Journal of Nonparametric Statistics
Vol. 00, No. 0, Month 2011, 1–25

An oracle inequality for penalised projection estimation of Lévy densities from high-frequency observations

Florian A. J. Ueltzhöfer[a,*] and Claudia Klüppelberg[a]

[a] Centre for Mathematical Sciences and TUM Institute for Advanced Study, Technische Universität München, Boltzmannstraße 3, 85748 Garching b. München, Germany

(*Received 8 April 2010; final version received 11 April 2011*)

We consider a multivariate Lévy process given by the sum of a Brownian motion with drift and an independent time-homogeneous pure jump process governed by a Lévy density. We assume that observation of a sample path takes place on an equidistant discrete time grid. Following Grenander's method of sieves, we construct families of non-parametric projection estimators for the restriction of a Lévy density to bounded sets away from the origin. Moreover, we introduce a data-driven penalisation criterion to select an estimator within a given family, where we measure the estimation error in an L^2-norm. We furthermore give sufficient conditions on the penalty such that an oracle inequality holds. As an application we prove adaptiveness for sufficiently smooth Lévy densities in some Sobolev space and explicitly derive the rate of convergence.

Keywords: Lévy density, Lévy process, non-parametric estimation, oracle inequality, adaptive model selection

AMS Subject Classification: 60G51, 60J25, 62G07, 62G20

1. Introduction

1.1. *Lévy processes*

When calibrating continuous-time stochastic models to data sampled on a discrete time grid, we inherently encounter the issue of separating continuous path components from purely discontinuous ones. Focusing on stochastic processes with stationary and independent increments, this translates to discriminating between the jumps, and Brownian motion and drift. Historically, special cases like the Wiener and the compound Poisson

* Corresponding author. Email: ueltzhoefer@ma.tum.de

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485252.2011.581375
http://www.informaworld.com

process and, more recently, general Lévy processes have received a lot of attention in various applied fields such as actuarial sciences, engineering, finance, geography, physics, and telecommunications.

In this paper, we consider a general d-dimensional Lévy process $X : [0,\infty[\times\Omega \to \mathbb{R}^d$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions, viz. the filtration $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and complete, that is, $\cap_{s>t}\mathscr{F}_s = \mathscr{F}_t$, and $\mathscr{F}_0$ contains all P-null sets. Throughout, we assume $X_0 = 0$ a. s. and the paths of X are càdlàg. For all $t \geq 0$, consequently, $X_{t-} := \lim_{s\uparrow t} X_s$ and $\Delta X_t := X_t - X_{t-}$ are well-defined. In addition, the law $\mathscr{L}(X_t)$ of X_t is fully determined by $\mathscr{L}(X_1)$. By the Lévy–Khintchine representation (cf. Theorem 8.1 of Sato (1999)), $\mathscr{L}(X_1)$ is infinitely divisible and uniquely specified by its characteristic triplet (b, σ^2, F), where $b \in \mathbb{R}^d$, $\sigma^2 \in \mathbb{R}^{d\times d}$ is symmetric positive semi-definite and F is a measure on $(\mathbb{R}^d_\circ, \mathscr{B}(\mathbb{R}^d_\circ))$, where $\mathbb{R}^d_\circ := \mathbb{R}^d \setminus \{0\}$ and $\mathscr{B}(\cdot)$ denotes the Borel σ-field of a set. F is called the Lévy measure of X and satisfies $\int(\|x\|^2 \wedge 1)F(\mathrm{d}x) < \infty$. Since the correspondence between Lévy processes and infinitely divisible distributions is one-to-one, (b, σ^2, F) is also called the generating triplet of X.

Furthermore, the Lévy–Itô decomposition (cf. Theorem 19.2 of Sato (1999)) relates the generating triplet to the components of X. Let $\mu : \mathscr{B}([0,\infty[\times\mathbb{R}^d_\circ) \times \Omega \to \mathbb{N}_0$ be given by

$$\mu(B, \omega) := \#\{s \geq 0 : (s, \Delta X_s(\omega)) \in B\}, \tag{1}$$

where $\#$ denotes cardinality. Then, μ forms a Poisson random measure on $([0,\infty[\times\mathbb{R}^d_\circ, \mathscr{B}([0,\infty[\times\mathbb{R}^d_\circ))$ with intensity measure ν given by $\nu(\mathrm{d}t, \mathrm{d}x) = \mathrm{d}tF(\mathrm{d}x)$. Moreover, let σ denote the Cholesky-triangle of σ^2. Then, there exists a d-dimensional Wiener process W independent of μ given by Equation (1) such that almost surely

$$X_t = bt + \sigma W_t + \int_0^t \int_{\{x:\|x\|>1\}} x\mathrm{d}\mu + \lim_{\varepsilon\downarrow 0} \int_0^t \int_{\{x:\varepsilon<\|x\|\leq 1\}} x\mathrm{d}(\mu - \nu) \quad (t \geq 0).$$

Certainly, various sub-classes of Lévy processes are interesting by its own right. Despite presence or absence of Brownian motion and drift, the jump component of a Lévy process is classified in terms of activity, sample path variation and existence of moments. If $F(\mathbb{R}^d_\circ) < \infty$, then the jump component is of finite activity, that is, a compound Poisson process, or else the jump component is of infinite activity, that is, the process jumps infinitely often on each compact time interval. Note that the jumps ΔX_t accumulate at the origin. Furthermore, the (small) jumps are absolutely summable if, and only if, $\int(\|x\| \wedge 1)F(\mathrm{d}x) < \infty$. In this case, the paths of the jump component are of finite variation. Otherwise, they are of infinite variation. Finally, a Lévy process has moments of order k if, and only if, F satisfies $\int_{\{x:\|x\|>1\}} \|x\|^k F(\mathrm{d}x) < \infty$ (cf. Sato 1999, Theorem 5.23). In other words, the existence of moments is determined by the (big) jumps only.

1.2. *Methodology*

We aim for inference on the Lebesgue derivative f of the Lévy measure F, that is, the Lévy density, which we always assume to exist. We focus on the method of penalised projection estimation. This estimation procedure can be interpreted as a combination of minimum contrast estimation (see, e. g., Birgé and Massart (1993, 1997, 1998)), the method of sieves (cf. Grenander (1981)), and penalisation techniques which trace back to

Whittaker (1923). In principle, the method can be outlined as follows. We assume that, on a compact domain of estimation away from the origin, the Lévy density f belongs to the corresponding L^2-space. Firstly, we choose a family of finite dimensional subspaces $\{\mathbb{S}_m \subseteq L^2 : m \in M\}$, called sieves. Secondly, for every $m \in M$ we construct estimators $\hat{f}_m \in \mathbb{S}_m$ for the orthogonal projection of f to $\mathbb{S}_m$. Finally, we choose $m^{\text{pen}} \in M$ using an empirical penalty. Combining these steps, we call $\hat{f}_{\text{pen}} := \hat{f}_{m^{\text{pen}}}$ a penalised projection estimator (PPE). We note that this method has also been applied in the context of density estimation of an i. i. d. sample and for the estimation of the intensity of finite Poisson random measures by Barron, Birgé and Massart (1999) and Reynaud-Bouret (2003), respectively.

Figueroa-López and Houdré (2006) presented the method outlined above for inference on univariate Lévy densities, assuming all jumps of X are observed. As an important ingredient of the method, oracle inequalities are proved to hold. A prototype for these inequalities is given by

$$E\|f - \hat{f}_{\text{pen}}\|_{L^2}^2 \leq C \inf_{m \in M} E\|f - \hat{f}_m\|_{L^2}^2, \tag{2}$$

where C is a finite positive constant. In other words, the asymptotic behaviour of $\hat{f}_{\text{pen}}$ can be derived from the so-called oracle, namely the optimal estimator within $\{\hat{f}_m : m \in M\}$.

Explicit observation of jumps, however, is known to remain theoretical in the context of Lévy processes. In practice, observation of process increments on a discrete time grid is commonly encountered. This more realistic observation scheme is our starting point. We derive a sufficient condition on penalties such that an oracle-type risk bound and an oracle inequality hold. As an application we consider smooth Lévy densities belonging to a Sobolev space $\mathcal{W}^{k,2}$ for an unknown degree of smoothness $k > 0$. We choose the sieves $\{\mathbb{S}_m : m \in M\}$ to consist of piecewise polynomials. These are known to provide good approximations to this type of smooth functions. By virtue of the oracle(-type) inequality, we derive an explicit rate of convergence for the corresponding penalised projection estimator. Particularly, as the length T of the observation interval $[0, T]$ tends to infinity, we obtain $E\|f - \hat{f}_{\text{pen}}\|_{L^2}^2 = O(T^{-2k/(2k+d)})$. This rate is a natural extension of the rate obtained in the univariate setting when all jumps are observed. Moreover, the achieved rate turns out to be optimal in the minimax sense.

Figueroa-López (2009) also develops a projection estimator for (univariate) smooth Lévy densities from observations on a discrete time grid, which achieves the same rate. However, there is an important distinction. Unlike the estimator of Figueroa-López (2009), knowledge of the degree of smoothness k is neither needed nor used in the construction of our estimator.

Comte and Genon-Catalot (2009, 2010, 2011) also based inference for Lévy processes on penalised contrast, which, again, achieves the same rate. Nevertheless, there are several differences. First and foremost, Comte and Genon-Catalot (2009, 2010, 2011) did not estimate the Lévy density f but $x \mapsto xf(x)$ (resp., $x \mapsto x^2 f(x)$ and $x \mapsto x^3 f(x)$) depending on additional assumptions. Second, their projection spaces are not finite dimensional. Last, they estimate $x \mapsto x^j f(x)$ for $j = 1, 2, 3$ arbitrarily close to zero. Certainly, dividing by the appropriate power of x yields an estimate for f. But if the transformed estimates are not truncated near zero, unsatisfactory effects arise. In particular, unbounded estimates with a pole at zero occur even in case of a compound Poisson process with a bounded Lévy density. Moreover, when Brownian motion is present, then estimates violating the defining property of Lévy measures, that is, $\int (\|x\|^2 \wedge 1) F(\mathrm{d}x) < \infty$, can result.

Early works in the literature on parametric and non-parametric inference for Lévy processes includes Rubin and Tucker (1959), Akritas (1982), and Basawa and Brockwell (1982). Numerous non-parametric and semiparametric approaches for the estimation of the characteristic triplet and, in particular, the Lévy density have been suggested recently. Besides the ones mentioned in the last paragraph, which are closely related to our work and will be discussed in detail in this work, we give an overview over the existing statistical work as far as it is known to us. An interesting collection of work is the special issue Gugushvili, Klaassen and Spreij (2010), where a wealth of interesting papers can be found with ample references to previous literature.

Special Lévy processes given by the sum of drift, compound Poisson process and α-stable process (including $\alpha = 2$ referring to Brownian motion) have been statistically estimated by Chen, Delaigle and Hall (2010), its restriction to drift, compound Poisson and Brownian motion by Gugushvili (2009a,b), and van Es, Gugushvili and Spreij (2007). In these papers the drift, the variance of the Brownian motion and the Poisson intensity are estimated parametrically, whereas the jump density of the compound Poisson process is estimated non-parametrically.

Watteel and Kulperger (2003), Neumann and Reiß (2009), and Kappus and Reiß (2010) estimate instead of the Lévy measure the canonical measure (cf. Billingsley (1995, (28.6))) for arbitrary Lévy processes with finite $4 + \delta$-moment. Estimated is the empirical characteristic function and the optimal characteristic triplet is found by minimising some distance measure.

Lévy processes have proved extremely useful in finance, where price processes are often modelled by an exponential Lévy process, or a stochastic volatility model is considered, whose volatility process is driven by a subordinator (a Lévy process with increasing sample paths). Problems like estimating the exponential Lévy model from option prices with different strikes, or estimating the stochastic volatility from the price process have attracted attention. These problems were considered in Belomestny and Reiß (2006) and Shimizu (2006, 2009). Moreover, estimating the latent stochastic volatility, and drawing conclusions about the jump behaviour of prices and volatilities have been studied in a number of important papers (usually on the more general class of Itô-semimartingales) by Jean Jacod and collaborators; we refer in particular to Aït-Sahalia and Jacod (2007). Based on results of this paper, tests have been developed to clarify, whether discretely sampled data originate from a continuous-time jump process or a diffusion (cf. Aït-Sahalia and Jacod (2009)). Finally, the theoretical results have been applied to real data in Jacod and Todorov (2010) and Jacod, Klüppelberg and Müller (2011).

The following is a brief outline of this paper. We specify the statistical problem and construct the penalised projection estimator in Section 2. We establish our main results — an oracle-type risk bound and a genuine oracle inequality — in Section 3. Moreover, we apply this risk bound to the estimation of Lévy densities belonging to some Sobolev space, where the sieves are chosen to consist of piecewise polynomials. Simulation results are summarised in Section 4. All proofs are given in the appendix.

2. Penalised projection estimation

2.1. *The statistical problem*

Let $X : [0, \infty[\times \Omega \to \mathbb{R}^d$ be a d-dimensional Lévy process with properties as specified in Section 1. Notably, recall that F denotes the Lévy measure of X and we assume that its Lebesgue derivative, the Lévy density f, exists.

For $n \in \mathbb{N}$ let $\Delta_n > 0$ be a *mesh size* and $T_n < \infty$ a *terminal observation time*. For simplicity, we suppose that T_n is an integer multiple of Δ_n. Our observation sample at stage n consists of a realisation of $X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{T_n}$. Equivalently, we observe the realised increments

$$\Delta_j^n X := X_{j\Delta_n} - X_{(j-1)\Delta_n} \quad (j = 1, \ldots, T_n/\Delta_n). \tag{3}$$

Based on an observed sample of Equation (3), our objective is to estimate the restriction $f_{|\mathbb{D}}$ of the Lévy density f to a set $\mathbb{D} \subseteq \mathbb{R}_\circ^d$ called *domain of estimation.* In the following, we shall denote both, the restricted and the unrestricted Lévy densitiy, as f.

As usual, the observed increments serve as a proxy for the latent jumps. We aim for high-frequency data, that is, we assume $\Delta_n \to 0$ as $n \to \infty$. The influence of Brownian motion and drift is then asymptotically negligible, since $\Delta_j^n W = O(\sqrt{\Delta_n})$ and $b(j\Delta_n - (j-1)\Delta_n) = O(\Delta_n)$. An infinite Lévy measure, however, implies infinitely many (small) jumps over any compact time interval. To circumvent problems with a possible singularity in zero, we always choose $\mathbb{D}$ away from zero such that there exists an $\varepsilon > 0$ with $\mathbb{D} \cap B_\varepsilon(0) = \emptyset$, where $B_\varepsilon(0)$ denotes the ε-ball centred at the origin. Moreover, we assume that $\mathbb{D}$ is a finite union of compact d-dimensional intervals.

2.2. *Projection estimation*

We assume that the Lévy density f is bounded outside every neighbourhood of the origin. This implies $\|f\|_{L^2(\mathbb{D})}^2 := \int_{\mathbb{D}} f^2(x)\mathrm{d}x < \infty$, hence $f \in L^2(\mathbb{D})$.

Let M be an auxiliary set to enumerate projection spaces. Furthermore, let $\{\mathbb{S}_m : m \in M\}$ be a family of finite-dimensional linear subspaces of $L^2(\mathbb{D})$, called sieves. The best approximation of f w. r. t. $(\mathbb{S}_m, \|\cdot\|_{L^2(\mathbb{D})})$ is given by its orthogonal projection, denoted by $\mathcal{P}_m f$. We denote the scalar product of a function $g \in L^2(\mathbb{D})$ and the Lévy density f by $F(g) := \int_{\mathbb{D}} g(x)f(x)\mathrm{d}x = \int_{\mathbb{D}} g(x)F(\mathrm{d}x)$ and set $d_m := \dim \mathbb{S}_m$. Let $\{g_{m,k} : k = 1, \ldots, d_m\}$ be an orthonormal basis of $\mathbb{S}_m$, then classical Hilbert space theory implies

$$\mathcal{P}_m f = \sum_{k=1}^{d_m} F(g_{m,k}) g_{m,k}.$$

For presentation purposes, we extend $g \in L^2(\mathbb{D})$ to $\mathbb{R}^d$, setting $g(x) = 0$ for all $x \notin \mathbb{D}$. If g is F-a. e. continuous, bounded, and vanishes in a neighbourhood of the origin, then we deduce $\lim_{\Delta \to 0} E[\Delta^{-1} g(X_\Delta)] = F(g)$ from Corollary 8.9 of Sato (1999).

The increments defined in Equation (3) are i. i. d. Thus, if $T_n \to \infty$ and $\Delta_n \to 0$, the estimator

$$\hat{F}^n(g) := \frac{1}{T_n} \sum_{j=1}^{T_n/\Delta_n} g(\Delta_j^n X)$$

of $F(g)$ based on $(\Delta_j^n X)_{j=1,\dots,T_n/\Delta_n}$ is consistent. This is a consequence of $0 \leq \mathrm{Var}[\hat{F}^n(g)] \leq E[(\hat{F}^n(g))^2] = T_n^{-1}(E[\Delta_n^{-1} g^2(X_{\Delta_n})]) \to 0$ as $n \to \infty$. Moreover, the estimator

$$\hat{f}_m^n := \sum_{k=1}^{d_m} \hat{F}^n(g_{m,k}) g_{m,k} \tag{4}$$

of $\mathcal{P}_m f$ is consistent as well. We call $\hat{f}_m^n$ the *projection estimator* of f w. r. t. $\mathbb{S}_m$. To our knowledge, this form of estimator traces back to Čencov (1962). Another representation of $\hat{f}_m^n$ as minimum contrast estimator (see, e. g., Definition 2 of Birgé and Massart (1993)) is shown in Figueroa-López (2009). We summarise relevant findings on $\hat{f}_m^n$ in the following Lemma. A proof is given in Appendix A.1.

Lemma 2.1 (minimum contrast estimator): *For $n \in \mathbb{N}$, let γ^n be an (empirical) contrast function defined by*

$$\gamma^n : L^2(\mathbb{D}) \to \mathbb{R};\ g \mapsto -\frac{2}{T_n} \sum_{j=1}^{T_n/\Delta_n} g(\Delta_j^n X) + \|g\|^2_{L^2(\mathbb{D})}. \tag{5}$$

Then, the projection estimator $\hat{f}_m^n$ defined in Equation (4) *and the minimum contrast estimator given by* $\arg\min_{g \in \mathbb{S}_m} \gamma^n(g)$ *coincide. Moreover,*

$$\gamma^n(\hat{f}_m^n) = -\|\hat{f}_m^n\|^2_{L^2(\mathbb{D})}. \tag{6}$$

The approach via contrast functions has the advantage that a specific orthonormal basis of $\mathbb{S}_m$ need not be chosen.

2.3. *Penalisation and model selection*

For a density estimator $\hat{f}$ of f we denote the *estimation risk* by $\|f - \hat{f}\|^2_{L^2(\mathbb{D})}$. We refer to its expectation as *mean squared error* (henceforth abbreviated MSE). For every $n \in \mathbb{N}$ we aim to select $m_n \in M$ such that the MSE of $\hat{f}_{m_n}^n$ is minimal. For flexibility, let $(M_n)_{n \in \mathbb{N}}$ be an increasing family such that $\cup_{n \in \mathbb{N}} M_n = M$. When choosing

$$m_n^\star := \operatorname*{arg\,min}_{m \in M_n} E\|f - \hat{f}_m^n\|^2_{L^2(\mathbb{D})},$$

then the projection estimator $\hat{f}_{m_n^\star}^n$ (henceforth shortened $\hat{f}_\star^n$) is called the *oracle*. Note, however, that $\hat{f}_\star^n$ is not computable without prior knowledge of f.

As a remedy, we invoke the method of model selection via penalisation (see, e. g., Barron et al. (1999) and references therein). Let $(\mathrm{pen}_n)_{n \in \mathbb{N}}$ be a family of mappings from $M \times \Omega$ to $[0, \infty[$. We call $(\mathrm{pen}_n)_{n\in\mathbb{N}}$ a *penalty on* $\{\mathbb{S}_m : m \in M\}$ if $\mathrm{pen}_n(m, \cdot)$ is measurable w. r. t. $\sigma(\Delta_j^n X : j = 1, \dots, T_n/\Delta_n)$ for all $n \in \mathbb{N}$ and $m \in M$. Note that this measurability condition is rather innocuous, requiring that pen_n depends only on the observable increments at stage n and can therefore be computed. In reality, given a penalty $(\mathrm{pen}_n)_{n\in\mathbb{N}}$ on $\{\mathbb{S}_m : m \in M\}$, we replace $m_n^\star$ by

$$m_n^{\mathrm{pen}} := \underset{m\in M_n}{\arg\min} \left\{ -\|\hat{f}_m^n\|_{L^2(\mathbb{D})}^2 + \mathrm{pen}_n(m) \right\},$$

and call $\hat{f}^n_{m_n^{\mathrm{pen}}}$ the PPE of f. For convenience of notation we set $\hat{f}^n_{\mathrm{pen}} := \hat{f}^n_{m_n^{\mathrm{pen}}}$, and also $\mathcal{P}_{\mathrm{pen}} f := \mathcal{P}_{m_n^{\mathrm{pen}}} f$.

Due to its relevance, again, we summarise the representation of the PPE $\hat{f}^n_{\mathrm{pen}}$ as a minimum penalised contrast estimator in the following Corollary to Lemma 2.1.

Corollary 2.2 (minimum penalised contrast estimator): *Let γ^n be the (empirical) contrast function defined in Equation* (5)*. Then, the PPE $\hat{f}^n_{\mathrm{pen}}$ coincides with the minimum penalised contrast estimator given by* $\arg\min_{g\in\mathbb{S}_{m_n^{\mathrm{pen}}}} \gamma^n(g)$*, where the data-driven model selection criterion is given by*

$$m_n^{\mathrm{pen}} := \underset{m\in M_n}{\arg\min} \left\{ \min_{g\in\mathbb{S}_m} \gamma^n(g) + \mathrm{pen}_n(m) \right\}. \tag{7}$$

Suitable penalties $(\mathrm{pen}_n)_{n\in\mathbb{N}}$ remain to be found.

3. Error bounds and adaptiveness

In this section we first establish our main results, that is, an oracle-type risk bound and a genuine oracle inequality (recall Equation (2)) for the PPE $\hat{f}^n_{\mathrm{pen}}$ introduced in Section 2. Then, we give an application, where rates of convergence for the MSE are explicit. We also show that the PPE achieves the best rate attainable in the framework of the application.

3.1. *Oracle inequality*

For general background on multi-index notation and weak derivatives, we refer to Brenner and Scott (1994, pp. 24–26). A d-tupel $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ is called a d-dimensional *multi-index*. The sum of its components, that is, $|\alpha| := \alpha_1 + \cdots + \alpha_d$ is called the *length* of α. The corresponding (weak) α-differential operator ∂^α is given by $\partial^\alpha := \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$. Note that a differentiable function is also weakly differentiable. In this case, the (strong) derivative and the weak derivative coincide.

For $k \in \mathbb{N}$ and $p \geq 1$ let $g : \mathbb{D} \to \mathbb{R}$ be locally integrable, denoted by $g \in L_{\mathrm{loc}}(\mathbb{D})$, and suppose that the weak α-derivative $\partial^\alpha g$ exists for every multi-index α with $|\alpha| \leq k$. Then, we call

$$\|g\|_{\mathcal{W}^{k,p}(\mathbb{D})} := \left(\sum_{|\alpha|\leq k} \|\partial^\alpha g\|_{L^p(\mathbb{D})}^p \right)^{1/p} \quad \text{and} \quad |g|_{\mathcal{W}^{k,p}(\mathbb{D})} := \left(\sum_{|\alpha|= k} \|\partial^\alpha g\|_{L^p(\mathbb{D})}^p \right)^{1/p}$$

Sobolev norm and *Sobolev semi-norm* of g, respectively. Additionally, we call

$$\mathcal{W}^{k,p}(\mathbb{D}) := \{g \in L_{\mathrm{loc}}(\mathbb{D}) : \|g\|_{\mathcal{W}^{k,p}(\mathbb{D})} < \infty\} \tag{8}$$

Sobolev space over $\mathbb{D}$ and refer to k as *degree of smoothness.*

Assumption 3.1 (regularity of projection spaces): *For all $m \in M$, we assume that there exists a finite partition $\mathcal{D}_m$ of disjoint d-dimensional intervals of $\mathbb{D}$ such that $g_{|D} \in \mathcal{W}^{1,\infty}(D)$ for all $D \in \mathcal{D}_m$ and all $g \in \mathbb{S}_m$.*

In the light of Mazja (1979, 3.1 Korollar 1.1), Assumption 3.1 is sufficient for the so-called "absolute continuity on lines" for $g \in \mathbb{S}_m$. Consequently, the (weak) partial derivative $\partial_i g_{|D}$ exists for all $g \in \mathbb{S}_m$, $D \in \mathcal{D}_m$ and $i \in \{1, \dots, d\}$. For our main results we require upper bounds for the range of $g \in \mathbb{S}_m$ and for the line integrals of $\partial_i g_{|D}$ along the coordinate axes for all $D \in \mathcal{D}_m$ and $i \in \{1, \dots, d\}$.

For all $x \in D$, therefore, let $\alpha_{D,x,i} : [\underline{\xi}_i(x), \overline{\xi}_i(x)] \to D$ be defined by $\alpha_{D,x,i}(\xi) := x+\xi e_i$, where e_i denotes the i-th standard unit vector in $\mathbb{R}^d$, $\overline{\xi}_i(x) := \sup\{\xi \geq 0 : x + \xi e_i \in D\}$, and $\underline{\xi}_i(x) := \inf\{\xi \leq 0 : x + \xi e_i \in D\}$. Then, $(g \circ \alpha_{D,x,i})' = \partial_i g_{|D} \circ \alpha_{D,x,i}$ exists under Assumption 3.1. To avoid tedious notation, we set

$$\|g'\|_{\alpha,L^1(\mathcal{D}_m)} := \max_{D\in\mathcal{D}_m} \max_{i=1,\dots,d} \sup_{x\in D} \|(g \circ \alpha_{D,x,i})'\|_{L^1([\underline{\xi}_i(x),\overline{\xi}_i(x)])}.$$

Consequently, for every $m \in M$ we set $\mathfrak{D}_m := \sup\{\|g\|^2_{L^\infty(\mathbb{D})} : g \in \mathbb{S}_m, \|g\|_{L^2(\mathbb{D})} = 1\}$, which is an upper bound, uniformly in $\mathbb{S}_m$, for (the square of) the range of g after normalisation. Likewise, we set $\mathfrak{D}'_m := \sup\{\|g'\|^2_{\alpha,L^1(\mathcal{D}_m)} : g \in \mathbb{S}_m, \|g\|_{L^2(\mathbb{D})} = 1\}$, which is an upper bound, uniformly in $\mathbb{S}_m$, for (the square of) the line integrals of $\partial_i g_{|D}$ along the coordinate axes, again, after normalisation.

In Appendix A.2, we prove the following remark.

Remark 1 : Under Assumption 3.1, the constants $\mathfrak{D}_m$ and $\mathfrak{D}'_m$ are finite for all $m \in M$.

Assumption 3.2 (polynomial family of sieves): *The family $\{\mathbb{S}_m : m \in M\}$ of projection spaces is assumed to be polynomial, that is, there exist $\zeta_1 > 0$ and $\zeta_2 \geq 0$ such that*

$$\forall l \in \mathbb{N} : \#\{m \in M : d_m = l\} \leq \zeta_1 l^{\zeta_2}. \tag{9}$$

We now present our main theorem that stands in line with Theorem 1 of Reynaud-Bouret (2003) and Theorem 4.1 of Figueroa-López and Houdré (2006). Theorem 3.3 justifies the empirical sieve selection via penalisation for our statistical problem, where (high-frequency) observation of realised increments $(\Delta_j^n X)_{j=1,\dots,T_n/\Delta_n}$ of a Lévy process X is available only. A proof is given in Appendix A.3.

Theorem 3.3 (oracle-type risk bound for PPE): *Let $(T_n)_{n\in\mathbb{N}}$ and $(\Delta_n)_{n\in\mathbb{N}}$ be sequences with values in $]0, \infty[$ such that $T_n \to \infty$ and $\sup_{n\in\mathbb{N}} T_n\Delta_n < \infty$ as $n \to \infty$. Furthermore, let $(\mathrm{pen}_n)_{n\in\mathbb{N}}$ be a penalty on a family of sieves $\{\mathbb{S}_m \subseteq L^2(\mathbb{D}) : m \in M\}$ such that there exist $c_1 > 1$ and $c_2, \dots, c_4 > 0$ with*

$$\mathrm{pen}_n(m) = \frac{c_1}{T_n} \sum_{k=1}^{d_m} \hat{F}^n(g^2_{m,k}) + c_2 \left(\frac{\mathfrak{D}_m}{T_n} \vee \frac{\mathfrak{D}^4_m}{T^3_n}\right) + c_3 \left(\frac{\mathfrak{D}'_m}{T_n} \vee \frac{\mathfrak{D}'^4_m}{T^3_n}\right) + c_4 \left(\frac{d_m}{T_n} \vee \frac{d^4_m}{T^3_n}\right) \tag{10}$$

and let $M_n := \{m \in M : \mathfrak{D}_m \le T_n\}$ *for* $n \in \mathbb{N}$. *Then, there exist positive constants* $K_1, K_2 < \infty$ *such that the PPE* $\hat{f}^n_{\text{pen}} = \hat{f}^n_{m_n^{\text{pen}}}$ *satisfies for all* $n \in \mathbb{N}$,

$$E\|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} \le K_1 \inf_{m \in M_n} \Big(\|f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + E[\text{pen}_n(m)] \Big) + \frac{K_2}{T_n}. \tag{11}$$

Let us briefly comment on the three last terms on the right-hand side of Equation (10). On the one hand, for a fixed model $m \in M$, we have $\mathfrak{D}_m T_n^{-1} \ge \mathfrak{D}_m^4 T_n^{-3}$ if, and only if, $T_n \ge \mathfrak{D}_m^{3/2}$. The analogue holds w. r. t. $\mathfrak{D}'_m$ and d_m. Thus, there exists a finite $N_m \in \mathbb{N}$ such that

$$\text{pen}_n(m) = \frac{c_1}{T_n} \sum_{k=1}^{d_m} \hat{F}^n(g^2_{m,k}) + c_2 \frac{\mathfrak{D}_m}{T_n} + c_3 \frac{\mathfrak{D}'_m}{T_n} + c_4 \frac{d_m}{T_n} \tag{12}$$

for all $n \ge N_m$. We note that this would match the penalty from part c) of Theorem 4.1 of Figueroa-López and Houdré (2006) if c_3 was equal to zero. On the other hand, for a fixed $n \in \mathbb{N}$, let $\bar{M}_n \subseteq M_n$ be the subset of models such that $T_n < \max(\mathfrak{D}_m^{3/2}, \mathfrak{D}_m'^{3/2}, d_m^{3/2})$ for all $m \in \bar{M}_n$. We interpret these models as 'more complex', for instance, in the sense of dimensionality. We observe that pen_n given in Equation (10) imposes higher penalties on models in $\bar{M}_n$ compared to the penalty given in Equation (12). Therefore, Equation (10) increases the probability for a less complex model $m_n^{\text{pen}} \in M_n \setminus \bar{M}_n$ to be chosen. Taking these findings into account, we readily derive the oracle inequality below, which we prove in Appendix A.4.

Corollary 3.4 (oracle inequality): *Let the prerequisites of Theorem 3.3 be satisfied, but let* $M_n := \{m \in M : \max(\mathfrak{D}_m^{3/2}, \mathfrak{D}_m'^{3/2}, d_m^{3/2}) \le T_n\}$. *If, additionally,*

$$\inf_{n \in \mathbb{N}} \inf_{m \in M_n} \frac{\sum_{k=1}^{d_m} E[\hat{F}^n(g^2_{m,k})]}{\mathfrak{D}_m + \mathfrak{D}'_m + d_m} > 0 \tag{13}$$

holds, then there exist positive constants $K_1, K_2 < \infty$ *such that the PPE* $\hat{f}^n_{\text{pen}} = \hat{f}^n_{m_n^{\text{pen}}}$ *and the oracle* $\hat{f}^n_\star = \hat{f}^n_{m_n^\star}$ *satisfy for all* $n \in \mathbb{N}$,

$$E\|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} \le K_1 E\|f - \hat{f}^n_\star\|^2_{L^2(\mathbb{D})} + \frac{K_2}{T_n}. \tag{14}$$

3.2. *Estimation of Sobolev-type smooth Lévy densities*

As an application we consider the estimation of a Sobolev-type smooth Lévy density $f \in \mathcal{W}^{k,2}(\mathbb{D})$, recall Equation (8), where the degree of smoothness $k \in \mathbb{N}$ is unknown. From approximation theory (see, e. g., Sections 2.9–10 of DeVore and Lorentz (1993)) and the theory of finite elements (see, e. g., Chapter 4 of Brenner and Scott (1994)) we know that piecewise polynomials have good approximation properties for Sobolev-type smooth functions. For $k \in \mathbb{N}$ let $\mathscr{P}_k := \{g \in \mathbb{R}[x] : \deg(g) \le k\}$ denote the ring of polynomials with degree less than or equal to k. Then, we define

$$\mathscr{P}_k^d := \left\{ g : \mathbb{R}^d \to \mathbb{R}; g(x_1, \dots, x_d) = \sum_{j=1}^{l} c_j \prod_{i=1}^{d} g_{j,i}(x_i) : l \in \mathbb{N}, c_j \in \mathbb{R}, g_{j,i} \in \mathscr{P}_k \right\}. \quad (15)$$

We choose $M := \mathbb{N}$ in this setting. For all $m \in M$ let $\mathcal{D}_m$ be a partition of $\mathbb{D}$, consisting of d-dimensional intervals, such that

$$\forall D \in \mathcal{D}_m : \quad \operatorname{diam} D = m^{-1} \operatorname{diam} \mathbb{D}, \quad (16)$$

where the *diameter of* D is defined by $\operatorname{diam} D := \max\{\|x_1 - x_2\|_\infty : x_1, x_2 \in D\}$. Then

$$\mathbb{S}_m^k := \{g \in L^2(\mathbb{D}) : g_{|D} \in \mathscr{P}_k^d \text{ for all } D \in \mathcal{D}_m\} \quad (17)$$

is called the space of *piecewise polynomials with degree less than or equal to k based on* $\mathcal{D}_m$. Certainly, $\{\mathbb{S}_m^k : m \in \mathbb{N}\}$ satisfies Assumptions 3.1 and 3.2, where ζ_1 and ζ_2 in Equation (9) can be chosen equal to one and zero, respectively. Nevertheless, we have to exclude some cases of degenerate partitions.

Assumption 3.5 (regularity of partitions): *In addition to Equation* (16)*, we assume that there exists a positive* $\rho \leq 1$ *such that for all* $m \in \mathbb{N}$ *and* $D \in \mathcal{D}_m$*, there exists a ball* $B_D \subseteq D$ *with* $\operatorname{diam} B_D \geq \rho \operatorname{diam} D$.

As a consequence of Theorem 4.4.20 of Brenner and Scott (1994), we have the following approximation result.

Proposition 3.6 (approximation error for Sobolev-type smooth functions): *Let* $\mathbb{D} \subseteq \mathbb{R}_\circ^d$ *satisfy Assumption 3.1 such that* $f \in \mathcal{W}^{k,2}(\mathbb{D})$ *for some* $k \in \mathbb{N}$*. Furthermore, for* $m \in \mathbb{N}$*, let* $\mathbb{S}_m^{k-1}$ *be defined by Equation* (17)*, where the underlying partition* $\mathcal{D}_m$ *satisfies Assumption 3.5. Then, there exists a positive constant* $C_{k,\rho} < \infty$ *such that*

$$\|f - \mathcal{P}_m f\|_{L^2(\mathbb{D})} \leq C_{k,\rho} |f|_{\mathcal{W}^{k,2}(\mathbb{D})} m^{-k}. \quad (18)$$

Remark 2: In the setting of Proposition 3.6 we observe that Equation (13) in Corollary 3.4 is satisfied for all sequences $(T_n)_{n\in\mathbb{N}}$ and $(\Delta_n)_{n\in\mathbb{N}}$ such that $T_n \to \infty$ as $n \to \infty$ and $\sup_{n\in\mathbb{N}} T_n \Delta_n < \infty$.

We prove Proposition 3.6 and Remark 2 in Appendices A.5 and A.6, respectively.

We now combine the risk bound (11) from Theorem 3.3 with the approximation result (18) from Proposition 3.6 to derive an explicit rate of convergence for the PPE estimating a Lévy density belonging to some Sobolev space. Our result stands in line with Corollary 5.1 of Figueroa-López and Houdré (2006). We show that it can be extended to the estimation of Lévy densities from observations on a discrete time grid. Without prior knowledge of the smoothness of f, in contrast to Proposition 3.5 of Figueroa-López (2009), we prove that the PPE based on the empirical sieve selection method given in Equation (10) achieves the optimal rate of convergence.

Theorem 3.7 (adaptive rate of convergence of PPE): *Let* $\mathbb{D} \subseteq \mathbb{R}_\circ^d$ *satisfy Assumption 3.1 such that* $f \in \mathcal{W}^{k,2}(\mathbb{D})$ *for some* $k \in \mathbb{N}$ *with* $k > d/4$*. Furthermore, let* $\{\mathbb{S}_m^{k-1} : m \in \mathbb{N}\}$ *be defined by Equation* (17)*, where the underlying partitions* $(\mathcal{D}_m)_{m\in\mathbb{N}}$

satisfy Assumption 3.5. Assume also that $(T_n)_{n\in\mathbb{N}}$ *and* $(\Delta_n)_{n\in\mathbb{N}}$ *be such that* $T_n \to \infty$ *as* $n \to \infty$ *and* $\sup_{n\in\mathbb{N}} T_n\Delta_n < \infty$*, and define* $M_n := \{m \in M : \mathfrak{D}_m \leq T_n\}$ *for* $n \in \mathbb{N}$*. Then, a penalty* $(\mathrm{pen}_n)_{n\in\mathbb{N}}$ *satisfying Equation* (10) *for some* $c_1 > 1$ *and* $c_2, \ldots, c_4 > 0$ *is sufficient for*

$$\sup_{n\in\mathbb{N}} T_n^{\frac{2k}{2k+d}} E\|f - \hat{f}^n_{\mathrm{pen}}\|^2_{L^2(\mathbb{D})} < \infty. \tag{19}$$

Moreover, let $a_1, a_2 > 0$ *be finite constants and denote by*

$$B(a_1, a_2) := \{g \in \mathcal{W}^{k,2}(\mathbb{D}) : \|g\|_{L^\infty(\mathbb{D})} \leq a_1 \ \textit{and}\ |g|_{\mathcal{W}^{k,2}(\mathbb{D})} \leq a_2\}$$

the ball of all Sobolev-type smooth functions with supremum norm and Sobolev semi-norm bounded by a_1 *and* a_2*, respectively. Then, additionally,*

$$\sup_{n\in\mathbb{N}} T_n^{\frac{2k}{2k+d}} \sup_{f\in B(a_1,a_2)} E\|f - \hat{f}^n_{\mathrm{pen}}\|^2_{L^2(\mathbb{D})} < \infty. \tag{20}$$

Remark 3 : The rate of convergence in Equations (19) and (20) is optimal in the minimax sense. In particular, $\liminf_{n\to\infty} T_n^{2k/(2k+d)} \inf_{\hat{f}^n} \sup_{f\in\mathcal{W}^{k,2}} E_f\|\hat{f}^n - f\|^2_{L^2(\mathbb{D})} > 0$ holds in analogy to Corollary 4.2 and Remark 4.3 of Figueroa-López (2009), where the infimum is taken over all estimators $\hat{f}_n$ that are $\sigma(\Delta^n_j X : j = 1, \ldots, T_n/\Delta_n)$-measurable.

We prove Theorem 3.7 and Remark 3 in Appendices A.7 and A.8, respectively.

4. Simulations

In this section we present our simulation results. We have implemented the PPE method based on piecewise quadratic polynomials. Recall that $M = \mathbb{N}$ and $\mathbb{S}_m = \mathbb{S}^{k-1}_m$ with $k = 3$ as defined in Equation (17). For a given domain of estimation $\mathbb{D}$ note that $\mathfrak{D}_m = 9m/\mathrm{vol}(\mathbb{D})$, $\mathfrak{D}'_m = 45m/\mathrm{vol}(\mathbb{D})$, and $d_m = 3m$. Consequently, $M_n = \{1, \ldots, \lfloor T_n\mathrm{vol}(\mathbb{D})/9\rfloor\}$. In addition, the penalty constants in Equation (10) are set to $c_1 = 2$, $c_2 = 1$, $c_3 = 0.1$, and $c_4 = 0.5$. Although in practice, the penalty constants could be tuned to give better estimates in instances where Brownian motion is clearly present, here, we use the same constants whether Brownian motion is present or not. In doing so, we intend to emphasise the effect of Brownian motion on the PPE and the asymptotic behaviour of the PPE.

As a comparison, we also implemented the estimation procedure described in secs. 6 and 7 of Comte and Genon-Catalot (2009, 2011), respectively. We denote this estimator by SCE, which indicates the sinus cardinal (basis). Moreover, any notation referring to the latter procedure will be appended by the label *SC*. Let g^* denote the Fourier transform of a function g and let φ denote the sinus cardinal, that is, $\varphi(x) = \sin(\pi x)/(\pi x)$ with $\varphi(0) = 1$. For $m_{\mathrm{sc}} > 0$ the corresponding SC-projection space is given by $\mathbb{S}^{\mathrm{sc}}_{m_{\mathrm{sc}}} = \{g \in L^2(\mathbb{R} : \mathrm{supp}(g^*) \in [-\pi m_{\mathrm{sc}}, \pi m_{\mathrm{sc}}])\}$. The set $\{\varphi_{m_{\mathrm{sc}},k} : k \in \mathbb{Z}\}$, where $\varphi_{m_{\mathrm{sc}},k}(x) = \sqrt{m_{\mathrm{sc}}}\varphi(m_{\mathrm{sc}}x - k)$, forms an orthonormal basis of $\mathbb{S}^{\mathrm{sc}}_{m_{\mathrm{sc}}}$. Note that m_{sc} plays the role of a bandwidth and is unrelated to the m of our method.

Depending on whether Brownian motion is absent or present, the corresponding SCEs of $x \mapsto g^{\rm sc}(x) = xf(x)$ and $x \mapsto p^{\rm sc}(x) = x^3 f(x)$ are given by

$$\hat{g}^{\rm sc}_{m_{\rm sc}} = \sum_{k\in\mathbb{Z}} \hat{a}^{\rm sc}_{m_{\rm sc},k}\varphi_{m_{\rm sc},k} \quad \text{and} \quad \hat{p}^{\rm sc}_{m_{\rm sc}} = \sum_{k\in\mathbb{Z}} \hat{b}^{\rm sc}_{m_{\rm sc},k}\varphi_{m_{\rm sc},k},$$

respectively, where

$$\hat{a}^{\rm sc}_{m_{\rm sc},k} = \frac{1}{T_n}\sum_{j=1}^{T_n/\Delta_n} \Delta^n_j X \varphi_{m_{\rm sc},k}(\Delta^n_j X) \text{ and } \hat{b}^{\rm sc}_{m_{\rm sc},k} = \frac{1}{T_n}\sum_{j=1}^{T_n/\Delta_n} (\Delta^n_j X)^3 \varphi_{m_{\rm sc},k}(\Delta^n_j X).$$

The contrast values for the SCEs are equal to $-\sum_{k\in\mathbb{Z}}(\hat{a}^{\rm sc}_{m_{\rm sc},k})^2$ and $-\sum_{k\in\mathbb{Z}}(\hat{b}^{\rm sc}_{m_{\rm sc},k})^2$, and the respective penalty functions are defined by

$$\text{pen}^{\rm sc}_n(m_{\rm sc}) = \frac{\kappa_{\rm sc} m_{\rm sc}}{T_n^2}\sum_{j=1}^{T_n/\Delta_n}(\Delta^n_j X)^2 \text{ and } \text{pen}^{\rm sc}_n(m_{\rm sc}) = \frac{\kappa_{\rm sc} m_{\rm sc}}{T_n^2}\sum_{j=1}^{T_n/\Delta_n}(\Delta^n_j X)^6.$$

In analogy to Comte and Genon-Catalot (2009, 2011), we truncate the infinite sum in the definition of $\hat{g}^{\rm sc}_{m_{\rm sc}}$ and $\hat{p}^{\rm sc}_{m_{\rm sc}}$ to $\{k : |k| \le 15\}$. In addition, $m_{\rm sc}$ is chosen from the set $\{0.1, 0.2, \ldots, 10\}$, and the constant in the penalties is set to $\kappa_{\rm sc} = 7.5$ if there is no Brownian motion and $\kappa_{\rm sc} = 3$ otherwise. As we are interested in the Lévy density itself, we transform the raw estimates $\hat{g}^{\rm sc}_{m_{\rm sc}}$ and $\hat{p}^{\rm sc}_{m_{\rm sc}}$ to $\hat{f}^{\rm sc}_{m_{\rm sc}}(x) = \hat{g}^{\rm sc}_{m_{\rm sc}}(x)/x$ and $\hat{f}^{\rm sc}_{m_{\rm sc}}(x) = \hat{p}^{\rm sc}_{m_{\rm sc}}/x^3$, respectively, and restrict them to the domain of estimation $\mathbb{D}$ from our method.

We simulated the following univariate models:

(i) a compound Poisson process with intensity 0.5 and exponentially distributed jumps with mean 1: $f(x) = 0.5e^{-x}\mathbb{1}_{\{x>0\}}$;
(ii) a superposition of (i) and Brownian motion with $\sigma = 0.5$;
(iii) a standard gamma process: $f(x) = x^{-1}e^{-x}\mathbb{1}_{\{x>0\}}$;
(iv) a superposition of (iii) and Brownian motion with $\sigma = 0.5$;
(v) a superposition of a bilateral gamma process with parameters $(\alpha^+, \beta) = (1, 1)$ and $(\alpha^-, \beta) = (0.7, 1)$ and Brownian motion with $\sigma = 0.5$: $f(x) = x^{-1}e^{-x}\mathbb{1}_{\{x>0\}} + x^{-1}e^{0.7x}\mathbb{1}_{\{x<0\}}$.

Note that the parameters of the processes are taken as in Comte and Genon-Catalot (2009, 2011). In all cases, we investigated the scenarios

(1) $T_1 = 2500$, $\Delta_1 = 0.05$ (50 000 observations), and
(2) $T_2 = 5000$, $\Delta_2 = 0.02$ (250 000 observations).

Furthermore, we choose $\mathbb{D} = [0.05, 10]$ in cases (i) and (iii), $\mathbb{D} = [0.25, 10]$ in cases (ii) and (iv), and $\mathbb{D} = [-10, -0.35] \cup [0.35, 10]$ in case (v).

As $f \in \mathcal{C}^\infty(\mathbb{D})$ in all cases (i–v), by Theorem 3.7 we expect the PPE based on piecewise quadratic polynomials to converge with rate $T^{-6/7}$. By Theorem 3.1 and Theorem 4.1 of Comte and Genon-Catalot (2009, 2011), respectively, we expect the SCE to converge with rates (i) $T^{-3/4}$, (ii) $T^{-7/8}$, (iii) $T^{-1/2}$, and (iv–v) $T^{-5/6}$ in the respective cases. We give a summary of the theoretical relative reductions corresponding to doubling T from scenario (1) to (2) in Table 1.

For the cases (ii) and (iv), moreover, we remark the probability for a purely Brownian increment to be bigger than the lower bound of $\mathbb{D}$ (0.25 in these cases) equals 1.27%

Table 1. Summary of asymptotic rates of convergence (rows 1 and 3) and relative reduction of the MSE (rows 2 and 4) as T doubles from scenario (1) to (2) for the PPE (rows 1 and 2) and the SCE (rows 3 and 4) corresponding to the estimation of f for a CPP–Exp(1) with rate 0.5 (column i), a superposition of (i) and Brownian motion with $\sigma = 0.5$ (column ii), a standard gamma process (column iii), a superposition of (iii) and Brownian motion with $\sigma = 0.5$ (column iv), and a superposition of a bilateral gamma(1,1;0.7,1) process and Brownian motion with $\sigma = 0.5$ (column v).

	(i)	(ii)	(iii)	(iv)	(v)
PPE					
Asymptotic rate	$T^{-6/7}$	$T^{-6/7}$	$T^{-6/7}$	$T^{-6/7}$	$T^{-6/7}$
Rel. reduction ($T_2 = 2T_1$)	44.8%	44.8%	44.8%	44.8%	44.8%
SCE					
Asymptotic rate	$T^{-3/4}$	$T^{-7/8}$	$T^{-1/2}$	$T^{-5/6}$	$T^{-5/6}$
Rel. reduction ($T_2 = 2T_1$)	40.5%	45.5%	29.3%	43.9%	43.9%

in scenario (1) and 0.02% in scenario (2). Therefore, we expect significant distortions of the PPEs caused by Brownian motion in scenario (1), whereas these effects should remarkably diminish in scenario (2). In case (v), we have chosen $\mathbb{D}$ further away from the origin such that $\min_{x\in\mathbb{D}}|x| = 0.35$. The probabilities that a purely Brownian increment falls into $\mathbb{D}$, hence, are reduced to 0.174% and $7.43\cdot 10^{-5}$%, respectively, in comparison to cases (ii) and (iv). Accordingly, we expect the impact of Brownian motion on the PPEs to be small in either scenario. We want to emphasise that the SCEs are based on all increments independent of their sizes. Hence, we do not expect a significant difference for the SCEs between cases (ii) and (iv) on the one hand, and case (v) on the other hand.

Results are given in Figure 1. Columns (a/b) correspond to the PPE, and columns (c/d) correspond to the SCE. Columns (a/c) show 50 estimated Lévy densities for scenario (1), and columns (b/d) show 50 estimated Lévy densities for scenario (2). On the y-axis, we restrict the plotted range to (i) $[0, 0.75]$, (ii) $[0, 1.5]$, (iii) $[0, 20]$, (iv) $[0, 6]$, and (v) $[0, 5]$. Near zero, some of the estimates fall out of this range and had to be truncated above. Nonetheless, all these cases are explicitly discussed below. Moreover, for the cases (iii–v) the Lévy densities and their estimates plotted over $\mathbb{D}$ are indistinguishable to the naked eye. However, there are notable differences over the range $\mathbb{D} \cap [-2, 2]$ which we present here. In addition, for each scenario, we calculated the empirical MSE, that is, the mean of the empirical squared error of each estimate ($\|f - \hat{f}\|^2_{L^2(\mathbb{D})}$; cf. the definition at the beginning of Section 2.3), and the mean of the estimated m and $m_{\rm sc}$ selected by penalisation. These are summarised in Table 2. In brackets, we give the standard deviation over 50 samples.

For (i), the pure compound Poisson process, we observe that all four plots exhibit high quality estimates with small variability. Near zero, the PPE follows the slope of the true Lévy density closely. The estimated values $\hat{f}^n_{\rm pen}(0.05)$ range between 0.36 and 0.50, and between 0.31 and 0.39 in scenarios (1) and (2), respectively. The conclusion that the true Lévy density is bounded (on $\mathbb{R}_\circ$) becomes obvious. For the SCE, this is not necessarily the case. The estimated values $\hat{f}^{\rm sc}(0.05)$ range between 0.79 and 1.25 in scenario (1), and between 1.54 and 1.95 in scenario (2). Compare these values with the true value $f(0.05) \approx 0.48$. Note also, the raw estimates $\hat{g}^{\rm sc}$ are, in general, non-zero

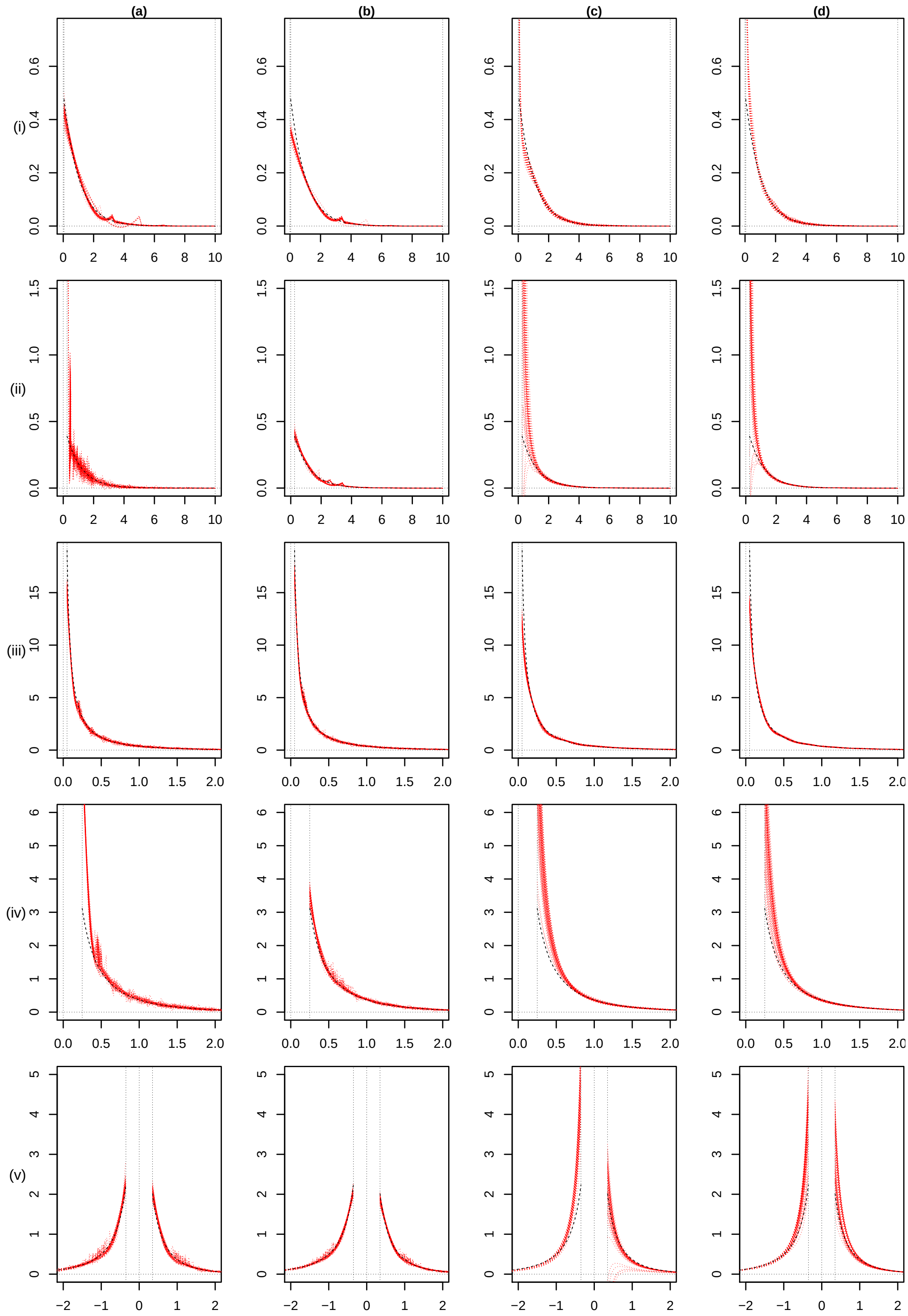

Figure 1. Estimation of f for a CPP–Exp(1) with intensity 0.5 (row i), a superposition of (i) and Brownian motion with $\sigma = 0.5$ (BM) (row ii), a standard gamma process (row iii), a superposition of (iii) and BM (row iv), and a superposition of a bilateral gamma(1,1;0.7,1) process and BM (row v). We present the true (dashed black) and 50 Lévy densities estimated (dotted red) by the PPE (columns a/b) and the SCE (column c/d), where $(T_n, \Delta_n) = (2500, 0.05)$ (columns a/c) and $(T_n, \Delta_n) = (5000, 0.02)$ (columns b/d).

Table 2. Summary of the estimation of f for a CPP–Exp(1) with rate 0.5 (row i), a superposition of (i) and Brownian motion with $\sigma = 0.5$ (row ii), a standard gamma process (row iii), a superposition of (iii) and Brownian motion with $\sigma = 0.5$ (row iv), and a superposition of a bilateral gamma(1,1;0.7,1) process and Brownian motion with $\sigma = 0.5$ (row v) by the PPE based on piecewise quadratic polynomials and the SCE.

X	(T_n, Δ_n)	$\overline{m_n^{\text{pen}}}$	$\overline{m_{\text{sc}}}$	$\overline{\text{se}}(\hat{f}^n_{\text{pen}})$	$\overline{\text{se}}(\hat{f}^{\text{sc}}_{m_{\text{sc}}})$	
(i)	$(2500, 0.05)$	2.92 (0.34)	0.96 (0.13)	0.876 (0.642)	8.065 (2.599)	$\times 10^{-3}$
	$(5000, 0.02)$	2.98 (0.14)	1.87 (0.35)	0.415 (0.209)	0.385 (0.271)	$\times 10^{-3}$
(ii)	$(2500, 0.05)$	43.34 (2.73)	0.47 (0.21)	0.752 (0.068)	1.527 (2.178)	
	$(5000, 0.02)$	3.50 (0.71)	0.55 (0.25)	0.007 (0.003)	5.397 (5.410)	$\times 10^{-1}$
(iii)	$(2500, 0.05)$	59.10 (4.40)	4.82 (0.41)	0.174 (0.052)	0.765 (0.133)	
	$(5000, 0.02)$	73.12 (8.68)	5.93 (0.31)	0.059 (0.018)	0.329 (0.053)	
(iv)	$(2500, 0.05)$	43.96 (4.54)	0.63 (0.27)	0.885 (0.091)	1.185 (0.747)	
	$(5000, 0.02)$	25.58 (5.40)	0.72 (0.24)	0.015 (0.005)	0.674 (0.488)	
(v)	$(2500, 0.05)$	26.36 (3.11)	0.29 (0.03)	0.137 (0.057)	5.679 (4.369)	$\times 10^{-1}$
	$(5000, 0.02)$	25.56 (3.00)	0.46 (0.23)	0.051 (0.012)	3.733 (2.003)	$\times 10^{-1}$

Notes: The empirical mean of the values for m chosen by penalisation, and the empirical MSE for each pair (T_n, Δ_n) are presented. Standard deviations over 50 samples are given within the brackets. The squared errors and their standard deviations are to be scaled by the factor in the last column.

at the origin. Without restriction to $\mathbb{D}$, therefore, the SCEs of f have a pole at zero, whereas $f(x) \to 0.5$ as $x \to 0$. In contrast, the SCEs are smoother than the PPEs further away from zero. Moreover, the empirical MSEs of the PPEs and SCEs reduce by 52.6%, and 95.2% on average, respectively. For comparison, we refer to the asymptotic values summarised in Table 1.

For (ii), the superposition of (i) and Brownian motion, we observe highly unstable estimates in columns (a), (c) and (d), and high quality estimates in column (b) only. The distortions in the former cases are due to Brownian motion. However, in the latter case, the PPE behaves similar to case (i), where Brownian motion was absent. In particular, the PPE benefits considerably from the smaller observation time lag Δ_2. For the SCE this is not the case, as all observed increments are taken into account independent of their sizes. The values $\hat{f}^n_{\text{pen}}(0.25)$ estimated by the PPE range between 4.47 and 6.01 in scenario (1), and between 0.36 and 0.48 in scenario (2). In contrast, the values $\hat{f}^{\text{sc}}(0.25)$ estimated by the SCE range between -1.46 and +17.0, and between -0.50 and +7.95 in scenarios (1) and (2), respectively. The true value is $f(0.25) \approx 0.39$. Note that the raw estimates $\hat{p}^{\text{sc}}$ are, in general, non-zero at the origin. Unrestricted, thus, the SCEs of f have a pole at zero, whereas $f(x) \to 0.5$ as $x \to 0$. Moreover, the defining property of Lévy densities, that is, $\int(|x|^2 \wedge 1)\hat{f}^{\text{sc}}(x)\mathrm{d}x < \infty$, is violated.

For (iii), the standard gamma process, we observe that all four plots exhibit high quality estimates with small variability. The empirical MSE of the PPE is slightly smaller than the corresponding MSE of the SCE as the PPEs follow the slope near zero slightly closer. Further away from zero, though, the SCEs are smoother than the PPEs. We observe the empirical MSEs of the PPEs and SCEs reduce by 66.1%, and 57.9% on average, respectively. Again, we refer to the asymptotical values summarised in Table 1 for comparison.

For (iv), the superposition of (iii) and Brownian motion, similar to (ii) we observe unstable estimates in columns (a), (c) and (d), and estimates of higher quality in column (b) only. Once more, we observe distortions in the former cases due to Brownian motion. However, in the latter case, the PPE behaves very similar to case (iii), where Brownian motion was absent. The PPE benefits considerably from the smaller observation time lag Δ_2, whereas the SCE does not. The values $\hat{f}^n_{\text{pen}}(0.25)$ estimated by the PPE range between 7.55 and 9.11 in scenario (1), and between 3.14 and 3.89 in scenario (2). In contrast, the values $\hat{f}^{\text{sc}}(0.25)$ estimated by the SCE range between 3.50 and 11.6 with mean 8.02 in scenario (1), and between 3.49 and 10.4 with mean 6.78 in scenario (2). Compare these values to the true value $f(0.25) \approx 3.12$. Note also, the raw estimates $\hat{p}^{\text{sc}}$ exhibit, in general, non-zero values at the origin for both scenarios (1) and (2). Analogously to case (ii), therefore, the unrestricted SCEs of f violate $\int(|x|^2 \wedge 1)\hat{f}^{\text{sc}}(x)\mathrm{d}x < \infty$. Furthermore, the empirical MSEs of the PPEs and SCEs reduce by 98.3%, and 57.9% on average, respectively. For comparison, once more, we refer to the asymptotical values in Table 1.

For (v), the superposition of a bilateral gamma process and Brownian motion, we chose $\mathbb{D}$ further away from the origin in comparison to cases (ii) and (iv). The PPE exhibits a reasonable empirical MSE in both scenarios (1) and (2) as compared to case (iii), where Brownian motion was absent. Moreover, the PPEs are not too large to be plotted and, hence, not truncated. Although one may expect estimates like those in case (iv), changing $\mathbb{D}$ yields estimates like those in case (iii). The influence of purely Brownian increments is lowered considerably in comparison to case (iv). As for the SCE, in scenario (1) the estimated values $\hat{f}^{\text{sc}}(-0.35)$ and $\hat{f}^{\text{sc}}(0.35)$ range between 3.21 and 6.36, and between -1.87 and $+3.25$, respectively. In scenario (2), the SCEs' corresponding values range between 2.04 and 5.49, and between 1.43 and 4.35, respectively. Compare these values to the true values $f(-0.35) \approx 2.24$ and $f(0.35) \approx 2.01$. We note that the SCE does not benefit significantly from the change of $\mathbb{D}$.

From a statisticians point of view, if Brownian motion is present, the choice of $\mathbb{D}$ appears to be crucial for a given scenario. In cases (ii) and (iv) above, if we choose a domain of estimation further away from the origin, e. g., $\mathbb{D} = [0.35, 10]$, the distortions observed in scenarios (ii-1) and (iv-1) vanish and the plots look similar to cases (i-1) and (iii-1), respectively, where Brownian motion was absent. A practicable method, therefore, is to estimate σ first, e. g., as presented in Mancini (2005). Then, assuming $\hat{\sigma} = \sigma$, we determine $\mathbb{D}$ such that the probability for purely Brownian increments to fall into $\mathbb{D}$ is very small.

Having said that, there exists another provision despite changing $\mathbb{D}$. Again for cases (ii) and (iv), we observe that the penalisation criterion chooses on average $m = 43.34$ and $m = 43.96$, respectively, in scenario (1), and $m = 3.50$ and $m = 25.58$ on average, respectively, in scenario (2). Although, the optimal m, that is, $m_n^\star$, increases with rate $T^{1/7}$ in these cases (cf. Proposition 3.5 of Figueroa-López (2009)), the estimated m chosen by penalisation, in fact, decreases from scenario (1) to (2). Obviously, the relatively large amount of purely Brownian increments just above the threshold of 0.25 causes the penalised contrast to favour large m in scenario (1). Since we partition the domain equidistantly, only a few increments remain for each partition cell where a jump of corresponding size occurred. This increases the variance of our estimator significantly. If we increase the constants $c_1, \ldots, c_4$ in our penalty, the influence of Brownian motion is decreased such that smaller m, that is, coarser partitions, resulting in a smaller empirical MSE are chosen. In summary, not only the right choice of the domain of estimation $\mathbb{D}$ but the right balance between $\mathbb{D}$ and the penalty constants $c_1, \ldots, c_4$ is crucial.

Acknowledgment

The authors wish to thank the referees and the associate editor for their detailed comments that helped to improve this paper significantly. It is also a pleasure to thank Jean Jacod for helpful discussions. The first author gratefully acknowledges support by the International Graduate School of Science and Engineering (IGSSE) of the Technische Universität München.

References

Aït-Sahalia, Y., and Jacod, J. (2007), "Volatility estimators for discretely sampled Lévy processes," *Ann. Statist.*, 35, 355–392.

Aït-Sahalia, Y., and Jacod, J. (2009), "Testing for jumps in a discretely observed process," *Ann. Statist.*, 37, 184–222.

Akritas, M.G. (1982), "Asymptotic theory for estimating the parameters of a Lévy process," *Ann. Inst. Statist. Math.*, 34, 259–280.

Barron, A.R., Birgé, L., and Massart, P. (1999), "Risk bounds for model selection via penalization," *Probab. Theory Relat. Fields*, 113, 301–413.

Basawa, I.V., and Brockwell, P.J. (1982), "Non-Parametric Estimation for Non-decreasing Lévy processes," *J. Roy. Statist. Soc. Ser. B*, 44, 262–269.

Belomestny, D., and Reiß, M. (2006), "Spectral calibration of exponential Lévy models," *Finance and Stochastics*, 10, 449–474, 10.1007/s00780-006-0021-5.

Billingsley, P. (1995), *Probability and Measure*, 3rd edition ed., New York: Wiley.

Birgé, L., and Massart, P. (1993), "Rates of convergence for minimum contrast estimators," *Probab. Theory Relat. Fields*, 97, 113–150.

Birgé, L., and Massart, P. (1997), "From Model Selection to adaptive estimation," in *Festschrift for Lucien LeCam: Research Papers in Probability and Statistics*, eds. D. Pollard, E. Torgersen and G. Yang, New York: Springer, pp. 55–87.

Birgé, L., and Massart, P. (1998), "Minimum contrast estimators on sieves: exponential bounds and rates of convergence," *Bernoulli*, 4, 329–375.

Brenner, S.C., and Scott, L.R. (1994), *The Mathematical Theory of Finite Element Methods*, New York: Springer.

Čencov, N.N. (1962), "Evaluation of an unknown distribution density from observations," *Soviet Math. Dokl.*, 3, 1559–1562.

Chen, S.X., Delaigle, A., and Hall, P. (2010), "Nonparametric estimation for a class of Lévy processes," *J. Econom.*, 157, 257–271.

Comte, F., and Genon-Catalot, V. (2009), "Nonparametric estimation for pure jump Lévy processes based on high frequency data," *Stoch. Proc. Appl.*, 119, 4088–4123.

Comte, F., and Genon-Catalot, V. (2010), "Nonparametric adaptive estimation for pure jump Lévy processes," *Ann. Inst. H. Poincaré Probab. Statist.*, 46, 595–617.

Comte, F., and Genon-Catalot, V. (2011), "Estimation for Lévy processes from high frequency data within a long time intervals," *Ann. Statist.*, 39, 803–837.

DeVore, R.A., and Lorentz, G.G. (1993), *Constructive Approximation*, Berlin: Springer.

van Es, B., Gugushvili, S., and Spreij, P. (2007), "A kernel type nonparametric density estimator for decompounding," *Bernoulli*, 13, 672–694.

Figueroa-López, J.E. (2009), "Non-parametric estimation for Lévy models based on discrete sampling," in *Optimality: The Third Erich L. Lehmann Symposium*, ed. J. Rojo, Vol. 57 of IMS Lecture Notes Monograph Series, Beachwood, OH: Institute of Mathematical Statistics, pp. 117–146.

Figueroa-López, J.E. (2011), "Sieve-based confidence intervals and bands for Lévy densities," *Bernoulli*, 17, 643–670.

Figueroa-López, J.E., and Houdré, C. (2006), "Risk bounds for the non-parametric estimation of Lévy processes," in *High Dimensional Probability: Proceedings of the Fourth International Conference*, eds. E. Giné, V. Kolchinskii, W. Li and J. Zinn, Vol. 51 of IMS Lecture Notes Monograph Series, Beachwood, OH: Institute of Mathematical Statistics, pp. 96–116.

Grenander, U. (1981), *Abstract Inference*, New York: Wiley.

Gugushvili, S. (2009a), "Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process," *J. Nonparametr. Statist.*, 21, 321–343.

Gugushvili, S., "Nonparametric inference for discretely sampled Lévy processes." arXiv:0908.3121v2 [math.ST] (2009b).

Gugushvili, S., Klaassen, C.A.J., and Spreij, P. (eds.), *Special Issue: Statistical Inference for Lévy Processes with Applications to Finance*, Vol. 64 of *Statistica Neerlandica*, pp. 255–366 (2010).

Jacod, J., Klüppelberg, C., and Müller, G. (2011), "Testing for non-correlation or for a functional relationship between price and volatility jumps.." Submitted for publication.

Jacod, J., and Todorov, V. (2010), "Do price and volatility jump together?," *Ann. Appl. Probab.*, 20, 1425–1469.

Kappus, J., and Reiß, M. (2010), "Estimation of the characteristics of a Lévy process observed at arbitrary frequency," *Statistica Neerlandica*, 64, 314–328.

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

Ledoux, M. (1997), "On Talagrand deviation inequalities for product measures," *ESAIM Probab. Statist.*, 1, 63–87.

Mancini, C. (2005), "Estimating the integrated volatility in stochastic volatility models with Lévy type jumps." Available at `http://www.cmap.polytechnique.fr/SEM/finance/2004-2005/mancini1.pdf`.

Massart, P. (2000), "About the constants in Talagrand's concentration inequalities for empirical processes," *Ann. Probab.*, 28, 863–884.

Mazja, W. (1979), *Einbettungssätze für Sobolewsche Räume (Teil 1)*, Leipzig: Teubner.

Neumann, M.H., and Reiß, M. (2009), "Nonparametric estimation for Lévy processes from low-frequency observations," *Bernoulli*, 15, 223–248.

Reynaud-Bouret, P. (2003), "Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities," *Probab. Theory Relat. Fields*, 126, 103–153.

Rubin, H., and Tucker, H.G. (1959), "Estimating the parameters of a differential process," *Ann. Math. Statist.*, 30, 641–658.

Rüschendorf, L., and Woerner, J.H.C. (2002), "Expansion of transition distributions of Lévy processes in small time," *Bernoulli*, 8, 81–96.

Sato, K., *Lévy Processes and Infinitely Divisible Distributions*, Cambridge: Cambridge University Press (1999).

Shimizu, Y. (2006), "Density estimation for Lévy measures for discretely observed diffusion processes with jumps," *J. Japan Statist. Soc.*, 36, 37–62.

Shimizu, Y. (2009), "Functional estimation for Lévy measures of semimartingales with Poissonian jumps," *J. Multivariate Anal.*, 100, 1073–1092.

Talagrand, M. (1996), "New concentration inequalities in product spaces," *Invent. Math.*, 126, 505–563.

Watteel, R.N., and Kulperger, R.J. (2003), "Nonparametric estimation of the canonical measure for infinitely divisible distributions," *J. Statist. Comput. Simulation*, 73, 525–542.

Whittaker, E.T. (1923), "On a new method of graduation," *Proc. Edinburgh Math. Soc.*, 2, 63–75.

Appendix A. Proofs

The proofs are based on the scheme of proofs of Figueroa-López and Houdré (2006) and Reynaud-Bouret (2003) given in the framework of estimating the intensity of infinite and finite Poisson random measures from explicit observation of the associated pure jump process, respectively. In turn, the latter credits Talagrand (1996), Ledoux (1997) and Massart (2000) for crucial ideas and results incorporated.

A.1. *Proof of Lemma 2.1*

Let $g \in \mathbb{S}_m$ and let $\{g_{m,k} : k = 1, \dots, d_m\}$ be an arbitrary orthonormal basis of $\mathbb{S}_m$. From classical algebra we know there exists a unique representation $g = \sum_{k=1}^{d_m} \alpha_k g_{m,k}$ with $\alpha \in \mathbb{R}^{d_m}$. We deduce that $\gamma^n(g) = -2\sum_{k=1}^{d_m} \alpha_k \hat{F}^n(g_{m,k}) + \sum_{k=1}^{d_m} \alpha_k^2$. Equivalently, we observe that

$$\gamma^n(g) = \sum_{k=1}^{d_m} (\alpha_k - \hat{F}^n(g_{m,k}))^2 - \sum_{k=1}^{d_m} (\hat{F}^n(g_{m,k}))^2.$$

Evidently, $\gamma^n(g)$ is minimal if, and only if, $\alpha_k = \hat{F}^n(g_{m,k})$ for all $k = 1, \dots, d_m$. Thus, $\hat{f}_m^n$ and $\arg\min_{g \in \mathbb{S}_m} \gamma^n(g)$ coincide and Equation (6) holds.

A.2. *Proof of Remark 1*

By Assumption 3.1, there exists a finite partition $\mathcal{D}_m$ of $\mathbb{D}$ so that for every $D \in \mathcal{D}_m$, $g_{|D} \in \mathcal{W}^{1,\infty}(D)$. Let $\{g_{m,k} : k = 1, \dots, d_m\}$ be an arbitrary orthonormal basis of $\mathbb{S}_m$. Then, $\mathfrak{D}_m = \|\sum_{k=1}^{d_m} g_{m,k}^2\|_{L^\infty(\mathbb{D})}$ by duality. As $\mathcal{W}^{1,\infty}(D) \subset L^\infty(D)$, $\|g_{m,k|D}^2\|_{L^\infty(D)} < \infty$ for each D, and $\sum_{k=1}^{d_m} \sum_{D \in \mathcal{D}_m} \|g_{m,k|D}^2\|_{L^\infty(D)}$ is a finite upper bound for $\mathfrak{D}_m$.

Let $x \in D$ and $i \in \{1, \dots, d\}$. Then, $\partial_i g_{m,k|D} \in L^\infty(D) \subset L^1(D)$ by assumption. Hence, $(g_{m,k} \circ \alpha_{D,x,i})' \in L^1([\underline{\xi}_i(x), \overline{\xi}_i(x)])$ and $x \mapsto \|(g_{m,k} \circ \alpha_{D,x,i})'\|_{L^1([\underline{\xi}_i(x), \overline{\xi}_i(x)])} \in L^\infty(D)$ by Fubini's theorem. Thus, $K_{m,k,D,i} := \sup_{x \in D} \|(g_{m,k} \circ \alpha_{D,x,i})'\|_{L^1([\underline{\xi}_i(x), \overline{\xi}_i(x)])} < \infty$ for each $k = 1, \dots, d_m$, $D \in \mathcal{D}_m$ and $i = 1, \dots, d$. Therefore, for arbitrary $g \in \mathbb{S}_m$,

$$\|g'\|_{\alpha, L^1(\mathcal{D}_m)} \leq \|g\|_{L^2(\mathbb{D})} \sum_{k=1}^{d_m} \sum_{D \in \mathcal{D}_m} \sum_{i=1}^{d} K_{m,k,D,i} < \infty.$$

Consequently, $\mathfrak{D}'_m < \infty$.

A.3. *Proof of Theorem 3.3*

We recall that there exists an $\varepsilon > 0$ with $\mathbb{D} \cap B_\varepsilon(0) = \emptyset$ and let $\eta \in]0, \varepsilon/2 \wedge 1[$. We decompose the Lévy process X as follows:

$$\begin{aligned} V_t &:= b_\eta t + \sigma W_t \qquad \text{with } b_\eta = b - \int_{\{x \in \mathbb{R}^d : \eta < \|x\| \leq 1\}} x F(\mathrm{d}x), \\ Y_t &:= \sum_{k=1}^{N_t} Z_k := \sum_{s \leq t} \Delta X_t \mathbb{1}_{]\eta, \infty[}(\|\Delta X_t\|), \qquad\qquad (t \geq 0) \\ R_t &:= X_t - V_t - Y_t, \end{aligned}$$

where V, Y and R are independent stochastic processes, W denotes a Wiener process, N denotes a Poisson process with rate $\lambda_\eta := F(\mathbb{R}^d_\circ \setminus B_\eta(0)) < \infty$, and $\{Z_k\}_{k \in \mathbb{N}}$ is an i. i. d. family of $\mathbb{R}^d$-valued random variables (independent of N) with law P_{Z_1} given by $P_{Z_1}(\mathrm{d}x) = \lambda_\eta^{-1} f(x) \mathbb{1}_{\mathbb{R}^d_\circ \setminus B_\eta(0)}(x) \mathrm{d}x$. We always assume $\lambda_\eta > 0$, since otherwise $f_{|\mathbb{D}} \equiv 0$ and the argument of proof would simplify considerably.

Recalling Equation (6) and Equation (7), we observe that $\gamma^n(\hat{f}^n_{\text{pen}}) + \text{pen}_n(m^{\text{pen}}_n) \leq \gamma^n(\mathcal{P}_m f) + \text{pen}_n(m)$ is satisfied for all $m \in M_n$. For all $g \in L^2(\mathbb{D})$ we denote $\upsilon^n(g) := \hat{F}^n(g) - F(g)$ and conclude that $\gamma^n(g) = \|f - g\|^2_{L^2(\mathbb{D})} - \|f\|^2_{L^2(\mathbb{D})} - 2\upsilon^n(g)$ holds. Therefore, we deduce in analogy to Lemma 7.1 of Figueroa-López and Houdré (2006) that

$$\begin{aligned} \|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} &\leq \|f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + 2\upsilon^n(\mathcal{P}_{\text{pen}} f - \mathcal{P}_m f) \\ &\quad + 2\|\mathcal{P}_{\text{pen}} f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} + \text{pen}_n(m) - \text{pen}_n(m^{\text{pen}}_n) \end{aligned} \tag{A1}$$

holds for every $m \in M_n$.

We derive further upper bounds for the right-hand side of Equation (A1) that hold on sets of large probability. At first, we focus on $\upsilon^n(g)$. Subsequently, we take $g = \mathcal{P}_{m'} f - \mathcal{P}_m f$ for $m' \in M_n$ arbitrary. For ease of notation, we introduce the (non-observable) auxiliary object $\tilde{F}^n(g) := T_n{}^{-1} \sum_{j=1}^{k_n} g(\Delta^n_j X) \mathbb{1}_{\{\Delta^n_j N = 1\}}$, where we abbreviate $k_n := T_n/\Delta_n$. We note that this quantity is well-defined as we extend $g \in L^2(\mathbb{D})$ to $\mathbb{R}^d$ by setting $g(x) = 0$ for all $x \notin \mathbb{D}$. Clearly, we have

$$\upsilon^n(g) \leq |\hat{F}^n(g) - \tilde{F}^n(g)| + |\tilde{F}^n(g) - E[\tilde{F}^n(g)]| + |E[\tilde{F}^n(g)] - F(g)|. \tag{A2}$$

We separately analyse the three summands on the right-hand side of Equation (A2).

i) By definition, the j-th summands of $\hat{F}^n(g)$ and $\tilde{F}^n(g)$ differ on the set $\Omega_j := \{\Delta^n_j X \in \mathbb{D}, \Delta^n_j N \neq 1\}$ only. For each $j = 1, \dots, k_n$, moreover, the difference is bounded by $\|g\|_{L^\infty(\mathbb{D})}$. Consequently, we observe that

$$|\hat{F}^n(g) - \tilde{F}^n(g)| \leq \|g\|_{L^\infty(\mathbb{D})} \Psi^n T_n^{-1}, \tag{A3}$$

where $\Psi^n := \sum_{j=1}^{k_n} \mathbb{1}_{\Omega_j}$.

ii) We apply Bernstein's inequality to $\tilde{F}^n(g)$. Particularly, we deduce for all $x > 0$ that

$$P(|\tilde{F}^n(g) - E[\tilde{F}^n(g)]| \geq x) \leq 2 \exp\left(-\frac{x^2 T_n}{2E[\tilde{F}^n(g^2)] + 2\|g\|_{L^\infty(\mathbb{D})} x/3} \right).$$

Subsequently, we invert the exponent on the right-hand side as a function of x and use the subadditivity of the root function. For arbitrary $a'_{m'} > 0$ we arrive at

$$P\left(|\tilde{F}^n(g) - E[\tilde{F}^n(g)]| \leq \sqrt{\frac{2a'_{m'} E[\tilde{F}^n(g^2)]}{T_n}} + \frac{a'_{m'} \|g\|_{L^\infty(\mathbb{D})}}{2T_n} \right) \geq 1 - 2e^{-a'_{m'}}. \tag{A4}$$

We also observe that

$$E[\tilde{F}^n(g^2)] = \Delta_n^{-1} E[g^2(\Delta_1^n X) \mathbb{1}_{\{\Delta_1^n N = 1\}}] = e^{-\lambda_\eta \Delta_n} \lambda_\eta E[g^2(\Delta_1^n V + \Delta_1^n R + Z_1)],$$

since N is a Poisson process independent of V, R and Z. Let h denote the density of the convolution of $P(V_{\Delta_n} + R_{\Delta_n} \in \cdot)$ with $F(\cdot \,|\, \mathbb{R}^d_\circ \setminus B_\eta(0))$. Then, we observe $\|h\|_{L^\infty(\mathbb{R}^d_\circ)} \le \|f\|_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))} < \infty$, where the finiteness follows as f is assumed to be bounded outside every neighbourhood of the origin. Thus,

$$E[\tilde{F}^n(g^2)] = e^{-\lambda_\eta \Delta_n} \int g^2(x) h(x) \mathrm{d}x \le \|f\|_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))} \|g\|^2_{L^2(\mathbb{D})}. \tag{A5}$$

We repeatedly apply the following equivalent inequalities that follow directly from the binomial identity. In particular, for all $x, y \in \mathbb{R}$ and arbitrary $a > 0$ we have

$$2xy \le a x^2 + a^{-1} y^2, \tag{A6}$$

$$(x+y)^2 \le (1+a)x^2 + (1+a^{-1})y^2. \tag{A7}$$

These inequalities can also be found in Figueroa-López and Houdré (2006, (7.10)). Invoking Equation (A6) for arbitrary $a_1 > 0$ we get

$$\sqrt{\frac{2a'_{m'} \|f\|_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))} \|g\|^2_{L^2(\mathbb{D})}}{T_n}} \le a_1 \|g\|^2_{L^2(\mathbb{D})} + \frac{a'_{m'} \|f\|_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))}}{2a_1 T_n}.$$

In summary, we derive that

$$|\tilde{F}^n(g) - E[\tilde{F}^n(g)]| \le a_1 \|g\|^2_{L^2(\mathbb{D})} + \frac{a'_{m'} \|f\|_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))}}{2a_1 T_n} + \frac{a'_{m'} \|g\|_{L^\infty(\mathbb{D})}}{2T_n} \tag{A8}$$

holds with probability greater than $1 - 2e^{-a'_{m'}}$.

iii) We note again that $E[\tilde{F}^n(g)] = \Delta_n^{-1} E[g(\Delta_1^n X) \mathbb{1}_{\{\Delta_1^n N = 1\}}]$. By virtue of Proposition 2.1 of Figueroa-López (2011) and a multivariate generalisation of Lemma 3.2 of Figueroa-López (2009) there exists a constant $\tilde{K}'_\eta < \infty$ such that

$$\sup_{n \in \mathbb{N}} \Delta_n^{-1} |E[\tilde{F}^n(g)] - F(g)| \le \left(\|g\|_{L^\infty(\mathbb{D})} + d \|g'\|_{\alpha, L^1(\mathcal{D}_m)} \right) \tilde{K}'_\eta. \tag{A9}$$

Since $\sup_{n \in \mathbb{N}} T_n \Delta_n < \infty$, there exists $K'_\eta := \tilde{K}'_\eta \sup_{n \in \mathbb{N}} T_n \Delta_n < \infty$ such that the same upper bound with $\tilde{K}'_\eta$ replaced by K'_η holds for $\sup_{n \in \mathbb{N}} T_n |E[\tilde{F}^n(g)] - F(g)|$.

The bounds on the right-hand side of Equation (A3), Equation (A8) and Equation (A9) are given in terms of $\|g\|_{L^\infty(\mathbb{D})}$ and $\|g'\|_{\alpha, L^1(\mathcal{D}_m)}$, where our interest is focused on $g = \mathcal{P}_{m'} f - \mathcal{P}_m f$ with $m, m' \in M_n$ arbitrary. Certainly, $\|\mathcal{P}_{m'} f - \mathcal{P}_m f\|_{L^\infty(\mathbb{D})} \le \|\mathcal{P}_{m'} f\|_{L^\infty(\mathbb{D})} + \|\mathcal{P}_m f\|_{L^\infty(\mathbb{D})}$. Furthermore, by definition, we have $\|\mathcal{P}_{m'} f\|_{L^\infty(\mathbb{D})} \le \sqrt{\mathfrak{D}_{m'}} \|\mathcal{P}_{m'} f\|_{L^2(\mathbb{D})}$ for all $m' \in M_n$. Due to Pythagoras' theorem, we have $\|\mathcal{P}_{m'} f\|_{L^2(\mathbb{D})} \le \|f\|_{L^2(\mathbb{D})}$ as well. Thus, applying Equation (A6) and Equation (A7) in analogy to Figueroa-López and Houdré (2006, p. 14), we deduce for arbitrary constants $a_2, a_3 > 0$ that

$$(a''_{m'} + \Psi^n) \|\mathcal{P}_{m'} f - \mathcal{P}_m f\|_{L^\infty(\mathbb{D})} \le \frac{a_2}{2} \mathfrak{D}_{m'} + \frac{a_3}{2} \mathfrak{D}_m + \frac{(a_2 + a_3)((a''_{m'})^2 + (\Psi^n)^2)}{a_2 a_3} \|f\|^2_{L^2(\mathbb{D})},$$

where $a''_{m'} = a'_{m'}/2 + K'_\eta$. For arbitrary $a_4, a_5 > 0$ replacing a_2 and a_3, respectively, we get an analogous inequality for $dK'_\eta \|(\mathcal{P}_{m'} f - \mathcal{P}_m f)'\|_{\alpha, L^1(\mathcal{D}_m)}$, where $\mathfrak{D}_{m'}$ and $\mathfrak{D}_m$ are replaced by $\mathfrak{D}'_{m'}$ and $\mathfrak{D}'_m$, respectively. We adapt the final step of Figueroa-López and Houdré (2006, p. 14). In particular, we fix $a'_{m'} = \tilde{a}_6 \sqrt{d_{m'}} \min(\|f\|^{-1}_{L^2(\mathbb{D})}, \|f\|^{-1}_{L^\infty(\mathbb{R}^d_\circ \setminus B_\eta(0))}) + \xi$ for an arbitrary $\tilde{a}_6 > 0$. Therefore, we deduce that for arbitrary $a_1, \ldots, a_6 > 0$ there exist a constant $K''_{\eta,a} < \infty$, a finite number $k_1 > 0$ (depending on f, ζ_1 and ζ_2 from Equation (9)), and a quadratic polynomial $h_1 : \xi \mapsto h_1(\xi)$ increasing on $[0, \infty[$ with $h_1(0) = 0$ such that

$$\begin{aligned} \|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} \le{}& \|f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + 2\|\mathcal{P}_{\text{pen}} f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} + 2a_1 \|\mathcal{P}_{\text{pen}} f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} \\ &+ \frac{1}{2T_n} \left(a_2 \mathfrak{D}_{m_n^{\text{pen}}} + a_3 \mathfrak{D}_m + a_4 \mathfrak{D}'_{m_n^{\text{pen}}} + a_5 \mathfrak{D}'_m + a_6 d_{m_n^{\text{pen}}} \right) \\ &+ \frac{(1 + (\Psi^n)^2) K''_{\eta,a} \|f\|^2_{L^2(\mathbb{D})}}{T_n} + \frac{h_1(\xi)}{T_n} + \text{pen}_n(m) - \text{pen}_n(m_n^{\text{pen}}). \end{aligned}$$

holds for all $m \in M_n$ with probability greater than $1 - k_1 e^{-\xi}$. Moreover, if we choose $a_1 \in]0, 1/4[$, then for an arbitrary $a_7 \geq 1 - 4a_1$ we observe that

$$\|f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + 2\|\mathcal{P}_{\text{pen}} f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} + 2a_1\|\mathcal{P}_{\text{pen}} f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} - \|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})}$$
$$\leq K^{(4)}\|f - \mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + (1 + a_7)\|\mathcal{P}_{\text{pen}} f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})} - K'''\|f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})}$$

holds in analogy to Figueroa-López and Houdré (2006, (7.17)) , where $K''' = 1 - 4a_1$ and $K^{(4)} = 1 + 4a_1$.

Next, we focus on $\|\mathcal{P}_{\text{pen}} f - \hat{f}^n_{\text{pen}}\|^2_{L^2(\mathbb{D})}$. For arbitrary $m' \in M_n$ we note that we have $\|\mathcal{P}_{m'} f - \hat{f}^n_{m'}\|^2_{L^2(\mathbb{D})} = \sum_{k=1}^{d_{m'}} (\upsilon^n(g_{m',k}))^2$. We recall Equation (A3), Equation (A8) and Equation (A9) and apply Equation (A7). Hence, we arrive at

$$\|\mathcal{P}_{m'} f - \hat{f}^n_{m'}\|^2_{L^2(\mathbb{D})} \leq 2\sum_{k=1}^{d_{m'}} \left|\tilde{F}^n(g_{m',k}) - E[\tilde{F}^n(g_{m',k})]\right|^2$$
$$+ 2\sum_{k=1}^{d_{m'}} \frac{1}{T_n^2}((K'_\eta + \Psi^n)\|g_{m',k}\|_{L^\infty(\mathbb{D})} + K'_\eta d\|g'_{m',k}\|_{\alpha,L^1(\mathcal{D}_m)})^2. \tag{A10}$$

For ease of notation, we denote the sums on the right-hand side of equation Equation (A10) by $S_1^{Equation\ (\text{A10})}$ and $S_2^{Equation\ (\text{A10})}$, respectively. We separately analyse $S_1^{Equation\ (\text{A10})}$ and $S_2^{Equation\ (\text{A10})}$.

a) By duality, we see that $S_1^{Equation\ (\text{A10})} = (\sup\{|\tilde{F}^n(g) - E[\tilde{F}^n(g)]| : g \in \mathbb{S}_{m'}, \|g\|_{L^2(\mathbb{D})} = 1\})^2$. Therefore, the concentration inequality established by Massart (2000, Theorem 1.3) can be applied. From Equation (A5) we conclude by virtue of the Cauchy–Schwarz inequality that $\sup\{\text{Var}[\tilde{F}^n(g)] : g \in \mathbb{S}_{m'}, \|g\|_{L^2(\mathbb{D})} = 1\} \leq \sqrt{\mathfrak{D}_{m'}}\|h\|_{L^2(\mathbb{D})}/T_n$. Then, for arbitrary $a'_{m'} > 0$ and $a_8 > 0$ we derive that

$$\sqrt{T_n S_1^{Equation\ (\text{A10})}} \leq (1 + a_8)\sqrt{\sum_{k=1}^{d_{m'}} E[\tilde{F}^n(g^2_{m',k})]}$$
$$+ \sqrt{8\sqrt{\mathfrak{D}_{m'}}\|h\|_{L^2(\mathbb{D})}a'_{m'}} + \left(3.5 + \frac{32}{a_8}\right)\sqrt{\frac{\mathfrak{D}_{m'}}{T_n}}a'_{m'}$$

holds for all $m' \in M_n$ with probability greater than $1 - \sum_{m' \in M_n} e^{-a'_{m'}}$. Moreover, Bernstein's inequality Equation (A4) implies for all $a_9 > 0$ and arbitrary $a''_{m'} > 0$ that

$$\frac{1}{1 + a_9}\sum_{k=1}^{d_{m'}} E[\tilde{F}^n(g^2_{m',k})] - \left(1 + \frac{1}{2a_9}\right)\frac{\mathfrak{D}_{m'}}{T_n}a''_{m'} \leq \sum_{k=1}^{d_{m'}} \tilde{F}^n(g^2_{m',k}) \leq \sum_{k=1}^{d_{m'}} \hat{F}^n(g^2_{m',k}) \tag{A11}$$

holds for all $m' \in M_n$ with probability greater than $1 - \sum_{m' \in M_n} e^{-a''_{m'}}$. The second inequality in Equation (A11) holds, since $\tilde{F}^n(g^2) \leq \hat{F}^n(g^2)$ for all $g \in L^2(\mathbb{D})$.

b) We invoke Equation (A6) and Equation (A7) repeatedly for arbitrary $a_{10}, a_{11}, a_{12} > 0$ and arrive at

$$S_2^{Equation\ (\text{A10})} \leq \frac{1}{2T_n^3}(a_{10}\mathfrak{D}^4_{m'} + a_{11}\mathfrak{D}'^4_{m'} + a_{12}d^4_{m'})$$
$$+ \frac{2(\Psi^n)^4}{T_n}\left(\frac{3}{a_{10}} + \frac{2}{a_{11}}\right) + \frac{8(K'_\eta)^4(1 + d^4)}{T_n}\left(\frac{6}{a_{10}} + \frac{2}{a_{11}} + \frac{1}{a_{12}}\right).$$

Certainly, $x + y \leq 2(x \vee y)$ holds for arbitrary $x, y > 0$. Adapting the final steps of Figueroa-López and Houdré (2006, p. 16 and (7.25)), we deduce that for arbitrary $a_1, \ldots, a_9 > 0$ there exist a constant $K^{(5)}_{\eta,a} < \infty$, a finite number $k_2 > 0$ (depending on f, ζ_1 and ζ_2 from Equation (9)), and a quadratic polynomial $h_2 : \xi \mapsto h_2(\xi)$

increasing on $[0,\infty[$ with $h_2(0)=0$ such that

$$\begin{aligned}
K'''\|f-\hat{f}^n_{\mathrm{pen}}\|^2_{L^2(\mathbb{D})} &\le K^{(4)}\|f-\mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + \frac{(1+a_1)}{T_n}\sum_{k=1}^{d_{m_n^{\mathrm{pen}}}} \hat{F}^n(g^2_{m_n^{\mathrm{pen}},k}) \\
&+ \left(\frac{a_2\mathfrak{D}_{m_n^{\mathrm{pen}}}}{T_n}\vee\frac{a_{10}\mathfrak{D}^4_{m_n^{\mathrm{pen}}}}{T_n^3}\right) + \left(\frac{a_4\mathfrak{D}'_{m_n^{\mathrm{pen}}}}{T_n}\vee\frac{a_{11}\mathfrak{D}'^4_{m_n^{\mathrm{pen}}}}{T_n^3}\right) + \left(\frac{a_6 d_{m_n^{\mathrm{pen}}}}{T_n}\vee\frac{a_{12}d^4_{m_n^{\mathrm{pen}}}}{T_n^3}\right) \\
&+ \frac{1}{2T_n}\left(a_3\mathfrak{D}_m + a_5\mathfrak{D}'_m\right) + \frac{1}{T_n}\left(1+(\Psi^n)^2+(\Psi^n)^4\right)K^{(5)}_{\eta,a}\|f\|^2_{L^2(\mathbb{D})} \qquad \text{(A12)} \\
&+ \frac{h_2(\xi)}{T_n} + \mathrm{pen}_n(m) - \mathrm{pen}_n(m_n^{\mathrm{pen}})
\end{aligned}$$

holds with probability greater than $1-k_2e^{-\xi}$.

In Equation (A12) we choose the (so-far) arbitrary constants $a_1, a_2, a_4, a_6, a_{10}, a_{11}$, and a_{12} appropriately depending on $c_1,\dots,c_4$ from Equation (10) to cancel "$-\mathrm{pen}_n(m_n^{\mathrm{pen}})$". Then, by Lemma 7.4 of Figueroa-López and Houdré (2006), and integration by parts, we deduce that there exists a $\delta>0$ such that

$$K'''E\|f-\hat{f}^n_{\mathrm{pen}}\|^2_{L^2(\mathbb{D})} \le K^{(4)}E\|f-\mathcal{P}_m f\|^2_{L^2(\mathbb{D})} + (1+\delta)E[\mathrm{pen}_n(m)] + \frac{K_n^{(6)}}{T_n},$$

where $K_n^{(6)} := k_2\int_0^\infty e^{-\xi}h_2(\xi)\mathrm{d}\xi + (1+E[(\Psi^n)^2]+E[(\Psi^n)^4])K^{(5)}_{\eta,a}\|f\|^2_{L^2(\mathbb{D})}$. We recall that $\Psi^n = \sum_{j=1}^{T_n/\Delta_n}\mathbb{1}_{\{\|\Delta_j^n V+\Delta_j^n R\|\ge\varepsilon,\Delta_j^n N=0\}\cup\{\Delta_j^n N\ge 2\}}$. By Corollary 3.2 of Rüschendorf and Woerner (2002), and by the choice of $\eta<\varepsilon/2$, we conclude that there is a finite constant $K_{\eta,\varepsilon}$ such that $P(\|\Delta_1^n V+\Delta_1^n R\|\ge\varepsilon)\le K_{\eta,\varepsilon}\Delta_n^2$. Additionally, we recall that $\Delta_j^n N$ has Poisson distribution with mean $\lambda_\eta\Delta_n$. Hence, $P(\Delta_j^n N\ge 2) = e^{-\lambda_\eta\Delta_n}\sum_{k=2}^\infty(\lambda_\eta\Delta_n)^k/k!$. Since increments of Lévy processes are independent and stationary, we deduce that Ψ^n has binomial distribution B_{k_n,ρ_n} with parameters $k_n = T_n/\Delta_n$ and $\rho_n \le K_{\eta,\varepsilon}\Delta_n^2 + \sum_{k=2}^\infty(\lambda_\eta\Delta_n)^k/k!$. Moreover, there exist polynomials h_3 and h_4 of order 2 and 4, respectively, such that $E[(\Psi^n)^2]\le h_3(k_n\rho_n)$ and $E[(\Psi^n)^4]\le h_4(k_n\rho_n)$ for all $n\in\mathbb{N}$. Since $\sup_{n\in\mathbb{N}}T_n\Delta_n<\infty$, by assumption, $\sup_{n\in\mathbb{N}}k_n\rho_n<\infty$. Therefore, $\sup_{n\in\mathbb{N}}K_n^{(6)}<\infty$. As $m\in M_n$ was arbitrary, Equation (11) follows.

A.4. *Proof of Corollary 3.4*

To show Equation (14) we need to prove that there exists a finite constant $K>0$ such that

$$KE\|\mathcal{P}_m f-\hat{f}^n_m\|^2_{L^2(\mathbb{D})} \ge E[\mathrm{pen}_n(m)]$$

for all $n\in\mathbb{N}$ and $m\in M_n = \{m\in M : \max(\mathfrak{D}_m^{3/2}, \mathfrak{D}_m'^{3/2}, d_m^{3/2})\le T_n\}$.

For arbitrary $m\in M_n$ we observe $E\|\mathcal{P}_m f-\hat{f}^n_m\|^2_{L^2(\mathbb{D})} = E[\sum_{k=1}^{d_m}(F(g_{m,k})-\hat{F}^n(g_{m,k}))^2]$ for every orthonormal basis $\{g_{m,k}: k=1,\dots,d_m\}$ of $\mathbb{S}_m$. By virtue of the binomial identity and the representation $\mathcal{P}_m f = \sum_{k=1}^{d_m}F(g_{m,k})g_{m,k}$, we arrive at

$$E\|\mathcal{P}_m f-\hat{f}^n_m\|^2_{L^2(\mathbb{D})} = F(\mathcal{P}_m f) - 2E[\hat{F}^n(\mathcal{P}_m f)] + E\left[\sum_{k=1}^{d_m}(\hat{F}^n(g_{m,k}))^2\right].$$

Since Lévy increments are independent and stationary, we obtain that

$$E\left[\sum_{k=1}^{d_m}(\hat{F}^n(g_{m,k}))^2\right] = \sum_{k=1}^{d_m}\frac{E[\hat{F}^n(g^2_{m,k})]}{T_n} + \left(1-\frac{\Delta_n}{T_n}\right)\sum_{k=1}^{d_m}(E[\hat{F}^n(g_{m,k})])^2.$$

From Equation (A9) we infer that there exists a linear functional $\kappa: L^2(\mathbb{D})\to\mathbb{R}$ such that $E[\hat{F}^n(g)] = F(g)+\kappa(g)\Delta_n$ for all $g\in L^2(\mathbb{D})$. Moreover, the operator norm of the restriction $\kappa_m := \kappa_{|\mathbb{S}_m}$ of κ to $\mathbb{S}_m$ satisfies $\|\kappa_m\|\le a\sqrt{\mathfrak{D}_m}+a'\sqrt{\mathfrak{D}'_m}$ for some constants $a, a'<\infty$ independent of m. We recall that $F(\mathcal{P}_m f) = \sum_{k=1}^{d_m}(F(g_{m,k}))^2$.

Therefore, we conclude

$$E\|\mathcal{P}_m f - \hat{f}_m^n\|_{L^2(\mathbb{D})}^2 = -F(\mathcal{P}_m f) - 2\kappa(\mathcal{P}_m f)\Delta_n + \sum_{k=1}^{d_m} \frac{E[\hat{F}^n(g_{m,k}^2)]}{T_n}$$
$$+ (1-\Delta_n T_n^{-1}) \sum_{k=1}^{d_m} (F(g_{m,k}) + \kappa(g_{m,k})\Delta_n)^2$$
$$= \sum_{k=1}^{d_m} \frac{E[\hat{F}^n(g_{m,k}^2)]}{T_n} + \sum_{k=1}^{d_m} \kappa^2(g_{m,k})\Delta_n^2$$
$$- \frac{\Delta_n}{T_n} \left(F(\mathcal{P}_m f) + 2\kappa(\mathcal{P}_m f)\Delta_n + \sum_{k=1}^{d_m} \kappa^2(g_{m,k})\Delta_n^2 \right).$$

In contrast, we recall that

$$E[\mathrm{pen}_n(m)] = c_1 T_n^{-1} \sum_{k=1}^{d_m} E[\hat{F}^n(g_{m,k}^2)] + T_n^{-1}(c_2\mathfrak{D}_m + c_3\mathfrak{D}_m' + c_4 d_m).$$

Subsequently, since $d_m^{3/2} \le T_n$ and $K' := \sup_{n\in\mathbb{N}} T_n\Delta_n < \infty$ implies $d_m\Delta_n \le K' d_m^{-1/2}$, we additionally observe that

$$\sum_{k=1}^{d_m} \kappa^2(g_{m,k})\Delta_n^2 \le d_m\Delta_n^2(a\sqrt{\mathfrak{D}_m} + a'\sqrt{\mathfrak{D}_m'})^2 \le K' d_m^{-1/2}\Delta_n(a\sqrt{\mathfrak{D}_m} + a'\sqrt{\mathfrak{D}_m'})^2.$$

Hence, $T_n(\sum_{k=1}^{d_m} \kappa^2(g_{m,k})\Delta_n^2)/(c_2\mathfrak{D}_m + c_3\mathfrak{D}_m' + c_4 d_m) = O(d_m^{-1/2})$. Additionally, by Phytagoras' theorem and Cauchy-Schwarz inequality, $F(\mathcal{P}_m f) \le F(\mathbb{D})\|f\|_{L^2(\mathbb{D})}$ independent of m. Moreover, $\kappa(\mathcal{P}_m f) = \sum_{k=1}^{d_m} F(g_{m,k})\kappa(g_{m,k})$. Thus, $\Delta_n\kappa(\mathcal{P}_m f)/(c_2\mathfrak{D}_m + c_3\mathfrak{D}_m' + c_4 d_m) = O(\Delta_n(\sqrt{\mathfrak{D}_m} + \sqrt{\mathfrak{D}_m'})) = O(\Delta_n^{2/3})$.

Consequently, if $K > c_1$, then $KE\|\mathcal{P}_m f - \hat{f}_m^n\|_{L^2(\mathbb{D})}^2 \ge E[\mathrm{pen}_n(m)]$ is equivalent to

$$\frac{\sum_{k=1}^{d_m} E[\hat{F}^n(g_{m,k}^2)]}{(c_2\mathfrak{D}_m + c_3\mathfrak{D}_m' + c_4 d_m)} \ge \frac{1 + O(d_m^{-1/2}) + O(\Delta_n)}{K - c_1}. \tag{A13}$$

Thus, Equation (13) is sufficient to ensure the existence of a finite constant K that satisfies Equation (A13) for all $n \in \mathbb{N}$ and $m \in M_n$.

A.5. *Proof of Proposition 3.6*

Let $E := [0,1]^d$ be the d-dimensional unit orthotope and for $k \in \mathbb{N}$ let $\mathscr{P}_{k-1}^d$ be given by Equation (15). We note that $\dim(\mathscr{P}_{k-1}^d) = \binom{d+k-1}{k-1}$. In Sections 3.5 and 3.6 of Brenner and Scott (1994) a basis $\mathscr{N}$ of the dual $(\mathscr{P}_{k-1}^d)^*$ of $\mathscr{P}_{k-1}^d$ is constructed inductively such that $\mathscr{N} \subseteq (\mathscr{C}^0(E))^*$. Then, $(E, \mathscr{P}_{k-1}^d, \mathscr{N})$ is a finite element (cf. Definition 3.1.1 of Brenner and Scott (1994)) satisfying the conditions of Theorem 4.4.4 of Brenner and Scott (1994).

We note that every d-dimensional interval is (geometrically speaking) a polyhedron. For every $m \in M$ and $D \in \mathcal{D}_m$, moreover, there exists a finite element $(D, \mathscr{P}_{k-1}^d(D), \mathscr{N}(D))$ which is affine-equivalent to $(E, \mathscr{P}_{k-1}^d, \mathscr{N})$ (cf. Definition 3.4.1 of Brenner and Scott (1994)). Hence, by virtue of Theorem 4.4.20 of Brenner and Scott (1994), there exists a $g \in \mathbb{S}_m^{k-1}$ such that for every $0 \le k' \le k$ there is a positive constant $c_{k,\rho} < \infty$ with

$$\sqrt{\sum_{D\in\mathcal{D}_m} \|f-g\|_{\mathcal{W}^{k',2}(D)}^2} \le c_{k,\rho}|f|_{\mathcal{W}^{k,2}(\mathbb{D})} m^{k'-k},$$

where we implicitly use Equation (16). The left-hand side reduces to $\|f-g\|_{L^2(\mathbb{D})}$ if $k' = 0$. Consequently, the validity of Equation (18) follows directly from the definition of the orthogonal projection, that is, the identity $\mathcal{P}_m f = \arg\min_{g\in\mathbb{S}_m} \|f-g\|_{L^2(\mathbb{D})}$.

A.6. *Proof of Remark* 2

For all $m \in M$ and $D \in \mathcal{D}_m$ Assumption 3.5 in combination with Equation (16) implies that the volume of D is bounded below and above by $\rho^{d-1}m^{-d}$ and m^{-d} times the volume $\mathrm{vol}(\mathbb{D})$ of $\mathbb{D}$, respectively. Therefore, the cardinality of $\mathcal{D}_m$ is bounded below and above by m^d and $m^d\rho^{1-d}$, respectively. Let $\{g_{1,l} : l = 1,\ldots,d_1\}$ be an orthonormal basis of $\mathscr{P}^d_{k-1}(\mathbb{D})$. Then, for every $D \in \mathcal{D}_m$ there exists a $y \in \mathbb{R}^d$ such that $\{g_{m,l} : l = 1,\ldots,d_1\}$ given by $g_{m,l}(x) := \mathrm{vol}(D)^{-1/2}\, g_{1,l}(m(x-y))$ is an orthonormal basis of $\mathscr{P}^d_{k-1}(D)$. Thus, we infer $m^{-d}d_m \in [d_1, d_1\rho^{1-d}]$ for all $m \in M$ from the cardinality of $\mathcal{D}_m$. Similarly, we infer $m^{-d}\mathfrak{D}_m \in [\mathfrak{D}_1, \mathfrak{D}_1\rho^{1-d}]$ for all $m \in M$ from the definition of $g_{m,l}$. Moreover, integration by substitution yields

$$\|(g_{m,k}\circ\alpha_{D,x,i})'\|^2_{L^1([\underline{\xi}_i(x),\overline{\xi}_i(x)])} = \frac{1}{\mathrm{vol}(D)}\|(g_{1,k}\circ\alpha_{\mathbb{D},m(x-y),i})'\|^2_{L^1([\underline{\xi}_i(m(x-y)),\overline{\xi}_i(m(x-y))])}.$$

Hence, $m^{-d}\mathfrak{D}'_m \in [\mathfrak{D}'_1, \mathfrak{D}'_1\rho^{1-d}]$ for all $m \in M$.

We recall that $E[\hat{F}^n(g)] = F(g) + \kappa(g)\Delta_n$ for all $g \in L^2(\mathbb{D})$ holds for the linear functional κ defined in the proof of Corollary 3.4. Since $g^2_{m,k}(X_{\Delta_n})$ is positive with non-zero probability, $E[\hat{F}^n(g^2_{m,k})] > 0$ for all k and m. W. l. o. g. we choose $g_{1,1} = (\int_{\mathbb{D}} \mathrm{d}x)^{-1/2}\mathbb{1}_{\mathbb{D}}(\cdot)$. Therefore, we conclude that $\sum_{k=1}^{d_m} g^2_{m,k} \ge (m^d/(\mathrm{vol}(\mathbb{D}))\mathbb{1}_{\mathbb{D}}(\cdot)$. Consequently, we arrive at

$$\sum_{k=1}^{d_m} E[\hat{F}^n(g^2_{m,k})] > \sum_{k=1}^{\#\mathcal{D}_m} E[\hat{F}^n(g^2_{m,k})] = \frac{m^d F(\mathbb{D})}{\mathrm{vol}(\mathbb{D})} + \sum_{k=1}^{\#\mathcal{D}_m}\kappa(g^2_{m,k})\Delta_n > 0.$$

By construction, $\|\sum_{k=1}^{\#\mathcal{D}_m} g^2_{m,k}\|_{L^2(\mathbb{D})} \le \rho^{1-d}m^d/\sqrt{\mathrm{vol}(\mathbb{D})}$. Analogously to the proof of Corollary 3.4, hence $|m^{-d}\sum_{k=1}^{\#\mathcal{D}_m}\kappa(g^2_{m,k})\Delta_n| = O(\Delta_n(\sqrt{\mathfrak{D}_m} + \sqrt{\mathfrak{D}'_m})) = O(\Delta_n^{2/3})$. Consequently, for every $\delta \in]0, F(\mathbb{D})/\mathrm{vol}(\mathbb{D})[$ the set

$$A_\delta := \left\{(n,m) \in \mathbb{N}^2 : m \in M_n, 0 < \frac{F(\mathbb{D})}{\mathrm{vol}(\mathbb{D})} + m^{-d}\sum_{k=1}^{\#\mathcal{D}_m}\kappa(g^2_{m,k})\Delta_n < \delta\right\}$$

is a finite subset of $\mathbb{N}^2$. Certainly,

$$\inf_{n\in\mathbb{N}}\ \inf_{m\in M_n,(n,m)\notin A_\delta} \frac{\sum_{k=1}^{d_m} E[\hat{F}^n(g^2_{m,k})]}{\mathfrak{D}_m + \mathfrak{D}'_m + d_m} > \frac{\delta\rho^{d-1}}{\mathfrak{D}_1 + \mathfrak{D}'_1 + d_1}.$$

In addition, the infimum over a finite set of strictly positive numbers is equal to the minimum and is strictly positive. Thus, Equation (13) holds as $\mathfrak{D}_1$, $\mathfrak{D}'_1$, d_1 and ρ are finite constants independent of m and n.

A.7. *Proof of Theorem 3.7*

We recall $\sup_{m\in\mathbb{N}} m^{-d}(d_m + \mathfrak{D}_m + \mathfrak{D}'_m) < \infty$ from the previous proof. Thus, Equation (10) implies $E[\mathrm{pen}_n(m)] \le K\|f\|_{L^\infty(\mathbb{R}^d_\circ\setminus B_\eta(0))}(m^d/T_n + m^{4d}/T_n^3)$ for a finite constant $K < \infty$ (independent of f), where $\eta > 0$ comes from the proof of Theorem 3.3. In combination with Equation (11) and Equation (18) we observe that there exist finite constants $K', K'' < \infty$ (monotone in $(\|f\|_{L^\infty(\mathbb{R}^d_\circ\setminus B_\eta(0))} \vee |f|_{\mathcal{W}^{k,2}(\mathbb{D})})$ and $\|f\|_{L^2(\mathbb{D})}$, respectively) such that

$$E\|f - \hat{f}^n_{\mathrm{pen}}\|^2_{L^2(\mathbb{D})} \le K' \inf_{m\in M_n}\left(m^{-2k} + \frac{m^d}{T_n} + \frac{m^{4d}}{T_n^3}\right) + \frac{K''}{T_n}.$$

Since $k > d/4$ by assumption, the order of the right-hand side is clearly minimised if $\limsup_{n\to\infty} m_n T_n^{-1/(2k+d)} < \infty$ and $\liminf_{n\to\infty} m_n T_n^{-1/(2k+d)} > 0$ as $n \to \infty$. Hence we proved Equation (19). Finally, the monotonicity of K' in $\|f\|_{L^\infty(\mathbb{R}^d_\circ\setminus B_\eta(0))} \vee |f|_{\mathcal{W}^{k,2}(\mathbb{D})}$ is sufficient for Equation (20) as $B(a_1, a_2)$ is compact in $\mathcal{W}^{k,2}(\mathbb{D})$.

A.8. *Proof of Remark* 3

We briefly outline how to adapt the proof of Theorem 4.1 of Figueroa-López (2009) from the univariate case. The underlying results, viz. Theorem 1.3 and equation (2.11) of Kutoyants (1998), are independent of dimension. We proceed to use the notation given on pp. 138–142 of Figueroa-López (2009).

Let $\alpha = k$ and let $\Theta_\alpha(L,\mathbb{D})$ denote the class of Lévy densities that belong to the Hölder space of $(k-1)$-times differentiable functions such that $g^{(k-1)}$ is Lipschitz continuous with Lipschitz constant $L < \infty$. In the definition of s_θ on p. 138, we replace the exponents of T, that is "$-\alpha/(2\alpha+1)$" and "$1/(2\alpha+1)$", by "$-\alpha/\gamma$" and "$1/\gamma$", respectively. Then, $s_\theta \in \Theta_\alpha(L,\mathbb{D})$ remains valid for all $|\theta| < \kappa^{-\alpha}$. We follow the proof on pp. 138–139 until the last two equations on p. 139 of Figueroa-López (2009), where the central limit theorem for Δ_T is shown. Integration by substitution in $\mathbb{R}^d$ with $u := \kappa^{-1}T^{1/\gamma}(x-x_0)$ yields $E|\Delta_T|^{2+\delta} = O(T^{1-\alpha(2+\delta)/\gamma-d/\gamma})$ for the Liapunov condition and $\mathrm{Var}[\Delta_T] = O(T^{1-2\alpha/\gamma-d/\gamma})$ as $T \to \infty$. Therefore, $\gamma = 2\alpha+d$ is necessary and sufficient to proceed the proof. All further conclusions drawn on pp.140–142 in the proofs of Theorem 4.1 and Corollary 4.2 of Figueroa-López (2009) remain valid in the multivariate setting if every exponent "$\ldots/(2\alpha+1)$" of T is replaced by "$\ldots/(2\alpha+d)$".

Finally, we note that, by Rademacher's theorem, every Lipschitz continuous function has a weak derivative bounded by L. Therefore, as $\mathbb{D}$ is compact, $g \in \Theta_k(L,\mathbb{D})$ implies that $g^{(k)} \in L^p(\mathbb{D})$ for all $p \geq 1$. Thus, $\Theta_\alpha(L,\mathbb{D}) \subseteq \mathcal{W}^{k,2}(\mathbb{D})$. Hence, in analogy to Remark 4.3 of Figueroa-López (2009), the rate $T_n^{-2k/(2k+d)}$ is proved to be minimax over the class of Lévy densities that belong to $\mathcal{W}^{k,2}(\mathbb{D})$.