (3.2), and pen^π is given by

$$\mathrm{pen}^\pi(\mathcal{S}_{k,m}) = \frac{2}{T^2} \sum_{i=1}^{n} \sum_{i,j} \hat{\varphi}_{i,j}^2 (X_{t_i} - X_{t_{i-1}}). \tag{5.5}$$

The quantity to be minimized in (5.4) is a discrete-time version of an unbiased estimator of the shifted risk $\mathbb{E}\,\|s - \hat{s}_m^\pi\|^2 - \|s\|^2$ (see [11, Section 5] for more details).

The following table shows the coverage probabilities for the interval $[a, b] = [0.001, 0.1]$ (based on 100 simulations):

$\delta \backslash T$	1 year	3 year	5 year	10 year
5 sec	.97 (m=40)	.99 (m=40)	.97 (m=40)	.97 (m=40)
	.98 (m=35)	.95 (m=25)	.80 (m=25)	
1 min.	.93 (m=40)	.94 (m=40)	.98 (m=40)	.87 (m=40)
	.97 (m=35)	.75 (m=25)	.60 (m=25)	.94 (m=50)

Overall the coverage probabilities of the confidence bands for $m = 40$ are good. In the case of the data-driven selected m, there are some values of m for which probabilities are quite low. Such cases occur (only) when the band does not contain the density very near $a = .001$. It seems more reasonable to take an average between different classes values of m which are reasonably close in terms of the quantity in (5.5).

To illustrate how close the estimated Lévy density is to the true Lévy density and the overall width of the confidence bands, Figure 1 shows the actual Lévy density (solid blue line), the mean of the penalized projection estimator (solid red line), and the means of the lower and upper 95%-confidence bands (dashed lines). All the means are computed using 100 confidence bands based on $\delta = 5$ seconds and time horizons of $T = 3$ and $T = 10$ years. The analog figures with a sampling time span of $\delta = 1$ minute are shown in Figure 2. In our empirical results (not shown here for the sake of space), we found that high-frequency data is crucial to estimate the Lévy density near the origin. For instance, the confidence bands near the origin do not perform well when taking 30 minutes observations in a time period of 10 years. The table below gives the estimated coverage probabilities on the interval $[0.005, .2]$ based on 30-minute returns:

$\delta\backslash T$	1 year	3 year	5 year	10 year
30 min	.34 (m=40)	.73 (m=40)	.87 (m=40)	.97 (m=40)
	.43 (m=10)	.71 (m=35)	.85 (m=35)	.97 (m=25)

Let us finish with two remarks. First, from an algorithmic point of view, the estimation for the variance Gamma model using penalized projection is not different from the estimation of the Gamma Lévy 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 [7]), there are well-documented issues with MLE (see for instance [21]). Finally, it worth pointing out that applying an efficient estimation method to a misspecified model could lead to quite undesirable results as it was illustrated in [11], where MLE was applied to a CGMY model (see [6]) with parameter values quite close to those of a Gamma process. The numerical experiments in [11] show that sometime a modestly efficient robust nonparametric method is preferably to a very efficient estimation method.

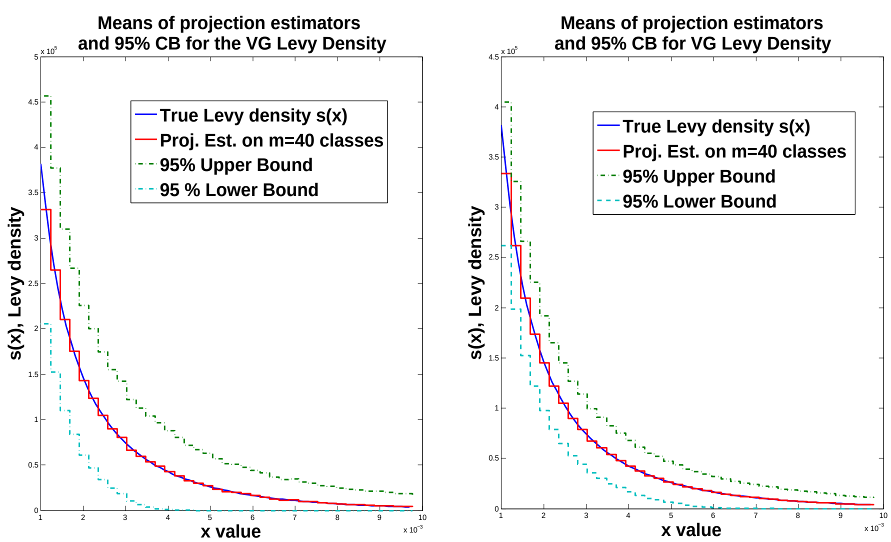

FIG 1. *Means of projection estimators and corresponding confidence bands for the VG model based on* 100 *simulations with a sampling time span of* $1/(252*6.5*60*12)$ *(about* 5 *seconds) during* 3 *years (left panel) and* 10 *years (right panel).*

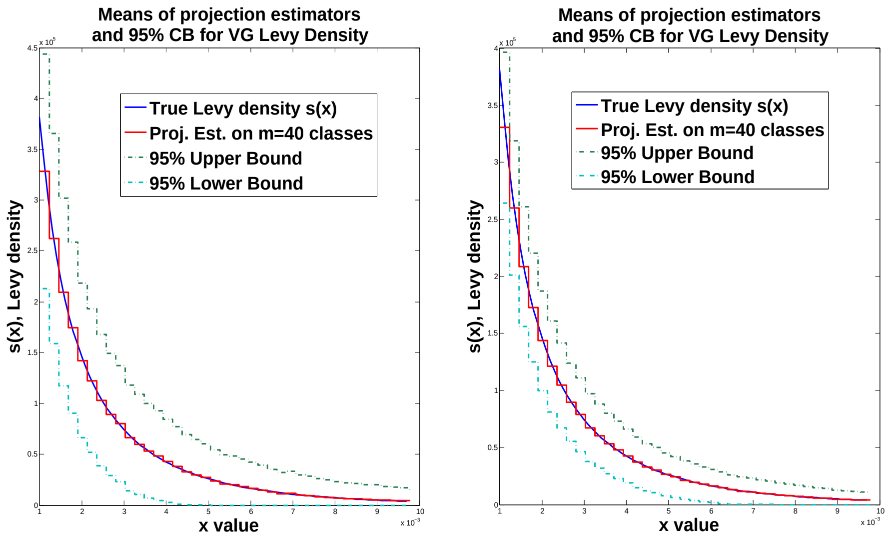

FIG 2. *Means of projection estimators and corresponding confidence bands for the VG model based on* 100 *simulations with a sampling time span of* $1/(252 * 6.5 * 60)$ *(about* 1 *minute) during* 3 *years (left panel) and* 10 *years (right panel).*

Appendix A: Proof of Proposition 2.1

Proof. Without lose of generality, we assume that $a > 0$. Consider the process

$$\widetilde{X}_t^{\varepsilon} := \int_0^t \int_{\mathbb{R}} x \mathbf{1}_{\{|x| \geq \varepsilon\}} \mu(dx, ds), \tag{A.1}$$

for $0 < \varepsilon < 1$, which is well-known to be a compound Poisson process with intensity of jumps $\lambda_\varepsilon := \nu(\{|x| \geq \varepsilon\})$ and jump distribution $\frac{1}{\lambda_\varepsilon} \mathbf{1}_{\{|x| \geq \varepsilon\}} \nu(dx)$. The remainder process, $X^\varepsilon := X - \widetilde{X}^\varepsilon$, is then a Lévy process with jumps bounded by ε. Concretely, X^ε has Lévy triplet $(\sigma^2, b_\varepsilon, \mathbf{1}_{\{|x| \leq \varepsilon\}} \nu(dx))$, where $b_\varepsilon = b - \int_{\varepsilon < |x| \leq 1} x \nu(dx)$. The following tail estimate will play an important role in the sequel:

$$\mathbb{P}\left(|X_t^\varepsilon| \geq z\right) \leq \exp\{\alpha z_0 \log z_0\} \exp\left\{\alpha z - \alpha z \log z\right\} t^{z\alpha}, \tag{A.2}$$

valid for an arbitrary, but fixed, positive real $\alpha \in (0, \varepsilon^{-1})$, and for any $t, z > 0$ such that $t < z_0^{-1} z$, where z_0 depends only upon α (see [22, Lemma 3.2] or [23, Section 26] for a proof).

Denote

$$A_y(t) := \frac{1}{t}\left\{\frac{1}{t}\mathbb{P}\left[X_t \geq y\right] - \nu([y, \infty))\right\},$$

which, for $\varepsilon < \frac{y}{2} \wedge 1$ and after conditioning on the number of jumps, can be written as follows

$$
\begin{aligned}
A_y(t) &= \frac{1}{t^2}\,\mathbb{E}\, f_y\left(X_t^{\varepsilon}\right) e^{-\lambda_\varepsilon t} + e^{-\lambda_\varepsilon t}\int_{|x|\geq\varepsilon} \frac{1}{t}\left\{\,\mathbb{E}\, f_y(X_t^{\varepsilon}+x) - f_y(x)\right\}\nu(dx)\\
&\quad - \frac{1-e^{-\lambda_\varepsilon t}}{t}\int_{x>y} f_y(x)\nu(dx) + e^{-\lambda_\varepsilon t}\sum_{n=2}^{\infty}\frac{(\lambda_\varepsilon)^n t^{n-2}}{n!}\,\mathbb{E}\, f_y(X_t^{\varepsilon}+\sum_{i=1}^{n}\xi_i),
\end{aligned}
$$

where $f_y(x) = \mathbf{1}_{x\geq y}$. The first term on the right hand side of the above expression is bounded uniformly for $y\in[a,b]$ and $t<t_0$, for certain $t_0(\alpha)>0$, because of (A.2) taking $z=a$ and $\alpha\in(2a^{-1},\varepsilon^{-1})$. The last two terms in the same expression are uniformly bounded in absolute value by $\nu(x\geq a)$ and $\nu(|x|\geq\varepsilon)^2$, respectively. We need to show that the second term is uniformly bounded. Define $B_y(t) := \int_{|x|\geq\varepsilon}\{\,\mathbb{E}\, f_y(X_t^{\varepsilon}+x) - f_y(x)\}\,\nu(dx)$. Clearly,

$$
\begin{aligned}
B_y(t) &:= \int_{y-\varepsilon}^{y}\mathbb{P}\left\{X_t^{\varepsilon}\geq y-x\right\}s(x)dx - \int_{y}^{y+\varepsilon}\mathbb{P}\left\{X_t^{\varepsilon}< y-x\right\}s(x)dx\\
&\quad + \int_{\{x<y-\varepsilon,|x|\geq\varepsilon\}}\mathbb{P}\left\{X_t^{\varepsilon}\geq y-x\right\}s(x)dx - \int_{y+\varepsilon}^{\infty}\mathbb{P}\left\{X_t^{\varepsilon}< y-x\right\}s(x)dx.
\end{aligned}
$$

Since s is bounded and integrable away from the origin, the last two terms in the expression for $B_y(t)$ can be bounded in absolute value by $\nu\{|x|\geq\varepsilon\}\,\mathbb{P}\left\{|X_t^{\varepsilon}|\geq\varepsilon\right\}$. Dividing by t, this converges to 0 in light of the well known limit

$$
\lim_{t\to 0}\frac{1}{t}\ \mathbb{P}\left(Z_t\geq z\right) = \nu([z,\infty)), \tag{A.3}
$$

valid for any Lévy process Z with Lévy measure ν and any point z of continuity of ν (see, e.g., Chapter 1 of Bertoin [2]). The other two terms can be bounded as follows:

$$
\begin{aligned}
&\left|\int_{y-\varepsilon}^{y}\mathbb{P}\left\{X_t^{\varepsilon}\geq y-x\right\}s(x)dx - \int_{y}^{y+\varepsilon}\mathbb{P}\left\{X_t^{\varepsilon}< y-x\right\}s(x)dx\right|\\
&\leq K_1\int_0^{\varepsilon}\mathbb{P}\left\{|X_t^{\varepsilon}|\geq u\right\}udu + K_0\left|\int_0^{\varepsilon}\mathbb{P}\left\{X_t^{\varepsilon}\geq u\right\}du - \int_0^{\varepsilon}\mathbb{P}\left\{X_t^{\varepsilon}<-u\right\}du\right|,
\end{aligned}
$$

where K_1 is the Lipschitz constant of s in D_0 and $K_0 := \sup_{x\in D_0}|s(x)|$. Next, applying Fubini, we can write the expression in the last line above as follows:

$$
K_1\frac{1}{2}\,\mathbb{E}\,\left\{(|X_t^{\varepsilon}|\wedge\varepsilon)^2\right\} + K_0\left|\,\mathbb{E}\, h(X_t^{\varepsilon})\right|,
$$

where $h(x) = x\mathbf{1}_{|x|\leq\varepsilon} - \varepsilon\mathbf{1}_{x<-\varepsilon} + \varepsilon\mathbf{1}_{x>\varepsilon}$. Using the formulas for the variance and mean of a Lévy process, we obtain that

$$
\sup_{0<t\leq 1}\frac{1}{t}\,\mathbb{E}\,\left\{(|X_t^{\varepsilon}|\wedge\varepsilon)^2\right\} \leq \sigma^2 + \int_{|x|\leq\varepsilon}x^2\nu(dx) + b_\varepsilon < \infty.
$$

Also,

$$\left|\frac{1}{t}\,\mathbb{E}\,h(X_t^{\varepsilon})\right| \leq \left|\frac{1}{t}\,\mathbb{E}\,X_t^{\varepsilon}\right| + \left|\frac{1}{t}\,\mathbb{E}\,X_t^{\varepsilon}\,\mathbf{1}_{\{|X_t^{\varepsilon}|>\varepsilon\}}\right| + \varepsilon\frac{1}{t}\,\mathbb{P}\left\{|X_t^{\varepsilon}|>\varepsilon\right\}.$$

The last term above converges to 0 by (A.2). The second term also vanishes since

$$\frac{1}{t}\left|\mathbb{E}\,X_t^{\varepsilon}\,\mathbf{1}_{\{|X_t^{\varepsilon}|>\varepsilon\}}\right| \leq \left\{\frac{1}{t}\,\mathbb{P}\left\{|X_t^{\varepsilon}|>\varepsilon\right\}\right\}^{1/2}\left\{\frac{1}{t}\,\mathbb{E}\,\left(X_t^{\varepsilon}\right)^2\right\}^{1/2} \to 0,$$

as $t \to 0$. Finally, using the formula for the mean of X_t^{ε}:

$$\lim_{t\to 0}\frac{1}{t}\,\mathbb{E}\,h(X_t^{\varepsilon}) \leq \lim_{t\to 0}\frac{1}{t}\left|\mathbb{E}\,X_t^{\varepsilon}\right| = |b_{\varepsilon}|\,.$$

We conclude that there exists a t_0 and $K > 0$ such that for $t \leq t_0$, $\sup_{y\in D}|B_y(t)|/t \leq K$. This finish the proof since all other terms in $A_y(t)$ can be easily bounded uniformly in D. □

Appendix B: Proofs of the point-wise CLT

Throughout this section, we shall use the orthonormal basis $\{\hat{\varphi}_{i,j}\}_{1\leq i\leq m, 0\leq j\leq k}$ of (1.4). We start our proof with following easy lemma.

Lemma B.1. *Suppose that φ has support $[c,d] \subset \mathbb{R}_+\backslash\{0\}$, where φ is continuous with continuous derivative. Then,*

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

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

Proof. The result is clear from the below identities

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

$$\int\varphi(x)\nu(dx) = \varphi(c)\nu\left([c,\infty)\right)) + \int_c^{\infty}\varphi'(u)\nu\left([u,\infty)\right)du,$$

which are standard consequences of Fubini. □

Our first result shows a CLT for $\hat{s}(x)$ centered at $\mathbb{E}\,\hat{s}(x)$. Let us remark that the fact that the Legendre polynomial Q_j is not constant for $j > 0$ poses some difficulty, since the relative position of x inside its class changes greatly with m.

Lemma B.2. *Under the notation and assumptions of Theorem 3.1, it follows that*

$$\frac{c_{T}}{b_{k,m_{T}}(x)}\left(\hat{s}_{T}(x) - \mathbb{E}\,\hat{s}_{T}(x)\right) \xrightarrow{\mathfrak{D}} \bar{\sigma}Z.$$

Proof. We apply a CLT version for row wise independent arrays of random variables (see e.g. the Corollary following Theorem 7.1.2 in [8]). Note that

$$S_T := \frac{c_T}{b_{m_T}} \left(\hat{s}_T(x) - \mathbb{E}\, \hat{s}_T(x) \right)$$
$$= \frac{c_T}{T b_{m_T}} \sum_i \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x) \left\{ \widetilde{\varphi}_{j,T}(X_{t_T^i} - X_{t_T^{i-1}}) - \mathbb{E}\, \widetilde{\varphi}_{j,T}(X_{\Delta_T^i}) \right\},$$

where $\widetilde{\varphi}_{j,T}(\cdot)$ is of the form

$$\sqrt{\frac{2j+1}{b_T - a_T}}\, Q_j \left(\frac{2 \cdot - (a_T + b_T)}{b_T - a_T} \right) \mathbf{1}_{[a_T, b_T)}(\cdot),$$

with a_T, b_T satisfying that $x \in [a_T, b_T)$ and $b_T - a_T = (b-a)/m_T$. In that case, $\bar{\sigma}_T^2 := \operatorname{Var} S_T$ is given by

$$\bar{\sigma}_T^2 := \frac{c_T^2}{T^2 b_{m_T}^2} \sum_i \sum_{j_1, j_2 = 0}^{k} \widetilde{\varphi}_{j_1,T}(x) \widetilde{\varphi}_{j_2,T}(x) \operatorname{Cov}\left(\widetilde{\varphi}_{j_1,T}(X_{\Delta_T^i}), \widetilde{\varphi}_{j_2,T}(X_{\Delta_T^i}) \right), \tag{B.1}$$

where we wrote $\Delta_T^i := t_T^i - t_T^{i-1}$. Let us analyze the above covariances, scaled by Δ_T^i. First, applying Lemma B.1, (1.5), and Proposition 2.1, there exists a $t_0 > 0$ and $K > 0$ such that whenever $\Delta < t_0$,

$$\left| \frac{1}{\Delta} \mathbb{E}\, \widetilde{\varphi}_{j_1,T}(X_\Delta) \widetilde{\varphi}_{j_2,T}(X_\Delta) - \int \widetilde{\varphi}_{j_1,T}(y) \widetilde{\varphi}_{j_2,T}(y) s(y) dy \right| \leq \frac{K\Delta}{b_T - a_T}.$$

Similarly, using also that $\left| \int \widetilde{\varphi}_{j,T}(y) s(y) dy \right| \leq \|s\|$, there exists a $t_0 > 0$ and $K > 0$ such that whenever $\Delta < t_0$,

$$\left| \frac{1}{\Delta} \mathbb{E}\, \widetilde{\varphi}_{j_1,T}(X_\Delta)\, \mathbb{E}\, \widetilde{\varphi}_{j_2,T}(X_\Delta) \right| \leq K\Delta.$$

Thus, using the assumption (iii) of Theorem 3.1,

$$\frac{1}{\Delta_T^i} \operatorname{Cov}\left(\widetilde{\varphi}_{j_1,T}(X_{\Delta_T^i}), \widetilde{\varphi}_{j_2,T}(X_{\Delta_T^i}) \right) = o_T(1) + \int \widetilde{\varphi}_{j_1,T}(x) \widetilde{\varphi}_{j_2,T}(y) s(y) dy,$$

where $o_T(1) \to 0$ uniformly in i, as $T \to \infty$. Thus, in view that $b_{m_T} \geq 1$, (1.5), and assumption (ii) of Theorem 3.1, $\bar{\sigma}_T^2 - \hat{\sigma}_T^2 \stackrel{T\to\infty}{\longrightarrow} 0$, where

$$\hat{\sigma}_T^2 := \frac{c_T^2}{T b_{m_T}^2} \sum_{j_1, j_2 = 0}^{k} \widetilde{\varphi}_{j_1,T}(x) \widetilde{\varphi}_{j_2,T}(x) \int \widetilde{\varphi}_{j_1,T}(y) \widetilde{\varphi}_{j_2,T}(y) s(y) dy.$$

Next, the continuity of s at x, assumption (ii) of Theorem 3.1, and the fact that the support of $\widetilde{\varphi}_{j,T}$ contains x and shrinks to 0, yield that

$$\lim_{T\to\infty} \frac{c_T^2}{T b_{m_T}^2} \sum_{j_1, j_2 = 0}^{k} \widetilde{\varphi}_{j_1,T}(x) \widetilde{\varphi}_{j_2,T}(x) \int \widetilde{\varphi}_{j_1,T}(y) \widetilde{\varphi}_{j_2,T}(y) (s(y) - s(x)) dy = 0.$$

This implies that $\lim_{T\to\infty}\hat{\sigma}_T^2 = \lim_{T\to\infty}\bar{\sigma}_T^2 = s(x)/(b-a)$, in view of condition (ii) and the definition of b_k. Finally, we consider the "standardized" sum $Z_T := S_T/\bar{\sigma}_T$. By the Corollary following Theorem 7.1.2 in [8], Z_T will converge to $\mathcal{N}(0,1)$ because

$$\sup_i \frac{c_T}{T\bar{\sigma}_T b_{m_T}} \sum_{j=0}^{k} \left| \widetilde{\varphi}_{j,T}(x)\widetilde{\varphi}_{j,T}(X_{t_T^i} - X_{t_T^{i-1}}) \right| \leq \frac{c_T m_T}{T\bar{\sigma}_T b_{m_T}(b-a)} \to 0,$$

as $T\to\infty$ in view of assumptions (i)-(ii) and the fact that $b_m \geq 1$. This implies the proposition since $\bar{\sigma}_T^2 \to s(x)(b-a)^{-1}$. □

The last step is to estimate the rate of convergence of the bias term:

Lemma B.3. *Under the notation and assumptions of Theorem 3.1,* $\mathbb{E}\,\hat{s}_T(x) - s(x) = o(b_{m_T}/c_T)$*, as* $T\to\infty$*, for any fixed* $x\in(a,b)$ *for which* $s(x)>0$*.*

Proof. We use the same notation as in the proof of Lemma B.2. Obviously,

$$\frac{c_T}{b_{m_T}} \left| \mathbb{E}\,\hat{s}_T(x) - s(x) \right| \leq \frac{1}{T}\sum_i \Delta_T^i A_T(\Delta_T^i),$$

where

$$A_T(\Delta) := \frac{c_T}{b_{m_T}} \left| \frac{1}{\Delta}\sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x)\,\mathbb{E}\,\widetilde{\varphi}_{j,T}(X_\Delta) - s(x) \right|.$$

Then, it suffices to show that $\max_i A_T(\Delta_T^i) \to 0$, as $T\to\infty$. Notice that

$$A_T(\Delta) \leq \frac{c_T}{b_{m_T}} \left| \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x) \left\{ \frac{1}{\Delta}\,\mathbb{E}\,\widetilde{\varphi}_{j,T}(X_\Delta) - \int \widetilde{\varphi}_{j,T}(y)s(y)dy \right\} \right|$$
$$+ \frac{c_T}{b_{m_T}} \left| \int \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x)\widetilde{\varphi}_{j,T}(y)(s(y)-s(x))dy \right|,$$

where we have used that $\int \widetilde{\varphi}_{j,T}(y)dy = \delta_0(j)$. We shall show that each of the two on the right hand side of the above inequality, that we denote $A_T^1(\Delta)$ and A_T^2, respectively, vanish as $T\to\infty$. Using (1.5), Lemma B.1, and Proposition 2.1, there exists a $K>0$ and $T_0>0$ such that for $T>T_0$,

$$A_T^1(\Delta_T^i) \leq K\frac{c_T\Delta_T^i}{b_{m_T}(b_T-a_T)} \leq K\frac{c_T m_T \bar{\pi}_T}{(b-a)} \to 0,$$

as $T\to\infty$ due to (i)-(iii). To dealt with the term A_T^2, we treat the two cases $\alpha=1$ and $\alpha>1$ separately. Suppose that $\alpha=1$. Using two times Cauchy-Schwartz (for summation and for integral) and that $\sum_{j=0}^{k}\widetilde{\varphi}_{j,T}^2(x) = b_{m_T}^2(x)/(b_T-a_T)$:

$$A_T^2 \leq \frac{c_T}{\sqrt{b_T-a_T}} \left\{ \sum_{j=0}^{k} \int_{a_T}^{b_T} (s(y)-s(x))^2 dy \right\}^{1/2} \leq K c_T(b_T-a_T),$$

for some constant $K < \infty$. In light of the assumption (iv) of Theorem 3.1, $A_T^2 \stackrel{T\to\infty}{\longrightarrow} 0$. Let is now assume that $\alpha > 1$. We first note that

$$\int \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x)\widetilde{\varphi}_{j,T}(y)(y-x)^{j'} dy = 0,$$

for $j' = 1, \ldots, k$. This is because the left hand side is $p^{\perp}(x)$, where $p^{\perp}(y)$ is the orthogonal projection of the function $p(y) := (y-x)^{j'}$ on $\mathcal{S}_{k,m_T}$, and clearly, $p^{\perp}(x) = p(x) = 0$. Also, by Taylor's Theorem,

$$s(y) - s(x) = \sum_{j'=1}^{r} \frac{s^{(j')}(x)}{j'!}(y-x)^{j'} + \int_x^y \left(s^{(r)}(v) - s^{(r)}(x)\right) \frac{(y-v)^{r-1}}{(r-1)!} dv,$$

where $r := \lfloor \alpha \rfloor$, the largest integer that is (strictly) smaller than α. Since $k \geq \alpha - 1$, we have that $k \geq r$ and

$$\begin{aligned}
&\int \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x)\widetilde{\varphi}_{j,T}(y)(s(y)-s(x))dy \\
&= \int \sum_{j=0}^{k} \widetilde{\varphi}_{j,T}(x)\widetilde{\varphi}_{j,T}(y) \int_x^y (s^{(r)}(v) - s^{(r)}(x)) \frac{(y-v)^{r-1}}{(r-1)!} dv dy.
\end{aligned}$$

Applying again two times Cauchy-Schwartz (for summation and for integral),

$$\begin{aligned}
A_T^2 &\leq \frac{c_T}{b_{m_T}} \sum_{j=0}^{k} |\widetilde{\varphi}_{j,T}(x)| \left| \int \widetilde{\varphi}_{j,T}(y) \int_x^y (s^{(r)}(v) - s^{(r)}(x)) \frac{(y-v)^{r-1}}{(r-1)!} dv dy \right| \\
&\leq \frac{c_T}{\sqrt{b_T - a_T}} \left\{ \sum_{j=0}^{k} \int_{a_T}^{b_T} \left\{ \int_x^y (s^{(r)}(v) - s^{(r)}(x)) \frac{(y-v)^{r-1}}{(r-1)!} dv \right\}^2 dy \right\}^{1/2}.
\end{aligned}$$

Finally, by the Holder condition (1.7), $A_T^2 \leq K c_T m_T^{-\alpha} \stackrel{T\to\infty}{\longrightarrow} 0$. □

Appendix C: Proofs of the uniform CLT.

In this part we show the results of Section 4. We recall that the estimators $\hat{s}_T^n$ are based on observation of the process at evenly-spaced times $\pi_T^n : t_0 = 0 < \cdots < t_n = T$. The time span between observations is $\delta^n := \delta_T^n := T/n$.

Let us first remark that under the assumption that $\sigma \neq 0$ or $\nu(\mathbb{R}) = \infty$, the distribution $F_t(x)$ is continuous for all $t > 0$ (see [23, Theorem 27.4]). In particular, $\{F_{\delta^n}(X_{t_i} - X_{t_{i-1}})\}_{i \leq n}$ is necessarily a random sample of uniform random variables and hence, Z_n^0 of (4.2) is indeed the standardized empirical process of a uniform random sample. Note also that

$$Z_n^0\left(F_{\delta^n}(x)\right) = n^{1/2}\left\{F^n(x) - F_{\delta^n}(x)\right\}, \quad \forall x \in \mathbb{R},$$

where $F^n := F^n_T$ is the empirical process of $\{X_{t_i} - X_{t_{i-1}} : i = 0, \ldots, n\}$. The following transformation will be useful in the sequel:

$$\mathcal{L}\,(x; m, \kappa, H) = \kappa \sum_{i=1}^{m} \sum_{j=0}^{k} \hat{\varphi}_{i,j}(x) \left\{ \hat{\varphi}_{i,j}(x_i) \left(H(x_i) - H(x_{i-1}) \right) \right. \\ \left. - \int_{x_{i-1}}^{x_i} \hat{\varphi}'_{i,j}(u) \left(H(u) - H(x_{i-1}) \right) du \right\},$$

where $\hat{\varphi}_{i,j}$ is the basis element in (1.4) and $H : \mathbb{R} \to \mathbb{R}$ is a locally integrable function. Note that if H is a function of bounded variation, then

$$\mathcal{L}\,(x; m, \kappa, H) = \kappa \sum_{i=1}^{m} \sum_{j=0}^{k} \hat{\varphi}_{i,j}(x) \int_{x_{i-1}}^{x_i} \hat{\varphi}_{i,j}(u) dH(u).$$

The following estimate follows easily from (1.5):

$$\sup_{x \in [a,b]} |\mathcal{L}(x; m, \kappa, H)| \leq K \cdot \kappa \cdot m \cdot \omega \left(H; [a,b], \frac{b-a}{m} \right), \tag{C.1}$$

where K is a constant (depending only on k) and ω is the modulus of continuity of H defined by

$$\omega(H; [a,b], \delta) = \sup \left\{ |H(u) - H(v)| : u, v \in [a,b], |u - v| < \delta \right\}.$$

Let us write the estimator (3.2) in terms of F^n_T as follows:

$$\hat{s}^n_T(x) := \sum_{i=1}^{m} \sum_{j=0}^{k} \hat{\beta}^{\pi^n_T}(\hat{\varphi}_{i,j}) \hat{\varphi}_{i,j}(x) = \mathcal{L}\left(x; m, \frac{n}{T}, F^n_T(\cdot)\right). \tag{C.2}$$

Note that $\mathbb{E}\,\hat{s}^n_T(x)$ admits a similar expression with F^n_T replaced by $F_{\delta^n_T}$. Thus, it follows that, a.s.,

$$Y^n_T(x) := \hat{s}^n_T(x) - \mathbb{E}\,\hat{s}^n_T(x) = \mathcal{L}\left(x; m, n^{1/2} T^{-1}, Z^0_n(F_{\delta^n}(\cdot)\right), \tag{C.3}$$

for all x. As it was explained in Section 4, one of the key idea of Bickel and Rosenblatt [3] approach consists of approximating Z^0_n by a Brownian bridge Z^0. To this end, we use the following result, which follows from the Komlós, Major, and Tusnády construction [19]:

Theorem C.1. *There exists a probability space* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, *equipped with a standard Brownian motion* $\widetilde{Z}$, *on which one can construct a version* $\widetilde{Z}^0_n$ *of* Z^0_n *such that*

$$\|\widetilde{Z}^0_n - \widetilde{Z}^0\|_{[0,1]} = O_p\left(n^{-1/2} \log n\right),$$

where $\widetilde{Z}^0(x) := \widetilde{Z}(x) - x\widetilde{Z}(1)$ *is the corresponding Brownian bridge.*

Since we are looking for the asymptotic distribution of $\sup_x |Y_T^n(x)|$, properly scaled and centered, we can work with the process $\widetilde{Z}_n^0$, instead of Z_n^0. Thus, with some abuse of notation, we drop the tilde in all the processes of Theorem C.1. The following is an easy estimate. Abusing again of the notation, the process ${}_0Y_T^n$ in the following lemma is actually the process resulting from replacing $Z_n^0(F_{\delta^n}(\cdot))$ in (C.3) by $\widetilde{Z}_n^0(F_{\delta^n}(\cdot)$.

Lemma C.2. *Let* ${}_0Y_T^n(x) = \mathcal{L}\left(x; m, n^{1/2}T^{-1}, Z^0(F_{\delta^n}(\cdot)\right)$. *Then, it follows that* $\|{}_0Y_T^n - Y_T^n\|_{[a,b]} = O_p\left(m \log n/T\right)$, *as* $n \to \infty$

Proof. Clearly, $\omega(H;[a,b],\delta) \le 2\|H\|_{[a,b]}$, for any process H. Thus, we get the result from (C.1) and Theorem C.1. □

As in [3], our approach is to devise successive approximations of ${}_0Y_T^n(x)$, denoted by ${}_1Y_T^n, \dots, {}_NY_T^n$, such that the asymptotic distribution of the supremum $\sup_{x\in[a,b]}\left|{}_NY_T^n(x)\right|$, properly centered and scaled by certain constants b_T^n and a_T^n is easy to determine, and such that the error of the successive approximations is negligible when multiplied by a_T^n. We proceed to carry out this program.

First, note that, since a Brownian bridge satisfies that $\{Z^0(x)\}_{x\le1} \overset{\mathfrak{D}}{=} \{Z^0(1-x)\}_{x\le1}$, we have

$$\{{}_0Y_T^n(x)\}_{x\in[a,b]} \overset{\mathfrak{D}}{=} \{{}_1Y_T^n(x)\}_{x\in[a,b]},$$

where ${}_1Y_T^n(x) := \mathcal{L}\left(x; m, n^{1/2}T^{-1}, Z^0(\bar{F}_{\delta^n}(\cdot)\right)$ and $\bar{F} := 1 - F$. The following is our first estimate:

Lemma C.3. *Suppose that the assumptions of Proposition 2.1 are satisfied. There exist constants K and $t_0 > 0$ such that if $T/n < t_0$, then*

$${}_2Y_T^n(x) = \mathcal{L}\left(x; m, n^{1/2}T^{-1}, Z(\bar{F}_{\delta^n}(\cdot)\right)$$

is such that

$$\|{}_1Y_T^n - {}_2Y_T^n\|_{[a,b]} \le K n^{-1/2}\left(\frac{mT}{n} \vee 1\right)|Z(1)|$$

for a constant $K < \infty$.

Proof. Clearly,

$${}_2Y_T^n(x) - {}_1Y_T^n(x) = \mathcal{L}\left(x; n, T, m, n^{1/2}T^{-1}, Z(1)\bar{F}_{\delta^n}(\cdot)\right)$$

Thus, by (C.1),

$$\|{}_1Y_T^n - {}_2Y_T^n\|_{[a,b]} \le K\frac{mn^{1/2}}{T}\omega\left(\bar{F}_{\delta^n}; [a,b], d_m\right)|Z(1)|,$$

where $d_m = (b-a)/m$. In view of Proposition 2.1, for n and T such that $T/n < t_0$, there are constants k and k' such that

$$|\bar{F}_{\delta^n}(u) - \bar{F}_{\delta^n}(v)| \le 2k(\delta^n)^2 + 2k'\delta^n m^{-1},$$

provided that $u, v \in [a,b]$ and $|v-u| < d_m$. □

Let us now work with ${}_2Y^n_{T}$. Because of the self-similarity of the Brownian motion, we have that

$$\{{}_2Y^n_{T}(x)\}_{x\in[a,b]} \stackrel{\mathfrak{D}}{=} \{{}_3Y^n_{T}(x)\}_{x\in[a,b]},$$

where

$${}_3Y^n_{T}(x) := \mathcal{L}\left(x; m, T^{-1/2}, Z(\frac{1}{\delta^n}\bar{F}_{\delta^n}(\cdot))\right).$$

The following estimate results from Lévy's modulus of continuity theorem.

Lemma C.4. *Let* ${}_4Y^n_{T}(x) = \mathcal{L}\left(x; m, T^{-1/2}, Z\left(\int_{\cdot}^{\infty} s(u)du\right)\right)$. *If* T_n *is such that* $\delta^n := \frac{T_n}{n} \to 0$, *then for* n *large enough,*

$$\|{}_3Y^n_{T_n} - {}_4Y^n_{T_n}\|_{[a,b]} \leq m \cdot O_p\left(n^{-1/2}\log^{1/2}\frac{n}{T_n}\right)$$

for a constant $K < \infty$.

Proof. It is not hard to see that there exists a constant K such that

$$\|{}_3Y^n_{T} - {}_4Y^n_{T}\| \leq KT^{-1/2}m \sup_{x\in[a,b]}\left|Z\left(\frac{1}{\delta^n}\bar{F}_{\delta^n}(x)\right) - Z\left(\int_x^{\infty} s(u)du\right)\right|$$

By Proposition 2.1, there exists a constants $k > 0$ and $t_0 > 0$ such that for all $0 < \delta < t_0$

$$\sup_{y\in D}\left|\frac{1}{\delta}\mathbb{P}\left[X_\delta \geq y\right] - \nu([y,\infty))\right| < k\,\delta. \tag{C.4}$$

Thus, there exists a constant $K > 0$ such that, for large enough n,

$$\|{}_3Y^n_{T} - {}_4Y^n_{T}\| \leq Kn^{-1/2}m\log^{1/2}\frac{n}{T_n}, \quad a.s.$$

□

We now note that

$$\left\{Z\left(\int_x^{\infty} s(u)du\right)\right\}_{x\in[a,b]} \stackrel{\mathfrak{D}}{=} \left\{\int_x^{\infty} s^{1/2}(u)dZ(u)\right\}_{x\in[a,b]},$$

and hence,

$$\{{}_4Y^n_{T}(x)\}_{x\in[a,b]} \stackrel{\mathfrak{D}}{=} \{{}_5Y^n_{T}(x)\}_{x\in[a,b]},$$

where

$${}_5Y^n_{T}(x) := \mathcal{L}\left(x; m, T^{-1/2}, \int_{\cdot}^{\infty} s^{1/2}(u)dZ(u)\right).$$

Using integration by parts, one can simplify ${}_5Y^n_{T}(x)$ as follows:

$${}_5Y^n_{T}(x) = T^{-1/2}\sum_{i=0}^{m}\sum_{j=0}^{k}\hat{\varphi}_{i,j}(x)\int_{x_{i-1}}^{x_i} s^{1/2}(u)\hat{\varphi}_{i,j}(u)dZ(u).$$

The following is the last estimate:

Lemma C.5. *Suppose that the Standing Assumptions 1 in Section 4 hold true. Let*

$$ {}_6Y^n_T(x) := (b-a)^{1/2} T^{-1/2} \sum_{i=0}^{m} \sum_{j=0}^{k} \hat{\varphi}_{i,j}(x) \int_{x_{i-1}}^{x_i} \hat{\varphi}_{i,j}(u) dZ(u). $$

Then, there exists a random variable M such that

$$ \| {}_6Y^n_T(\cdot) - (b-a)^{1/2} s^{-1/2}(\cdot) {}_5Y^n_T(\cdot) \| \le M T^{-1/2}. $$

Proof. Let $q(x) = s^{1/2}(x)$ and $c = (b-a)^{1/2}$. Using integration by parts, we have

$$ \begin{aligned} H_{i,j}(x) &:= s^{-1/2}(x) \int_{x_{i-1}}^{x_i} s^{1/2}(u) \hat{\varphi}_{i,j}(u) dZ(u) - \int_{x_{i-1}}^{x_i} \hat{\varphi}_{i,j}(u) dZ(u) \\ &= q^{-1}(x) \left\{ \hat{\varphi}_{i,j}(x_i) \left(q(x_i) - q(x) \right) Z(x_i) - \hat{\varphi}_{i,j}(x_{i-1}) \left(q(x_{i-1}) - q(x) \right) Z(x_{i-1}) \right\} \\ &\quad - q^{-1}(x) \int_{x_{i-1}}^{x_i} \left\{ \hat{\varphi}'_{i,j}(u) \left(q(u) - q(x) \right) - \hat{\varphi}_{i,j}(u) q'(u) \right\} Z(u) du. \end{aligned} $$

Since $q^{-1}(\cdot)$ and $q'(\cdot)$ are bounded on $[a,b]$, there exists a constant K such that

$$ \sup_{x \in [x_{i-1}, x_i]} |H_{i,j}(x)| \le K m^{-1/2} \sup_{u \in [x_{i-1}, x_i]} |Z(u)|. $$

Thus,

$$ \begin{aligned} \| {}_6Y^n_T(\cdot) - c s^{-1/2}(\cdot) {}_5Y^n_T(\cdot) \| &\le \left(\frac{T}{b-a} \right)^{-1/2} \sum_{i=0}^{m} \sum_{j=0}^{k} \sup_{x \in [x_{i-1}, x_i]} |H_{i,j}(x) \hat{\varphi}_{i,j}(x)| \\ &\le K T^{-1/2} \sup_{u \in [a,b]} |Z(u)|. \end{aligned} $$

□

The latter approximation ${}_6Y^n_T$ is simple enough for trying to determine its asymptotic distribution (appropriately centered and scaled). Indeed,

$$ M(T,n,m) := \sup_{x \in [a,b]} |{}_6Y^n_T(x)| \overset{\mathfrak{D}}{=} T^{-1/2} m^{1/2} \max_{1 \le j \le m} \{ \zeta^{(k)}_m \}, \tag{C.5} $$

where $\{\zeta^{(k)}_j\}_i$ are independent copies of the r.v. $\zeta^{(k)}$ defined in (4.3). The following result obtains the asymptotic distributions of $\bar{M}_m := \max_{1 \le j \le m} \{ \zeta^{(k)}_j \}$, for the cases of $k=0$ and $k=1$.

Lemma C.6. *Let a_n and b_n as in (4.5-4.6). Then, the following limits hold*

$$ \lim_{m \to \infty} \mathbb{P} \left(\max_{1 \le j \le m} \{ \zeta^{(0)}_m \} \le \frac{y}{a_{m_n}} + b_{m_n} \right) = e^{-2e^{-y}}, \tag{C.6} $$

$$ \lim_{m \to \infty} \mathbb{P} \left(2^{-1} \max_{1 \le j \le m} \{ \zeta^{(1)}_m \} \le \frac{y}{a_{m_n}} + b_{m_n} \right) = e^{-4e^{-y}}. \tag{C.7} $$

for all $y \in \mathbb{R}_+$.

Proof. The limit (C.6) follows from the well-known identity

$$\lim_{m\to\infty} m\left(1-\Phi\left(u_m(y)\right)\right)=e^{-y}, \tag{C.8}$$

where Φ is the normal distributio and $u_m(y)=y/a_m+b_m$. Indeed, for large enough m, the probability in (C.6) can be written as follows

$$\left(2\Phi\left(u_m(y)\right)-1\right)^m=\left(1-\frac{2m\left(1-\Phi(u_m(y))\right)}{m}\right)^m \longrightarrow e^{-2e^{-y}}.$$

To handle the case $k=1$, we embed the problem into the theory of multivariate extreme values (see e.g. [16]). Consider independent copies $\{\mathbf{V_i}\}_i$ of the following vector of jointly standard Gaussian variables

$$\mathbf{V}:=\left(\frac{1}{2}Z_0+\frac{\sqrt{3}}{2}Z_1,\frac{1}{2}Z_0-\frac{\sqrt{3}}{2}Z_1\right)'. \tag{C.9}$$

Since $\zeta^{(1)}=|Z_0|+\sqrt{3}|Z_1|$, we can see that

$$\begin{aligned}&\left\{2^{-1}\max_{1\le j\le m}\{\zeta_m^{(1)}\}\le\frac{y}{a_m}+b_m\right\}\\ &=\left\{\max_{i\le m}\mathbf{V}_i\le\hat{\mathbf{a}}_m^{-1}\mathbf{y}+\hat{\mathbf{b}}_m,\min_{i\le m}\mathbf{V}_i\ge-\hat{\mathbf{a}}_m^{-1}\mathbf{y}-\hat{\mathbf{b}}_m\right\},\end{aligned}$$

where $\mathbf{y}:=(y,y)'$, $\hat{\mathbf{b}}_m:=(b_m,b_m)'$, and $\hat{\mathbf{a}}_m:=(a_m,a_m)'$, and all operations are point-wise. Then, (C.7) will follow from the following identity

$$\begin{aligned}\lim_{m\to\infty}\mathbb{P}&\left(\max_{1\le i\le m}\mathbf{V}_i\le\hat{\mathbf{a}}_m^{-1}\mathbf{y}+\hat{\mathbf{b}}_m,\min_{1\le i\le m}\mathbf{V}_i\ge-\hat{\mathbf{a}}_m^{-1}\mathbf{z}-\hat{\mathbf{b}}_m\right)\\ &=e^{-e^{-y_1}-e^{-y_2}-e^{-z_1}-e^{-z_1}},\end{aligned} \tag{C.10}$$

for any $\mathbf{y}=(y_1,y_2)'$ and $\mathbf{z}=(z_1,z_2)'$. To show (C.10), first note that the probability in there can be written as follows

$$A_n:=\{\,\mathbb{P}\left(-u_n(z_1)\le V_1\le u_n(y_1),-u_n(z_2)\le V_2\le u_n(y_2)\right)\}^n,$$

where $\mathbf{V}:=(V_1,V_2)'$ is defined in (C.9) and $u_n(x):=x/a_n+b_n$. Let

$$\bar{\mathbf{F}}_n(y,z;X,Y):=\mathbb{P}\left(X\ge u_n(y),Y\ge u_n(z)\right),\quad \bar{F}_n(y;X):=\mathbb{P}\left(X\ge u_n(y)\right),$$

where X and Y represent random variables. We recall the following results valid for any jointly normal variables X and Y and arbitrary y and z (see Example 5.3.1 in [16]):

$$\lim_{n\to\infty}n\bar{\mathbf{F}}_n(y,z;X,Y)=0,\quad \lim_{n\to\infty}n\bar{F}_n(y;X)=e^{-y}.$$

Then, (C.10) follows once we note that $A_n^{1/n}$ can be written as follows:

$$A_n^{1/n} = 1 - \frac{1}{n}\left\{n\bar{F}_n(z_1;V_1) + n\bar{F}_n(z_2;V_2) + n\bar{F}_n(y_1;-V_1) + n\bar{F}_n(y_2;-V_2)\right.$$
$$\left. -n\bar{\mathbf{F}}_n(z_1,z_2;V_1,V_2) - n\bar{\mathbf{F}}_n(y_1,z_2;-V_1,V_2) - n\bar{\mathbf{F}}_n(z_1,y_2;V_1,-V_2)\right\}.$$

□

In view of (C.5), the following are easy consequence of the above lemma:

$$\lim_{n\to\infty} \mathbb{P}\left(T_n^{1/2} m_n^{-1/2} \sup_{x\in[a,b]} |{}_6Y_{T_n}^n(x)| \le \frac{y}{a_{m_n}} + b_{m_n}\right) = e^{-2e^{-y}}, \quad \text{(C.11)}$$

$$\lim_{n\to\infty} \mathbb{P}\left(2^{-1} T_n^{1/2} m_n^{-1/2} \sup_{x\in[a,b]} \left|{}_6Y_{T_n}^n(x)\right| \le \frac{y}{a_{m_n}} + b_{m_n}\right) = e^{-4e^{-y}}, \quad \text{(C.12)}$$

valid for all $y \in \mathbb{R}_+$, $T_n > 0$, and m_n such that $m_n \to \infty$. We are ready to show our main Theorem in Section 4:

Proof of Theorem 4.1. The idea is to use the following simple observations. Let $\mathcal{L}_n$ be a functional on $D[a,b]$ such that

$$|\mathcal{L}_n(\omega_1) - \mathcal{L}_n(\omega_2)| \le M_n \|\omega_1 - \omega_2\|, \quad \text{(C.13)}$$

and let A_n, B_n processes with values on $D[a,b]$ such that $\|A_n - B_n\| = o_p(1/M_n)$. Then, if $\mathcal{L}_n(A_n)$ converges in distribution to F, then $\mathcal{L}_n(B_n)$ converges to F as well. Throughout this proof,

$$\mathcal{L}_n(\omega) := a_{m_n}\left\{\kappa \cdot \frac{c}{d} \cdot \frac{T_n^{1/2}}{m_n^{1/2}} \cdot \sup_{x\in[a,b]} |s^{-1/2}(x)\omega(x)| - b_{m_n}\right\},$$

which satisfies the Lipschitz condition (C.13) with $M_n = \frac{\kappa c}{d} a_{m_n} T_n^{1/2}/m_n^{1/2}$. From Lemma C.5, in order for (C.12) to hold with ${}_6Y_{T_n}^n$ replaced by ${}_5Y_{T_n}^n$, it suffices that

$$\lim_{n\to\infty} \frac{T_n^{1/2}}{m_n^{1/2}} a_{m_n} T_n^{-1/2} = \lim_{n\to\infty}\left(\frac{2\log m_n}{m_n}\right)^{1/2} = 0,$$

which is obvious since $m_n \to \infty$. Since ${}_4Y_{T_n}^n$ has the same law as ${}_5Y_{T_n}^n$, (C.12) holds for ${}_4Y_{T_n}^n$ as well. In the light of Lemma C.4, (C.12) will hold for ${}_3Y_{T_n}^n$ (and hence, for ${}_2Y_{T_n}^n$ as well) since

$$\lim_{n\to\infty} \frac{T_n^{1/2}}{m_n^{1/2}} a_{m_n} m_n n^{-1/2} \log^{1/2}\frac{n}{T_n} = c \lim_{n\to\infty}\left(m_n \log m_n \cdot \frac{T_n}{n} \log\frac{n}{T_n}\right)^{1/2} = 0,$$

which follows from condition (ii) in the statement of Theorem 4.1. Similarly, in view of Lemma C.3, (C.12) will hold for ${}_1Y_{T_n}^n$ (and hence, for ${}_0Y_{T_n}^n$ as well) since

$$\lim_{n\to\infty} \frac{T_n^{1/2} a_{m_n} n^{-1/2}}{m_n^{1/2}}\left(\frac{m_n T_n}{n} \vee 1\right) = 0.$$

Indeed,the above expression is upper bounded by $\left(\frac{T_n m_n}{n}\right)^{1/2} \frac{\log^{1/2} m_n}{m_n}$, which converge to 0 because of assumption (i) and $m_n \to \infty$. Finally, in the light of Lemma C.2, in order for (C.12) to hold for $Y^n_{T_n}$, it suffices that

$$\lim_{n\to\infty} \frac{T_n^{1/2}}{m_n^{1/2}} a_{m_n} \frac{m_n}{T_n} \log n = 0,$$

which follows from assumption (ii) in the statement of Theorem 4.1. □

Proof of Corollary 4.2. Using the same reasoning in the proof of Theorem 3.1, it turns out that

$$\sup_{x\in[a,b]} |\,\mathbb{E}\,\hat{s}^n_{T_n}(x) - s(x)| \leq K \left(\frac{m_n T_n}{n} \vee m_n^{-\alpha} \right),$$

for an absolute constant K. As in the proof of Theorem 4.1, to show (4.9), it suffices that

$$\lim_{n\to\infty} \frac{T_n^{1/2}}{m_n^{1/2}} a_{m_n} \left(\frac{m_n T_n}{n} \vee m_n^{-\alpha} \right) = 0,$$

which holds in light of the assumption (iii) in the statement of Corollary 4.2. □

References

[1] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113:301–413, 1999.

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

[3] P.J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095, 1973.

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

[5] Brillinger. An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.*, 75:545–547, 1969.

[6] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, pages 305–332, April 2002.

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

[8] K.L. Chung. *A course in Probability Theory*. Academic Press, 2001.

[9] R. Cont and P. Tankov. *Financial modelling with Jump Processes*. Chapman & Hall, 2004.

[10] J.E. Figueroa-López. *Nonparametric estimation of Lévy processes with a view towards mathematical finance*. PhD thesis, Georgia Institute of Technology, April 2004. http://etd.gatech.edu. No. etd-04072004-122020.

[11] J.E. Figueroa-López. Nonparametric estimation for Lévy models based on discrete-sampling. *IMS Lecture Notes-Monograph Series. Optimality: The Third Erich L. Lehmann Symposium*, 57:117–146, 2009.

[12] J.E. Figueroa-López. Nonparametric estimation of time-changed Lévy models under high-frequency data. *Advances in Applied Probability*, 41(4), 2009.

[13] J.E. Figueroa-López. *Jump-diffusion models driven by Lévy processes*. Springer, 2010. To appear in Handbook of Computational Finance. Jin-Chuan Duan, James E. Gentle, and Wolfgang Hardle (eds.).

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

[15] J.E. Figueroa-López and C. Houdré. Small-time expansions for the transition distributions of Lévy processes. *Stochastic Processes and Their Applications*, 119:3862–3889, 2009.

[16] J. Galambos. *The asymptotic theory of extreme order statistics*. Krieger, 1987.

[17] U. Grenander. *Abstract Inference*. John Wiley & Sons, 1981.

[18] P. Hall. Effect of bias estimation on coverage accuracy of bootstrap confidence interval for a probability density. *Annals of Statistics*, (22):675–694, 1992.

[19] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV'-s, and the sample DF. I. *Z. Wahrsch. Verw. Gebiete*, 32:111–131, 1975.

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

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

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

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

[24] E. Seneta. Fitting the variance-gamma model to financial data. *Journal of Applied Probability*, 41A:177–187, 2004.

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