NONPARAMETRIC ESTIMATION FOR PURE JUMP LÉVY PROCESSES BASED ON HIGH FREQUENCY DATA.

F. COMTE[1] AND V. GENON-CATALOT[1]

ABSTRACT. In this paper, we study nonparametric estimation of the Lévy density for pure jump Lévy processes. We consider n discrete time observations with step Δ. The asymptotic framework is: n tends to infinity, $\Delta = \Delta_n$ tends to zero while $n\Delta_n$ tends to infinity. First, we use a Fourier approach ("frequency domain"): this allows to construct an adaptive nonparametric estimator and to provide a bound for the global $\mathbb{L}^2$-risk. Second, we use a direct approach ("time domain") which allows to construct an estimator on a given compact interval. We provide a bound for $\mathbb{L}^2$-risk restricted to the compact interval. We discuss rates of convergence and give examples and simulation results for processes fitting in our framework. February 3, 2009

Keywords. Adaptive nonparametric estimation. High frequency data. Lévy processes. Projection estimators.
AMS Classification. 62G05-62M05-60G51.

1. INTRODUCTION

Let $(L_t, t \geq 0)$ be a real-valued Lévy process, *i.e.* a process with stationary independent increments. We assume that the characteristic function of L_t has the form:

$$\psi_t(u) = \mathbb{E}(\exp iuL_t) = \exp\left(t \int_{\mathbb{R}} (e^{iux} - 1)n(x)dx\right), \tag{1}$$

where the Lévy density $n(.)$ satisfies $\int_{\mathbb{R}} |x| \wedge 1 \; n(x)dx < \infty$. Under these assumptions, the process (L_t) is of pure jump type, with no drift component, has finite variation on compacts (see *e.g.* Bertoin, 1996, Chap. 1). The distribution of (L_t) is therefore completely specified by the knowledge of $n(.)$ which describes the jumps behavior.

In this paper, we consider the nonparametric estimation of $n(.)$ based on a discrete observation of the sample path with sampling interval Δ. Our estimation procedure is therefore based on the random variables $(Z_k = Z_k^{\Delta} = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \ldots, n)$ which are independent, identically distributed, with common characteristic function $\psi_\Delta(u)$. Since the problem reduces to estimation from an i.i.d. sample, statistical inference for discrete observations of Lévy processes may appear standard. However, difficulties arise due to specific features. First, the exact distribution of the increment Z_k is most often hardly tractable. Second, one is interested in the Lévy density $n(.)$ and the relationship between $n(.)$ and the distribution of the r.v.'s Z_k is not straightforward (see examples, below). This is why statistical approaches often rely on the simple link between $n(.)$ and the characteristic function ψ_Δ. Illustrations of this approach can be found in Watteel and Kulperger (2003), Jongbloed and van der Meulen (2006), Neumann and Reiss (2007), Jongbloed *et al.* (2005) for the related problem of Lévy-driven Ornstein-Uhlenbeck processes or Comte and Genon-Catalot (2008).

For what concerns the sampling interval, it is now classical in statistical inference for discretely observed continuous time processes to distinguish two points of view. In the low frequency point of view, it is assumed that the sampling interval Δ is kept fixed while the number n of

[1] Université Paris Descartes, MAP5, UMR CNRS 8145.

observations tends to infinity. This is the assumption done in the above references. On the other hand, the high frequency data point of view is naturally well fitted when the underlying model is continuous in time. It consists in assuming that the sampling interval $\Delta = \Delta_n$ tends to 0 as n tends to infinity. This is the point of view adopted in this paper. Moreover, in order to make our results comparable to those obtained for low frequency data, we also assume that the total length time of observation, $n\Delta_n$, tends to infinity with n.

As in Comte and Genon-Catalot (2008), we strengthen the assumption on $n(.)$ into:

(H1) $\int_{\mathbb{R}} |x|n(x)dx < \infty$,

and focus on the estimation of the function

$$g(x) = xn(x).$$

By (H1), derivating ψ_Δ yields:

$$g^*(u) = \int e^{iux}g(x)dx = -i\frac{\psi'_\Delta(u)}{\Delta\psi_\Delta(u)}. \tag{2}$$

In the framework of low frequency data, this relation suggests to estimate g^* by using empirical estimators of $\psi'_\Delta(u)/\Delta$ and $\psi_\Delta^{-1}(u)$. Then, g can be recovered adaptively by Fourier methods. The fact that the denominator $\psi_\Delta(u)$ has to be estimated makes the study difficult (see Comte and Genon-Catalot, 2008). Now, for high frequency data, the above relation is written as:

$$-i\frac{\psi'_\Delta(u)}{\Delta} = g^*(u) + g^*(u)(\psi_\Delta(u) - 1) = \frac{1}{\Delta}\mathbb{E}(Z_k e^{iuZ_k}). \tag{3}$$

Since $\psi_\Delta(u) - 1$ tends to 0 as Δ tends to 0, $\psi_\Delta(u)$ needs not be estimated and $g^*(u)$ may be estimated by the empirical estimator

$$\hat{\theta}_\Delta(u)/\Delta = \hat{g}^*(u) := \frac{1}{n\Delta}\sum_{k=1}^{n} Z_k e^{iuZ_k}, \tag{4}$$

As a basic consequence of (3), under (H1), the empirical measure

$$\hat{\mu}_n(dx) = \frac{1}{n\Delta}\sum_{k=1}^{n} Z_k \delta_{Z_k}(dx) \tag{5}$$

is a consistent estimator of the measure $g(x)dx$ (δ_z denotes the Dirac measure at z). This allows to study the nonparametric estimation of g by two approaches. On the one hand, using (4), we proceed to Fourier inversion, introducing a cutoff parameter that is adaptively selected. This construction yields a global estimator. On the other hand, relying directly on the property of (5), we are able to apply the penalized projection method classically used to estimate densities (see Massart (2007)). In this way, we obtain an estimator of g on a compact set. Note that the penalized projection method is applied in Figueroa-Lopez and Houdré (2006) to estimate the Lévy density $n(.)$ from a continuous time observation of the sample path (L_t) throughout a time interval $[0, T]$. They obtain theoretical results on the rates of convergence on which we can rely as a benchmark of comparison. Since we build our estimators on discrete data, our results have the advantage of giving concrete estimators that can be easily implemented.

In Section 2, we give our assumptions and preliminary results concerning empirical estimators based on (Z_k). The rate of convergence $\sqrt{n\Delta}$ is obtained under the condition $n\Delta^3 = o(1)$ on the sampling interval. For the nonparametric estimation of g, we assume that g belongs also to $\mathbb{L}^2(\mathbb{R})$. Section 3 is devoted to estimation of g by Fourier methods. We construct a collection $(\hat{g}_m, m = 1, \ldots, m_n)$ of estimators using the frequency domain. The bound for the $\mathbb{L}^2$-risk of an estimator $\hat{g}_m$ (Proposition 3.1) allows to deduce rates of convergence in Sobolev classes of regularity (Proposition 3.2). Afterwards, we define a penalty in order to select adaptively the best estimator of the collection (Theorem 3.1). In Section 4, we construct another collection of estimators $(\tilde{g}_m)$ by using projection subspaces of $\mathbb{L}^2(A)$ where A is a compact subset of $\mathbb{R}$. We follow the same scheme. First, we study the risk bound of an estimator $\tilde{g}_m$ before selection (Proposition 4.2). Then, we define a penalty and obtain the risk bound for the adaptive estimator (Theorem 4.1). We deduce the rate of convergence on Besov classes of regularity (Corollary 4.1). For both methods, as it is usual for high frequency data, constraints on the sampling interval appear and are discussed. Section 5 discusses rates on examples. Section 6 illustrates and compares the methods through simulations. Section 7 contains some conclusions and possible extensions. Proofs (not given in the main text) are gathered in the Appendix.

2. Preliminary results.

2.1. **Framework.** Recall that the Lévy process (L_t) satisfying (1) is observed at n discrete instants $t_k = k\Delta$, $k = 1, \dots, n$, with regular sampling interval and our estimation procedure is based on the random variables $(Z_k = Z_k^{\Delta} = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \dots, n)$ which are independent, identically distributed, with common characteristic function $\psi_\Delta(u)$. We assume that, as n tends to infinity, $\Delta = \Delta_n$ tends to 0 and $n\Delta_n$ tends to infinity so that the observations are $(Z_k = Z_k^n = Z_k^{\Delta_n}, k = 1, \dots, n)$. Nevertheless, to avoid cumbersome notations, we omit the sub- or super-script n everywhere.
For the estimation of $g(x) = xn(x)$, (H1) and the following additional assumptions are required.

- (H2)(p) For p integer, $\int_{\mathbb{R}} |x|^{p-1}|g(x)|dx < \infty$.
- (H3) The function g belongs to $\mathbb{L}^2(\mathbb{R})$.
- (H4) $M_2 := \int x^2 g^2(x)dx < +\infty$.

Assumptions (H1) and (H2)(p) are moment assumptions for Z_1 (see Proposition 2.2 below). Under (H1), (H2)(p) for $p > 1$ implies (H2)(k) for $k \le p$. The required value of p is given in each proposition or theorem.

Noting that

$$\|g\|_1^2 := (\int |g(x)|dx)^2 \le \int (1+|x|)^2 g^2(x)dx \int \frac{dx}{(1+|x|)^2},$$

we see that (H3) and (H4) imply (H1). If g is decreasing, the distribution of Z_1 is self-decomposable and g is called the canonical function (see Barndorff-Nielsen and Shephard (2001) and Sato (1999, chap.3.15 p.90)).

Under (H1), let us introduce (see (2)-(3))

$$\theta_\Delta(u) = \mathbb{E}(Z_1 e^{iuZ_1}) = -i\psi'_\Delta(u) = \Delta g^*(u)\psi_\Delta(u). \tag{6}$$

and

$$\hat{\theta}_\Delta(u) = \frac{1}{n}\sum_{k=1}^{n} Z_k e^{iuZ_k}.$$

As $\theta_\Delta(u)/\Delta = g^*(u) + g^*(u)(\psi_\Delta(u) - 1)$ and $\psi_\Delta(u) = 1 + O(\Delta)$, a simple estimator of $g^*(u)$ is given by (4). We first state a proposition useful for the sequel.

Proposition 2.1. *Denote by P_Δ the distribution of Z_1 and define $\mu_\Delta(dx) = \Delta^{-1}xP_\Delta(dx)$ and $\mu(dx) = g(x)dx$. Under* (H1)*, the distribution μ_Δ has a density h_Δ given by*

$$h_\Delta(x) = \int g(x-y)P_\Delta(dy) = \mathbb{E}g(x - Z_1).$$

And μ_Δ weakly converges to μ as Δ tends to 0.

Proof. Note that

$$\int \mathbb{E}|g(x-Z_1)|dx = \mathbb{E}\int |g(x-Z_1)|dx = \int |g(x)|dx < +\infty.$$

Thus $\mathbb{E}|g(x-Z_1)| < +\infty$ a.e. (dx), which implies that $\mathbb{E}(g(x-Z_1))$ is a.e. well defined. Equation (6) states that

$$\mu^*_\Delta = \mu^* P^*_\Delta.$$

Hence, $\mu_\Delta = \mu \star P_\Delta$ where $\star$ denotes the convolution product. This gives the result. □

Note that, although P_Δ may have no density, under (H1), μ_Δ always has. Note also that the Lévy measure can always be obtained as a limit: for every fixed $a > 0$, $(1/\Delta)P_\Delta(dx)$ converges vaguely on $|x| > a$ as $\Delta \to 0$ to $n(x)dx$, see e.g. Bertoin (1996, p. 39, ex. 5.1). Assumption (H1) ensures the stronger result of Proposition 2.1.

2.2. **Limit theorems and inequalities.** In this section, we study some properties illustrating the framework of high frequency in the context of pure jump Lévy processes. In particular, a condition on the sampling interval is exhibited.

First, we give some properties of the moments of Z_1 and of empirical moments associated with the observations: Proposition 2.2 shows that the moments of Z_1 have all the same rate of convergence with respect to Δ; Theorem 2.1 gives inequalities and a central limit theorem for empirical moments.

Proposition 2.2. *Let $p \geq 1$ integer. Under (H2)(p), $\mathbb{E}|Z_1|^p < \infty$. For $1 \leq \ell \leq p$,*

$$\mathbb{E}(Z_1^\ell) = \Delta\, m_\ell + o(\Delta)$$

where

$$m_\ell = \int_{\mathbb{R}} x^{\ell-1} g(x)dx = \int_{\mathbb{R}} x^\ell n(x)dx. \tag{7}$$

More precisely, if $p \geq 2$, $\mathbb{E}(Z_1) = \Delta m_1$, $\mathbb{E}(Z_1^2) = \Delta m_2 + \Delta^2 m_1^2$. And more generally, if $p \geq \ell$,

$$\mathbb{E}(Z_1^\ell) = \Delta m_\ell + \sum_{j=2}^{\ell} \Delta^j c_j,$$

where the c_j' are explicitly expressed as functions of the $m_j, j \leq \ell$.

Moreover, under (H1), $\mathbb{E}(|Z_1|) \leq 2\Delta\|g\|_1$.

Proof. By the assumption, the exponent of the exponential in (1) is p times differentiable and, for $\ell = 1, \ldots, p$,

$$\frac{d^\ell}{du^\ell}\left(\int_{\mathbb{R}} (e^{iux}-1)n(x)dx)\right) = i^\ell \int_{\mathbb{R}} x^{\ell-1} e^{iux} g(x)dx). \tag{8}$$

By differentiating ψ_Δ and using an elementary induction, we get the result.

Using the classical decomposition $Z_1 = Z_1^+ + Z_1^-$, we compute $\mathbb{E}(Z_1^+)$. By Proposition (2.1),

$$\mathbb{E}(Z_1^+) = \Delta \int_0^{+\infty} \mathbb{E}(g(z-Z_1))dz = \Delta\mathbb{E}\left(\int_{-Z_1}^{+\infty} g(x)dx\right) \leq \Delta\|g\|_1.$$

The computation of $\mathbb{E}(Z_1^-)$ is analogous, and the result follows from $|Z_1| = Z_1^+ + Z_1^-$. □

Theorem 2.1.
- *If $p\ell$ is even,* (H2)($p\ell$) *and* (H2)(2ℓ) *hold,*

$$\mathbb{E}\left(\left|\frac{1}{n\Delta}\sum_{k=1}^{n} Z_k^\ell - \mathbb{E}(Z_1^\ell)\right|^p\right) \leq C_p\left(\frac{1}{(n\Delta)^{p-1}} + \frac{1}{(n\Delta)^{p/2}}\right). \tag{9}$$

- *Assume (H2)(4ℓ). If n tends to infinity and Δ tends to 0 in such a way that $n\Delta$ tends to infinity and $n\Delta^3$ tends to 0, then*

$$\sqrt{n\Delta}\left(\frac{1}{n\Delta}\sum_{k=1}^{n} Z_k^\ell - m_\ell\right) \to \mathcal{N}(0, m_{2\ell})$$

in distribution.

Then, we give inequalities useful to evaluate bias and variance terms for the sequel and a result concerning the behavior of $\hat{\theta}_\Delta(u)/\Delta$ as a pointwise estimator.

Proposition 2.3. *Under* (H1)*, we have:*

$$|\psi_\Delta(u) - 1| \leq |u|\Delta\|g\|_1, \tag{10}$$

$$|\Delta^{-1}\theta_\Delta(u) - g^*(u)| \leq |u|\Delta\|g\|_1^2, \tag{11}$$

Under (H1) *and* (H2)($2p$)*, for $p \geq 1$,*

$$\Delta^{-2p}\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^{2p}) \leq \frac{C_p}{(n\Delta)^p}. \tag{12}$$

Note that for $p = 1$, (12) is a simple variance inequality:

$$\Delta^{-2}\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2) \leq \frac{1}{n\Delta}(m_2 + \Delta m_1^2) = \frac{1}{n\Delta^2}\mathbb{E}(Z_1^2). \tag{13}$$

Theorem 2.2. *Under* (H1) *and* (H2)(2)*, if $n\Delta^3 = o(1)$, $\sqrt{n\Delta}(\hat{\theta}_\Delta(u)/\Delta - g^*(u))$ converges in finite-dimensional distributions to the process $X(u) = \int e^{iux}x\sqrt{n(x)}dB(x), u \in \mathbb{R}$, where B is a Brownian motion indexed by $\mathbb{R}$.*

It would be interesting to develop this study further and obtain a stronger form of Central Limit Theorem (CLT) for the process $\sqrt{n\Delta}(\hat{\theta}_\Delta(u)/\Delta - g^*(u))$. Note that, in the context of low frequency data, Jongbloed and van der Meulen (2006) build a parametric minimum distance estimator relying on a strong CLT for the empirical characteristic function.

3. Estimation of g by Fourier methods

Recall that u^* is the Fourier transform of the function u defined as $u^*(y) = \int e^{iyx}u(x)dx$, and denote by $\|u\|$, $< u, v >$, $u \star v$ the quantities

$$\|u\|^2 = \int |u(x)|^2 dx,$$

$$< u, v >= \int u(x)\overline{v}(x)dx \text{ with } z\overline{z} = |z|^2 \text{ and } u \star v(x) = \int u(y)\bar{v}(x-y)dy.$$

Moreover, for any integrable and square-integrable functions u, u_1, u_2, the following holds:

$$(u^*)^*(x) = 2\pi u(-x) \text{ and } \langle u_1, u_2\rangle = (2\pi)^{-1}\langle u_1^*, u_2^*\rangle. \tag{14}$$

3.1. **Definition of a collection of estimators.** In this paragraph, we present a collection of estimators $(\hat{g}_m)$, indexed by a positive parameter m that will below be subject to constraints for adaptiveness results. Three distinct constructions give rise to this class of estimators, each having its own interest for interpretation, implementation or theoretical aspects. We start with the simple cutoff approach. We have at our disposal an estimator of g^* given by (4). For taking the inverse Fourier transform of $\hat{\theta}_\Delta/\Delta$, since this function is not integrable, we are led to set:

$$\hat{g}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixu}\frac{\hat{\theta}_\Delta(u)}{\Delta}du, \tag{15}$$

for a positive cutoff parameter m. In other words, $\hat{g}_m^* = (\hat{\theta}_\Delta/\Delta)\mathbf{1}_{[-\pi m,\pi m]}$. Introducing

$$\varphi(x) = \frac{\sin(\pi x)}{\pi x} \text{ (with } \varphi(0) = 1), \tag{16}$$

a simple integration leads to

$$\hat{g}_m(x) = \frac{m}{n\Delta}\sum_{k=1}^{n} Z_k\varphi(m(Z_k - x)).$$

Therefore $\hat{g}_m$ may be interpreted as a kernel estimator with kernel φ and bandwidth $1/m$. Formula (15) allows to study the $\mathbb{L}^2$-risk of $\hat{g}_m$ for all m. We need to introduce

$$g_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}g^*(u)du,$$

which is such that $g_m^* = g^*\mathbf{1}_{[-\pi m,\pi m]}$.

Proposition 3.1. *Assume that* (H2)*(2)*- (H3)-(H4) *hold, then for all positive m,*

$$\mathbb{E}(\|g - \hat{g}_m\|^2) \le \|g - g_m\|^2 + 2[\mathbb{E}(Z_1^2/\Delta)]\frac{m}{n\Delta} + \frac{\|g\|_1^2}{\pi}\Delta^2\int_{-\pi m}^{\pi m} u^2|g^*(u)|^2du.$$

Proof. We have $\|\hat{g}_m - g\|^2 = \|\hat{g}_m^* - g^*\|^2/(2\pi)$, and thus (see (4) and (6)),

$$\begin{aligned}
\|\hat{g}_m - g\|^2 &= \frac{1}{2\pi}[\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta})\mathbf{1}_{[-\pi m,\pi m]} + (\frac{\theta_\Delta}{\Delta} - g^*)\mathbf{1}_{[-\pi m,\pi m]} - g^*\mathbf{1}_{[-\pi m,\pi,m]^c}\|^2] \\
&\le \frac{1}{\pi}(\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta})\mathbf{1}_{[-\pi m,\pi m]}\|^2 + \|(\frac{\theta_\Delta}{\Delta} - g^*)\mathbf{1}_{[-\pi m,\pi m]}\|^2) \\
&\quad + \frac{1}{2\pi}\|g^*\mathbf{1}_{[-\pi m,\pi,m]^c}\|^2.
\end{aligned}$$

The last term is exactly $\|g - g_m\|^2$. For the second term, using (3) and (10), we have

$$\|(\frac{\theta_\Delta}{\Delta} - g^*)\mathbf{1}_{[-\pi m,\pi m]}\|^2 = \|(\psi_\Delta - 1)g^*\mathbf{1}_{[-\pi m,\pi m]}\|^2 \le \Delta^2\|g\|_1^2\int_{-\pi m}^{\pi m} u^2|g^*(u)|^2du.$$

Lastly, (13) yields

$$\mathbb{E}(\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta})\mathbf{1}_{[-\pi m,\pi m]}\|^2) = \int_{-\pi m}^{\pi m}\Delta^{-2}\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2)du \le \frac{2\pi m\mathbb{E}(Z_1^2)}{n\Delta^2}.$$

By gathering the three bounds, we obtain the result. □

3.2. **Rates of convergence.** Let us study the rates implied by Proposition 3.1. For that purpose, consider classical classes of regularity for g, defined by

$$\mathcal{C}(a,L)=\left\{g\in(\mathbb{L}^1\cap\mathbb{L}^2)(\mathbb{R}),\ \int(1+u^2)^a|g^*(u)|^2du\leq L\right\}.$$

We obtain the following result:

Proposition 3.2. *Assume that* (H2)*(2)-*(H3)-(H4) *hold and that g belongs to $\mathcal{C}(a,L)$. If $n\to+\infty$, $\Delta\to 0$ and $n\Delta^2\leq 1$, we have*

$$\mathbb{E}(\|g-\hat{g}_m\|^2)\leq O((n\Delta)^{-2a/(2a+1)}).$$

If $a\geq 1$, then it is enough to have $n\Delta^3=O(1)$ (instead of $n\Delta^2\leq 1$).

Proof. We know that

$$\|g-g_m\|^2=\frac{1}{2\pi}\int_{|u|\geq\pi m}|g^*(u)|^2du\leq\frac{L}{2\pi}(\pi m)^{-2a}.$$

Thus, the compromise between $\|g-g_m\|^2$ and $m/(n\Delta)$(first two terms in the risk bound of Proposition 3.1) is obtained for $m=(n\Delta)^{1/(2a+1)}$ and leads to the rate $(n\Delta)^{-2a/(2a+1)}$.
There remains to study the term $\Delta^2\int_{-\pi m}^{\pi m}u^2|g^*(u)|^2du$, which is a bias term due to the high frequency framework and must be made negligible. As $g\in\mathcal{C}(a,L)$, we find

$$\int_{-\pi m}^{\pi m}u^2|g^*(u))|^2du\leq Lm^{2(1-a)_+}.$$

If $a\geq 1$, under the condition $n\Delta^3=O(1)$, $\Delta^2=O(1/(n\Delta))$. So the order of the risk bound is $(n\Delta)^{-2a/(2a+1)}$.

If $a\in(0,1)$, we must have at least $\Delta^2m^{2(1-a)}\leq m^{-2a}$. Hence, $\Delta^2m^2\leq 1$. This is achieved for $n\Delta^2\leq 1$ as $m\leq n\Delta$. The order of the risk bound is again $(n\Delta)^{-2a/(2a+1)}$. □

Remark 3.1. *1. If g is analytic i.e. belongs to a class*

$$\mathcal{A}(\gamma,Q)=\{f,\int(e^{\gamma x}+e^{-\gamma x})^2|f^*(x)|^2dx\leq Q\},$$

then the risk is of order $O(\ln(n\Delta)/(n\Delta))$ (choose $m=O(\ln(n\Delta))$).

2. If g is regular enough, ($a\geq 1$), we find the constraint $n\Delta^3=O(1)$ exhibited in Section 2. If not ($a\in(0,1)$), we must strengthen the constraint on Δ to get the optimal rate of convergence (see the examples below).

3.3. **Adaptive estimator.** Now, we have to select adaptively a relevant bandwidth m. For this, it is convenient to show that the estimators $\hat{g}_m$ are projection estimators, obtained as minimizers of a projection contrast. For positive m, consider the following closed subspace of $\mathbb{L}^2(\mathbb{R})$

$$S_m\quad=\quad\{h\in\mathbb{L}^2(\mathbb{R}),\mathrm{supp}(h^*)\subset[-\pi m,\pi m]\}.$$

For $h\in\mathbb{L}^2(\mathbb{R})$, let h_m denote its orthogonal projection on S_m. A noteworthy property of S_m is that h_m is characterized by the fact that $h_m^*=h^*\mathbf{1}_{[-\pi m,\pi m]}$. Hence,

$$\|h-h_m\|^2=\frac{1}{2\pi}\int_{|x|\geq\pi m}|h^*(x)|^2dx.$$

Moreover, for $t \in S_m$, $t(x) = (1/2\pi)\int_{-\pi m}^{\pi m} e^{-iux} t^*(u)du$, and

$$|t(x)| \leq \frac{1}{2\pi}\left(\int_{-\pi m}^{\pi m} |t^*(u)|^2 du \int_{-\pi m}^{\pi m} |e^{iux}|^2 du\right)^{1/2}.$$

Thus

$$\forall t \in S_m, \quad \|t\|_\infty \leq \sqrt{m}\|t\|. \tag{17}$$

Let, for $t \in S_m$,

$$\gamma_n(t) = \|t\|^2 - \frac{1}{\pi}\int \frac{\hat{\theta}_\Delta(u)}{\Delta} t^*(-u)du = \|t\|^2 - 2\langle \hat{g}_m, t\rangle. \tag{18}$$

Evidently,

$$\hat{g}_m = \arg\min_{t \in S_m} \gamma_n(t),$$

and $\gamma_n(\hat{g}_m) = -\|\hat{g}_m\|^2$. Using (15) and (16), we have

$$\|\hat{g}_m\|^2 = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \left|\frac{\hat{\theta}_\Delta(u)}{\Delta}\right|^2 du = \frac{m}{n^2\Delta^2}\sum_{1\leq k,\ell\leq n} Z_k Z_\ell \varphi(m(Z_k - Z_\ell)).$$

Finally, it is interesting to stress that the space S_m is generated by an orthonormal basis, the sinus cardinal basis, given by:

$$\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j), \quad j \in \mathbb{Z}$$

where φ is defined by (16) (see Meyer (1990), p.22). This can be seen noting that:

$$\varphi^*_{m,j}(x) = \frac{e^{ixj/m}}{\sqrt{m}} \mathbf{1}_{[-\pi m, \pi m]}(x). \tag{19}$$

As above, we use that $\varphi_{m,j}(x) = (1/2\pi)\int_{-\pi m}^{\pi m} e^{iux}\varphi^*_{m,j}(-u)du$ to obtain

$$\sum_{j\in\mathbb{Z}} \varphi^2_{m,j}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |e^{iux}|^2 du = m.$$

Therefore, a third formulation of $\hat{g}_m$ is

$$\hat{g}_m = \sum_{j\in\mathbb{Z}} \hat{a}_{m,j}\varphi_{m,j} \text{ where } \hat{a}_{m,j} = \frac{1}{2\pi\Delta}\int \hat{\theta}_\Delta(u)\varphi^*_{m,j}(-u)du = \frac{1}{n\Delta}\sum_{k=1}^n Z_k \varphi_{m,j}(Z_k).$$

Due to the explicit formula (15), even if S_m is not finite-dimensional, we need not truncate the series. Nevertheless, the introduction of the basis is crucial for the proof. Using the development on $(\varphi_{m,j})_j$, we also have

$$\|\hat{g}_m\|^2 = \sum_{j\in\mathbb{Z}} |\hat{a}_{m,j}|^2.$$

For $h \in \mathbb{L}^2(\mathbb{R})$, its orthogonal projection h_m on S_m can be written as

$$h_m = \sum_{j\in\mathbb{Z}} a_{m,j}(h)\varphi_{m,j} \quad \text{with } a_{m,j}(h) = \langle h, \varphi_{m,j}\rangle.$$

We consider a collection $(S_m, m = 1, \ldots, m_n)$ where m_n is restricted to satisfy $m_n \leq n\Delta$.

As it is usual, we select adaptively the value as follows:

$$\hat{m} = \arg\min_{m\in\{1,\ldots,m_n\}} (\gamma_n(\hat{g}_m) + \text{pen}(m)) \text{ with } \text{pen}(m) = \kappa\left(\frac{1}{n\Delta}\sum_{k=1}^n Z_k^2\right)\frac{m}{n\Delta}.$$

We shall denote by

$$\text{pen}_{th}(m) = \mathbb{E}(\text{pen}(m)) = \kappa(\mathbb{E}(Z_1^2)/\Delta)\frac{m}{n\Delta}.$$

Then we can prove

Theorem 3.1. *Assume that* (H2)(8)-(H3)-(H4) *are fulfilled, that n is large and Δ is small with $n\Delta$ tends to infinity when n tends to infinity. Then*

$$\mathbb{E}(\|g-\hat{g}_{\hat{m}}\|^2) \leq C \inf_{m\in\{1,\ldots,m_n\}} \left(\|g-g_m\|^2 + \text{pen}_{th}(m)\right) + \frac{C'\Delta^2}{2\pi}\int_{-\pi m_n}^{\pi m_n} u^2|g^*(u)|^2 du + \frac{C\text{''} \ln^2(n\Delta)}{n\Delta}.$$

If g belongs to a class of regularity $\mathcal{C}(a,L)$, with unknown a and L, the estimator is automatically such that

$$\mathbb{E}(\|g-\hat{g}_{\hat{m}}\|^2) \leq C\left[(n\Delta)^{-2a/(2a+1)} + \Delta^2 m_n^{2(1-a)_+} + \frac{C\text{''} \ln^2(n\Delta)}{n\Delta}\right].$$

If either ($a \geq 1$, $n\Delta^3 = O(1)$) or ($0 < a < 1$ and $n\Delta^2 = O(1)$), then

$$\mathbb{E}(\|g-\hat{g}_{\hat{m}}\|^2) = O((n\Delta)^{-2a/(2a+1)}).$$

Remark 3.2. *In Comte and Genon-Catalot (2008), it is assumed that Δ is fixed and that $|\psi_\Delta(x)| \asymp c(1+x^2)^{-b\Delta/2}$. Using a different estimator and a different penalty (since ψ_Δ has to be estimated), we obtained that the $\mathbb{L}^2$-risk of the adaptive estimator of g automatically attains the rate $O((n\Delta)^{-2a/(2b\Delta+2a+1)})$, when g belongs to $\mathcal{C}(a,L)$: it appears that the exponent in the rate effectively depends on Δ. But it coincides with the exponent obtained here when $\Delta \to 0$ or $b = 0$ (compound Poisson process).*

4. Estimation of g on a compact set.

4.1. **Time domain point of view.** In this section, we intend to proceed without Fourier inversion and directly use the fact that $(1/(n\Delta))\sum_{k=1}^n Z_k\delta_{Z_k} = \hat{\mu}_n$ converges weakly to $\mu(dx) = g(x)dx$. Recall that, for any function t such that t^* is compactly supported,

$$\gamma_n(t) = \|t\|^2 - \frac{2}{2\pi}\langle\frac{\hat{\theta}_\Delta}{\Delta}, t^*\rangle.$$

Since $\hat{\theta}_\Delta/\Delta$ is the Fourier Transform of $\hat{\mu}_n$, we now consider, with the same notation and for any compactly supported function t,

$$\gamma_n(t) = \|t\|^2 - 2\langle\hat{\mu}_n, t\rangle = \|t\|^2 - \frac{2}{n\Delta}\sum_{k=1}^n Z_k t(Z_k).$$

More precisely, we fix a compact set $A \subset \mathbb{R}$ and focus on the estimation of $g_A := g\mathbf{I}_A$. In other words, the estimation is performed in the "time domain" instead of previously, the "frequency domain". We consider a family $(\Sigma_m, m \in \mathcal{M}_n\}$ of finite dimensional linear subspaces of $\mathbb{L}^2(A)$: $\Sigma_m = \text{span}\{\varphi_\lambda, \lambda \in \Lambda_m\}$ where $\text{card}(\Lambda_m) = D_m$ is the dimension of Σ_m. The set $\{\varphi_\lambda, \lambda \in \Lambda_m\}$ denotes an orthonormal basis of Σ_m. We shall denote by $\|f\|_A^2 = \int_A f^2(u)du$ for any function f.

For $m \geq 1$, we define

$$\tilde{g}_m = \arg\min_{t\in\Sigma_m} \gamma_n(t). \tag{20}$$

4.2. **Projection spaces and their fundamental properties.** We consider projection spaces satisfying

(M1) $(\Sigma_m)_{m\in\mathcal{M}_n}$ is a collection of finite-dimensional linear sub-spaces of $\mathbb{L}^2(A)$, with dimension D_m such that $\forall m \in \mathcal{M}_n, D_m \le n\Delta$. For all m, functions in Σ_m are of class C^1 in A, and, satisfy

$$\exists \Phi_0 > 0, \forall m \in \mathcal{M}_n, \forall t \in S_m, \|t\|_\infty \le \Phi_0\sqrt{D_m}\|t\|_A, \quad \text{and} \quad \|t'\|_A \le \Phi_0 D_m \|t\|_A. \tag{21}$$

where $\|t\|_\infty = \sup_{x\in A}|t(x)|$.

(M2) $(\Sigma_m)_{m\in\mathcal{M}_n}$ is a collection of nested models, all embedded in a space $\mathcal{S}_n$ belonging to the collection ($\forall m \in \mathcal{M}_n, \Sigma_m \subset \mathcal{S}_n$). We denote by N_n the dimension of $\mathcal{S}_n$: $\dim(\mathcal{S}_n) = N_n$ ($\forall m \in \mathcal{M}_n, D_m \le N_n \le n\Delta$).

Inequality (21) is often referred to as the *norm connection* property of the projection spaces and is the basic tool to obtain the adequate order of the risk bound. It follows from Lemma 1 in Birgé and Massart (1998), that (21) is equivalent to

$$\exists \Phi_0 > 0, \|\sum_{\lambda\in\Lambda_m} \varphi_\lambda^2\|_\infty \le \Phi_0^2 D_m. \tag{22}$$

Functions of the spaces Σ_m are considered as functions on $\mathbb{R}$ equal to zero outside A.

Here are the examples we have in view, and that we describe with $A = [0,1]$ for simplicity. They satisfy assumptions (M1) and (M2).

[T] *Trigonometric spaces*: they are generated by $\varphi_0 = 1_{[0,1]}$, $\varphi_j(x) = \sqrt{2}\cos(2\pi jx)\mathbf{1}_{[0,1]}(x)$ and $\varphi_{j+m+1}(x) = \sqrt{2}\sin(2\pi jx)\mathbf{1}_{[0,1]}(x)$ for $j = 1,\dots,m$ }, $D_m = 2m+1$ and $\mathcal{M}_n = \{1,\dots,[n\Delta/2]-1\}$.

[W] *Dyadic wavelet generated spaces* with regularity $r \ge 2$ and compact support, as described e.g. in Härdle *et al.* (1998). The generating basis is of cardinality $D_m = 2^{m+1}$ and $m \in \mathcal{M}_n = \{1, 2, \dots, [\ln(n\Delta)/2] - 1\}$.

4.3. **Integrated risk on a compact set.** Now, we have

$$\tilde{g}_m = \sum_{\lambda\in\Lambda_m} \tilde{a}_\lambda\varphi_\lambda \text{ with } \tilde{a}_\lambda = \frac{1}{n\Delta}\sum_{k=1}^n Z_k\varphi_\lambda(Z_k). \tag{23}$$

Let g_m denote the orthogonal projection of g on Σ_m, now given by $g_m = \sum_{\lambda\in\Lambda_m} a_\lambda\varphi_\lambda$ with $a_\lambda = (1/\Delta)\mathbb{E}(Z_1\varphi_\lambda(Z_1))$. At this stage, note that the "time domain approach" differs from the "frequency domain approach" only through the projection spaces.

A useful decomposition of the contrast is

$$\gamma_n(t) - \gamma_n(s) = \|t-g\|^2 - \|s-g\|^2 - 2\nu_n(t-s) - 2R_n(t-s), \tag{24}$$

where we set

$$\nu_n(t) = \frac{1}{n\Delta}\sum_{k=1}^n (Z_k t(Z_k) - \mathbb{E}(Z_1 t(Z_1))), \tag{25}$$

and

$$R_n(t) = \frac{1}{\Delta}\mathbb{E}(Z_1 t(Z_1)) - \int t(x)g(x)dx. \tag{26}$$

We can prove the following propositions:

Proposition 4.1. *Let $t \in \Sigma_m$ and assume that* (H1) *and* (H3) *hold.*
1) If $L := \int u^2 |g^*(u)|^2 du < +\infty$, then

$$|R_n(t)| \le \Delta \|t\|_A \|g\|_1 L^{1/2}/\sqrt{2\pi}.$$

2) If g is bounded, $|R_n(t)| \le C\Phi_0 \|t\|_A \Delta D_m$ where C depends on $\|g\|_1$, $\|g\|$ and A.
3) Otherwise:

$$|R_n(t)| \le C\Phi_0 \|t\|_A (\sqrt{\Delta D_m} + \Delta D_m), \tag{27}$$

where C depends on $\|g\|_1$, $\|g\|$ and A.

Proposition 4.2. *Assume that* (H1)-(H2)(2)-(H3) *hold. We consider $\tilde{g}_m$, an estimator of g defined by (23) on a space Σ_m and denote by g_m the orthogonal projection of g on Σ_m. Then*

$$\mathbb{E}(\|\tilde{g}_m - g\|_A^2) \le 3\|g - g_m\|_A^2 + 16\Phi_0[\mathbb{E}(Z_1^2)/\Delta]\frac{D_m}{n\Delta} + K\rho_{m,\Delta}, \tag{28}$$

where K depends on m_1, m_2 and g and $\rho_{m,\Delta} = \Delta^2$ if $\int u^2|g^*(u)|^2 du < +\infty$, $\rho_{m,\Delta} = \Delta^2 D_m^2$ if g is bounded. Otherwise $\rho_{m,\Delta} = \Delta D_m$ if $n\Delta^2 \le 1$.

As for Proposition 3.1, we draw the consequences of Proposition 4.2 on the rate of convergence of the risk bound.

In the setting of this section, the regularity of g_A must be described by using classical Besov spaces on compact sets. Let us recall that the Besov space $\mathcal{B}_{\alpha,2,\infty}([0,1])$ is defined by:

$$\mathcal{B}_{\alpha,2,\infty}([0,1]) = \{f \in \mathbb{L}^2([0,1]), \quad |f|_{\alpha,2} := \sup_{t>0} t^{-\alpha}\omega_r(f,t)_2 < +\infty\}$$

where $r = [\alpha] + 1$ ([.] denotes the integer part), and $\omega_r(f,t)_2$ is called the r-th modulus of smoothness of a function $f \in \mathbb{L}^2(A)$. Note that $|f|_{\alpha,2}$ is a semi-norm with usual associated norm $\|f\|_{\alpha,2} = \|f\|_2 + |f|_{\alpha,2}$, $\|f\|_2 = \left(\int |f|^2(x)dx\right)^{1/2}$. For details, we refer to DeVore and Lorentz (1993, p.54-57).

Heuristically, a function in $\mathcal{B}_{\alpha,2,\infty}([0,1])$ can be seen as square integrable and $[\alpha]$-times differentiable with derivative of order $[\alpha]$ having a Hölder property of order $\alpha - [\alpha]$.

Proposition 4.3. *Consider $A = [0,1]$ and Σ_m a space in collection* [T] *or* [W]. *Assume that* (H1), (H2)(2) *and* (H3) *hold. Let $g \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, $D_m = (n\Delta)^{1/(2\alpha+1)}$ and $\Delta = n^{-a}$ with $a \in (0,1)$.*

- *If $\int u^2|g^*(u)|^2 du < +\infty$, choose $a \ge \alpha/(3\alpha+1)$,*
- *If g is bounded, choose $a \ge (\alpha+1)/(3\alpha+2)$,*
- *Otherwise, choose $a \ge 1/2$.*

Then $\mathbb{E}(\|g - \tilde{g}_m\|_A^2) \le K(n\Delta)^{-2\alpha/(2\alpha+1)}$.

Proof. It is well known (see DeVore and Lorentz (1993)) that, if Σ_m is a space of [T] or [W], and if $g \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, then $\|g - g_m\|_{[0,1]}^2 \le CD_m^{-2\alpha}$. The usual compromise between $D_m^{-2\alpha}$ and $D_m/(n\Delta)$ leads to the best choice $D_m = O((n\Delta)^{1/(2\alpha+1)})$. Therefore, the first two terms in (28) have rate $O((n\Delta)^{-2\alpha/(2\alpha+1)})$. Now, we search for the choice of $\Delta = n^{-a}$ such that $\rho_{m,\Delta} \le (n\Delta)^{-2\alpha/(2\alpha+1)}$. If $\int u^2|g^*(u)|^2du < +\infty$, $\rho_{m,\Delta} = \Delta^2$ and we find $a \ge \alpha/(3\alpha+1)$. If g is bounded, $\rho_{m,\Delta} = \Delta^2 D_m^2$ and we find $a \ge (\alpha+1)/(3\alpha+2)$. Otherwise, $\rho_{m,\Delta} = \Delta D_m$ and we find $a \ge 1/2$. $\square$

Note that $a \ge \alpha/(3\alpha+1)$ and $a \ge (\alpha+1)/(3\alpha+2)$ holds for any $\alpha \ge 0$ if $a \ge 1/3$ (hence $n\Delta \le n^{2/3}$), and $a \ge 1/2$ implies $n\Delta \le n^{1/2}$.

4.4. **Adaptive result.** Now, to get an adaptive result, we need to define a penalty function pen(.) and set

$$\tilde{m} = \arg\min_{m\in\mathcal{M}_n} (\gamma_n(\tilde{g}_m) + \mathrm{pen}(m))$$

Let

$$\mathrm{pen}(m) = \frac{\kappa}{n\Delta}\sum_{i=1}^{n} Z_i^2 \frac{D_m}{n\Delta}, \quad \mathrm{pen}_{th}(m) = \mathbb{E}(\mathrm{pen}(m)) = \kappa\mathbb{E}(Z_1^2/\Delta)\frac{D_m}{n\Delta}.$$

Here too, we use the same notation pen(m), pen$_{th}(m)$ as above, although the definitions differ. Then the following theorem holds:

Theorem 4.1. *Assume that assumptions* (H1)-(H2)(12)-(H3) *and* (M1)-(M2) *are fulfilled. Consider a nested collection of models and the estimator $\tilde{g}_{\tilde{m}}$, then*

$$\mathbb{E}(\|g-\tilde{g}_{\tilde{m}}\|_A^2) \le C \inf_{m\in\mathcal{M}_n}\left(\|g-g_m\|_A^2 + \mathrm{pen}_{th}(m)\right) + C\rho_{n,\Delta} + \frac{C'}{n\Delta},$$

where $\rho_{n,\Delta} = \Delta^2$ if $\int u^2|g^*(u)|^2du < +\infty$, $\rho_{n,\Delta} = \Delta^2 N_n^2$ if g is bounded. Otherwise, $\rho_{n,\Delta} = \Delta N_n$.

Remark 4.1. *The moment condition of order 12 in Theorem 4.1 can be weakened into a condition of order 8 for basis [T], which is bounded.*

The constant κ is numerical and must be calibrated by preliminary simulations (see Section 6).

Then the following (standard) rate is obtained:

Corollary 4.1. *Let the S_m's be D_m-dimensional linear spaces in collections* [T] *or* [W]. *Assume moreover that g belongs to $\mathcal{B}_{\alpha,2,\infty}([0,1])$ with $r > \alpha > 0$ and $\Delta = n^{-a}$ with $a \in [1/3, 1[$ if $\int u^2|g^*(u)|^2du < +\infty$, $a \in [3/5, 1[$ if g is bounded, and otherwise, $a \in [2/3, 1[$. Then, under the assumptions of Theorem 4.1,*

$$\mathbb{E}(\|g-\tilde{g}_{\tilde{m}}\|^2) = O\left((n\Delta)^{-\frac{2\alpha}{2\alpha+1}}\right). \tag{29}$$

Remark 4.2. *The bound r on α stands for the regularity of the basis functions for collection [W]. For the trigonometric collection [T], no upper bound for the regularity α is required.*

Proof. The result is a straightforward consequence of the results of DeVore and Lorentz (1993) and of Lemma 12 of Barron, Birgé and Massart (1999). They imply that, if $g \in \mathcal{B}_{\alpha,2,\infty}([0,1])$ for some $\alpha > 0$, then $\|g - g_m\|$ is of order $D_m^{-\alpha}$ in the collections [T] and [W]. Thus the infimum in Theorem 4.1 is reached for $D_{m_n} = O([(n\Delta)^{1/(1+2\alpha)}])$, which is less than $n\Delta$ for $\alpha > 0$. We know that the collection of models is such that $N_n \le n\Delta$. Thus, if $\int u^2|g^*(u)|^2du < +\infty$, $\Delta^2 \le 1/(n\Delta)$ holds for $\Delta = n^{-a}$ if $a \in [1/3, 1[$. If g is bounded, $\Delta^2 N_n^2 \le 1/(n\Delta)$ holds if $\Delta^2(n\Delta)^2 \le 1/(n\Delta)$ which gives $a \in [3/5, 1[$. Otherwise, $N_n\Delta \le 1/(n\Delta)$ holds for $\Delta = n^{-a}$ if $a \in [2/3, 1[$. Unfortunately, this also implies that $n\Delta \le n^{2/3}$ in the first case, $n\Delta \le n^{2/5}$ in the second case and $n\Delta \le n^{1/3}$ in the third case. Then, we find the standard nonparametric rate of convergence $(n\Delta)^{-2\alpha/(1+2\alpha)}$. □

Remark 4.3. *Figueroa-López and Houdré (2006) investigate the nonparametric estimation of $n(.)$ from a continuous observation $(L_t)_{t\in[0,T]}$. They use projection methods and penalization to obtain estimators with rate $O(T^{-2\alpha/(2\alpha+1)})$ on a Besov class $\mathcal{B}_{\alpha,2,\infty}([0,1])$. Thus, our result can be compared to theirs since our rate is $O((n\Delta)^{-2\alpha/(2\alpha+1)})$.*

5. Rates of convergence on examples.

The discussion on rates of convergence is different according to the estimation method. We give some illustrating examples.

5.1. Rates for the Fourier method on examples.

Below, we compute with more precision the risk bound given by Theorem 3.1 on specific examples.

Example 1. Compound Poisson processes.

Let $L_t = \sum_{i=1}^{N_t} Y_i$, where (N_t) is a Poisson process with constant intensity c and (Y_i) is a sequence of i.i.d. random variables with density f independent of the process (N_t). Then, (L_t) is a compound Poisson process with characteristic function (1) with $n(x) = cf(x)$. Assumptions (H1)-(H2)(p) are equivalent to $\mathbb{E}(|Y_1|^p) < \infty$. Assumption (H3) is equivalent to $\int_{\mathbb{R}} x^2 f^2(x)dx < \infty$, which holds for instance if $\sup_x f(x) < +\infty$ and $\mathbb{E}(Y_1^2) < +\infty$. The distribution of $Z_1 = L_\Delta$ is:

$$P_\Delta(dz) = P_{Z_1}(dz) = e^{-c\Delta}\left(\delta_0(dz) + \sum_{n\geq 1} f^{*n}(z)\frac{(c\Delta)^n}{n!}dz\right). \tag{30}$$

Hence,

$$\mu_\Delta(dz) = e^{-c\Delta}\left(czf(z) + c^2\Delta z\sum_{n\geq 2}\frac{c^{n-2}\Delta^{n-2}}{n!}f^{*n}(z)dz\right) \tag{31}$$

Now as f is any density and $g(x) = cxf(x)$, any type of rate can be obtained.

We summarize in Table 1 the rates obtained for several examples that we test below by simulation experiments.

Density f	Gaussian $\mathcal{N}(0,1)$	Exponential $\mathcal{E}(1)$	Uniform $\mathcal{U}([0,1])$
$g(x)(= cxf(x)) =$	$cxe^{-x^2/2}/\sqrt{2\pi}$	$cxe^{-x}\mathbf{1}_{\mathbb{R}^+}(x)$	$cx\mathbf{1}_{[0,1]}(x)$
$g^*(x) =$	$cixe^{-x^2/2}$	$c/(1-ix)^2$	$c\dfrac{e^{ix}-1-ixe^{ix}}{x^2}$
$\int_{\|x\|\geq\pi m}\|g^*(x)\|^2dx =$	$O(me^{-\pi^2m^2})$	$O(m^{-3})$	$O(m^{-1})$
$\int_{\|x\|\leq\pi m_n}x^2\|g^*(x)\|^2dx =$	$O(1)$	$O(1)$	$O(m_n)$
Constraint on Δ	$n\Delta^3\leq 1$	$n\Delta^3\leq 1$	$n\Delta^2\leq 1$
Selected $m =$	$m = \sqrt{\ln(n\Delta)}/\pi$	$m = O((n\Delta)^{1/4})$	$m = O((n\Delta)^{1/2})$
Rate $=$	$O(\dfrac{\sqrt{\ln(n\Delta)}}{n\Delta})$	$O((n\Delta)^{-3/4}))$	$O((n\Delta)^{-1/2}))$

Table 1. Choice of m and rates in three compound Poisson examples.

It is worth stressing that the rates obtained in Table 1 are the same as the ones obtained for fixed Δ in Comte and Genon-Catalot (2008). This is because $b = 0$ for compound Poisson models (see Remark 3.2). The fixed Δ case requires a more complicated estimator, but no additional constraint on Δ, while here, the estimator is simpler, under small Δ.

For instance, for $\Delta = n^{-a}$, with $a \in [1/3, 1[$, the best risk is of order $\ln^{1/2}(n)/n^{2/3}$ in the Gaussian case and of order $n^{-1/2}$ in the exponential case. In the uniform case for $\Delta = n^{-a}$ and now

$a \in [1/2, 1[$, the best risk is of order $n^{-1/4}$.

Example 2. The Lévy gamma process. Let $\alpha > 0, \beta > 0$. The Lévy Gamma process (L_t) with parameters (β, α) is a subordinator such that, for all $t > 0$, L_t has distribution Gamma with parameters $(\beta t, \alpha)$, *i.e.* has density:

$$\frac{\alpha^{\beta t}}{\Gamma(\beta t)} x^{\beta t - 1} e^{-\alpha x} 1_{x \geq 0}. \tag{32}$$

The characteristic function of Z_1 is equal to:

$$\psi_\Delta(u) = \left(\frac{\alpha}{\alpha - iu} \right)^{\beta \Delta}. \tag{33}$$

The Lévy density is $n(x) = \beta x^{-1} e^{-\alpha x} \mathbf{1}_{\{x>0\}}$ so that $g(x) = \beta e^{-\alpha x} \mathbf{1}_{\{x>0\}}$ satisfies our assumptions. We have: $g^*(x) = \beta/(\alpha - ix)$. Table 2 gives the risk bound and auxiliary quantities.

Example 2. (continued) More generally, we consider the Lévy process (L_t) with parameters (δ, β, c) and Lévy density

$$n(x) = c x^{\delta - 1/2} x^{-1} e^{-\beta x} 1_{x>0}.$$

For $\delta > 1/2$, $\int_0^{+\infty} n(x)dx < +\infty$, and we recover compound Poisson processes. For $0 < \delta \leq 1/2$, $\int_0^{+\infty} n(x)dx = +\infty$ and $g(x) = xn(x)$ belongs to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$. This includes the case $\delta = 1/2$ of the Lévy Gamma process. We have:

$$g^*(u) = c \frac{\Gamma(\delta + 1/2)}{(\beta - iu)^{\delta + 1/2}}.$$

It follows from Table 2 that for $\Delta = n^{-a}$, with $a \in [1/2, 1[$, the best risk is of order $n^{-\delta/(2\delta+1)}$.

Example 3. The variance Gamma stochastic volatility model. This model was introduced by Madan and Seneta (1990).

Let (W_t) be a Brownian motion, and let (V_t) be an increasing Lévy process (subordinator), independent of (W_t). Assume that the observed process is

$$L_t = W_{V_t}.$$

We have

$$\psi_\Delta(u) = \mathbb{E}(e^{iuL_\Delta}) = \mathbb{E}(e^{-\frac{u^2}{2} V_\Delta}) = \left(\frac{\alpha}{\alpha + \frac{u^2}{2}} \right)^{\Delta \beta}.$$

The Lévy measure of (L_t) is equal to:

$$n_L(x) = \beta (2\alpha)^{1/4}) |x|^{-1} \exp\left(-(2\alpha)^{1/2} |x|\right).$$

We can compute the density of $L_\Delta = Z_1$ which is a variance mixture of Gaussian distributions with mixing distribution Gamma $\Gamma(\beta\Delta, \alpha)$:

$$\begin{aligned} f_{Z_1}(x) &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} v^{\beta\Delta - 3/2} e^{-\frac{1}{2}(x^2/v + 2\alpha v)} \frac{\alpha^{\beta\Delta}}{\Gamma(\beta\Delta)} dv \\ &= \frac{2}{\sqrt{2\pi}} \frac{\alpha^{\beta\Delta}}{\Gamma(\beta\Delta)} \left(\frac{(2\alpha)^{1/2}}{|x|}\right)^{\frac{1}{2} - \beta\Delta} K_{\beta\Delta - \frac{1}{2}}((2\alpha)^{1/2}|x|) \end{aligned}$$

where K_ν is the modified Bessel function (third kind) with index ν (see *e.g.* Lebedev).

Now with $\tilde{\alpha} = (2\alpha)^{1/2}$, $\tilde{\beta} = \beta(2\alpha)^{1/4}$,

$$g(x) = \tilde{\beta}\exp(-\tilde{\alpha}x)\mathbf{1}_{x\geq 0} - \tilde{\beta}\exp(\tilde{\alpha}x)\mathbf{1}_{x<0} \Rightarrow g^*(x) = \frac{2i\tilde{\alpha}\tilde{\beta}x}{\tilde{\alpha}^2 + x^2}.$$

Example 3 (continued). The variance Gamma stochastic volatility model is a special case of bilateral Gamma process (see Küchler and Tappe (2008), Comte and Genon-Catalot (2008)). Consider the Lévy process L_t with characteristic function

$$\psi_t(u) = \left(\frac{\alpha}{\alpha - iu}\right)^{\beta t}\left(\frac{\alpha'}{\alpha' + iu}\right)^{\beta' t}$$

and Lévy density

$$n(x) = |x|^{-1}(\beta e^{-\alpha x}\mathbf{1}_{(0,+\infty)}(x) + \beta' e^{-\alpha|x|}\mathbf{1}_{(-\infty,0)}(x)).$$

Rates are given in Table (2).

Process	Example 2	Ex.2 (continued) $\delta \in]0,1/2[$	Example 3 (continued)
$g^*(x) =$	$\dfrac{\beta}{\alpha - ix}$	$c\dfrac{\Gamma(\delta+1/2)}{(\beta - iu)^{\delta+1/2}}$	$\dfrac{\beta}{\alpha - ix} - \dfrac{\beta'}{\alpha' - ix}$
$\int_{\lvert x\rvert\geq \pi m} \lvert g^*(x)\rvert^2 dx =$	$O(1/m)$	$O(1/m^{2\delta})$	$O(1/m)$
$\int_{\lvert x\rvert\leq \pi m_n} x^2\lvert g^*(x)\rvert^2 dx =$	$O(m_n)$	$O(m_n^{2-2\delta})$	$O(m_n)$
Constraint on Δ	$n\Delta^2 \leq 1$	$n\Delta^2 \leq 1$	$n\Delta^2 \leq 1$
Selected $m =$	$O((n\Delta)^{1/2})$	$O((n\Delta)^{1/(2\delta+1)})$	$O((n\Delta)^{1/2})$
Rate (small Δ)	$O((n\Delta)^{-1/2})$	$O((n\Delta)^{-2\delta/(2\delta+1)})$	$O((n\Delta)^{-1/2})$
Rate (fixed Δ) (see [5] (2008))	$O((n\Delta)^{-1/(2\beta\Delta+1)})$	$O([\ln(n\Delta)]^{-2\delta})$	$O((n\Delta)^{-1/(4\beta\Delta+1)})$

TABLE 2. Choice of m and rates in examples 2, 2 (continued), 3 (continued).

The rates of the last lime in Table 2 come from Comte and Genon-Catalot (2008). As announced before (see Remark 3.2), the rates for small Δ are different from the rates with fixed Δ, and they are in all cases better. Moreover, the estimation strategy in simpler and more complete (see Remark 4.2 in Comte and Genon-Catalot (2008)). The price to pay is the constraint on Δ.

5.2. **Rates for the estimation on a compact set.** In all the examples above, it is possible to find a compact set A such that g is of class C^∞ on A.

Due to Corollary 4.1, for all $\alpha > 0$, $\mathbb{E}(\|g - \tilde{g}_{\hat{m}}\|_A^2) = O((n\Delta)^{-2\alpha/(2\alpha+1)})$.

For the conditions under which this rate arises, three possibilities happen:

(1) for the compound Poisson process with Gaussian and exponential density, we have $\int u^2|g^*(u)|^2 du < +\infty$,
(2) for the compound Poisson process with uniform density f, the Lévy Gamma process and the bilateral Lévy Gamma process, we have $\int u^2|g^*(u)|^2 du = +\infty$ and g is bounded.
(3) For the Lévy-δ (see Example 2 (continued)), $\int u^2|g^*(u)|^2 du = +\infty$ and g is not bounded.

Choosing $\Delta = n^{-a}$ (see Corollary 4.1), in the first case, the best rate corresponding to $\alpha \to +\infty$ is of order $O(n^{-2/3})$, for the second case, of order $O(n^{-2/5})$ and for the third case of order $O(n^{-1/3})$.

To conclude this section we show in Table 3 the best rate that can be obtained on each example according to the method, either Fourier method (Sinus Cardinal basis) or the time domain method (Trigonometric basis). The winner of the challenge is always the trigonometric basis. This is because the limit $\alpha \to +\infty$ is considered for the latter basis only. However, on simulations, the global method performs better.

Process	Sinus Cardinal basis	Trigonometric basis
Poisson-Gaussian	$\ln^{1/2}(n)n^{-2/3}$	$n^{-2/3}$
Poisson-Exp.	$n^{-1/2}$	$n^{-2/3}$
Poisson-Unif.	$n^{-1/4}$	$n^{-2/5}$
Lévy-Gamma	$n^{-1/4}$	$n^{-2/5}$
Lévy-δ	$n^{-\delta/(2\delta+1)}, \delta \in (0, 1/2)$	$n^{-1/3}$
Bilateral Gamma	$n^{-1/4}$	$n^{-2/5}$

TABLE 3. Comparison of best possible rates with the two methods.

In all cases, rates measured as powers of n are very slow. As will be illustrated in the simulations, the important value is $n\Delta$, that should be large enough. This means that Δ cannot be too small in order to keep a reasonable number n of observations. This is why, in our simulations, we have not always taken Δ^2 smaller that $1/(n\Delta)$.

6. Simulations

We provide in this section simulation results. We have implemented the estimation method for two bases: the sinus cardinal basis of Section 3 and the trigonometric basis of Section 4. We simulated Lévy processes chosen among the examples given in Section 5. Precisely,

(1) A compound Poisson process with Gaussian $\mathcal{N}(0,1)$ Y_i's, $g(x) = cx\exp(-x^2/2)/\sqrt{2\pi}$.
(2) A compound Poisson process with Exponential $\mathcal{E}(1)$ Y_i's, $g(x) = cxe^{-x}\mathbf{1}_{x>0}$.
(3) A compound Poisson process with Uniform $\mathcal{U}([0,1])$ Y_i's, $g(x) = cx\mathbf{1}_{[0,1]}(x)$.
(4) A Lévy-Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$, $g(x) = \beta\exp(-\alpha x)\mathbf{1}_{x>0}$,
(5) A Lévy-Gamma process with parameters $(\alpha, \beta) = (1, 1)$,
(6) A Bilateral Lévy-Gamma process with parameters $(\alpha, \beta) = (\alpha', \beta') = (2, 0.2)$, $g(x) = \beta\exp(-\alpha x)\mathbf{1}_{x\geq 0} - \beta'\exp(\alpha' x)\mathbf{1}_{x<0}$,
(7) A Bilateral Lévy-Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$ and $(\alpha', \beta') = (1, 1)$

After preliminary experiments, the constant κ is taken equal to 7.5 for the sinus cardinal basis and to 1 for the trigonometric basis. The cutoff $\hat{m}$ is chosen among 100 equispaced values between 0 and 10. The dimension $D_{\hat{m}}$ is chosen among 80 values between 1 and 80. We used in both cases the expression of the estimators using the coefficients on the basis. In the sinus cardinal case, this avoids high dimensional matrices manipulations, but the series have to be truncated (we keep coefficients $\hat{a}_{m,j}$ for $|j| \leq K_n$ and we take $K_n = 15$).

Results are given in Figures 1 and 2. We give confidence bands for our estimators by plotting 50 estimated curves on the same figure. The first two columns give estimation results with the sinus cardinal basis for $n = 5000, \Delta = 0.2$ $(n\Delta = 1000)$ and $n = 50000, \Delta = 0.05$ $(n\Delta = 2500)$. The third columns concerns the trigonometric basis for $n = 50000, \Delta = 0.05$.

It is clear from the first two columns that increasing $n\Delta$ improves the result by showing a thinner confidence band. Comparing the last two columns amounts to comparing the performance of the two bases. It appears that the sinus cardinal must be preferred because the trigonometric basis has very important edge effects for highly dissymmetric densities: see in particular the exponential-Poisson, and the Gamma case, which start with a peak and end at zero.

On top of each graph in Figures 1 and 2, we give the mean of the selected values for $\hat{m}$ (sinus cardinal basis) or for $D_{\tilde{m}}$ (trigonometric basis) with the associated standard deviation in parentheses. Various values are chosen by the estimation procedure, and in each case, the standard deviation exhibits a reasonable variability. This is an indication that the constants in the penalties are adequately chosen: too small constants κ imply very unstable choices for the same model, while greater κ's quickly lead to null standard deviations for 50 sample paths. Note also that the higher the regularity of g, the smaller the selected $\hat{m}$'s and $D_{\tilde{m}}$'s (which is coherent with orders as $D_{\tilde{m}} = O(n^{1/(2\alpha+1)})$ for a regularity α). The uniform-Poisson case involves larger values for $\hat{m}$ than the two other Poisson cases, for instance.

At last, let us remark that, when $\int u^2|g^*(u)|^2du < +\infty$, we need not use a C^1-basis in the second approach. For instance, a histogram basis or a piecewise polynomial basis can also be implemented. But in practice, it is not possible to know if the condition is fulfilled or not.

7. Concluding remarks

In this paper, we have investigated in the high frequency framework the nonparametric estimation of the Lévy density $n(\cdot)$ of a pure jump Lévy process under assumption (H1). This paper complements a previous one (Comte and Genon-Catalot (2008)) where the low frequency framework was treated. The estimation of $n(.)$ is done through the estimation of the function $g(x) = xn(x)$. In here, we use two kinds of bases. On the one hand, the sinus cardinal basis, which is of classical use in deconvolution provides a global estimation. On the other hand, finite dimensional bases satisfying (21) provide an estimation of g restricted to a compact set. In each approach, an adaptive estimator is built which reaches automatically the classical best rate that can be achieved on a prescribed class of regularity. The estimators can easily be implemented. Especially in the case of the sinus cardinal basis, the method allows the automatic (and adaptive) choice of a cutoff value m in Fourier inversion, a point that was unsolved in several previous references (quoted in the introduction).

There remain several open problems: mainly, how can the method be extended to more general Lévy processes having a drift term and under a weaker assumption on $n(\cdot)$ such as $\int x^2 n(x)dx < +\infty$ (see Neumann and Reiss (2009)).

8. Appendix: Proofs

8.1. **Proof of Theorem 2.1.** We apply Rosenthal's inequality recalled in Appendix (see (51)):

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=1}^{n} Z_k^\ell - \mathbb{E}(Z_1^\ell)\right|^p\right) &\leq \frac{C(p)}{n^p}\left(\sum_{k=1}^{n}\mathbb{E}[|Z_k^\ell - \mathbb{E}(Z_1^\ell)|^p] + \left(\sum_{k=1}^{n}\mathbb{E}[(Z_k^\ell - \mathbb{E}(Z_1^\ell))^2]\right)^{p/2}\right)\\
&\leq \frac{C(p)}{n^p}(n\mathbb{E}[|Z_1^\ell - \mathbb{E}(Z_1^\ell)|^p] + n^{p/2}(\mathbb{E}[(Z_1^\ell - \mathbb{E}(Z_1^\ell))^2])^{p/2})\\
&\leq \frac{C'(p)}{n^p}(n\mathbb{E}(Z_1^{p\ell}) + n^{p/2}[\mathbb{E}(Z_1^{2\ell})]^{p/2})\\
&\leq C"(p)\left(\frac{\Delta}{n^{p-1}} + \left(\frac{\Delta}{n}\right)^{p/2}\right).
\end{aligned}$$

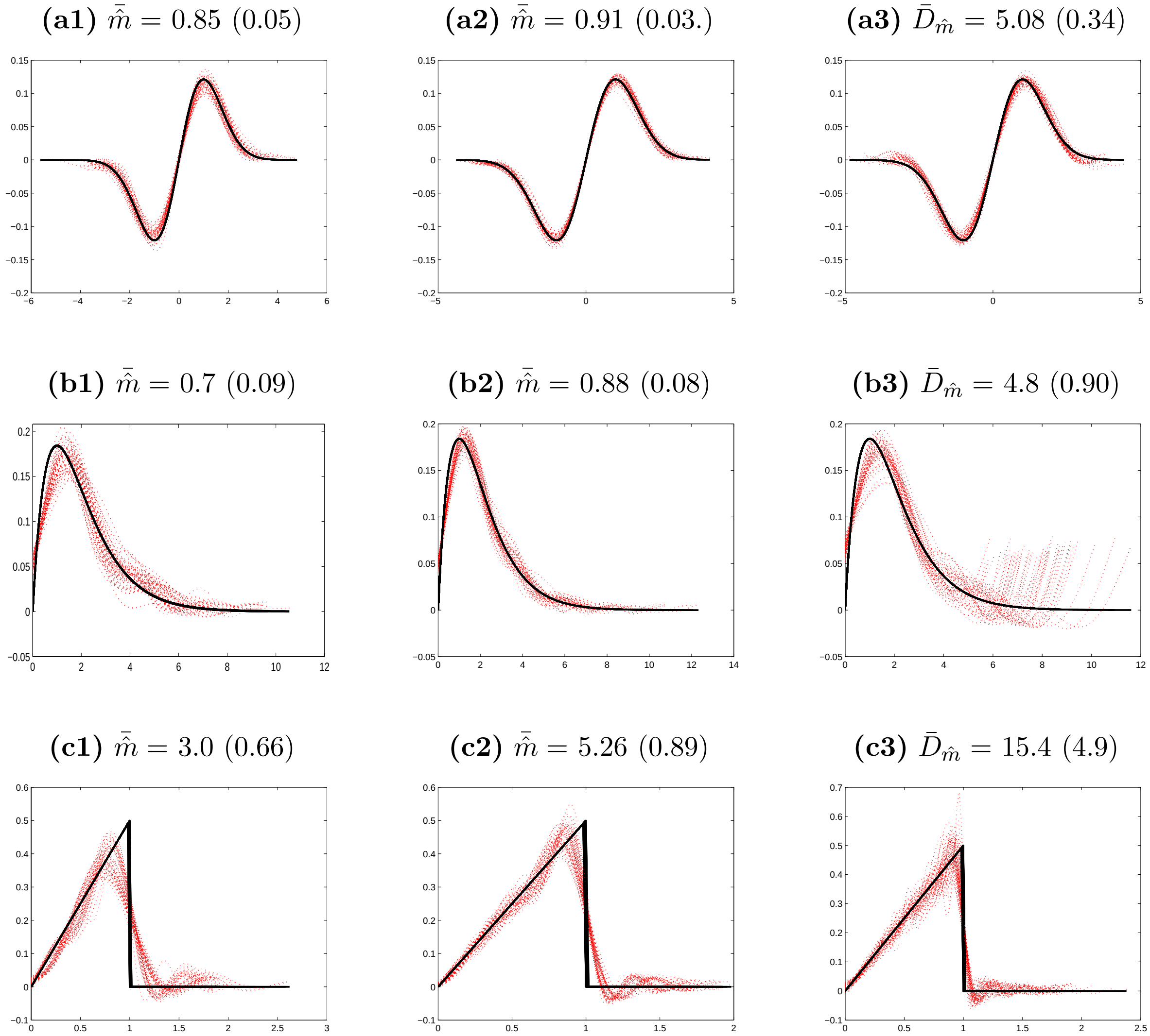

FIGURE 1. Confidence band for the estimation of g for a compound Poisson process with Gaussian (first line), Exponential $\mathcal{E}(1)$ (second line), and uniform $\mathcal{U}([0,1])$ (third line) Y_i's, $c = 0.5$. True (bold black line) and 50 estimated curves (dotted red), left $\Delta = 0.2$ $n = 5000$, Sinus Cardinal Basis; center, $\Delta = 0.05$, $n = 5.10^4$, Sinus Cardinal Basis; right $\Delta = 0.05$, $n = 5.10^4$, trigonometric basis.

We have

$$\frac{1}{n\Delta}\sum_{k=1}^{n} Z_k^\ell - m_\ell = \frac{1}{n\Delta}\left(\sum_{k=1}^{n}(Z_k^\ell - \mathbb{E}(Z_1^\ell)\right) + \frac{1}{\Delta}\mathbb{E}(Z_1^\ell) - m_\ell.$$

First note that, by Proposition 2.2,

$$\frac{1}{\Delta}\mathbb{E}(Z_1^\ell) - m_\ell = \Delta O(1).$$

Using that $\sqrt{n\Delta}\Delta = (n\Delta^3)^{1/2} = o(1)$, we see that the bias term tends to 0.

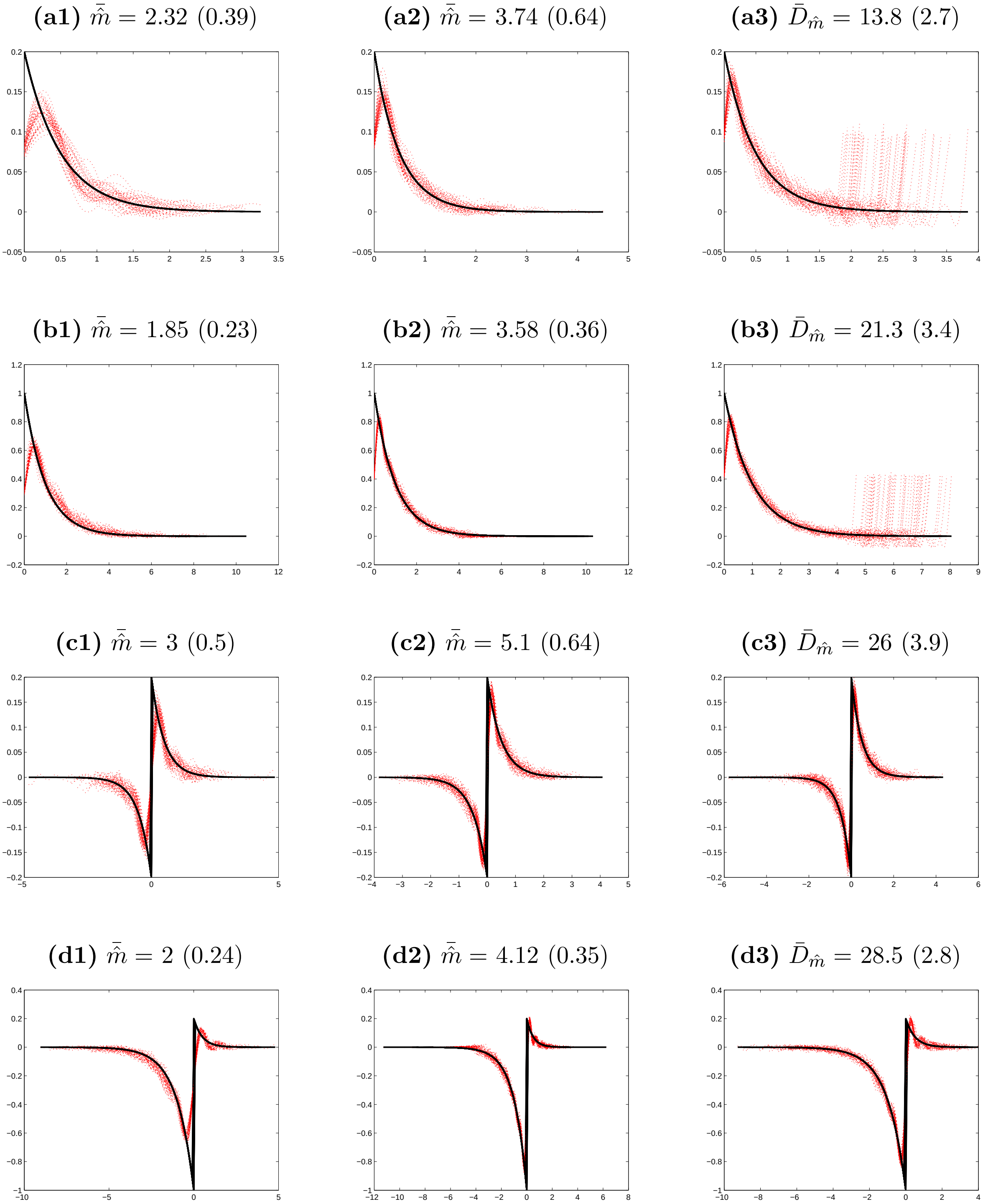

FIGURE 2. Confidence band for the estimation of g for a Lévy Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$ (first line), $(\alpha, \beta) = (1, 1)$ (second line), a bilateral Lévy Gamma process with parameters $(\alpha, \beta) = (\alpha', \beta') = (2, 0.2)$ (third line) and a bilateral Lévy Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$, $(\alpha', \beta') = (1, 1)$. True (bold black line) and 50 estimated curves (dotted red), left $\Delta = 0.2$ $n = 5000$, Sinus Cardinal Basis; center, $\Delta = 0.05$, $n = 5.10^4$, Sinus Cardinal Basis; right $\Delta = 0.05$, $n = 5.10^4$, trigonometric basis.

Let us introduce the centered i.i.d. random variables

$$\xi_k = \frac{1}{\sqrt{n\Delta}}(Z_k^\ell - \mathbb{E}(Z_1^\ell)).$$

We have

$$n\mathbb{E}(\xi_k^2) = \frac{1}{\Delta}(\mathbb{E}(Z_k^{2\ell}) - (\mathbb{E}(Z_1^\ell))^2) = m_{2\ell} + o(1).$$

And

$$n\mathbb{E}(\xi_k^4) \leq Cn\frac{1}{n^2\Delta^2}(\mathbb{E}(Z_k^{4\ell}) + (\mathbb{E}(Z_1^\ell)^4)) = \frac{1}{n\Delta}(m_{4\ell} + O(1)),$$

which tends to 0. Hence, the result. □

8.2. **Proof of Proposition 2.3.** By the Taylor formula,

$$\psi_\Delta(u) - 1 = u\psi_\Delta^{'}(c_u u) = iu\Delta\psi_\Delta(c_u u)g^*(c_u u),$$

for some $c_u \in (0,1)$. This gives (10) and thus (11).
For the other bound with $p = 1$, note that

$$\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2) = \frac{1}{n}\mathrm{Var}(Z_1 \exp(iuZ_1)) \leq \frac{1}{n}\mathbb{E}(Z_1^2).$$

Inequality (13)follows.

For $p \geq 1$, we apply Rosenthal's inequality recalled in Appendix (see (51)):

$$\begin{aligned}\mathbb{E}\left(\mathbb{E}(|\hat{\theta}_\Delta(u))) - \theta_\Delta(u)|^{2p})\right) \quad &\leq \quad \frac{C(2p)}{n^{2p}}\left(\sum_{k=1}^n \mathbb{E}[|Z_k e^{iuZ_k} - \mathbb{E}(Z_k e^{iuZ_k})|^{2p}]\right. \\ &\qquad \left. + \left(\sum_{k=1}^n \mathbb{E}|Z_k e^{iuZ_k} - \mathbb{E}(Z_k e^{iuZ_k})|^2]\right)^p\right) \\ &\leq \quad \frac{C'(2p)}{n^{2p}}(n\mathbb{E}(Z_1^{2p}) + n^p(\mathbb{E}(Z_1^2))^p).\end{aligned}$$

We conclude using Proposition 2.2 and $p \geq 1$. □

8.3. **Proof of Theorem 2.2.** We have

$$\sqrt{n\Delta}(\hat{\theta}_\Delta(u)/\Delta - g^*(u)) = \sqrt{n\Delta}\left(\frac{\hat{\theta}_\Delta(u) - \theta_\Delta(u)}{\Delta} + g^*(u)(\psi_\Delta(u) - 1)\right).$$

Hence, the bias term in $\sqrt{n\Delta}(\hat{\theta}_\Delta(u)/\Delta - g^*(u))$ is of order $\sqrt{n\Delta}\Delta = \sqrt{n\Delta^3}$. This explains the condition $n\Delta^3 = o(1)$ in Theorem 2.2. There remains to study $X_n(u) = \sqrt{n\Delta}(\Delta^{-1}\hat{\theta}_\Delta(u) - \Delta^{-1}\theta_\Delta(u))$.

Using (5), we have,

$$X_n(u) = \int e^{iux}\sqrt{n\Delta}(\hat{\mu}_n(dx) - \mu_\Delta(dx)) = \sum_{k=1}^n X_{k,n}(u),$$

with

$$X_{k,n}(u) = \frac{1}{\sqrt{n\Delta}}(Z_k e^{iuZ_k} - \theta_\Delta(u)).$$

Consider, for any integer $l \geq 1$, $u_1, u_2, \ldots, u_l \in \mathbb{R}$. We need to prove that the random vector $X_n = (X_n(u_1), X_n(u_2), \ldots, X_n(u_l))'$ (with values in $\mathbb{C}^l$) converges in distribution to $\mathcal{N}_l(0, V)$ where the covariance matrix V is given by:

$$V_{j,j'} = \int e^{i(u_j - u_{j'})x}x^2 n(x)dx.$$

Let us set $g_1 = g$ and for $\ell \geq 1$,

$$g_\ell(x) = x^{\ell-1} g(x) = x^\ell n(x). \tag{34}$$

Since the random variables $X_{k,n}(u)$ are independent, identically distributed and centered, it is enough to check that, for all (u, v)

$$n\mathbb{E}(X_{k,n}(u)\bar{X}_{k,n}(v)) \to V(u,v) = \int e^{i(u-v)x} x^2 n(x)dx = g_2^*(u-v), \tag{35}$$

and

$$n\mathbb{E}(|X_{k,n}(u)|^4) \to 0. \tag{36}$$

We have:

$$n\mathbb{E}(X_{k,n}(u)\bar{X}_{k,n}(v)) = \frac{1}{\Delta}(\mathbb{E}(Z_k^2 e^{i(u-v)Z_k}) - \theta_\Delta(u)\bar{\theta}_\Delta(v))$$

and

$$\mathbb{E}(Z_1^2 e^{i(u-v)Z_1}) = -\psi_\Delta^{''}(u-v)).$$

Computing $\psi_\Delta^{''}$, we get:

$$\psi_\Delta^{''}(u) = -\Delta\psi_\Delta(u)[g_2^*(u) + \Delta(g^*(u))^2].$$

Since $\theta_\Delta(u)\bar{\theta}_\Delta(v) = \Delta^2 g^*(u) g^*(-v)\psi_\Delta(u)\psi_\Delta(-v)$,

$$n\mathbb{E}(X_{k,n}(u)\bar{X}_{k,n}(v)) = \psi_\Delta(u-v)[g_2^*(u-v) + \Delta O(1)] = g_2^*(u-v) + o(1),$$

which gives (35).

Moreover,

$$n\mathbb{E}(|X_{k,n}(u)|^4) \leq \frac{8}{n\Delta^2}(\mathbb{E}(Z_k^4) + |\theta_\Delta(u)|^4) = \frac{8}{n\Delta}(m_4 + o(1)), \tag{37}$$

which implies (36).□

8.4. **Proof of Theorem 3.1.** The proof is given in two steps. We define, for some b, $0 < b < 1$,

$$\Omega_b := \left\{ \left| \frac{(1/n\Delta)\sum_{k=1}^n Z_k^2}{\mathbb{E}(Z_1^2/\Delta)} - 1 \right| \leq b \right\},$$

so that $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2) = \mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{I}_{\Omega_b}) + \mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{I}_{\Omega_b^c})$.

Step 1. Study of $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{I}_{\Omega_b})$. For notational convenience, let us define, for $t \in S_m$:

$$\nu_n(t) = \frac{1}{2\pi}\int \frac{\hat{\theta}_\Delta(u) - \theta_\Delta(u)}{\Delta} t^*(-u)du \tag{38}$$

$$R_n(t) = \frac{1}{2\pi}\int (\psi_\Delta(u) - 1) g^*(u) t^*(-u)du. \tag{39}$$

We have used the same notation as in (25) and (26) but the interpretation is different. Note that $\nu_n = \bar{\nu}_n$ and $R_n = \bar{R}_n$ so that they are both real valued. Then (24) holds.

We must split ν_n into two terms. With k_n to be defined later on, let

$$\theta_\Delta^{(1)}(x) = \mathbb{E}\left(Z_1 \mathbf{I}_{|Z_1| \leq k_n\sqrt{\Delta}} e^{ixZ_1}\right) \quad \text{and} \quad \theta_\Delta^{(2)}(x) = \mathbb{E}\left(Z_1 \mathbf{I}_{|Z_1| > k_n\sqrt{\Delta}} e^{ixZ_1}\right)$$

and $\hat{\theta}_\Delta^{(1)}(x)$ and $\hat{\theta}_\Delta^{(2)}(x)$ their empirical counterparts. We define

$$\nu_n^{(1)}(t) = \frac{1}{2\pi\Delta}\int (\hat{\theta}_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)) t^*(-u)du$$

and

$$\nu_n^{(2)}(t) = \frac{1}{2\pi\Delta}\int (\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u))t^*(-u)du.$$

The definition of $\hat{g}_{\hat{m}}$ implies that

$$\gamma_n(\hat{g}_{\hat{m}}) + \text{pen}(\hat{m}) \le \gamma_n(g_m) + \text{pen}(m) \tag{40}$$

where g_m denotes the orthogonal projection of g on S_m.
Then, using (24) yields that, for all $m = 1, \dots, m_n$,

$$\begin{aligned}
\|\hat{g}_{\hat{m}} - g\|^2 &\le \|g - g_m\|^2 + \text{pen}(m) + 2\nu_n^{(1)}(g_m - \hat{g}_{\hat{m}}) - \text{pen}(\hat{m}) \\
&\quad + 2R_n(g_m - \hat{g}_{\hat{m}}) + 2\nu_n^{(2)}(g_m - \hat{g}_{\hat{m}}) \\
&\le \|g - g_m\|^2 + \text{pen}(m) + \frac{3}{8}\|g_m - \hat{g}_{\hat{m}}\|^2 + 8\sup_{t\in S_m+S_{\hat{m}},\|t\|=1}[\nu_n^{(1)}(t)]^2 - \text{pen}(\hat{m}) \\
&\quad + 8\sup_{t\in S_{m_n},\|t\|=1}[R_n(t)]^2 + 8\sup_{t\in S_{m_n},\|t\|=1}[\nu_n^{(2)}(t)]^2 \\
&\le (1+\frac{3}{4})\|g - g_m\|^2 + \text{pen}(m) + \frac{3}{4}\|\hat{g}_{\hat{m}} - g\|^2 \\
&\quad + 8\left(\sup_{t\in S_m+S_{\hat{m}},\|t\|=1}[\nu_n^{(1)}(t)]^2 - p(m,\hat{m})\right)_+ + 8p(m,\hat{m}) - \text{pen}(\hat{m}) \\
&\quad + 8\sup_{t\in S_{m_n},\|t\|=1}[R_n(t)]^2 + 8\sup_{t\in S_{m_n},\|t\|=1}[\nu_n^{(2)}(t)]^2.
\end{aligned}$$

The function $p(m,m')$ plugged in the last inequality is fixed by applying Talagrand's inequality (see Lemma 9.1) to $\nu_n^{(1)}$, which yields the following result:

Proposition 8.1. *Under the Assumptions of Theorem 3.1, define*

$$p(m,m') = 4\mathbb{E}(Z_1^2/\Delta)\frac{m \vee m'}{n\Delta}, \tag{41}$$

then

$$\mathbb{E}(\sup_{t\in S_m+S_{\hat{m}},\|t\|=1}[\nu_n^{(1)}(t)]^2 - p(m,\hat{m}))_+ \le \sum_{m'=1}^{m_n}\mathbb{E}(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2 - p(m,m'))_+ \le \frac{C}{n\Delta},$$

where C is a constant.

Now, on Ω_b, the following inequality holds (by bounding the indicator by 1), for any choice of κ:

$$\forall m,\ (1-b)\text{pen}_{th}(m) \le \text{pen}(m) \le (1+b)\text{pen}_{th}(m). \tag{42}$$

Therefore

$$\begin{aligned}
\frac{1}{4}\|\hat{g}_{\hat{m}} - g\|^2\mathbf{I}_{\Omega_b} &\le \frac{7}{4}\|g - g_m\|^2 + (1+b)\text{pen}_{th}(m)\mathbf{I}_{\Omega_b} + 8\left(\sup_{t\in S_m+S_{\hat{m}},\|t\|=1}[\nu_n^{(1)}(t)]^2 - p(m,\hat{m})\right)_+ \\
&\quad + (8p(m,\hat{m}) - (1-b)\text{pen}_{th}(\hat{m}))\mathbf{I}_{\Omega_b} \\
&\quad + 8\sup_{t\in S_{m_n},\|t\|=1}[R_n(t)]^2 + 8\sup_{t\in S_{m_n},\|t\|=1}[\nu_n^{(2)}(t)]^2.
\end{aligned}$$

The constant κ is now chosen such that

$$\forall m, m' \in \{1,\dots,m_n\},\ \ 8p(m,m') \le (1-b)(\text{pen}_{th}(m) + \text{pen}_{th}(m')),$$

that is $\kappa \geq 32/(1-b)$. In view of (41), this gives the choices

$$\mathrm{pen}_{th}(m) = \frac{32}{1-b}\mathbb{E}(Z_1^2/\Delta)\frac{m}{n\Delta} \text{ and } \mathrm{pen}(m) = \frac{32}{1-b}\frac{1}{n\Delta}\sum_{i=1}^{n} Z_i^2 \frac{m}{n\Delta}.$$

It follows that

$$\begin{aligned}\frac{1}{4}\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{I}_{\Omega_b} \quad &\leq \quad \frac{7}{4}\|g - g_m\|^2 + 2\mathrm{pen}_{th}(m) \\ &+8\sum_{m'=1}^{m_n}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2 - p(m,m')\right)_+ \\ &+8\sup_{t\in S_{m_n},\|t\|=1}[R_n(t)]^2 + 8\sup_{t\in S_{m_n},\|t\|=1}[\nu_n^{(2)}(t)]^2.\end{aligned}$$

Using (39) and (10), we get

$$\sup_{t\in S_{m_n},\|t\|=1} R_n^2(t) \leq C\Delta^2\int_{-\pi m_n}^{\pi m_n} u^2|g^*(u)|^2du. \tag{43}$$

For $\nu_n^{(2)}(t)$, we write

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in S_{m_n},\|t\|=1}[\nu_n^{(2)}(t)]^2\right) \quad &\leq \quad \frac{1}{2\pi\Delta^2}\int_{-\pi m_n}^{\pi m_n}\mathbb{E}|\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)|^2du \\ &\leq \quad \frac{\mathbb{E}(Z_1^2\mathbf{I}_{|Z_1|>k_n\sqrt{\Delta}})m_n}{n\Delta^2} \\ &\leq \quad \frac{\mathbb{E}(Z_1^4)m_n}{nk_n^2\Delta^3} = \frac{[\mathbb{E}(Z_1^4)/\Delta]m_n}{nk_n^2\Delta^2} \leq \frac{[\mathbb{E}(Z_1^4)/\Delta]}{k_n^2\Delta}\end{aligned}$$

since $m_n \leq n\Delta$. We know that $[\mathbb{E}(Z_1^4)/\Delta]$ is bounded. If $k_n^2 \geq Cn/\ln^2(n\Delta)$, then the above term is of order $\ln^2(n\Delta)/(n\Delta)$.

Then we obtain that, for all $m \in \{1,\dots,m_n\}$,

$$\mathbb{E}\left(\|\hat{g}_{\hat{m}} - g\|^2\mathbf{I}_{\Omega_b}\right) \leq 7\|g-g_m\|^2 + 8\mathrm{pen}_{th}(m) + \frac{C_1}{n\Delta} + C_2\Delta^2\int_{-\pi m_n}^{\pi m_n}u^2|g^*(u)|^2du + C_3\frac{\ln^2(n\Delta)}{n\Delta}.$$

Step 2. Study of $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2\mathbf{I}_{\Omega_b^c})$.

The strategy is different. Using (24) and (40) yields that, $\forall m \in \{1,\dots,m_n\}$,

$$\|\hat{g}_{\hat{m}} - g\|^2 \quad \leq \quad \|g-g_m\|^2 + \mathrm{pen}(m) + 2\nu_n(g_m - \hat{g}_{\hat{m}}) - \mathrm{pen}(\hat{m}) + 2R_n(g_m - \hat{g}_{\hat{m}})$$

$$\leq \quad \|g-g_m\|^2 + \mathrm{pen}(m) + \frac{1}{4}\|g_m - \hat{g}_{\hat{m}}\|^2 \tag{44}$$

$$+8\sup_{t\in S_{m_n},\|t\|=1}[\nu_n(t)]^2 + 8\sup_{t\in S_{m_n},\|t\|=1}[R_n(t)]^2. \tag{45}$$

Now we apply inequality (43) to $R_n(t)$ and the Parseval formula for $\nu_n(t)$, and get

$$\begin{aligned}\frac{1}{2}\|\hat{g}_{\hat{m}} - g\|^2 \quad &\leq \quad \frac{3}{2}\|g-g_m\|^2 + \mathrm{pen}_{th}(m) + [\mathrm{pen}(m) - \mathbb{E}(\mathrm{pen}(m))] \\ &+\frac{4}{\pi\Delta^2}\int_{-\pi m_n}^{\pi m_n}|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2du + C'\Delta^2\int_{-\pi m_n}^{\pi m_n}u^2|g^*(u)|^2du.\end{aligned}$$

Therefore, using that $\mathrm{pen}_{th}(m) = \mathbb{E}(\mathrm{pen}(m))$, we obtain

$$\mathbb{E}\left((\mathrm{pen}(m) - \mathrm{pen}_{th}(m))\mathbf{1}_{\Omega_b^c}\right) \le \left\{\mathbb{E}\left[\left(\frac{1}{n\Delta}\sum_{k=1}^{n}(Z_k^2 - \mathbb{E}(Z_1^2)\right)^2\right]\right\}^{1/2} (\mathbb{P}(\Omega_b^c))^{1/2}, \tag{46}$$

and we find

$$\begin{aligned}\frac{1}{2}\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2\mathbf{1}_{\Omega_b^c}) \quad &\le \quad \left(\frac{3}{2}\|g\|^2 + \mathrm{pen}_{th}(m) + C"\Delta^2 m_n^2\|g\|^2\right)\mathbb{P}(\Omega_b^c)\\ &+\mathbb{E}^{1/2}\left[\left(\frac{1}{n\Delta}\sum_{k=1}^{n}(Z_k^2 - \mathbb{E}(Z_1^2)\right)^2\right]\mathbb{P}^{1/2}(\Omega_b^c)\\ &+\mathbb{E}^{1/2}\left((\frac{4}{\pi\Delta^2}\int_{-\pi m_n}^{\pi m_n}|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2du)^2\right)\mathbb{P}^{1/2}(\Omega_b^c).\end{aligned}$$

Then we apply (9) of Theorem 2.1 with $\ell = 2$ and get for $p \ge 2$:

$$\mathbb{E}\left(\left|\frac{1}{n\Delta}\sum_{k=1}^{n}Z_k^2 - \mathbb{E}(Z_1^2)\right|^p\right) \le C_p\left(\frac{1}{n\Delta}\right)^{p/2}.$$

Thus, by taking $p = 2$,

$$\mathbb{E}^{1/2}\left((\frac{1}{n\Delta}\sum_{i=1}^{n}(Z_i^2 - \mathbb{E}(Z_i^2))^2\right) \le \frac{C}{\sqrt{n\Delta}}.$$

Applying (12) for $p = 2$ gives

$$\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^4) \le \frac{C\Delta^2}{n^2}.$$

Thus

$$\begin{aligned}\mathbb{E}\left((\frac{4}{\pi\Delta^2}\int_{-\pi m_n}^{\pi m_n}|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2du)^2\right) \quad &\le \quad \frac{16}{\pi^2\Delta^4}(2\pi m_n)\int_{-\pi m_n}^{\pi m_n}\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^4du)\\ &\le \quad C'\frac{m_n^2}{\Delta^4}\frac{\Delta^2}{n^2} \le C'\end{aligned}$$

as $m_n \le n\Delta$. We obtain:

$$\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2\mathbf{1}_{\Omega_b^c}) \le C\left(1 + n^2\Delta^4\right)\mathbb{P}(\Omega_b^c) + C'(1 + \frac{1}{\sqrt{n\Delta}})\mathbb{P}^{1/2}(\Omega_b^c). \tag{47}$$

Lastly, if follows from the Markov inequality that

$$\begin{aligned}\mathbb{P}(\Omega_b^c) \quad &\le \quad \frac{1}{b^p}\mathbb{E}\left(\left|\frac{(1/n\Delta)\sum_{k=1}^n Z_k^2}{\mathbb{E}(Z_1^2/\Delta)} - 1\right|^p\right)\\ &\le \quad \frac{1}{(\mathbb{E}(Z_1^2/\Delta)b)^p}\mathbb{E}\left(\left|\frac{1}{n\Delta}\sum_{k=1}^{n}Z_i^2 - \mathbb{E}(Z_1^2/\Delta)\right|^p\right).\end{aligned}$$

We find that, if $\mathbb{E}(|Z_1|^{2p}) < +\infty$ and $p \ge 2$,

$$\mathbb{P}(\Omega_b^c) \le \frac{C_p}{(\mathbb{E}(Z_1^2/\Delta)b)^p}\frac{1}{(n\Delta)^{p/2}}. \tag{48}$$

Therefore, using (47) and the above inequality, if we take $p = 4$ (i.e. $\mathbb{E}(Z_1^8) < \infty$), we get

$$\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2\mathbf{1}_{\Omega_b^c}) \le C/(n\Delta).$$

This ends step 2 and the proof of Theorem 3.1. □

8.5. **Proof of Proposition 8.1.** Here we apply the Talagrand (see Lemma 9.1) Inequality to the class

$$\mathcal{F} = \{f_t, t \in S_m + S_{m'}\} \text{ where } f_t(z) = \frac{z\mathbf{I}_{|z|\le k_n\sqrt{\Delta}}}{2\pi\Delta}\int_{-\pi(m\vee m')}^{\pi(m\vee m')} e^{ixz}t^*(-x)dx.$$

In that case, $\nu_n^{(1)}(t) = (1/n)\sum_{k=1}^n (f_t(Z_k) - \mathbb{E}(f_t(Z_k)))$. We have to find the three quantities M, H, v.

Let $m" = m \vee m'$, and note that $S_m + S_{m'} = S_{m"}$. Using Inequality (17),

$$\sup_{z\in\mathbb{R}} |f_t(z)| \le \frac{k_n}{2\pi\sqrt{\Delta}}\sup_{z\in\mathbb{R}}|2\pi t(z)| \le \frac{k_n\|t\|_\infty}{\sqrt{\Delta}} \le \frac{k_n\sqrt{m"}}{\sqrt{\Delta}} := M.$$

Clearly,

$$\mathbb{E}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2\right) \le \frac{1}{2\pi\Delta^2}\int_{-\pi m"}^{\pi m"}\mathbb{E}|\hat\theta_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)|^2du \le \frac{\mathbb{E}(Z_1^2)m"}{n\Delta^2}.$$

Thus we set

$$H^2 = \frac{\mathbb{E}(Z_1^2)m"}{n\Delta^2}.$$

The most delicate term is v.

$$\begin{aligned}\mathrm{Var}(f_t(Z_1)) &\le \frac{1}{4\pi^2\Delta^2}\mathbb{E}\left(\iint Z_1^2\mathbf{I}_{|Z_1|\le k_n\sqrt{\Delta}}e^{i(x-y)Z_1}t^*(-x)t^*(y)dxdy\right)\\ &= \frac{1}{4\pi^2\Delta^2}\iint p_\Delta^*(x-y)t^*(-x)t^*(y)dxdy,\end{aligned}$$

where

$$p_\Delta^*(x) = \mathbb{E}(Z_1^2\mathbf{I}_{|Z_1|\le k_n\sqrt{\Delta}}e^{ixZ_1}).$$

Using that $t = \sum_{j\in\mathbb{Z}} t_j\varphi_{m",j}$ with $\|t\|^2 = \sum_{j\in\mathbb{Z}} t_j^2 = 1$,

$$\begin{aligned}\mathrm{Var}(f_t(Z_1)) &\le \frac{1}{4\pi^2\Delta^2}\sum_{j,k\in\mathbb{Z}} t_jt_k\iint p_\Delta^*(x-y)\varphi_{m",j}^*(-x)\varphi_{m",k}^*(y)dxdy\\ &\le \frac{1}{4\pi^2\Delta^2}\left(\sum_{j,k\in\mathbb{Z}}\left|\iint p_\Delta^*(x-y)\varphi_{m",j}^*(-x)\varphi_{m",k}^*(y)dxdy\right|^2\right)^{1/2},\end{aligned}$$

Now, using Proposition 2.1, we have

$$p_\Delta^*(x) = \Delta\int z\mathbf{I}_{|z|\le k_n\sqrt{\Delta}}e^{ixz}\mathbb{E}(g(z-Z_1))dz.$$

This implies that (see (H4))

$$\begin{aligned}\int |p_\Delta^*(z)|^2dz &\le 2\pi\int|p_\Delta(z)|^2dz = 2\pi\Delta^2\int z^2\mathbf{I}_{|z|\le k_n\sqrt{\Delta}}\mathbb{E}^2(g(z-Z_1))dz\\ &\le 2\pi\Delta^2\mathbb{E}\left(\int z^2\mathbf{I}_{|z|\le k_n\sqrt{\Delta}}g^2(z-Z_1)dz\right)\\ &\le 4\pi\Delta^2\mathbb{E}\left(\int(x^2+Z_1^2)g^2(x)dx\right) = 4\pi\Delta^2\left(M_2 + \mathbb{E}(Z_1^2)\|g\|^2\right).\end{aligned}$$

Therefore, it follows

$$\begin{aligned}\mathrm{Var}(f_t(Z_1)) &\leq \frac{1}{4\pi^2\Delta^2}\left(\iint_{[-\pi m",\pi m"]^2}|p_\Delta^*(x-y)|^2dxdy\right)^{1/2}\\ &\leq \frac{1}{4\pi^2\Delta^2}(2\pi m")^{1/2}(\int|p_\Delta^*(z)|^2dz)^{1/2}\\ &\leq \frac{\sqrt{m"}}{\sqrt{2}\pi\Delta}\left(M_2+\|g\|^2\mathbb{E}(Z_1^2)\right)^{1/2}:=v.\end{aligned}$$

Applying Lemma 9.1 yields, for $\epsilon^2=1/2$ and $p(m,m')$ given by (41) yields

$$\mathbb{E}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2-p(m,m')\right)_+\leq C_1\left(\frac{\sqrt{m"}}{n\Delta}e^{-C_2\sqrt{m"}}+\frac{k_n^2m"}{n^2\Delta}e^{-C_3\sqrt{n}/k_n}\right)$$

as $p(m,m')=4H^2$. We choose

$$k_n=\frac{C_3}{4}\frac{\sqrt{n}}{\ln(n\Delta)},$$

and as $m\leq n\Delta$, we get

$$\mathbb{E}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2-p(m,m')\right)_+\leq C_1'\left(\frac{\sqrt{m"}}{n\Delta}e^{-C_2\sqrt{m"}}+\frac{1}{(\Delta n)^4\ln^2(n\Delta)}\right).$$

Therefore

$$\sum_{m'=1}^{m_n}\mathbb{E}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2-p(m,m')\right)_+\leq C_1'\left(\frac{\sum_{m'=1}^{n\Delta}\sqrt{m"}e^{-C_2\sqrt{m"}}}{n\Delta}+\frac{1}{(n\Delta)^3\ln^2(n\Delta)}\right).$$

As $C_2xe^{-C_2x}$ is decreasing for $x\geq 1/C_2$, and its maximum is $1/(eC_2)$, we get

$$\begin{aligned}\sum_{m'=1}^{m_n}\sqrt{m"}e^{-C_2\sqrt{m"}} &\leq \sum_{\sqrt{m'}\leq 1/C_2}(eC_2)^{-1}+\sum_{\sqrt{m'}\geq 1/C_2}\sqrt{m'}e^{-C_2\sqrt{m'}}\\ &\leq \frac{1}{eC_2^3}+\sum_{m'=1}^{\infty}\sqrt{m'}e^{-C_2\sqrt{m'}}<+\infty.\end{aligned}$$

I follows that

$$\sum_{m'=1}^{m_n}\mathbb{E}\left(\sup_{t\in S_m+S_{m'},\|t\|=1}[\nu_n^{(1)}(t)]^2-p(m,m')\right)_+\leq\frac{C}{n\Delta}$$

and Proposition 8.1 is proved. □

8.6. **Proof of Proposition 4.1.** First, we know that $R_n(t)=(1/2\pi)\int(\psi_\Delta-1)g^*(u)t^*(-u)$ and thus, if $\int u^2|g^*(u)|du<+\infty$, if follows from (10) that

$$R_n^2(t)\leq\frac{\Delta^2\|g\|_1^2}{(2\pi)^2}\left(\int|ug^*(u)t^*(-u)|du\right)^2\leq\frac{\Delta^2\|g\|_1^2}{(2\pi)^2}\int u^2|g^*(u)|^2du\int|t^*(-u)|^2du.$$

Noting that $\int|t^*(-u)|^2du=2\pi\|t\|^2=2\pi\|t\|_A^2$ gives 1).

For the two other cases, using Proposition 2.1, we have, for t a function with support $[a,b]$:

$$\frac{1}{\Delta}\mathbb{E}(Z_1t(Z_1))=\int_a^b t(z)\mathbb{E}g(z-Z_1)dz=\mathbb{E}(\int_{a-Z_1}^{b-Z_1}t(x+Z_1)g(x)dx).$$

Thus $R_n(t) = \mathbb{E}(\int_{a-Z_1}^{b-Z_1} t(x+Z_1)g(x)dx - \int_a^b t(x)g(x)dx)$.

On $(|Z_1| > b-a)$, $[a-Z_1, b-Z_1] \cap [a,b] = \emptyset$ and

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{|Z_1|>b-a}\left|\int_{a-Z_1}^{b-Z_1} t(x+Z_1)g(x)dx - \int_a^b t(x)g(x)dx\right|\right) &\leq 2\|t\|_\infty\|g\|_1\frac{\mathbb{E}(|Z_1|)}{b-a}\\ &\leq \frac{4\Phi_0\|g\|_1^2\sqrt{D_m}\Delta\|t\|_A}{b-a}.\end{aligned}$$

On $(|Z_1| \leq b-a)$, $[a-Z_1, b-Z_1] \cap [a,b] \neq \emptyset$. Assume for instance that $0 \leq Z_1 \leq b-a$.

$$\begin{aligned}&\int_{a-Z_1}^{b-Z_1} t(x+Z_1)g(x)dx - \int_a^b t(x)g(x)dx\\ =\ &\int_{a-Z_1}^{a} t(x+Z_1)g(x)dx + \int_a^{b-Z_1}(t(x+Z_1)-t(x))g(x)dx - \int_{b-Z_1}^b t(x)g(x)dx.\end{aligned}$$

To study the middle term, we use the fact that t is C^1 on $[a,b]$.

$$\begin{aligned}T_1 &:= \mathbb{E}\left(\mathbf{1}_{0\leq Z_1\leq b-a}\int_a^{b-Z_1}(t(x+Z_1)-t(x))g(x)dx\right)\\ &= \mathbb{E}\left(Z_1\mathbf{1}_{0\leq Z_1\leq b-a}\int_a^{b-Z_1}\int_0^1 t'(x+uZ_1)dug(x)dx\right)\\ &= \mathbb{E}\left(Z_1\mathbf{1}_{0\leq Z_1\leq b-a}\int_0^1(\int_a^{b-Z_1} t'(x+uZ_1)g(x)dx)du\right)\end{aligned}$$

An application of the Cauchy-Schwarz inequality yields

$$|T_1| \leq \mathbb{E}|Z_1|\|t'\|_A\|g\| \leq 2\Phi_0\|g\|_1\|g\|\|t\|_A\Delta D_m.$$

Next,

$$T_2 := \mathbb{E}\left(\mathbf{1}_{0\leq Z_1\leq b-a}\int_{a-Z_1}^a t(x+Z_1)g(x)dx\right).$$

Here we distinguish between 2) and 3). If g is bounded (case 2)), then

$$|T_2| \leq \|t\|_\infty\|g\|_\infty\mathbb{E}(|Z_1|) \leq 2\Phi_0\|g\|_1\|t\|_A\Delta\sqrt{D_m}.$$

Otherwise (case 3)), using the Cauchy-Schwarz inequality again,

$$\begin{aligned}|T_2| &\leq \mathbb{E}(\sqrt{Z_1^+})\|t\|_\infty\|g\| \leq \sqrt{\mathbb{E}(|Z_1|)}\Phi_0\sqrt{D_m}\|t\|_A\|g\|\\ &\leq \sqrt{2}\Phi_0\|t\|_A\sqrt{\|g\|_1}\|g\|\sqrt{D_m\Delta}.\end{aligned}$$

The same bound holds for the last term.

The same study can be done for $a-b \leq Z_1 \leq 0$. Joining all terms, we find that, if g is bounded

$$|R_n(t)| \leq C\Phi_0\|t\|_A\Delta D_m.$$

Otherwise,

$$|R_n(t)| \leq C'\Phi_0\|t\|_A(\sqrt{\Delta D_m} + \Delta D_m).$$

The constants C and C' depend on a, b, $\|g\|_1$ and $\|g\|$. $\square$

8.7. **Proof of Proposition 4.2.** Relation (24) still holds with ν_n and R_n respectively defined by (25) and (26). As for any $t \in \Sigma_m$, $\|t-g\|^2 = \|t-g\|_A^2 + \|g\|_{A^c}^2$, and we get

$$\gamma_n(t) - \gamma_n(s) = \|t-g\|_A^2 - \|s-g\|_A^2 - 2\nu_n(t-s) - 2R_n(t-s).$$

Writing that $\gamma_n(\tilde{g}_m) - \gamma_n(g) \le \gamma_n(g_m) - \gamma_n(g)$, we get

$$\|\tilde{g}_m - g\|_A^2 \le \|g_m - g\|_A^2 + 2\nu_n(\tilde{g}_m - g_m) + 2R_n(\tilde{g}_m - g_m).$$

We have

$$2\nu_n(\tilde{g}_m - g_m) \le \frac{1}{8}\|\tilde{g}_m - g_m\|_A^2 + 8 \sup_{t\in\Sigma_m, \|t\|_A=1} [\nu_n(t)]^2,$$

and the analogous inequality for R_n. Using that $\|\tilde{g}_m - g_m\|_A^2 \le 2\|g-g_m\|_A^2 + 2\|\tilde{g}_m - g\|_A^2$ and some algebra yields:

$$\frac{1}{2}\|\tilde{g}_m - g\|_A^2 \le \frac{3}{2}\|g_m - g\|_A^2 + 8 \sup_{t\in\Sigma_m, \|t\|_A=1} [\nu_n(t)]^2 + 8 \sup_{t\in\Sigma_m, \|t\|_A=1} [R_n(t)]^2.$$

We have:

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in\Sigma_m, \|t\|_A=1} [\nu_n(t)]^2\right) &\le \sum_{\lambda\in\Lambda_m} \mathbb{E}([\nu_n(\varphi_\lambda)]^2) = \sum_{\lambda\in\Lambda_m} \frac{1}{n\Delta^2}\mathrm{Var}(Z_1\varphi_\lambda(Z_1)) \\
&\le \mathbb{E}(Z_1^2 \sum_\lambda \varphi_\lambda^2(Z_1))\frac{1}{n\Delta^2} = [\mathbb{E}(Z_1^2)/\Delta]\frac{\Phi_0 D_m}{n\Delta}. \qquad (49)
\end{aligned}$$

Now, we conclude using Inequality (49) and Proposition 4.1. □

8.8. **Proof of Theorem 4.1.** The proof of Theorem 4.1 is close to the proof of Theorem 3.1. Hence we focus mainly on the differences. Note that ν_n defined in (25) can be written as

$$\nu_n(t) = \frac{1}{n}\sum_{k=1}^n (f_t(Z_k) - \mathbb{E}(f_t(Z_1)))$$

with f_t now given by $f_t(z) = zt(z) = z\mathbf{I}_{z\in A}t(z)$, since t has compact support A. As in step 1 of Theorem 3.1, we are led to the inequality:

$$\begin{aligned}
\frac{1}{2}\|\tilde{g}_{\tilde{m}} - g\|_A^2 \mathbf{I}_{\Omega_b} &\le \frac{3}{2}\|g-g_m\|_A^2 + 2\mathrm{pen}_{th}(m) \\
&\quad + 8\sum_{m'\in\mathcal{M}_n}\left(\sup_{t\in\Sigma_m+\Sigma_{m'}, \|t\|_A=1} [\nu_n(t)]^2 - p(m,m')\right)_+ \\
&\quad + 8 \sup_{t\in\mathcal{S}_n, \|t\|_A=1} [R_n(t)]^2,
\end{aligned}$$

with $8p(m,m') \le (1-b)(\mathrm{pen}_{th}(m) + \mathrm{pen}_{th}(m'))$, for all $m \in \mathcal{M}_n$.

It follows from Proposition (4.1) that

$$\sup_{t\in\mathcal{S}_n, \|t\|_A=1} [R_n(t)]^2 \le K\rho_{n,\Delta}.$$

Proposition 8.2. *Under the Assumptions of Theorem 4.1, define*

$$p(m,m') = 4\mathbb{E}(Z_1^2/\Delta)\frac{D_m \vee D_{m'}}{n\Delta}, \qquad (50)$$

then

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_A=1}[\nu_n(t)]^2-p(m,m')\right)_+\leq\frac{C}{n\Delta},$$

where C is a constant.

For the study of $\mathbb{E}(\|\tilde g_{\tilde m}-g\|_A^2\mathbf{1}_{\Omega_b^c})$, as in step 2 above, we have the inequality analogous to (44):

$$\frac{1}{2}\|\hat g_{\hat m}-g\|^2\quad\leq\quad\frac{3}{2}\|g_A-g_m\|^2+\mathrm{pen}(m)+8\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^2+8\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[R_n(t)]^2.$$

The bound for $\mathbb{P}(\Omega_b^c)$ is given by (48). Proposition 4.1 applies to bound $[R_n(t)]^2$ by $C\rho_{n,\Delta}$.
Then we have again

$$\mathrm{pen}(m)\mathbf{1}_{\Omega_b^c}\leq\mathrm{pen}_{th}(m)+(\mathrm{pen}(m)-\mathrm{pen}_{th}(m))\mathbf{1}_{\Omega_b^c}.$$

The same bound holds also for the term $\mathbb{E}[(\mathrm{pen}(m)-\mathbb{E}(\mathrm{pen}(m)))\mathbf{1}_{\Omega_b^c}]$. We apply inequality (46).
It remains to study the term $\mathbb{E}(\sup_{t\in\mathcal{S}_n}[\nu_n(t)]^2\mathbf{1}_{\Omega_b^c})$. We use

$$\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^2\mathbf{1}_{\Omega_b^c}\right)\leq\left(\mathbb{E}\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^4\right)^{1/2}\mathbb{P}^{1/2}(\Omega_b^c).$$

Denote by $(\varphi_\lambda)_{\lambda\in\Lambda_n}$ an orthonormal basis of $\mathcal{S}_n$, $|\Lambda_n|=N_n$. We have

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^4\right)&=\mathbb{E}\left[\left(\sum_{\lambda\in\Lambda_n}\nu_n^2(\varphi_\lambda)\right)^2\right]\\&\leq N_n\sum_{\lambda\in\Lambda_n}\mathbb{E}\left\{\left(\frac{1}{n\Delta}\sum_{k=1}^n(Z_k\varphi_\lambda(Z_k)-\mathbb{E}(Z_k\varphi_\lambda(Z_k)))\right)^4\right\}\\&\leq\frac{KN_n}{(n\Delta)^4}\sum_{\lambda\in\Lambda_n}\left[n\mathbb{E}[(Z_1\varphi_\lambda(Z_1))^4]+\left(n\mathbb{E}(Z_1^2\varphi_\lambda^2(Z_1))\right)^2\right],\end{aligned}$$

where the last inequality follows from the Rosenthal Inequality (51).
If the basis is bounded, $\varphi_\lambda^2\leq B$, $\forall\lambda$, as for instance basis [T] ($B=2$), we find

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^4\right)&\leq\frac{KN_n^2B^2}{(n\Delta)^4}\left[n\mathbb{E}(Z_1^4/\Delta)\Delta+n^2\mathbb{E}^2(Z_1^2/\Delta)\Delta^2\right]\\&\leq\frac{K'N_n^2}{(n\Delta)^2}\leq K'\end{aligned}$$

using $N_n\leq n\Delta$.

In the general case, we use that $\sum_\lambda \varphi_\lambda^4(x) \leq \|\varphi_\lambda\|_\infty^2 \sum_\lambda \varphi_\lambda^2(x)$ and $\|\sum_\lambda \varphi_\lambda^2\|_\infty \leq \Phi_0^2 N_n$ and $\|\varphi_\lambda\|_\infty^2 \leq \Phi_0^2 N_n$, so that

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^4\right) &\leq \frac{KN_n}{(n\Delta)^4}\left[\Phi_0^4 N_n^2 n\mathbb{E}(Z_1^4/\Delta)\Delta + n^2\mathbb{E}^2\left(\sum_{\lambda\in\Lambda_n}(Z_1^2/\Delta)\varphi_\lambda^2(Z_1)\right)\Delta^2\right]\\ &\leq \frac{KN_n}{(n\Delta)^4}\left[\Phi_0^4 N_n^2 n\mathbb{E}(Z_1^4/\Delta)\Delta + n^2\Phi_0^4 N_n^2\mathbb{E}^2(Z_1^2/\Delta)\Delta^2\right]\\ &\leq \frac{K"N_n^3}{(n\Delta)^2} \leq K"(n\Delta)\end{aligned}$$

using $N_n \leq n\Delta$.

Using (48), we obtain $\mathbb{E}\left(\sup_{t\in\mathcal{S}_n,\|t\|_A=1}[\nu_n(t)]^2\mathbf{I}_{\Omega_b^c}\right) \leq C/(n\Delta)$ if $\mathbb{P}(\Omega_b^c) \leq 1/(n\Delta)^2$ which holds for $p=4$ and $\mathbb{E}(Z_1^8) < +\infty$ in the first case (bounded basis). In the general case, we need $\mathbb{P}(\Omega_b^c) \leq 1/(n\Delta)^3$ and thus $p=6$ and $\mathbb{E}(Z_1^{12}) < +\infty$.

8.9. **Proof of Proposition 8.2.** Again, we apply the Talagrand (see Lemma 9.1) Inequality to the class

$$\mathcal{F} = \{f_t, t \in \Sigma_m + \Sigma_{m'}\} \text{ where } f_t(z) = \frac{z\mathbf{I}_{z\in A}t(z)}{\Delta}.$$

We obtain similarly to (49)

$$H^2 = [\mathbb{E}(Z_1^2)/\Delta]\Phi_0(D_m \vee D_{m'})/(n\Delta) \text{ and } M = b_A\Phi_0\sqrt{D_m \vee D_{m'}}/\Delta,$$

where $b_A = \sup_{z\in A}|z|$. Lastly, we find

$$\begin{aligned}\mathrm{Var}\left(\frac{Z_1}{\Delta}t(Z_1)\right) &\leq \mathbb{E}(Z_1^2t^2(Z_1))/\Delta^2 = \frac{1}{\Delta}\int zt^2(z)\mathbb{E}(g(z-Z_1))dz\\ &\leq \frac{b_A\|t\|_\infty}{\Delta}\mathbb{E}\left(\int|t(z)g(z-Z_1)|dz\right)\\ &\leq \frac{b_A\Phi_0(D_m\vee D_{m'})^{1/2}}{\Delta}\mathbb{E}\left(\|t\|\int g^2(z-Z_1)dz\right)^{1/2}\\ &\leq \frac{2b_A\Phi_0(D_m\vee D_{m'})^{1/2}\|g\|}{\Delta}.\end{aligned}$$

We denote by $v = C(D_m \vee D_{m'})^{1/2}/\Delta$ with $C = 2\Phi_0 b_A\|g\|$.

Then we get

$$\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_A=1}[\nu_n(t)]^2 - p(m,m')\right)_+ \leq C_1'\left(\frac{\sqrt{D_m\vee D_{m'}}}{n\Delta}e^{-C_2\sqrt{D_m\vee D_{m'}}} + \frac{1}{n\Delta}\exp(-\sqrt{n\Delta})\right).$$

Therefore, as $D_m \leq n\Delta$, as above

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_A=1}[\nu_n(t)]^2 - p(m,m')\right)_+ \leq \frac{C}{n\Delta}.$$

This ends the proof of Proposition 8.2. □

9. Appendix

The Talagrand inequality. The following result follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 9.1. *(Talagrand Inequality) Let $Y_1, \dots, Y_n$ be independent random variables, let $\nu_{n,Y}(f) = (1/n)\sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1+2\epsilon^2)H^2\Big]_+ \quad \leq \quad \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}} + \frac{98M^2}{K_1 n^2 C^2(\epsilon^2)} e^{-\frac{2K_1 C(\epsilon^2)\epsilon}{7\sqrt{2}}\frac{nH}{M}}\right),$$

with $C(\epsilon^2) = \sqrt{1+\epsilon^2} - 1$, $K_1 = 1/6$, and

$$\sup_{f\in\mathcal{F}} \|f\|_\infty \leq M, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|\Big] \leq H, \ \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{k=1}^n \mathrm{Var}(f(Y_k)) \leq v.$$

By standard density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \mapsto \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

The Rosenthal inequality. (see *e.g.* Hall and Heyde (1980, p.23)) Let $(X_i)_{1\leq i\leq n}$ be n independent centered random variables, such that $\mathbb{E}(|X_i|^p) < +\infty$ for an integer $p \geq 1$. Then there exists a constant $C(p)$ such that

$$\mathbb{E}\left(\left|\sum_{i=1}^n X_i\right|^p\right) \leq C(p)\left(\sum_{i=1}^n \mathbb{E}(|X_i|^p) + \left(\sum_{i=1}^n \mathbb{E}(X_i^2)\right)^{p/2}\right). \tag{51}$$

References

[1] Barndorff-Nielsen O.E. and Shephard N. (2001). Modelling by Lévy processes for financial econometrics. In: *Lévy processes. Theory and Applications* (Barndorff-Nielsen O.E., Mikosch T., Resnick S.L.), 283-318.

[2] Barron, A.R., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields.* **97**, 113-150.

[3] Bertoin J. (1996). *Lévy processes.* Cambridge University Press.

[4] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.

[5] Comte, F. and Genon-Catalot, V. (2008). Nonparametric adaptive estimation for pure jump Lévy processes. Preprint MAP5 2008-12.

[6] DeVore, R.A. and Lorentz, G.G. (1993). *Constructive approximation.* Springer-Verlag.

[7] Figueroa-López J.E. and Houdré C. (2006). Risk bounds for the nonparametric estimation of Lévy processes. IMS Lecture Notes-Monograph Series High dimensional probability **51**, 96-116.

[8] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its applications.* Academic press, Inc. (London) LTD.

[9] Jongbloed, G. and van der Meulen, F. H. (2006) Parametric estimation for subordinators and induced OU processes. *Scand. J. Statist.* **33**, 825-847.

[10] Jongbloed, G., van der Meulen, F. H. and van der Vaart, A. W. (2005) Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **11**, 759-791.

[11] Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications.* Lecture Notes in Statistics, 129. Springer-Verlag, New York.

[12] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.

[13] Küchler, U. and Tappe, S. (2008) Bilateral Gamma distributions and processes in financial mathematics. *Stochastic Processes and Their Applications*, **118**, 261-283.

[14] Lebedev, N.N. (1972). *Special functions and their applications.* Revised edition, translated from the Russian and edited by Richard A. Silverman. Dover Publications, Inc., New York.

[15] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag, Berlin, Ergebnisse der Mathematik und ihrer Grenzgebiete. **3**. Folge, 23.

[16] Madan, D.B. and Seneta, E. (1990) The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business* **63**, No. 4, 511-524.

[17] Massart, P. (2007) *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. With a foreword by Jean Picard. Lecture Notes in Mathematics, 1896. Springer, Berlin

[18] Meyer, Y. (1990) *Ondelettes et opérateurs.* I. Hermann, Paris.

[19] Neumann, M. and Reiss, M. (2007). Nonparametric estimation for Lévy processes from low-frequency observations. *Working paper*, ArXiv:0709.2007v1. To appear in *Bernoulli.*

[20] Sato K.I. (1999). *Lévy processes and infinitely divisible distributions.* Cambridge Studies in Advanced Mathematics, 68. Cambridge University Press, Cambridge.

[21] Watteel R.N. and Kulperger R.J. (2003). Nonparametric estimation of the canonical measure for infinitely divisible distributions. *Journal of Statistical Computation and Simulation*, **73**, 7, 525-542.