

Technische Universität München

ZENTRUM MATHEMATIK

**Non-parametric estimation of Lévy
densities from observations on a
discrete time grid**

Diplomarbeit

von

Florian Alexander Johann Ueltzhöfer

Themensteller / Betreuer: Prof. Dr. Claudia Klüppelberg

Prof. Dr. Jean Jacod

Abgabetermin:

2. November 2009

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Brunnthal, den 2. November 2009

Acknowledgment

I would like to take the opportunity to thank all people who guided and accompanied me in the last months.

First and foremost, I want to express my gratitude to my supervisor and mentor, Prof. Dr. Claudia Klüppelberg. This thesis would not have been realised without her valuable advice, demanding expectations and full support. The exquisite working environment combined with many valuable consultations made the last months an enjoyable experience. Additionally, I want to thank her for acquainting me with Prof. Dr. Jean Jacod and making my trip to Paris possible.

I also want to thank Prof. Dr. Jean Jacod for taking his time during his stay in Munich in May and for welcoming me in Paris cordially. Working with him was always very inspiring and gave me a great impression how research can be. I am genuinely looking forward to continue working with Prof. Dr. Claudia Klüppelberg and Prof. Dr. Jean Jacod as a doctoral candidate and I do not doubt that I made the right decision to stay on at university.

I want to thank Peter Hepperger for helpful hints concerning the theory of “finite elements”. Sometimes, small things can have a high leverage.

I would also like to thank my friend Christina for her companionship during our whole student life and, especially, for sharing the experience of the last six months.

I must not forget to thank my parents for their unlimited support, not only in the last months but throughout my whole time as a student. Particularly, I wish to express my gratitude to my dad, who encouraged me to reject other opportunities and pursue my studies in university in the upcoming three years.

Last but never least, I send my deep gratefulness to my beloved girlfriend Ren Hui (人惠) for her unconditional love and patience. No matter if we spend our day together in Munich or are separated from each other by 10 000 km and more, I can always rely on her sincere support.

Contents

1	Introduction	1
2	Preliminaries	5
2.1	General notation	5
2.2	Lévy processes	6
2.3	Non-parametric projection estimation	8
3	Penalised projection estimation in the continuous time framework	15
3.1	Penalisation and projection space selection	15
3.2	Oracle inequality for penalised projection estimation	20
3.3	Estimation of smooth univariate Lévy densities	21
4	Projection estimation in the discrete time framework	27
4.1	Observation of Lévy processes on discrete grids	27
4.2	The critical mesh	29
5	Projection estimation away from zero	33
5.1	A uniform bound for $\Delta_\tau f$ (technical result)	33
5.2	Brownian motion with compound Poisson jumps	39
5.3	Brownian motion with α -stable jumps	41
5.4	General Lévy processes	50
6	Penalisation in the discrete time framework	59
6.1	Penalised projection estimation in DT	59
6.2	Oracle inequality for penalised projection estimation	60
6.3	Estimation of smooth multivariate Lévy densities	75
7	A simulation study	79
7.1	Brownian motion with compound Poisson jumps	79
7.2	Brownian motion with α -stable jumps	95

Contents

1 Introduction

The past months, in particular since July 2007, are more or less characterised by the worst worldwide financial and economic crisis since 1929. Comments on the causes and backgrounds were made from many different kind of people – politicians, economists and mathematicians – to just name a few. Headlines like “Did math formula cause financial crisis?” and “The formula that killed Wall Street” are found frequently. All these headlines refer to the *Gaussian copula* introduced in Li [27].

In a press release from 1st April 2009, the Deutsche Mathematiker-Vereinigung (DMV) represented by Bargel and Wenzel retorted the blame. In particular, Bargel and Wenzel [1] claims that many mathematical terms are misunderstood and that it is not enough to employ a qualified mathematician in a bank if he (or she) is ignored in the decision process of the management. ([1] [...], dass viele mathematische Begriffe falsch verstanden werden. Und es hilft [...] nicht, einen kompetenten Mathematiker in einer Bank zu beschäftigen [...], ihn bei wichtigen Vorstandentscheidungen aber außen vor zu lassen.”) However, Bargel and Wenzel [1] also explains that new models are formed using Lévy processes but that these are not established as a business standard yet. ([1] “Neue Modelle werden mittels sogenannter Lévy-Prozesse gestaltet, [...] Allerdings haben sie sich noch nicht als Standard in der Praxis etabliert.”)

A more vigorous appeal can be found in Szpiro [38] published two weeks earlier. On the one hand, Szpiro [38] explains the mechanism of the incorrect usage of Li’s formula in simple words. On the other hand, Paul Embrechts is recited who warned already in 2001 that the guileless usage of simple risk valuation could cause a crisis and even destabilize an economy . ([38] “[...] , dass die arglose Verwendung simpler Risikobeurteilungen eine Krise heraufbeschwören und sogar eine Wirtschaft destabilisieren könne [...]”)

Naturally, the howl for regulation of the financial markets is strong in these times and many discussions are held on the question of how to prevent the financial industry from repeating their mistakes. However, it is evident from the simple call for “better” models that many business men in the banking and finance sector and also politicians are still likely to confuse mathematical models and reality. Thus, the claim for *worst-case methods* (cf. Föllmer [21]), where one compares multiple types of models and acts in compliance with the worst case, should strengthen.

Stable models were first proposed by Mandelbrot [30]. However, these models are extreme in the sense that stable distributions (excluding the Gaussian) do not

1 Introduction

have a finite variance. The use of more general Lévy processes in mathematical finance goes back to Eberlein and Keller [13]. In the last decade, many subclasses of Lévy processes were introduced and investigated, e. g. the *normal inverse Gaussian* (cf. Barndorff-Nielsen [2]), the *CGMY* (cf. Carr, Geman, Madan and Yor [9]) and the *variance Gamma* model (cf. Carr, Madan and Chang [10]). Recently, Cariboni and Schoutens [8] investigated Lévy processes in credit risk.

Clearly, the models to be used can be quite versatile. However, we recall that, despite model risk, there is also calibration risk, which cannot be neglected. Taking different parametric models into account (like the ones mentioned above) and making decision in compliance with all of them, is certainly a huge leap forward. Nevertheless, including non-parametric (Lévy) models allows for additional so-called stylised effects. Taking such models into account as well, we diversify some of the risk of relying on a collection of models neglecting some effects all together. Figueroa-López and Houdré [18, 19] proposed a non-parametric estimation method for the *Lévy density* of a Lévy process based on the observation of the corresponding *Poisson random measure*. Figueroa-López [17] started to enhance this method to the case, where the Lévy process is observed on a discrete time grid only. Since high-frequency data is commonly available for most (liquid) assets, the approach seems to be promising to be used in mathematical finance. However, the frequency of observations to be made remains vague, and so does the effect of the presence of a Brownian motion on the quality of the estimate in a neighbourhood of the origin.

In the following, we investigate these issues. Let us describe the outline of the thesis. Chapter 2 is dedicated to recall the fundamentals of Lévy processes and to give a brief introduction into the (Hilbert space theory based) *projection estimation* method for Lévy densities. In Chapter 3 thereafter, we develop further methodology for the estimation in the *continuous time (CT) framework*, where we explicitly assume that we observe the Poisson random measure associated with a concerning Lévy process. Upon this assumption we describe a *penalisation method* used as a data-driven criterion to select an appropriate projection space for the estimation method introduced in Section 2.3. Moreover, we give convergence rates for the (combined) *penalised projection estimation* method in the case of estimating so-called Besov-type smooth Lévy densities.

In Chapter 4, we adapt the methods of projection estimation to the *discrete time (DT) framework*, where we assume that we observe the Lévy process at equidistant points in time only. Furthermore, we connect the mean squared errors of *projection estimation* in CT and DT and impose conditions on the observation grid such that the additional error from discretisation becomes asymptotically negligible. Then in Chapter 5, we analyse the validity of the conditions mentioned above in the case of estimating the restriction of the Lévy density to a Borel

set away from the origin. Firstly, we deal with the case of the composition of a Brownian motion with drift and compound Poisson jumps. Secondly, we deal with the case of the composition of a Brownian motion with drift and an α -stable jump component. Lastly, we analyse the general case of uni- and multivariate Lévy processes.

Chapter 6 is dedicated to the penalisation method in the discrete time framework. We develop an oracle inequality for the mean squared error of penalised projection estimation in DT and connect this with an approximation result for multivariate Lévy densities belonging to some *Sobolev space*. Thereupon, we come up with a similar result for the convergence rates as shown for the continuous time framework. Finally, Chapter 7 provides an explicit numerical study of the different cases, analysed theoretically before.

1 Introduction

2 Preliminaries

In the following, we introduce briefly some general notation used throughout this thesis and recall the fundamentals of Lévy processes. Thereafter, we present the non-parametric estimation method for Lévy densities based on Figueroa-López and Houdré [18, 19].

2.1 General notation

Throughout this thesis, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_0^+}, P)$ be a *filtered probability space* upon which all random variables and stochastic processes are assumed to be defined on. The filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_0^+}$ is assumed to satisfy the *usual conditions*, i. e. it is *right-continuous* and *complete*.

For an arbitrary (in our case normed) space \mathbb{X} , we denote its *Borel σ -field* by $\mathcal{B}(\mathbb{X})$ and write $\lambda_{\mathbb{X}}$ for the *Lebesgue measure* on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. The corresponding *Lebesgue spaces* of (equivalence classes of) functions that are q -th power integrable for some $q \in [1, \infty[$ or essentially bounded in the case of $q = \infty$, we denote by $L^q(\mathbb{X})$. We denote the norm in an L^q -space by $\|\cdot\|_q$. Additionally, in a finite dimensional real space, e. g. \mathbb{R}^d , we denote the *Euclidean norm* by $\|\cdot\|_2$ and the *maximum norm* by $\|\cdot\|_{\infty} := \max_{k=1, \dots, d} |\cdot|_k$ as well. Moreover, we abbreviate the Borel σ -field of the Euclidean space by $\mathcal{B}^d := \mathcal{B}(\mathbb{R}^d)$. Moreover, for a *left-hand limit*, or *limit from below*, we write $\lim_{s \nearrow t} := \lim_{s \rightarrow t; s < t}$ and for a *right-hand limit*, or *limit from above*, we write $\lim_{s \searrow t} := \lim_{s \rightarrow t; s > t}$.

For a random variable $X : \Omega \rightarrow \mathbb{R}^d$, we denote its *law under P* by $P_X := \mathcal{L}(X)$, its *(cumulative) distribution function* by $F_X := F_{P_X}$, its *tail distribution function* by $\bar{F}_X := 1 - F_X$ and its *characteristic function* by $\hat{F}_X := \hat{F}_{P_X}$. If P_X has a *(probability) density*, we denote it accordingly by $F'_X := F'_{P_X}$. If two random variables X_1, X_2 have the same distribution, we write $X_1 \stackrel{d}{=} X_2$. Due to its remarkably importance, we use special identifiers for the univariate *standard normal distribution* $\mathcal{N}(0, 1)$. In particular, φ denotes the standard normal density, Φ its (cumulative) distribution function and $\bar{\Phi}$ its tail distribution function.

We denote the set of *positive integers* by \mathbb{N} and the set of *non-negative integers* by \mathbb{N}_0 . Analogously, we set $\mathbb{R}^+ :=]0, \infty[$, $\mathbb{R}^- :=]-\infty, 0[$ and $\mathbb{R}_0^+ := [0, \infty[$. We refer to the *List of Notation* and the *List of Abbreviations* at the end of this thesis for further notation and abbreviations.

2.2 Lévy processes

The class of Lévy processes is central to the theory of stochastic processes. Below, we give a definition of Lévy processes that fits best into our framework and recall the close connection between the former and infinitely divisible distributions. Then we give expositions of the Lévy-Khintchine representation and the Lévy-Itô decomposition. We refer to the monographs Sato [35] and Bertoin [4] for further reading. An introduction to the use of Lévy processes in Finance is provided in Schoutens [37]. Additionally, the latter gives an at length illustration using real market data.

Definition 2.2.1 (Càdlàg paths) *A stochastic processes*

$$X : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}^d; (t, \omega) \mapsto X_t(\omega) = (X_{1,t}(\omega), \dots, X_{d,t}(\omega)) \quad (2.1)$$

is said to have càdlàg paths if

- (i) $\forall t \in \mathbb{R}_0^+ : \lim_{s \searrow t} X_s = X_t$ (continue à droite),
- (ii) $\forall t \in \mathbb{R}_0^+ : \exists X_{t-} := \lim_{s \nearrow t} X_s$ (limitée à gauche).

For every stochastic process X with càdlàg paths, we denote $\Delta X_t := X_t - X_{t-}$.

Definition 2.2.2 (Lévy process) *An \mathbb{R}^d -valued stochastic process X adapted to the filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_0^+}$ is called a Lévy process w. r. t. $(\mathcal{F}_t)_{t \in \mathbb{R}_0^+}$ if*

- (i) $X_0 = 0$ a. s.,
- (ii) $\forall s, t \in \mathbb{R}_0^+ : s \leq t \Rightarrow X_t - X_s$ is independent of \mathcal{F}_s ,
- (iii) $\forall r \in \mathbb{R}_0^+ \forall s, t \in \mathbb{R}_0^+ : X_{t+r} - X_t \stackrel{d}{=} X_{s+r} - X_s$,
- (iv) X has càdlàg paths.

Definition 2.2.3 (Infinitely divisible distributions) *Let Q be a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$, then Q is called infinitely divisible if*

$$\forall n \in \mathbb{N} : \exists Q_n : Q = Q_n^{*n}, \quad (2.2)$$

where Q_n is a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$ and $*$ denotes the convolution operator, i. e.

$$\forall B \in \mathcal{B}^d : Q_1 * Q_2(B) = \int_{\mathbb{R}^d} Q_1(B - x) Q_2(dx) \quad (2.3)$$

for probability measures Q_1, Q_2 on $(\mathbb{R}^d, \mathcal{B}^d)$.

The correspondence of Lévy processes and infinitely divisible distributions was one of the essential results of Paul Lévy (cf. Sato [35, Theorem 7.10]). On the one hand, let X be a Lévy process, then $Q_t := \mathcal{L}(X_t)$ is infinitely divisible for all $t \in \mathbb{R}_0^+$. On the other hand, let \tilde{Q} be an infinitely divisible distribution, then there is a Lévy process \tilde{X} with $\mathcal{L}(\tilde{X}_1) = \tilde{Q}$.

Lévy-Khintchine representation

In addition to the concept described above, an analysis of the infinitely divisible distributions' characteristic functions discloses another useful representation of all Lévy processes. We define the *punctured Euclidean space* by

$$(\mathbb{R}_\circ^d, \|\cdot\|) := (\mathbb{R}^d \setminus \{0\}, \|\cdot\|_2),$$

and denote its Borel σ -field by \mathcal{B}_\circ^d .

Definition 2.2.4 (Lévy measure) *A measure ν on $(\mathbb{R}_\circ^d, \mathcal{B}_\circ^d)$ is called a Lévy measure if*

$$\int_{\mathbb{R}_\circ^d} (\|x\|^2 \wedge 1) \nu(dx) < \infty. \quad (2.4)$$

Let Q be an infinitely divisible distribution on $(\mathbb{R}^d, \mathcal{B}^d)$. Then, by a well-known result from Lévy and Khintchine (cf. Sato [35, Theorem 8.1]), there exist unique parameters $\gamma \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ symmetric positive semi-definite and a unique Lévy measure ν on $(\mathbb{R}_\circ^d, \mathcal{B}_\circ^d)$ such that for all $z \in \mathbb{R}^d$

$$\hat{F}_Q(z) = \exp \left(i\langle \gamma, z \rangle - \frac{\langle z, \Sigma z \rangle}{2} + \int_{\mathbb{R}_\circ^d} (e^{i\langle z, x \rangle} - 1 - i\langle z, x \rangle \mathbb{1}_{[0,1]}(\|x\|)) \nu(dx) \right) \quad (2.5)$$

is the characteristic function of Q . Let (γ, Σ, ν) be as before, then there is a unique Lévy process X such that $E[e^{i\langle z, X_1 \rangle}]$ is given by (2.5). The triplet is called *Lévy-Khintchine (or generating) triplet* of X . The truncation point of the small jumps is chosen arbitrarily as equal to one.

Lévy-Itô decomposition

The so-called *Lévy-Itô decomposition* provides another insight into the interrelation between an infinitely divisible distribution, its generating triplet and the corresponding Lévy process.

Definition 2.2.5 (Poisson random measure) *Let χ be a σ -finite measure on the space $(\mathbb{R}_0^+ \times \mathbb{R}_\circ^d, \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_\circ^d))$. A mapping $J : \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_\circ^d) \times \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$ is called Poisson random measure (PRM) on $(\mathbb{R}_0^+ \times \mathbb{R}_\circ^d, \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_\circ^d))$ with intensity measure χ if*

2 Preliminaries

- (i) $J(B, \cdot) \sim \text{Poi}(\chi(B))$ for all $B \in \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_0^d)$,
- (ii) the random variables $J(B_1, \cdot), \dots, J(B_n, \cdot)$ are independent for pairwise disjoint sets $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_0^d)$,
- (iii) $J(\cdot, \omega)$ is a measure on $(\mathbb{R}_0^+ \times \mathbb{R}_0^d, \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_0^d))$ for every $\omega \in \Omega$.

Let X be a Lévy process with generating triplet (γ, Σ, ν) and $\Sigma^{1/2} \in \mathbb{R}^{d \times d}$ be the Cholesky triangle of Σ . Then, by a result from Lévy and Itô (cf. Sato [35, Theorem 19.2]), there exist an $\Omega' \subseteq \Omega$ with $P(\Omega') = 1$, a standard Brownian motion

$$W : \mathbb{R}_0^+ \times \Omega' \rightarrow \mathbb{R}^d; (t, \omega) \mapsto W_t(\omega) \quad (2.6)$$

and a Poisson random measure

$$J : \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_0^d) \times \Omega' \rightarrow \mathbb{R}; (B, \omega) \mapsto \#\{t \in \mathbb{R}_0^+ : (t, \Delta X_t(\omega)) \in B\} \quad (2.7)$$

on $(\mathbb{R}_0^+ \times \mathbb{R}_0^d, \mathcal{B}(\mathbb{R}_0^+ \times \mathbb{R}_0^d))$ with intensity measure $\lambda_{\mathbb{R}_0^+} \times \nu$ such that almost surely

$$\left. \begin{aligned} X_t = & \gamma t + \Sigma^{1/2} W_t \\ & + \left\{ \iint_{[0,t] \times \{x: \|x\| > 1\}} x J(ds, dx) + \lim_{\varepsilon \searrow 0} \iint_{[0,t] \times \{x: \varepsilon < \|x\| \leq 1\}} x (J(ds, dx) - ds\nu(dx)) \right\} \end{aligned} \right\} \quad (2.8)$$

where W is independent from $X - W$.

2.3 Non-parametric projection estimation

This section is dedicated to introduce the non-parametric estimation method for Lévy densities based on Figueroa-López [15] and Figueroa-López and Houdré [18]. The main background ideas come from Reynaud-Bouret [33] and Kutoyants [26], who considered the problem of estimating the intensity measure of Poisson random measures with finite activity and Barron, Birgé and Massart [3], Birgé and Massart [5] and Ibragimov and Hasminskii [22], who dealt with the problem of density estimation based on i. i. d. random variables.

Continuous time framework and assumptions

Let X be a Lévy process with generating triplet (γ, Σ, ν) . In the continuous time framework (CT), we always assume that we explicitly observe the PRM J given by (2.7).

Assumption 2.3.1 (Existence of a Lebesgue density)

Throughout, the Lévy measure ν is assumed to be absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}_\circ^d, \mathcal{B}_\circ^d)$. We denote its Radon-Nikodym derivative by $\nu' : \mathbb{R}_\circ^d \rightarrow \mathbb{R}_\circ^+$.

Below, we present the projection estimation approach from [18] for the Lévy density ν' restricted to a Borel set $\mathbb{D} \in \mathcal{B}_\circ^d$ called *domain of estimation*. The results shall be our motivation for transferring the methods from the continuous time framework, where estimation is based on the observation of the PRM J , to a framework, where we base our estimators solely on the observation of the increments of X on a discrete time grid. This will be done in Chapter 4. We start with a simple explicit example.

Example 2.3.2 (Brownian motion with compound Poisson jumps)

Let X be the composition of a standard Brownian motion W and a fully compensated compound Poisson process (independent of W) with rate $\lambda = 10$ and jump sizes $Z_k \sim \chi_4^2$. The Lévy-Itô decomposition simplifies to

$$X_t = W_t + \sum_{k=1}^{N_t} Z_k - \mathbb{E}[N_t] \mathbb{E}[Z_1],$$

where the Poisson process N , the Brownian motion W and the family $\{Z_k\}_{k \in \mathbb{N}}$ are independent. A sample path $X(\omega)$ on the time horizon $[0, 100]$ is shown on the left-hand side of Figure 2.1. Assume, we choose $\mathbb{D} =]0, 20]$. Then we construct an estimator for the restriction of ν' to \mathbb{D} using the information provided by

$$\{J(B) : B \in \mathcal{B}([0, 100] \times \mathbb{D})\}.$$

The right-hand side of Figure 2.1 shows the corresponding jumps $(t, \Delta X_t(\omega))$. The expected value for the number of jumps was $\mathbb{E}[N_{100}] = 1000$, whereas in this sample, $N_{100}(\omega) = 1006$ jumps occurred.

Apparently, the Lévy density may have a singularity at the origin in the light of Definition 2.2.4. Therefore, we have to find an appropriate but general enough space to allow estimation in a neighbourhood of the origin as well as away from it. In the following, estimation of the Lévy density is based on Hilbert space theory. Let μ be a known measure on $(\mathbb{R}_\circ^d, \mathcal{B}_\circ^d)$ equivalent to the Lebesgue measure on this space. Under Assumption 2.3.1, the Lévy measure is, again, absolutely continuous w. r. t. μ . We denote the corresponding Radon-Nikodym derivative by $p := p_\mu$.

Assumption 2.3.3 (Square-integrability of p) The density p is assumed to be bounded on the domain of estimation \mathbb{D} and to belong to $L^2(\mathbb{D}, \mu)$, i. e.

$$\|p\|_\mu^2 := \int_{\mathbb{D}} p^2(x) \mu(dx) < \infty. \tag{2.9}$$

Brownian motion with compound Poisson χ_4^2 -jumps

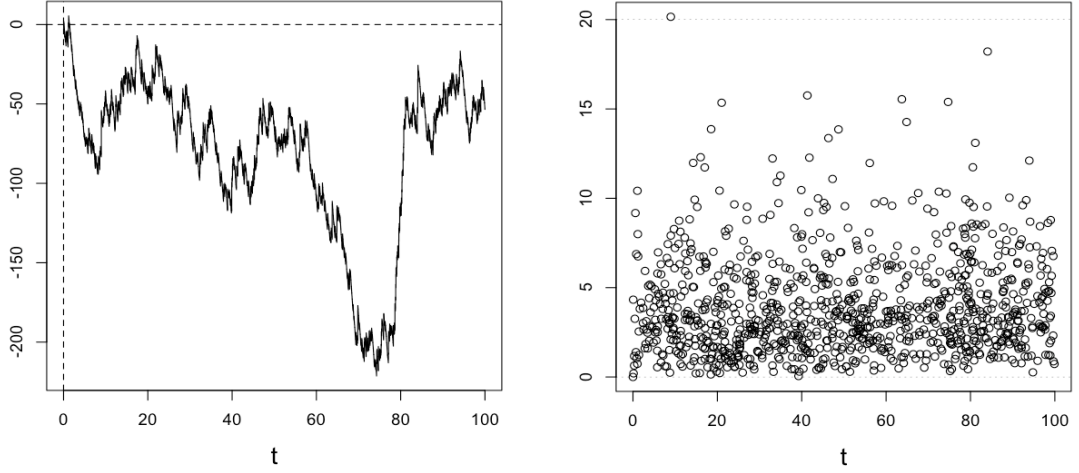


Figure 2.1: We present the sample path $X(\omega)$ of a Brownian motion plus fully compensated compound Poisson jumps coming from a χ_4^2 distribution (*left*) and the corresponding jumps $(t, \Delta X_t(\omega))$ (*right*).

In the case we choose \mathbb{D} away from zero, we let μ be the Lebesgue measure, in general. In this case, once the density is bounded on \mathbb{D} , Assumption 2.3.1 is sufficient for Assumption 2.3.3 in the light of Definition 2.2.4. Nevertheless, if we choose a different *reference measure* μ rather than the Lebesgue measure, then the appearing densities have a close interrelation. In particular, we have $\nu' = p\mu'$, where μ' denotes the Lebesgue density of μ . In the following, we make use of this density relation to derive an estimator for the Lebesgue density ν' . More precisely, given an estimator \hat{p} for the μ -density of ν , a natural estimator for the Lévy density is given by $\hat{\nu}' = \hat{p}\mu'$.

First, let us illustrate the effects of a non-Lebesgue reference measure using the instance of a univariate α -stable Lévy process.

Example 2.3.4 (Non-Lebesgue reference measure) *Let X be an \mathbb{R} -valued Lévy process with generating triplet (γ, σ^2, ν) and*

$$\nu(dx) = \frac{C_1}{x^{1+\alpha}} \mathbb{1}_{\mathbb{R}^+}(x)dx + \frac{C_2}{|x|^{1+\alpha}} \mathbb{1}_{\mathbb{R}^-}(x)dx$$

with $\alpha \in]0, 2[$ and $C_1, C_2 \in \mathbb{R}_0^+$ with $C_1 + C_2 > 0$. Further, let $\mathbb{D} \in \mathcal{B}_o$ be bounded and $\mu(dx) = |x|^{-2\alpha-2}dx$. Then the Lévy density w. r. t. the reference measure μ satisfies

$$p(x) = C_1 x^{\alpha+1} \mathbb{1}_{\mathbb{R}^+}(x) + C_2 |x|^{\alpha+1} \mathbb{1}_{\mathbb{R}^-}(x)$$

by virtue of the density identity $\nu' = p\mu'$. We conclude that

$$\int_{\mathbb{D}} p^2(x) \mu(dx) = \int_{\mathbb{D}} (C_1^2 \mathbb{1}_{\mathbb{R}^+}(x) + C_2^2 \mathbb{1}_{\mathbb{R}^-}(x)) dx < \infty.$$

Hilbert space projection

In the following, let $\mathbb{D} \in \mathcal{B}_\circ^d$ be a domain of estimation and μ a reference measure such that Assumption 2.3.3 holds. We recall that $L^2(\mathbb{D}, \mu)$ is a Hilbert space w. r. t. the inner product $\langle \cdot, \cdot \rangle_\mu$ derived from the norm $\|\cdot\|_\mu$ defined in (2.9). This is done by using the *polarisation identity* $2\langle u, v \rangle = \|u + v\|^2 - \|u\|^2 - \|v\|^2$, which defines a scalar product for every norm that satisfies the parallelogram law.

Definition 2.3.5 (μ -orthogonal projection) *Let $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ be a finite dimensional linear subspace. Then*

$$\mathcal{P}_{\mathbb{S}}p := \arg \min_{q \in \mathbb{S}} \|q - p\|_\mu. \quad (2.10)$$

is called the (μ -orthogonal) projection of p on \mathbb{S} .

The μ -orthogonal projection is the best approximation in \mathbb{S} . Another representation of the μ -orthogonal projection is derived from standard Hilbert space algebra. Let $\{f_k : k = 1, \dots, d_{\mathbb{S}}\}$ be a μ -orthonormal basis of \mathbb{S} . Then,

$$\mathcal{P}_{\mathbb{S}}p = \sum_{k=1}^{d_{\mathbb{S}}} \langle p, f_k \rangle_\mu f_k. \quad (2.11)$$

In general, it is necessary to know the μ -density p in order to calculate its projection. Hence, a good estimator for $\mathcal{P}_{\mathbb{S}}p$ is the most we can expect.

We denote the *integral of a function $f \in L^1(\mathbb{D}, \nu)$ w. r. t. the Lévy measure ν* by

$$\nu(f) := \int_{\mathbb{D}} f(x) \nu(dx). \quad (2.12)$$

We note that, by virtue of Definition 2.2.4, a sufficient condition for the integrability is, for instance, f being bounded, Borel measurable and vanishing in a neighbourhood of the origin.

Lemma 2.3.6 (Integral representation of “ $\langle \cdot, p \rangle_\mu$ ”) *The inner product of a function f and the μ -density p appearing in (2.11) is in fact an integral w. r. t. the Lévy measure ν , i. e. $\langle f, p \rangle_\mu = \nu(f)$.*

2 Preliminaries

Proof:

Clearly, we see that

$$\langle f, p \rangle_\mu = \int_{\mathbb{D}} f(x)p(x)\mu(dx) = \int_{\mathbb{D}} f(x)\nu(dx) = \nu(f)$$

holds for every $f \in L^2(\mathbb{D}, \mu)$. □

By virtue of Lemma 2.3.6, we rewrite (2.11). In particular, we see that

$$\mathcal{P}_{\mathbb{S}}p = \sum_{k=1}^{d_{\mathbb{S}}} \nu(f_k)f_k \quad (2.13)$$

holds for every μ -orthonormal basis (μ -ONB) $\{f_k : k = 1, \dots, d_{\mathbb{S}}\}$. Further, we recall that $\lambda_{\mathbb{R}^d} \times \nu$ is the intensity measure of the PRM J , motivating the following definition.

Definition 2.3.7 (Projection estimator in the CT framework)

Let $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ be a finite dimensional linear projection space and $T > 0$. Furthermore, let $\{f_k : k = 1, \dots, d_{\mathbb{S}}\}$ be a μ -ONB of \mathbb{S} , then for every $k = 1, \dots, d_{\mathbb{S}}$

$$\hat{\nu}(f_k) := \hat{\nu}(f_k; T) := \frac{1}{T} \iint_{[0, T] \times \mathbb{D}} f_k(x)J(dt, dx) \quad (2.14)$$

is called the integral estimator (in CT) for $\nu(f_k)$, and

$$\hat{p}_{\mathbb{S}} := \hat{p}(\cdot; T, \mathbb{S}) := \sum_{k=1}^{d_{\mathbb{S}}} \hat{\nu}(f_k)f_k(\cdot) \quad (2.15)$$

is called the projection estimator (in CT) for p on \mathbb{S} based on the observation of the Lévy process X – and the PRM J – on the time horizon $[0, T]$.

Remark 2.3.8 (Well-definedness of the projection estimator)

Figueroa-López and Houdré [18, Remark 2.2] shows that (2.15) does not depend on the choice of a particular μ -ONB of \mathbb{S} . This is proved by showing that the projection estimator is equivalent to a minimum contrast estimator. The concept is used equivalently in Birgé and Massart [5] and is based on classical Hilbert space theory only.

Proposition 2.3.9 (Unbiased integral estimation) *The integral estimator in CT $\hat{\nu}(\cdot)$ for $\nu(\cdot)$ defined in (2.14) is unbiased.*

Proof:

By definition of the PRM J in (2.7), we conclude that

$$\mathbb{E} [\hat{\nu}(f)] = \mathbb{E} \left[T^{-1} \iint_{[0,T] \times \mathbb{D}} f(x) J(dt, dx) \right] = \int_{\mathbb{D}} f(x) p(x) \mu(dx) = \nu(f) \quad (2.16)$$

holds for every $f \in \mathbb{S}$. □

Example 2.3.10 (Continuation of Example 2.3.2)

Figure 2.2 shows the empirical histogram of the jumps of the sample path $X(\omega)$ from Example 2.3.2. The bars are scaled such that the area underneath equals the empirical intensity of the Poisson random measure, i. e. $J([0, T] \times \mathbb{R}_o^d, \omega)/T = 10.06$ in our sample.

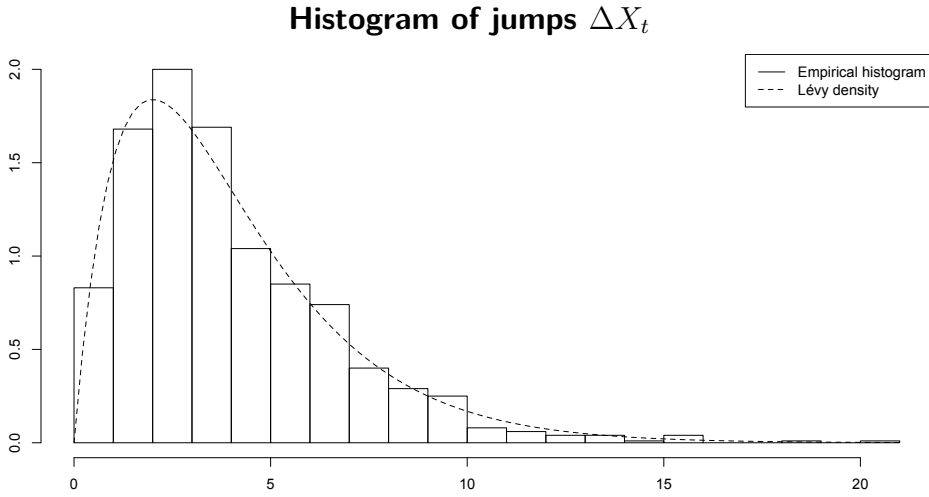


Figure 2.2: An empirical histogram of the jumps $\Delta X_t(\omega)$ occurred on the time horizon $[0, 100]$ as shown on the right-hand side of Figure 2.1 is presented. The bars are scaled such that the area underneath equals the empirical intensity $N_T(\omega)/T = 10.06$ (solid). Moreover, we present the Lévy density of $\nu(\cdot) = 10\chi_4^2(\cdot)$ of X (dashed).

We consider projection spaces $\mathbb{S}_1, \mathbb{S}_2 \subseteq L^2(\mathbb{D}, \mu)$ to illustrate the procedure of projection estimation. We recall that we have chosen $\mathbb{D} =]0, 20]$ as domain of estimation. We choose μ to be identical to the Lebesgue measure on \mathbb{D} . Since the density of the χ_4^2 distribution is bounded, finite and square-integrable, we observe that Assumption 2.3.3 is satisfied. Moreover, we choose \mathbb{S}_1 and \mathbb{S}_2 such that they have same dimension, e. g. $d_1 = d_2 = 10$. To construct \mathbb{S}_1 , let $D_l :=]2(l-1), 2l]$ for all $l = 1, \dots, 10$. Then let \mathbb{S}_1 be the projection space consisting of all functions $f : \mathbb{D} \rightarrow \mathbb{R}$, where the restriction $f|_{D_l}$ of f to every D_l is constant. An ONB

2 Preliminaries

of \mathbb{S}_1 is given by $\{1/\sqrt{2}\mathbb{1}_{D_l}(\cdot) : l = 1, \dots, 10\}$. Additionally, let \mathbb{S}_2 be the projection space consisting of all polynomials $f : \mathbb{D} \rightarrow \mathbb{R}$ with maximum degree 9. An ONB for \mathbb{S}_2 is given by the translated and scaled Legendre polynomials on \mathbb{D} .

Projection estimation in continuous time framework

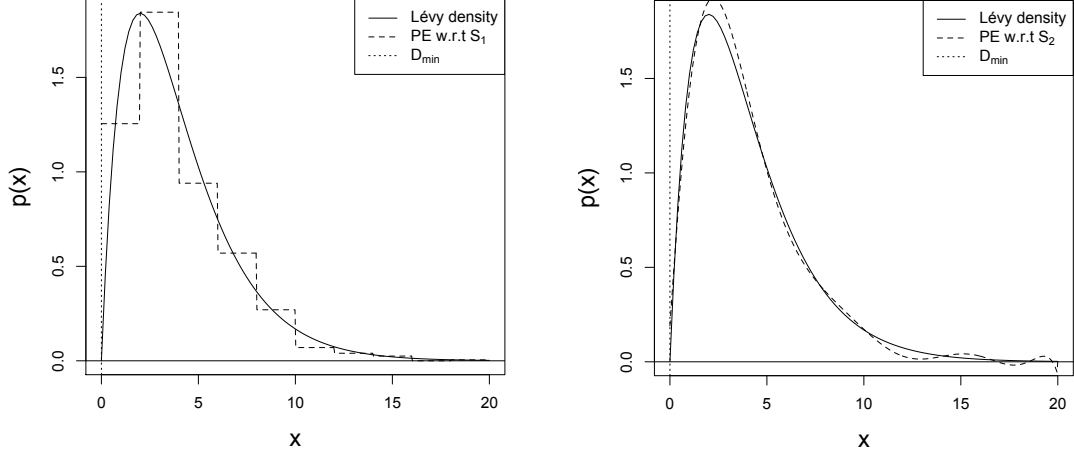


Figure 2.3: We present the projection estimate (*dashed*) w. r. t. the projection space \mathbb{S}_1 consisting of all piecewise constant functions on the partition $(D_l)_{l=1, \dots, 10}$ of \mathbb{D} (*left*) and w. r. t. \mathbb{S}_2 consisting of all polynomials on \mathbb{D} with maximum degree 9 (*right*) for the restriction of the Lévy density ν' of X to $\mathbb{D} =]0, 20]$, where $\nu = 10\chi_4^2$ (*solid*). The estimates are based on the jumps $\Delta X_t(\omega)$ occurred on the time horizon $[0, 100]$ as shown in Figure 2.2.

The results for our sample are shown in Figure 2.3 and Table 2.1. An advantage of $\hat{p}_{\mathbb{S}_2}$ is its intrinsically given smoothness compared to the discontinuous step function estimator $\hat{p}_{\mathbb{S}_1}$. Additionally, Table 2.1 shows that its squared error $\|p - \hat{p}_{\mathbb{S}_2}\|_{\mu}^2$ is much smaller than $\|p - \hat{p}_{\mathbb{S}_1}\|_{\mu}^2$. However, a major structural disadvantage is that $\hat{p}_{\mathbb{S}_2}(x)$ becomes negative for some $x \in \mathbb{D}$.

	SE	AE
\mathbb{S}_1	0.8008	2.087
\mathbb{S}_2	0.04279	0.673

Table 2.1: We present the empirical squared error $\|p - \hat{p}_{\mathbb{S}}\|_{\mu}^2$ (*SE*) and the empirical absolute error $\|p - \hat{p}_{\mathbb{S}}\|_{L^1(\mathbb{D}, \mu)}$ (*AE*) for the projection estimators based on \mathbb{S}_1 and \mathbb{S}_2 from Example 2.3.10.

In summary, we suggested a projection estimator for the restriction $p|_{\mathbb{D}}$ of the μ -density p to \mathbb{D} . From Table 2.1, we also see that an adequate choice for the projection space $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ is crucial for this method.

3 Penalised projection estimation in the continuous time framework

In this chapter, we develop further methodology for the projection estimation method introduced in Section 2.3. In Section 3.1, we decompose the mean squared error of projection estimation using Hilbert space theory to derive a data-driven criterion to select a projection space from a family of finite dimensional linear subspaces of $L^2(\mathbb{D}, \mu)$. In Section 3.2, we analyse the quality of this method using results from Reynaud-Bouret [33] and Figueroa-López and Houdré [19, Chapter 4]. Finally, Section 3.3 is dedicated to the approximation properties of certain families of projection spaces using results from DeVore and Lorentz [12].

Throughout, let X be a Lévy process with generating triplet (γ, Σ, ν) satisfying Assumption 2.3.1. Further, let $\mathbb{D} \in \mathcal{B}_0^d$ be a domain of estimation and μ a reference measure such that Assumption 2.3.3 holds. We remain in the framework of estimating the restriction $p|_{\mathbb{D}}$ of the μ -density p to \mathbb{D} based on the explicit observation of the PRM J .

3.1 Penalisation and projection space selection

So far, we did not address the issue of how to select an appropriate projection space $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$. Example 2.3.10 showed that the quality of the projection estimator is directly connected to the choice of \mathbb{S} . By Definition 2.3.5, from the Pythagoras identity follows that the mean squared error satisfies

$$\mathbb{E} \|p - \hat{p}\|_{\mu}^2 = \|p - \mathcal{P}_{\mathbb{S}}p\|_{\mu}^2 + \mathbb{E} \|\mathcal{P}_{\mathbb{S}}p - \hat{p}\|_{\mu}^2. \quad (3.1)$$

We call the first term on the right-hand side of (3.1) *squared μ -bias* and the second term *μ -variance* of projection estimation. Certainly, with respect to another projection space $\mathbb{S}' \subseteq L^2(\mathbb{D}, \mu)$ such that $\mathbb{S} \subseteq \mathbb{S}'$, we observe a decrease in the μ -bias, i. e.

$$\|p - \mathcal{P}_{\mathbb{S}'}p\|_{\mu} \leq \|p - \mathcal{P}_{\mathbb{S}}p\|_{\mu}.$$

However, we will see that, simultaneously, the μ -variance increases.

Families of projection spaces

In the following, we use the auxiliary set M as a way to enumerate our projection spaces. Let $\{\mathbb{S}_m : m \in M\}$ be a collection of finite dimensional linear subspaces of the Hilbert space $L^2(\mathbb{D}, \mu)$. For $m \in M$, we set

$$D_m := \sup\{\|f\|_\infty^2 : f \in \mathbb{S}_m, \|f\|_\mu^2 = 1\} \quad (3.2)$$

and

$$d_m := \dim \mathbb{S}_m \quad (3.3)$$

and let $\{f_{m,k} : k = 1, \dots, d_m\}$ be a μ -ONB of \mathbb{S}_m . Figueroa-López and Houdré [18, Remark 3.2] proves that $D_m = \|\sum_{k=1}^{d_m} f_{m,k}^2\|_\infty$ holds for every μ -ONB of \mathbb{S}_m . To avoid tedious notation, we set

$$\hat{p}_m := \hat{p}(\cdot; T, \mathbb{S}_m) \text{ and } \mathcal{P}_m p := \mathcal{P}_{\mathbb{S}_m} p.$$

Assumption 3.1.1 (Polynomial family of projection spaces)

For the sake of simplicity, we assume throughout that our collection of projection spaces $\{\mathbb{S}_m : m \in M\}$ is polynomial, i. e. there exist $a > 0$ and $b \geq 0$ such that

$$\forall n \in \mathbb{N} : \#\{m \in M : d_m = n\} \leq an^b. \quad (3.4)$$

Example 3.1.2 (Piecewise polynomials) Let $\mathbb{D} \subseteq \mathbb{R}^+$ be a compact interval and $M = \mathbb{N}$. For $k \in \mathbb{N}$, a possible collection of projection spaces is given by the family $\{\mathbb{S}_m : m \in M\}$, where \mathbb{S}_m denotes the space of piecewise polynomials of maximum degree k based on the partition of \mathbb{D} into pairwise disjoint intervals D_1, \dots, D_m , where $\mu(D_l) = \mu(\mathbb{D})/m$ for all $l = 1, \dots, m$. In the following, we call the partition $\mathcal{D}_m = \{D_l : l = 1, \dots, m\}$ the regular partition of \mathbb{D} into m classes.

Moreover, since $d_m = m(k+1)$, we observe that for every $n \in \mathbb{N}$ there is at most one $m \in M$ with $d_m = n$. Hence, the described collection is polynomial.

Mean squared error of projection estimation

In an optimal way, we aim for selecting that space out of our collection such that the projection estimator w. r. t. the chosen space minimizes the mean squared error (MSE). For $T > 0$, we restrict ourselves to the subfamily $\{\mathbb{S}_m : m \in M_T\}$, where

$$M_T := \{m \in M : D_m \leq T\}.$$

3.1 Penalisation and projection space selection

Definition 3.1.3 (Oracle space and estimator) *Let $T > 0$ and M_T as above. Then the space $\mathbb{S}_\star := \mathbb{S}_{m_\star}$ is called oracle space, where*

$$m_T^\star := \arg \min_{m \in M_T} \mathbb{E} \|p - \hat{p}_m\|_\mu^2, \quad (3.5)$$

and $\hat{p}_\star := \hat{p}(\cdot; T, \mathbb{S}_\star)$ is called oracle.

It is obviously not feasible to determine m_T^\star without prior knowledge of the density p . Therefore, we aim for establishing a method to dynamically select a “good” projection space using no more than the observation of the sample path $\{X_t(\omega) : t \in [0, T]\}$ – and explicitly $\{J(B, \omega) : B \in \mathcal{B}([0, T] \times \mathbb{D})\}$ – on the time horizon $[0, T]$. Therein, we estimate the μ -density p using the projection estimator from Definition 2.3.7. In a first step, we give a deeper analysis of (3.1) and derive an observable statistic that is an unbiased estimator for the MSE.

Proposition 3.1.4 (Mean squared error of integral estimation)

The mean squared error of the integral estimator in CT $\hat{\nu}(\cdot)$ for $\nu(\cdot)$ defined in (2.14) is given by

$$\text{m. s. e.} [\hat{\nu}(f)] = \frac{1}{T} \nu(f^2). \quad (3.6)$$

Proof:

We showed in Proposition 2.3.9 that $\hat{\nu}(f)$ is unbiased. Hence, the MSE is determined by its variance only. From Sato [35, Proposition 19.5] follows (3.6). \square

Proposition 3.1.5 (Mean squared error of projection estimation)

The mean squared error of the projection estimator in CT \hat{p}_m for the μ -density p is given by

$$\text{m. s. e.} [\hat{p}_m] = \|p\|_\mu^2 - \mathbb{E} \|\hat{p}_m\|_\mu^2 + \frac{2}{T} \sum_{k=1}^{d_m} \nu(f_{m,k}^2). \quad (3.7)$$

Proof:

First, we recall decomposition (3.1) of the MSE into squared μ -bias and μ -variance. By the Phytagoras identity, again, we conclude

$$\|p - \mathcal{P}_m p\|_\mu^2 = \|p\|_\mu^2 - \|\mathcal{P}_m p\|_\mu^2 \quad (3.8)$$

for the μ -bias term. Finally, since $\mathbb{E}[\hat{p}_m] = \mathcal{P}_m p$, we find that

$$-\|\mathcal{P}_m p\|_\mu^2 = \mathbb{E} \|\hat{p}_m - \mathcal{P}_m p\|_\mu^2 - \mathbb{E} \|\hat{p}_m\|_\mu^2. \quad (3.9)$$

Thus, (3.7) follows directly from the combination of (3.1), (3.8) and (3.9), since Proposition 3.1.4 implies $\mathbb{E} \|\mathcal{P}_{\mathbb{S}_m} p - \hat{p}\|_\mu^2 = T^{-1} \sum_{k=1}^{d_m} \nu(f_{m,k}^2)$. \square

Corollary 3.1.6 (Mean squared error of integral estimation)

The mean squared error of the projection estimator in CT \hat{p}_m for the μ -density p is given by

$$\text{m. s. e.}[\hat{p}_m] = \|p\|_\mu^2 + \mathbb{E} \left[-\|\hat{p}_m\|_\mu^2 + \frac{2}{T} \sum_{k=1}^{d_m} \hat{\nu}(f_k^2) \right]. \quad (3.10)$$

Proof:

The identity $\mathbb{E}[\hat{\nu}(f)] = \nu(f)$ from Proposition 2.3.9, valid for all $f \in L^1(\mathbb{D}, \mu)$, connects (3.10) directly to (3.7). \square

Penalised projection estimation

We see that the mean squared error of projection estimation is an affine function of the expectation of the observable statistic

$$-\|\hat{p}_m\|_\mu^2 + \frac{2}{T} \sum_{k=1}^{d_m} \hat{\nu}(f_k^2). \quad (3.11)$$

This enables us to set up a data-driven criterion, commonly known as *model selection via penalisation*, to select a projection space. Figueroa-López adapted this methodology mainly from Barron et al. [3] and Reynaud-Bouret [33]. However, from the point of view of a statistician, it is favourable to keep the complexity of a selected model small. It will turn out to be useful, to explicitly include D_m and d_m into the penalty for this purpose, motivating us to use general penalties instead of restricting ourselves to (3.11).

Definition 3.1.7 (Penalty) *A random variable*

$$\text{pen} : \mathbb{R}_0^+ \times \Omega \rightarrow (\mathbb{R}_0^+)^M; (T, \omega) \mapsto (\text{pen}_T(m, \omega))_{m \in M}$$

is called penalty on $\{\mathbb{S}_m : m \in M\}$ if it is adapted to the filtration generated by the Lévy process X .

In the continuous time framework, the penalties used are always of form

$$\text{pen}_T(m) = \frac{c_1}{T} \sum_{k=1}^{d_m} \hat{\nu}(f_{m,k}^2) + \frac{c_2 D_m}{T} + \frac{c_3 d_m}{T} \quad (3.12)$$

for some finite constants $c_1 > 1$ and $c_2, c_3 \geq 0$. The case of $c_1 = 2$ and $c_2 = c_3 = 0$ corresponds to (3.11). However, as mentioned before, in some settings it is favourable to impose $c_2, c_3 > 0$.

Definition 3.1.8 (Penalised projection estimator in CT) Let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$. Then $\mathbb{S}_{\text{pen}} := \mathbb{S}_{m_T^{\text{pen}}}$ is called penalised projection space, where

$$m_T^{\text{pen}} := \arg \min_{m \in M_T} \left\{ -\|\hat{p}_m\|_{\mu}^2 + \text{pen}_T(m) \right\}, \quad (3.13)$$

and $\hat{p}_{\text{pen}} := \hat{p}(\cdot; T, \mathbb{S}_{\text{pen}})$ is called penalised projection estimator (PPE) w. r. t. pen .

Example 3.1.9 (Continuation of Example 2.3.10)

We illustrate the effect of penalisation for our sample path $X(\omega)$ from Examples 2.3.2 and 2.3.10. Let $\{\mathbb{S}_m : m \in M\}$ be the family of piecewise constant functions on $\mathbb{D} =]0, 20]$ as introduced in Example 3.1.2, and $\text{pen}^{(1)}$ and $\text{pen}^{(2)}$

Penalised projection estimation in continuous time framework

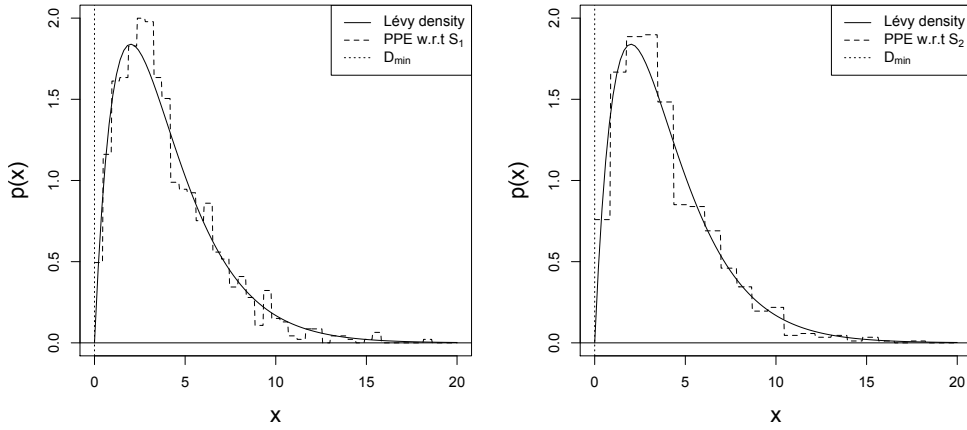


Figure 3.1: We present the PPE (dashed) $\hat{p}_{\text{pen}^{(1)}}$ w. r. t. $\text{pen}^{(1)}$ yielding \mathbb{S}_{43} as penalised projection space (left) and $\hat{p}_{\text{pen}^{(2)}}$ w. r. t. $\text{pen}^{(2)}$ yielding \mathbb{S}_{23} as penalised projection space (right) for the restriction of the Lévy density ν' of X to $\mathbb{D} =]0, 20]$, where $\nu = 10\chi_4^2$ (solid). The estimates are based on the jumps $\Delta X_t(\omega)$ occurred on the time horizon $[0, 100]$ as shown in Figure 2.2.

be of form (3.12) with $c_1^{(1)} = 2$, $c_2^{(1)} = c_3^{(1)} = 0$ and $c_1^{(2)} = 2$, $c_2^{(2)} = c_3^{(2)} = 1$. Figure 3.1 shows the PPE with respect to the penalties $\text{pen}^{(1)}$ and $\text{pen}^{(2)}$, while Table 3.1 compares the complexity of the (penalised) projection spaces, the squared and the absolute errors. Moreover, the squared error is decomposed into squared μ -bias and μ -variance. Again, from the statisticians' point of view, the estimator with less complexity, i. e. $\hat{p}_{\text{pen}^{(2)}}$, is favourable and more visually pleasing. Nevertheless, Table 3.1 shows that the squared error of $\hat{p}_{\text{pen}^{(1)}}$ is less than the squared error of $\hat{p}_{\text{pen}^{(2)}}$ for this sample. Moreover, Table 3.1 shows that the error reduction is due to a high decrease in the squared μ -bias that is not compensated by the small increase in μ -variance.

3 Penalised projection estimation in the continuous time framework

	m_T^{pen}	SE	AE	SB	MVar
pen ⁽¹⁾	43	0.1936	1.267	0.05503	0.1386
pen ⁽²⁾	23	0.3130	1.380	0.1816	0.1314

Table 3.1: We present the empirical best penalised projection space (m_T^{pen}), the squared error $\|p - \hat{p}_{\text{pen}}\|_{\mu}^2$ (SE) and the absolute error $\|p - \hat{p}_{\text{pen}}\|_{L^1(\mathbb{D}, \mu)}$ (AE) for the penalised projection estimators based on the penalties from Example 3.1.9. We further decompose the squared error into squared μ -bias (SB) and empirical μ -variance (MVar).

Recapitulating, we derived a criterion to dynamically select an estimator \hat{p}_{pen} out of the collection of projection estimators $\{\hat{p}_m : m \in M_T\}$. Nevertheless, we still have to analyse how effective this criterion is in practice.

3.2 Oracle inequality for penalised projection estimation

Figuroa-López and Houdré [18, 19] explicitly analyse the variance of the PPE (in CT) \hat{p}_{pen} in comparison with all estimators $\{\hat{p}_m : m \in M_T\}$. They derive a so-called *oracle inequality* linking the mean squared error of the PPE to the MSE of the oracle \hat{p}_* from Definition 3.1.3.

Figuroa-López and Houdré [19, Chapter 4] gives a full overview of oracle inequalities depending on different forms of penalties and proves the corresponding theorem in detail. We note that these results are one-to-one correspondent to Reynaud-Bouret [33, Theorem 1].

Theorem 3.2.1 (Figuroa-López and Houdré [19, Theorem 4.1 b / c])

Let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$ such that there are $c_1 > 1$ and $c_2, c_3 > 0$, with either

$$(i) \text{pen}_T(m) \geq c_1 T^{-1} \sum_{k=1}^{d_m} \hat{\nu}(f_{m,k}^2) + c_2 T^{-1} D_m + c_3 T^{-1} d_m \text{ or}$$

$$(ii) \text{pen}_T(m) \geq c_1 T^{-1} \sum_{k=1}^{d_m} \hat{\nu}(f_{m,k}^2) \text{ and both}$$

$$(a) \inf_{m \in M} D_m^{-1} \sum_{k=1}^{d_m} \nu(f_{m,k}^2) > 0 \text{ and}$$

$$(b) \inf_{m \in M} d_m^{-1} D_m > 0$$

being satisfied for all $m \in M_T = \{m \in M : D_m \leq T\}$. Then the mean squared error of the penalised projection estimator from Definition 3.1.8 satisfies

$$\mathbb{E} \|p - \hat{p}_{\text{pen}}\|_{\mu}^2 \leq C_1 \inf_{m \in M_T} \left(\|p - \mathcal{P}_m p\|_{\mu}^2 + \mathbb{E}[\text{pen}_T(m)] \right) + \frac{C_2}{T} \quad (3.14)$$

for some finite constants $C_1, C_2 > 0$.

3.3 Estimation of smooth univariate Lévy densities

Corollary 3.2.2 (Oracle inequality for PPE in CT) *Let pen be a penalty on the family $\{\mathbb{S}_m : m \in M\}$ of projection spaces such that*

$$\text{pen}_T(m) = c_1 T^{-1} \sum_{k=1}^{d_m} \hat{\nu}(f_{m,k}^2),$$

for $c_1 > 1$ and conditions (iia) and (iib) of Theorem 3.2.1 hold. Then the PPE \hat{p}_{pen} from Definition 3.1.8 satisfies

$$\mathbb{E} \|p - \hat{p}_{\text{pen}}\|_{\mu}^2 \leq C_1 \mathbb{E} \|p - \hat{p}_{\star}\|_{\mu}^2 + \frac{C_2}{T} \quad (3.15)$$

for some finite constants $C_1, C_2 > 0$, where \hat{p}_{\star} is the oracle from Definition 3.1.3.

Proof:

We connect (3.15) with (3.14) in the light of Corollary 3.1.6. \square

For some classes of μ -densities we are able to explicitly determine the oracle space for each density p and time horizon $[0, T]$. Moreover, in these cases it is possible to determine the rate of convergence in $\hat{p}_{\star} \rightarrow p$ for $T \rightarrow \infty$. Then Theorem 3.2.1 enables us to show that the PPE adapts the properties of the oracle, despite neither the real density p nor the oracle spaces $(m_T^*)_{T>0}$ are known.

3.3 Estimation of smooth univariate Lévy densities

In this section, we combine our results from Sections 3.1 and 3.2 for the case of estimating a smooth Lévy density in the continuous time framework. We recall decomposition (3.1) of the MSE of the projection estimator in squared μ -bias and μ -variance. We assume that $\mathbb{D} \in \mathcal{B}_o$ is compact and set

$$\|p\|_{\mathbb{D}, \infty} := \sup_{x \in \mathbb{D}} |p(x)| < \infty,$$

which is finite by virtue of Assumption 2.3.3. Then Proposition 3.1.4 implies that

$$\mathbb{E} \|\mathcal{P}_m p - \hat{p}\|_{\mu}^2 = \frac{1}{T} \sum_{k=1}^{d_m} \int_{\mathbb{D}} f_{m,k}^2(x) p(x) \mu(dx) \leq \frac{d_m \|p\|_{\mathbb{D}, \infty}}{T} \quad (3.16)$$

holds true for every projection space $\mathbb{S}_m \subseteq L^2(\mathbb{D}, \mu)$. It remains to analyse the bias term $\|p - \mathcal{P}_m p\|_{\mu}$. However, this requires additional assumptions. Below, we follow the concept Figueroa-López [15] adapted from Reynaud-Bouret [33] and Barron et al. [3]. Nevertheless, we emphasise that the main ideas are adapted from DeVore and Lorentz [12].

3 Penalised projection estimation in the continuous time framework

Definition 3.3.1 (Forward difference) *The k -th forward difference $\Delta_h^k(f, \cdot)$ of a function f is recursively defined by*

$$\Delta_h^k(f, x) := \Delta_h^{k-1}(\Delta_h(f, \cdot), x)$$

and

$$\Delta_h(f, x) := f(x + h) - f(x).$$

Definition 3.3.2 (Besov space over L^q) *Let $r \in \mathbb{R}^+$ and $q \geq 2$. Then the space $\mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)$ of all functions $f \in L^q(\mathbb{D}, \mu)$ such that*

$$|f|_{\mathcal{B}_\infty^{r,q}} := \sup_{\delta > 0} \frac{1}{\delta^r} \sup_{h \in]0, \delta]} \|\Delta_h^{k+1}(f, \cdot)\|_q < \infty, \quad (3.17)$$

where $k = \lfloor r \rfloor$, is called Besov space over $L^q(\mathbb{D}, \mu)$ with degree of smoothness r .

The subscript ‘‘infinity’’ indicates the use of the supremum over $\delta > 0$. DeVore and Lorentz [12] define Besov spaces $\mathcal{B}_{q'}^{r,q}$ w. r. t. arbitrary q' -norms. However, this is not necessary for our purposes.

For the class of Besov-type smooth functions, however, we obtain bounds for the approximation error, when using piecewise polynomials as described in Example 3.1.2. The following proposition is directly derived from DeVore and Lorentz [12, Chapter 2 (10.1)].

Proposition 3.3.3 (Approximation error for Besov-type functions)

Let $\mathbb{D} \in \mathcal{B}_\circ$ be a compact interval such that $\mu(\mathbb{D}) < \infty$ and assume that the restriction $p|_{\mathbb{D}}$ of the Lévy density p to \mathbb{D} satisfies $p \in \mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)$ for some $r \in \mathbb{R}^+$ and $q \geq 2$. Further, let $k = \lfloor r \rfloor$ and $\mathbb{S}_m \subseteq L^q(\mathbb{D}, \mu)$ be the space of piecewise polynomials of maximum degree k based on a partition \mathcal{D}_m of \mathbb{D} such that

$$\forall D \in \mathcal{D}_m : \mu(D) \leq m^{-1} \mu(\mathbb{D}).$$

Then there is a finite constant $c(r) < \infty$ such that

$$\inf_{f \in \mathbb{S}_m} \|p - f\|_q \leq c(r) |p|_{\mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)} m^{-r} < \infty. \quad (3.18)$$

Corollary 3.3.4 (Bias of projection estimation) *The μ -bias of projection estimation is bounded above by*

$$\|p - \mathcal{P}_m p\|_\mu \leq c(r) |p|_{\mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)}^{1/2-1/q+r} m^{-r} < \infty, \quad (3.19)$$

in the setting of Proposition 3.3.3.

3.3 Estimation of smooth univariate Lévy densities

Proof:

Hölder's inequality, i. e. $\|f\|_2 \leq \|f\|_q \|1\|_{q'}$ for $1/q + 1/q' = 1/2$, links (3.19) to (3.18), since $\|1\|_{q'} = (\int_{\mathbb{D}} \mu(dx))^{1/q'}$ in our case. \square

Although the Besov spaces seem to be of a rather abstract nature, they appear to be one of the most general classes of functions, for which (3.19) holds. For instance, every *Lipschitz (or Hölder) continuous function* belongs to a particular Besov space.

Example 3.3.5 (Hölder continuous functions) *Let $\mathbb{D} \in \mathcal{B}_\circ$ and $f \in L^q(\mathbb{D}, \mu)$ for some $q \geq 2$. Assume there exist $k \in \mathbb{N}$ and $\eta \in]0, 1]$ such that f is k -times differentiable with*

$$\forall h > 0 : \left\| \Delta_h \left(f^{(k)}, \cdot \right) \right\|_q \leq Kh^\eta$$

for some finite constant $K < \infty$. The k -th derivative is said to be Hölder continuous (or Lipschitz continuous if $\eta = 1$). Then we have $f \in \mathcal{B}_\infty^{k+\eta, q}(\mathbb{D}, \mu)$ by virtue of the Taylor argument $f^{(k)}(x) = \Delta_h^k(f, x)h^{-k} + O(h)$.

In view of (3.16) and (3.19), we are now able to distinguish the oracle space for the non-parametric projection estimation of Besov-type smooth Lévy densities and calculate its rate of convergence in terms of the degree of smoothness r as T tends to infinity. Thereafter, with Theorem 3.2.1, we show that the penalised projection estimator is self-adapting to the degree of smoothness, achieving the same rate of convergence without prior knowledge of r .

Theorem 3.3.6 (Rate of convergence of the oracle) *Let X be a univariate Lévy process with generating triplet (γ, σ^2, ν) . Further, let $\mathbb{D} \in \mathcal{B}_\circ$ be a compact interval, μ be a reference measure such that $\mu(\mathbb{D}) < \infty$ and the restriction $p|_{\mathbb{D}}$ of p to \mathbb{D} belongs to the Besov space $\mathcal{B}_\infty^{r, q}(\mathbb{D}, \mu)$ for some $r > 0$ and $q \geq 2$. Let $M = \mathbb{N}$ and, for each $m \in M$, let $\mathcal{S}_m \subseteq L^2(\mathbb{D}, \mu)$ be the space of piecewise polynomials of maximum degree $k = \lfloor r \rfloor$ based on the regular partition of \mathbb{D} into m classes (see Example 3.1.2). Then the oracle space satisfies $m_T^* \sim T^{1/(2r+1)}$ as $T \rightarrow \infty$ and the projection estimator satisfies*

$$\sup_{T>0} T^{\frac{2r}{2r+1}} \mathbf{E} \|p - \hat{p}_\star\|_\mu^2 < \infty.$$

Proof:

As the preliminaries of Corollary 3.3.4 are met, there exists a finite positive constant $K_1 < \infty$ such that

$$\|p - \mathcal{P}_m p\|_\mu^2 \leq K_1 m^{-2r},$$

and, since $d_m = (k + 1)m$, there exists another finite constant $K_2 < \infty$ such that (3.16) implies

$$\mathbf{E} \|\mathcal{P}_m p - \hat{p}_m\|_\mu^2 \leq K_2 \frac{m}{T}.$$

3 Penalised projection estimation in the continuous time framework

Therefore, we have m. s. e. $[\hat{p}_m] \leq K_3(m^{-2r} + mT^{-1})$ for a finite constant $K_3 < \infty$, where the right-hand side is clearly minimized for $m_T \sim T^{1/(2r+1)}$ as $T \rightarrow \infty$. Thus, we have

$$\mathbb{E}\|p - \hat{p}_\star\|_\mu^2 \leq K_3 \left(\left[T^{\frac{1}{2r+1}} \right]^{-2r} + \left[T^{\frac{1}{2r+1}} \right] T^{-1} \right) = O\left(T^{\frac{-2r}{2r+1}}\right)$$

completing the proof. \square

Theorem 3.3.7 (Rate of convergence of the PPE) *Given the setup of Theorem 3.3.6, let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$ with*

$$\text{pen}_T(m) = \frac{c_1}{T} \sum_{k=1}^{d_m} \hat{v}(f_{m,k}^2) + \frac{c_2 D_m}{T} + \frac{c_3 d_m}{T}$$

for some finite constants $c_1 > 1$, $c_2, c_3 \geq 0$. Then

$$\sup_{T>0} T^{\frac{2r}{2r+1}} \mathbb{E}\|p - \hat{p}_{\text{pen}}\|_\mu^2 < \infty \quad (3.20)$$

holds true for the penalised projection estimator from Definition 3.1.8.

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

$$B(a_1, a_2) := \left\{ f \in \mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu) : \|f\|_{\mathbb{D},\infty} \leq a_1 \text{ and } |f|_{\mathcal{B}_\infty^{r,q}} \leq a_2 \right\}$$

the space of all Besov-type smooth functions with supremum and Besov-quasi-norm bounded by a_1 and a_2 , respectively. Then, additionally,

$$\sup_{T>0} T^{\frac{2r}{2r+1}} \sup_{p \in B(a_1, a_2)} \mathbb{E}\|p - \hat{p}_{\text{pen}}\|_\mu^2 < \infty \quad (3.21)$$

holds true for the PPE w. r. t. the penalty pen.

Proof:

As we explicitly allow for $c_2 = c_3 = 0$, we show first that the conditions (ia) and (iib) of Theorem 3.2.1 are met. We recall that \mathbb{D} is assumed to be a compact interval. W. l. o. g., let $\mathbb{D} = [D_{\min}, D_{\max}] \subseteq \mathbb{R}^+$ and μ be the Lebesgue measure. Then, for fixed $m \in M$, we have

$$\sqrt{\frac{m}{D_{\max} - D_{\min}}} \mathbb{1}_{[D_{\min}, D_{\min} + \frac{D_{\max} - D_{\min}}{m}]}(\cdot) \in \{f \in \mathbb{S}_m : \|f\|_\mu = 1\},$$

hence $D_m \geq m/(D_{\max} - D_{\min})$. Thus

$$\inf_{m \in M} \frac{D_m}{d_m} \geq \inf_{m \in M} \frac{1}{(D_{\max} - D_{\min})(k+1)} > 0,$$

3.3 Estimation of smooth univariate Lévy densities

showing that condition (iib) holds true. Additionally,

$$D_m \leq \frac{(k+1)^2 m}{D_{\max} - D_{\min}}$$

holds. Finally, as

$$\sum_{l=1}^{d_m} f_{m,l}^2(\cdot) \geq \frac{m}{D_{\max} - D_{\min}} \mathbb{1}_{\mathbb{D}}(\cdot),$$

we deduce that

$$\inf_{m \in M} \frac{\sum_{l=1}^{d_m} \nu(f_{m,l}^2)}{D_m} \geq \frac{\nu(\mathbb{D})}{(k+1)^2} > 0,$$

showing that condition (iia) holds true.

Thus, there exist finite constants $C_1, C_2 < \infty$ such that

$$\mathbb{E} \|p - \hat{p}_{\text{pen}}\|_{\mu}^2 \leq C_1 \inf_{m \in M_T} \left\{ \|p - \mathcal{P}_m p\|_{\mu}^2 + \mathbb{E}[\text{pen}_T(m)] \right\} + \frac{C_2}{T}.$$

Similarly as in the proof of Theorem 3.3.6, we deduce from Corollary 3.3.4 that there is another finite constant $K_1 = K_1(|p|_{\mathcal{B}_{\infty}^{r,q}}) < \infty$ such that

$$\|p - \mathcal{P}_m p\|_{\mu}^2 \leq K_1 m^{-2r}.$$

Moreover, since $d_m = (k+1)m$ and $D_m \leq \mu(\mathbb{D})^{-1}(k+1)^2 m$ hold, we derive from (3.16) that we have

$$\mathbb{E}[\text{pen}_T(m)] \leq K_2 \frac{m}{T}$$

for a finite constant $K_2 = K_2(\|p\|_{\mathbb{D},\infty}) < \infty$. Therefore, the rate of convergence $T^{2r/(2r+1)}$ in (3.20) follows equivalently to the proof of Theorem 3.3.6.

Moreover, since $B(a_1, a_2)$ is a compact subspace of $\mathcal{B}_{\infty}^{r,q}(\mathbb{D}, \mu)$, therefore there exists a uniform bound $K' < \infty$ for $\sup_{f \in B(a_1, a_2)} \{K_1(f), K_2(f)\}$. Hence, the finiteness in (3.21) is clear by virtue of (3.20). \square

Remark 3.3.8 *We prove Theorem 3.3.7 under weaker assumptions than Figueroa-López and Houdré [19, Corollary 5.1]. In particular, we explicitly allow for $c_2 = 0$ and $c_3 = 0$. Therefore, we have validated property (3.20) for the penalty*

$$\text{pen}_T(m) = \frac{2}{T} \sum_{k=1}^{d_m} \hat{\nu}(f_{m,k}^2), \quad (3.22)$$

when used for penalised projection estimation of Besov-type smooth Lévy densities. We recall that Corollary 3.1.6 shows that the mean squared error of projection estimation is an affine function of the expected value of $-\|\hat{p}_m\|_{\mu}^2 + \text{pen}_T(m)$, in the case of penalty (3.22). In this sense, (3.22) is unbiased.

Again, from the statisticians' point of view, it can be favourable to use higher penalties, in order to keep the complexity of the estimate small. Theorem 3.3.7 shows that we can do so without decreasing the rate of convergence of our estimator.

3 Penalised projection estimation in the continuous time framework

4 Projection estimation in the discrete time framework

The aim of this chapter is to base the method of projection estimation on observations on a discrete time grid. First, we introduce our observation scheme. Then we analyse the mean squared error of the resulting estimators. In comparison to the MSE derived in Section 3.1, we give conditions on the observation grid such that the additional error from discretisation becomes asymptotically negligible.

Let X be an \mathbb{R}^d -valued Lévy process with Lévy-Khintchine triplet (γ, Σ, ν) and $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$. Throughout this chapter, we assume that every $f \in \mathbb{S}$ is integrable and square integrable w. r. t. the Lévy measure ν .

4.1 Observation of Lévy processes on discrete grids

First, let us introduce the observation scheme on a discrete time grid.

Definition 4.1.1 (Discretely observed Lévy process) *Let $\tau > 0$. We call the grid $\tau\mathbb{N}_0 := \{k\tau : k \in \mathbb{N}_0\}$ the regular grid with mesh size τ . The k -th observed increment on the regular grid $\tau\mathbb{N}_0$ of X is given by*

$$X_{\Delta_k^\tau} := X_{k\tau} - X_{(k-1)\tau}. \quad (4.1)$$

Moreover, we call the observed process

$$X^\tau : \mathbb{R}_0^+ \times \Omega; (t, \omega) \mapsto X_t^\tau(\omega) = X_{\tau\lfloor t/\tau \rfloor}(\omega) = X_0(\omega) + \sum_{k=1}^{\lfloor t/\tau \rfloor} X_{\Delta_k^\tau}(\omega) \quad (4.2)$$

discretisation of X w. r. t the mesh size τ .

Figure 4.1 illustrates the effect of discretisation. Although the Lévy-Itô decomposition (2.8) gives an insight into the components of a Lévy process, separate observation of its continuous and its jump part remains theoretical. In particular, the Poisson integrals related to the PRM J are impossible to observe and jumps cannot be foreseen. We conclude that the proposed estimators from Section 2.3 and Chapter 3 remain theoretical as well. Below, we replace them by their pendants based on the discretisation X^τ .

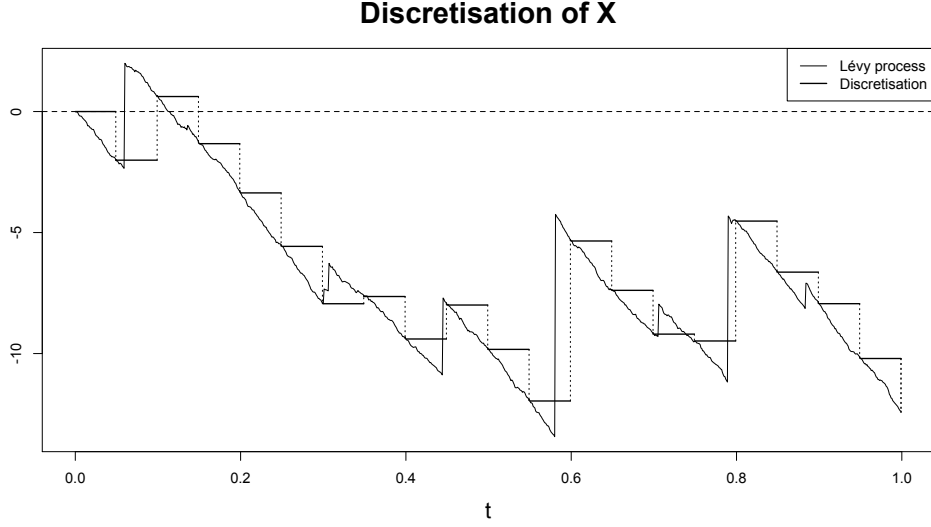


Figure 4.1: We present the sample path $X(\omega)$ from Example 2.3.2 on the time horizon $[0, 1]$ and the discretisation $X^\tau(\omega)$ of X w. r. t. the mesh size $\tau = 0.05$.

Definition 4.1.2 (Projection estimator in DT framework)

Let $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ be a finite dimensional linear projection space, $\tau > 0$ and $T \geq \tau$. Furthermore, let $\{f_k : k = 1, \dots, d_{\mathbb{S}}\}$ be a μ -ONB of \mathbb{S} and denote $T'_\tau := \tau \lfloor T/\tau \rfloor$. Then, for every $k = 1, \dots, d_{\mathbb{S}}$,

$$\tilde{\nu}(f_k) := \tilde{\nu}(f_k; \tau, T) := \frac{1}{T'_\tau} \sum_{l=1}^{\lfloor T/\tau \rfloor} f_k(X_{\Delta_l^\tau}), \quad (4.3)$$

is called the integral estimator (in DT) for $\nu(f_k)$, and

$$\tilde{p}_{\mathbb{S}} := \tilde{p}(\cdot; \tau, T, \mathbb{S}) := \sum_{k=1}^{d_{\mathbb{S}}} \tilde{\nu}(f_k) f_k(\cdot) = \frac{1}{T'_\tau} \sum_{k=1}^{d_{\mathbb{S}}} \sum_{l=1}^{\lfloor T/\tau \rfloor} f_k(X_{\Delta_l^\tau}) f_k(\cdot). \quad (4.4)$$

is called the projection estimator (in DT) for p on \mathbb{S} based on the observation of the discretisation X^τ on the time horizon $[0, T]$.

In Chapter 3, the rate of convergence of the penalised projection estimator over a set of projection spaces $\{\mathbb{S}_m : m \in M\}$ was analysed in four steps. Firstly, given $m \in M$, the proof of Proposition 3.1.5 shows that the μ -variance of the projection estimator w. r. t. the μ -orthogonal projection $\mathcal{P}_m p$ of the μ -density p is of order $O(d_m T^{-1})$. Secondly, for Besov-type smooth Lévy densities, Corollary 3.3.4

verifies that the μ -bias is of order $O(d_m^{-r})$, where r stands for the degree of smoothness. Thirdly, combining the two results, Theorem 3.3.6 made the oracle and its rate of convergence in mean squared sense readily available. Lastly, for an adequate choice of a penalty, Theorem 3.3.7 transfers the result for the oracle to the PPE by virtue of the oracle inequality from Theorem 3.2.1. In the discrete time framework, steps 1 and 4 have to be revised. The former is done in Section 4.2 and Chapter 5, the latter is done in Chapter 6.

4.2 The critical mesh

In the following, we focus on deriving conditions such that the μ -variance term of projection estimation in DT is of same order as in CT, i. e.

$$\mathbb{E}\|\mathcal{P}_{\mathbb{S}}p - \tilde{p}_{\mathbb{S}}\|_{\mu}^2 = O(d_m T^{-1}).$$

Mean squared error of integral estimation

Definition 4.2.1 (Discretisation bias) *Let $f \in L^1(\mathbb{D}, \nu)$, then*

$$\Delta_{\tau}f := \frac{1}{\tau} \mathbb{E}[f(X_{\tau})] - \nu(f) \quad (4.5)$$

is called its discretisation bias w. r. t. X^{τ} .

Since increments of a Lévy process are, by definition, stationary and independent, we deduce that

$$\Delta_{\tau}f = \text{Bias}[\tilde{\nu}(f)] = \mathbb{E}[\tilde{\nu}(f) - \nu(f)] \quad (4.6)$$

holds true for all $T > 0$.

Making use of our notation, the following proposition (cf. Figueroa-López and Houdré [20, p. 6]) shows the close relation between the MSEs of $\tilde{\nu}(f)$ and $\hat{\nu}(f)$ on the one hand, and $\Delta_{\tau}f$ on the other hand.

Proposition 4.2.2 (Mean squared error of integral estimation)

The mean squared error of the integral estimator in DT $\tilde{\nu}(\cdot)$ for $\nu(\cdot)$ defined in (4.3) is bounded by

$$\text{m. s. e.}[\tilde{\nu}(f)] \leq (\Delta_{\tau}f)^2 + \frac{1}{T'_{\tau}}\Delta_{\tau}f^2 + \frac{1}{T'_{\tau}}\nu(f^2), \quad (4.7)$$

where $T'_{\tau} = \tau \lfloor T/\tau \rfloor$.

4 Projection estimation in the discrete time framework

Proof:

Figueroa-López [17, p. 5] claims

$$\text{Var} [\tilde{\nu}(f)] = \frac{1}{T'_\tau} \left(\frac{1}{\tau} \mathbb{E} [f^2(X_\tau)] \right) - \frac{1}{[T/\tau]} \left(\frac{1}{\tau} \mathbb{E} [f(X_\tau)] \right)^2.$$

This can be validated using the standard formula $\text{Var} X = \mathbb{E}[X^2] - EX^2$. Moreover, we conclude that

$$\begin{aligned} \text{m. s. e.} [\tilde{\nu}(f)] &= \text{Bias} [\tilde{\nu}(f)]^2 + \text{Var} [\tilde{\nu}(f)] \\ &\leq (\Delta_\tau f)^2 + \frac{1}{T'_\tau} \left(\frac{1}{\tau} \mathbb{E} [f^2(X_\tau)] - \nu(f^2) + \nu(f^2) \right) \end{aligned}$$

holds. Then (4.7) is a direct consequence. \square

The last term in (4.7) is in principal the MSE of $\hat{\nu}(f)$, whereas the additional risk of estimation is incorporated in the discretisation bias of the function f and its squared pendant f^2 .

Bias conditions and conclusion

There are various papers dealing more or less directly with the limiting behaviour of $\Delta_\tau f$. [...] In our case, it is sufficient to note that

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \mathbb{E}[f(X_\tau)] = \nu(f),$$

i. e. $\lim_{\tau \rightarrow 0} \Delta_\tau f = 0$, holds for every function f that is ν -a. e. continuous, bounded and vanishing in a neighbourhood of the origin (cf. Sato [35, Corollary 8.9]).

For simplicity, we always assume in the following that the mesh size τ is dividing the time span T , i. e. $T'_\tau = T$.

Condition 4.2.3 (Bias condition BC ϑ / $\overline{\text{BC}}\vartheta$) Let $\vartheta > 0$ and $f \in L^2(\mathbb{D}, \mu)$. We say that f satisfies the (exclusive) bias condition BC ϑ if

$$\Delta_\tau f = O(\tau^{\vartheta'}) \text{ and } \Delta_\tau f^2 = O(\tau^{\vartheta'}) \quad \text{as } \tau \rightarrow 0 \quad (4.8)$$

for every $\vartheta' \in [0, \vartheta[$. Moreover, we say that f satisfies the (inclusive) bias condition $\overline{\text{BC}}\vartheta$ if (4.8) holds for $\vartheta' = \vartheta$, additionally. Finally, for a finite dimensional projection space $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$, we say that \mathbb{S} satisfies BC ϑ or $\overline{\text{BC}}\vartheta$ if the respective bias condition is satisfied for all $f \in \mathbb{S}$.

To simplify notation, we consider the case of an inclusive bias condition being valid in the following two theorems only. We can do this without loss of generality. Assume f satisfies an exclusive bias condition BC ϑ , thus the inclusive bias condition $\overline{\text{BC}}\vartheta'$ is valid for all $\vartheta' < \vartheta$.

Theorem 4.2.4 (Critical mesh (variant 1)) *Let $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ be a projection space that satisfies the inclusive bias condition $\overline{\text{BC}}^\vartheta$ for $\vartheta > 0$. Then a mesh size of $\tau_T = O(T^{-1/(2\vartheta)})$ as $T \rightarrow \infty$ is sufficient for*

$$\sup_{T>0} \frac{\text{m. s. e.}[\tilde{\nu}(f)]}{\text{m. s. e.}[\hat{\nu}(f)]} < \infty$$

for all $f \in \mathbb{S}$.

Proof:

We deduce from Proposition 3.1.4 and Proposition 4.2.2 that it is sufficient to prove $\sup_{T>0} T(\Delta_{\tau_T} f)^2 < \infty$ and $\sup_{T>0} \Delta_{\tau_T} f^2 < \infty$. As the inclusive bias condition $\overline{\text{BC}}^\vartheta$ holds true, (4.8) is equivalent to $\sup_{\tau>0} \tau^{-\vartheta} \Delta_\tau f < \infty$. Over and above, the required rate of convergence for the mesh size τ_T implies $\sup_{T>0} T\tau_T^{2\vartheta} < \infty$. Hence, we deduce

$$\sup_{T>0} T(\Delta_{\tau_T} f)^2 = \sup_{T>0} T\tau_T^{2\vartheta} (\tau_T^{-\vartheta} \Delta_{\tau_T} f)^2 \leq \sup_{T>0} T\tau_T^{2\vartheta} \left(\sup_{\tau>0} \tau^{-\vartheta} \Delta_\tau f \right)^2 < \infty.$$

Moreover, (4.8) for $\vartheta > 0$ implies $\lim_{T \rightarrow \infty} \Delta_{\tau_T} f^2 = 0$, by definition. \square

Theorem 4.2.5 (Critical mesh (variant 2)) *Let $\mathbb{S} \subseteq L^2(\mathbb{D}, \mu)$ be a projection space that satisfies the inclusive bias condition $\overline{\text{BC}}^\vartheta$ for $\vartheta > 0$. Then a mesh size of $\tau_T = o(T^{-1/(2\vartheta)})$ as $T \rightarrow \infty$ is sufficient that*

$$\text{m. s. e.}[\tilde{\nu}(f)] = \text{m. s. e.}[\hat{\nu}(f)] + o(T^{-1})$$

as $T \rightarrow \infty$ holds for all $f \in \mathbb{S}$.

Proof:

Analogously to Theorem 4.2.4, it is sufficient to prove $\lim_{T \rightarrow \infty} T(\Delta_{\tau_T} f)^2 = 0$ and $\lim_{T \rightarrow \infty} \Delta_{\tau_T} f^2 = 0$. Since the inclusive bias condition $\overline{\text{BC}}^\vartheta$ holds true, (4.8) is equivalent to $\sup_{\tau>0} \tau^{-\vartheta} \Delta_\tau f < \infty$. Additionally, the required rate of convergence for the mesh size τ_T implies $\lim_{T \rightarrow \infty} T\tau_T^{2\vartheta} = 0$. Hence, we deduce

$$\lim_{T \rightarrow \infty} T(\Delta_{\tau_T} f)^2 = \lim_{T \rightarrow \infty} T\tau_T^{2\vartheta} (\tau_T^{-\vartheta} \Delta_{\tau_T} f)^2 \leq \lim_{T \rightarrow \infty} T\tau_T^{2\vartheta} \left(\sup_{\tau>0} \tau^{-\vartheta} \Delta_\tau f \right)^2 = 0.$$

Again, (4.8) for $\vartheta > 0$ is sufficient for $\lim_{T \rightarrow \infty} \Delta_{\tau_T} f^2 = 0$, by definition. \square

By virtue of Theorem 4.2.5, we are able to adapt Theorem 3.3.6 to the discrete time framework.

4 Projection estimation in the discrete time framework

Theorem 4.2.6 (Estimation of smooth Lévy densities in DT) *Let X be a univariate Lévy process with generating triplet (γ, σ^2, ν) . Further, let $\mathbb{D} \in \mathcal{B}_\circ$ be a compact interval, μ be a reference measure such that $\mu(\mathbb{D}) < \infty$ and the restriction $p|_{\mathbb{D}}$ of the μ -density p to \mathbb{D} belongs to the Besov space $\mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)$ for some $r > 0$ and $q \geq 2$. Let $M = \mathbb{N}$ and, for each $m \in M$, let $\mathbb{S}_m \subseteq L^2(\mathbb{D}, \mu)$ be the space of piecewise polynomials of maximum degree $k = \lfloor r \rfloor$ based on the regular partition of \mathbb{D} into m classes (see Example 3.1.2). We assume that all $f \in \mathbb{S}_m$ satisfy the inclusive bias condition $\overline{\text{BC}}\vartheta$ for a $\vartheta > 0$. Then a mesh size of $\tau_T = o(T^{-1/(2\vartheta)})$ as $T \rightarrow \infty$ is sufficient to ensure that the oracle space satisfies $m_T^* \sim T^{1/(2r+1)}$ and that the projection estimator in the discrete time framework satisfies*

$$\sup_{T>0} T^{\frac{2r}{2r+1}} \mathbb{E} \|p - \tilde{p}_\star\|_\mu^2 < \infty.$$

Proof:

Equivalently to the proof of Theorem 3.3.6, there is a finite constant $K_1 < \infty$ such that

$$\|p - \mathcal{P}_m p\|_\mu^2 \leq K_1 m^{-2r}.$$

Since Theorem 4.2.5 implies m. s. e. $[\tilde{\nu}(f)] = \text{m. s. e.}[\hat{\nu}(f)] + o(T^{-1})$, there exists a finite constant $\tilde{K}_2 < \infty$ such that

$$\mathbb{E} \|\mathcal{P}_m p - \hat{p}_m\|_\mu^2 \leq \tilde{K}_2 \frac{m}{T}.$$

The remainder of the proof of Theorem 3.3.6 can be adapted one-to-one. Firstly, we have m. s. e. $[\hat{p}_m] \leq \tilde{K}_3(m^{-2r} + mT^{-1})$ for some finite constant $\tilde{K}_3 < \infty$, where the right-hand side is clearly minimized for $m_T \sim T^{1/(2r+1)}$ as $T \rightarrow \infty$. Thus, we have

$$\mathbb{E} \|p - \hat{p}_\star\|_\mu^2 \leq \tilde{K}_3 (\lfloor T^{1/(2r+1)} \rfloor^{-2r} + \lfloor T^{1/(2r+1)} \rfloor T^{-1}) = O\left(T^{\frac{-2r}{2r+1}}\right)$$

completing the proof. □

Remark 4.2.7 *In Chapter 6, we further analyse the penalisation method in the discrete time framework, in order to adapt Theorem 3.3.7, additionally. So far, prior knowledge of the degree of smoothness r in Theorem 4.2.6 is necessary to distinguish the oracle space $\mathbb{S}_\star \in \{\mathbb{S}_m : m \in M_T\}$ and to achieve the proposed rate of convergence.*

5 Projection estimation away from zero

In this chapter, we analyse the validity of bias conditions BC^ϑ in the case of selected Lévy processes. Throughout, let X be a Lévy process with generating triplet (γ, Σ, ν) . We choose $\mathbb{D} \in \mathcal{B}_\circ^d$ bounded and away from the coordinate axes and let μ be the Lebesgue measure on \mathbb{D} .

5.1 A uniform bound for $\Delta_\tau f$ (technical result)

To analyse the discretisation bias for arbitrary functions f can become quite tedious. Without loss of generality, we can restrict our further analysis to the case of indicator functions due to the following technical results.

Univariate framework

Let $d = 1$. In the framework of Definition 4.2.1, let $y \in \mathbb{R}_\circ$ and set

$$f = \begin{cases} \mathbb{1}_{[y, \infty[} & \forall y > 0, \\ \mathbb{1}_{]-\infty, y]} & \forall y < 0. \end{cases}$$

Then $\Delta_\tau f$ as defined in (4.5) is equal to

$$\Delta_\tau y := \Delta_\tau f = \begin{cases} \tau^{-1}P(X_\tau \geq y) - \nu([y, \infty[) & \forall y > 0, \\ \tau^{-1}P(X_\tau < y) - \nu(]-\infty, y]) & \forall y < 0. \end{cases} \quad (5.1)$$

Definition 5.1.1 (Discrete transition deviation) *The discrete transition deviation on the regular grid with mesh size $\tau > 0$ is defined by $\Delta_\tau^* := \sup_{y \in \mathbb{D}} |\Delta_\tau y|$.*

Lemma 5.1.2 (Uniformly bounded bias) *Let $\mathbb{D} \in \mathcal{B}_\circ$ be compact, away from zero and assume there are finitely many disjoint intervals $D_1, \dots, D_n \subseteq \mathbb{D}$ such that $\mathbb{D} = \cup_{k=1}^n D_k$ holds. Then there exists a finite constant $K < \infty$ such that, for all $\tau > 0$ and for every bounded function f , we have*

$$|\Delta_\tau f| \leq K \left(\|f\|_\infty + \sum_{k=1}^n \|f|_{D_k}\|_1 \right) \Delta_\tau^*, \quad (5.2)$$

5 Projection estimation away from zero

where we assume that the restriction $f|_{D_k}$ of f to D_k is absolutely continuous for all $k = 1, \dots, n$.

Example 5.1.3 (Disconnected domain of estimation) *An example for a disconnected domain of estimation in the univariate case is the union of an interval on the positive and another interval on the negative half-line, i. e.*

$$\mathbb{D} = [-D_{\max}, -D_{\min}] \cup [D_{\min}, D_{\max}]$$

for $0 < D_{\min} < D_{\max} < \infty$.

Proof of Lemma 5.1.2:

For $k \in \{1, \dots, n\}$, there are $c_k, C_k \in \mathbb{R}$ with $]c_k, C_k[\subseteq D_k \subseteq [c_k, C_k]$. W.l.o.g. assume $\forall k : D_k \subseteq \mathbb{R}^+$. The Fubini argument from Figueroa-López [17, Lemma 3.2] remains valid. Specifically, for any measure Q ,

$$\begin{aligned} \int f(x) \mathbb{1}_{D_k} Q(dx) &= \int_{D_k} \left(f(c_k) + \int_{c_k}^x f'_{|D_k}(u) du \right) Q(dx) \\ &= f(c_k) Q(D_k) + \iint_{[c_k, C_k] \times [u, C_k]} f'_{|D_k}(u) Q(dx) du \\ &= f(c_k) Q(D_k) + \int_{c_k}^{C_k} f'_{|D_k}(u) Q([u, C_k]) du. \end{aligned}$$

In particular, for the measures $P_{X_\tau} := P(X_\tau \in \cdot)$ and ν we have

$$\begin{aligned} |\Delta_\tau f \mathbb{1}_{D_k}| &= \left| \frac{\mathbb{E}[f(X_\tau) \mathbb{1}_{D_k}]}{\tau} - \nu(f \mathbb{1}_{D_k}) \right| \\ &= \left| f(c_k) \left(\frac{P_{X_\tau}(D_k)}{\tau} - \nu(D_k) \right) + \int_{D_k} f'_{|D_k}(u) \left(\frac{P_{X_\tau}([u, C_k])}{\tau} - \nu([u, C_k]) \right) du \right|. \end{aligned}$$

Since

$$\forall u \in D_k : \left| \frac{P_{X_\tau}([u, C_k])}{\tau} - \nu([u, C_k]) \right| = |\Delta_\tau u - \Delta_\tau C_k| \leq 2\Delta_\tau^*,$$

we have $|\Delta_\tau f \mathbb{1}_{D_k}| \leq 2(\|f\|_\infty + \|f'_{|D_k}\|_1) \Delta_\tau^*$. Consequently,

$$|\Delta_\tau f| \leq \sum_{k=1}^n |\Delta_\tau f \mathbb{1}_{D_k}| \leq 2n(\|f\|_\infty + \sum_{k=1}^n \|f'_{|D_k}\|_1) \Delta_\tau^*.$$

□

5.1 A uniform bound for $\Delta_r f$ (technical result)

Sobolev spaces

Before we give a multivariate analogue for Lemma 5.1.2, we adapt some notation for partial derivative calculus as presented in Brenner and Scott [7, p. 24–26]. For the sake of easier comprehension, we write vectors in bold font throughout this and the following subsection.

Definition 5.1.4 (Multi-index notation and partial derivatives)

A d -tuple $\mathbf{k} \in \mathbb{N}_0^d$ is called multi-index. The length of a multi-index \mathbf{k} is given by $|\mathbf{k}| := \sum_{j=1}^d k_j$. For every sufficiently smooth function f , we abbreviate its usual (pointwise) partial derivative by

$$\partial^{\mathbf{k}} f := \partial_1^{k_1} \cdots \partial_d^{k_d} f.$$

Making use of this notation, we define the notion of weak derivatives for *locally integrable* functions, i. e. all functions $f : \mathbb{D} \rightarrow \mathbb{R}$ defined on a domain $\mathbb{D} \in \mathcal{B}^d$ such that $f \in L^1(K)$ for every compact subset $K \subseteq \mathbb{D}$.

Definition 5.1.5 (Weak derivation) Let $f : \mathbb{D} \rightarrow \mathbb{R}$ be a locally integrable function and $\mathbf{k} \in \mathbb{N}_0^d$ be a multi-index. Assume, there exists a locally integrable function $g : \mathbb{D} \rightarrow \mathbb{R}$ such that

$$\int_{\mathbb{D}} g(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} = (-1)^{|\mathbf{k}|} \int_{\mathbb{D}} f(\mathbf{x}) \partial^{\mathbf{k}} \phi(\mathbf{x}) d\mathbf{x}$$

holds for every function $\phi \in \mathcal{C}^\infty(\mathbb{D})$ with compact support. Then $\partial_{\mathbf{w}}^{\mathbf{k}} f := g$ is called weak (\mathbf{k} -)derivative of f .

Remark 5.1.6 (Weak derivation of sufficiently smooth functions)

Brenner and Scott [7, Proposition 1.2.7] proves that the weak \mathbf{k} -derivative of every $\mathcal{C}^{|\mathbf{k}|}$ -function coincides with the usual (partial) derivative. In the following, we use the same differentiation symbol (dropping the “w”) referring to both concepts, as appropriate.

Moreover, we define the notion of Sobolev spaces as presented in Brenner and Scott [7, Definition 1.3.1].

Definition 5.1.7 (Sobolev space over L^q) Let $r \in \mathbb{N}_0$ and $q \geq 1$. Then the space $\mathcal{W}^{r,q}(\mathbb{D}, \mu)$ of all functions $f \in L^q(\mathbb{D}, \mu)$ such that

$$\|f\|_{\mathcal{W}^{r,q}} := \left(\sum_{|\mathbf{k}| \leq r} \|\partial^{\mathbf{k}} f\|_{L^q(\mathbb{D}, \mu)}^q \right)^{1/q} < \infty$$

5 Projection estimation away from zero

is called Sobolev space over $L^q(\mathbb{D}, \mu)$ with degree of smoothness r . Furthermore, we define the Sobolev semi-norm by

$$|f|_{\mathcal{W}^{r,q}} := \left(\sum_{|\mathbf{k}|=r} \|\partial^{\mathbf{k}} f\|_{L^q(\mathbb{D}, \mu)}^q \right)^{1/q}.$$

For our purposes, the following Lemma (cf. Mazja [31, §3.1. Korollar 1.1] gives a parsimonious criterion for the *absolute continuity on lines* of functions $f : \mathbb{D} \rightarrow \mathbb{R}$.

Lemma 5.1.8 (Absolute continuity on lines for Sobolev functions)

Let $\mathbb{D} \subseteq \mathbb{R}^d$ and $f \in \mathcal{W}^{1,q}(\mathbb{D})$ for a $q \geq 1$. Then for almost every line

$$l(\mathbf{x}, i) := \{\mathbf{x} + h\mathbf{e}_i : h \in \mathbb{R}\},$$

where \mathbf{e}_i denotes the i -th standard basis vector, we have that the (univariate) function

$$g : \{h \in \mathbb{R} : \mathbf{x} + h\mathbf{e}_i \in l(\mathbf{x}, i) \cap \mathbb{D}\} \rightarrow \mathbb{R}; h \mapsto g(h) := f(\mathbf{x} + h\mathbf{e}_i)$$

is absolutely continuous and $\partial_i f(\mathbf{x} + h\mathbf{e}_i) = g'(h)$ almost everywhere.

Multivariate framework

In the multivariate version, we adapt notation from Kallsen and Tankov [24]. For $\mathbf{a} = (a_1, \dots, a_d), \mathbf{b} = (b_1, \dots, b_d) \in (\mathbb{R} \cup \{\pm\infty\})^d$, we write $\mathbf{a} \leq \mathbf{b}$ if $a_k \leq b_k$ for all $k = 1, \dots, d$. In this case, let $]\mathbf{a}, \mathbf{b}[$ denote an open d -dimensional interval of $(\mathbb{R} \cup \{\pm\infty\})^d$ defined by

$$]\mathbf{a}, \mathbf{b}[:=]a_1, b_1[\times \dots \times]a_d, b_d[.$$

Similarly, we define $[\mathbf{a}, \mathbf{b}]$, $]\mathbf{a}, \mathbf{b}]$ and $[\mathbf{a}, \mathbf{b}[$. Furthermore, we make use of a specific interval. In particular, let $\mathbf{y} = (y_1, \dots, y_d) \in (\mathbb{R} \setminus \{0\})^d$ then we denote the d -dimensional tail interval by

$$\mathcal{T}(\mathbf{y}) := \{\mathbf{x} \in \mathbb{R}^d : |x_k| \geq |y_k| \text{ and } \text{sgn}(x_k) = \text{sgn}(y_k) \forall k = 1, \dots, d\}.$$

Set $f := \mathbb{1}_{\mathcal{T}(\mathbf{y})}$, then $\Delta_\tau f$ as defined in (4.5) is equal to

$$\Delta_\tau \mathbf{y} := \Delta_\tau f = \tau^{-1} P(\mathbf{X}_\tau \in \mathcal{T}(\mathbf{y})) - \nu(\mathcal{T}(\mathbf{y})). \quad (5.3)$$

Definition 5.1.9 (Discrete transition deviation) The discrete transition deviation on the regular grid with mesh size $\tau > 0$ is defined by $\Delta_\tau^* := \sup_{\mathbf{y} \in \mathbb{D}} |\Delta_\tau \mathbf{y}|$.

5.1 A uniform bound for $\Delta_\tau f$ (technical result)

Lemma 5.1.10 (Uniformly bounded bias) *Let $\mathbb{D} \in \mathcal{B}((\mathbb{R} \setminus \{0\})^d)$ be compact and away from the origin and assume there are finitely many disjoint d -dimensional intervals $D_1, \dots, D_n \subseteq \mathbb{D}$ such that $\mathbb{D} = \bigcup_{i=1}^n D_i$ holds. Then, for all $\tau > 0$ and for every bounded function f such that the restriction $f|_{D_i}$ of f to D_i belongs to $\mathcal{W}^{1,q}(D_i)$ for every $i = 1, \dots, n$, where $q \geq 1$ independent of i , there exists a finite constant K such that*

$$|\Delta_\tau f| \leq K \left(\|f\|_\infty + \sum_{k=1}^d \sum_{i=1}^n \|\partial_k f|_{D_i}\|_\infty \right) \Delta_\tau^*. \quad (5.4)$$

Proof:

W. l. o. g. assume for all $i = 1, \dots, n$ that $D_i \subseteq (\mathbb{R}^+)^d$.

There exist $\mathbf{c}, \mathbf{C} \in (\mathbb{R}^+)^d$ with $] \mathbf{c}, \mathbf{C} [\subseteq D_i \subseteq [\mathbf{c}, \mathbf{C} [$. The Fubini argument from Figueroa-López [17, Lemma 3.2] remains valid. By virtue of Lemma 5.1.8, it holds that

$$\forall \mathbf{x} \in [\mathbf{c}, \mathbf{C} [: f(\mathbf{x}) = f(\mathbf{c}) + \sum_{k=1}^d \int_{c_k}^{x_k} \partial_k f|_{D_i}(\mathbf{x}_{<k}, u_k, \mathbf{c}_{>k}) du_k,$$

where $\mathbf{x}_{<k} = (x_1, \dots, x_{k-1})$ and $\mathbf{c}_{>k} = (c_{k+1}, \dots, c_d)$. Specifically, let Q be a σ -finite measure on $(\mathbb{R}^d, \mathcal{B}^d)$, then

$$\begin{aligned} & \int f(\mathbf{x}) \mathbb{1}_{D_i}(\mathbf{x}) Q(d\mathbf{x}) \\ &= \int_{D_i} \left(f(\mathbf{c}) + \sum_{k=1}^d \int_{c_k}^{x_k} \partial_k f|_{D_i}(\mathbf{x}_{<k}, u_k, \mathbf{c}_{>k}) du_k \right) Q(d\mathbf{x}) \\ &= f(\mathbf{c}) Q([\mathbf{c}, \mathbf{C} [) \\ & \quad + \sum_{k=1}^d \int_{c_k}^{C_k} \int_{c_1}^{C_1} \cdots \int_{c_{k-1}}^{C_{k-1}} \partial_k f|_{D_i}(\mathbf{x}_{<k}, u_k, \mathbf{c}_{>k}) Q(d\mathbf{x}_{<k}, [u_k, C_k] \times [\mathbf{c}_{>k}, \mathbf{C}_{>k} [) du_k. \end{aligned}$$

Although we deal with tail distributions rather than distribution functions in the sequel, we have an equation similar to Kallsen and Tankov [24, Definition 2.1]. In due form, we have

$$Q([\mathbf{a}, \mathbf{b} [) = \sum_{\mathbf{v} \in \{a_1, b_1\} \times \dots \times \{a_d, b_d\}} (-1)^{n(\mathbf{v})} Q(\mathcal{T}(\mathbf{v})),$$

where $n(\mathbf{v}) := \#\{k : v_k = b_k\}$.

5 Projection estimation away from zero

In particular, for the measures $P_{\mathbf{X}_\tau} := P(\mathbf{X}_\tau \in \cdot)$ and ν , we have

$$\begin{aligned} & |\Delta_\tau f \mathbb{1}_{D_i}| \\ &= \left| \tau^{-1} \mathbb{E}[f(\mathbf{X}_\tau) \mathbb{1}_{D_i}] - \nu(f \mathbb{1}_{D_i}) \right| \\ &\leq f(\mathbf{c}) \left| \tau^{-1} P_{\mathbf{X}_\tau}([\mathbf{c}, \mathbf{C}]) - \nu([\mathbf{c}, \mathbf{C}]) \right| \\ &\quad + \sum_{k=1}^d \int_{c_k}^{C_k} \partial_k f|_{D_i}(\boldsymbol{\xi}_{<k}(u_k), u_k, \mathbf{c}_{>k}) \left| \tau^{-1} P_{\mathbf{X}_\tau}([\mathbf{c}_{\hat{k}}(u_k), \mathbf{C}]) - \nu([\mathbf{c}_{\hat{k}}(u_k)]) \right| du_k, \end{aligned}$$

where $\mathbf{c}_{\hat{k}}(u_k) := (\mathbf{c}_{<k}, u_k, \mathbf{c}_{>k})$ and $\boldsymbol{\xi}_{<k}(u_k) \in [c_{<k}, C_{<k}[$ by virtue of the mean value theorem.

On the one hand, for all $\mathbf{c}_{\hat{k}}(u_k) \in D_i$, we have

$$\begin{aligned} & \left| \tau^{-1} P_{\mathbf{X}_\tau}([\mathbf{c}_{\hat{k}}(u_k), \mathbf{C}]) - \nu([\mathbf{c}_{\hat{k}}(u_k), \mathbf{C}]) \right| \\ &= \left| \sum_{\mathbf{v} \in \{c_1, C_1\} \times \dots \times \{u_k, C_k\} \times \dots \times \{c_d, C_d\}} (-1)^{n(\mathbf{v})} \Delta_\tau \mathbf{v} \right| \\ &\leq 2^d \Delta_\tau^*, \end{aligned}$$

on the other hand, we have

$$\int_{c_k}^{C_k} \partial_k f|_{D_i}(\boldsymbol{\xi}_{<k}(u_k), u_k, \mathbf{c}_{>k}) du_k \leq \left(\sup_{\mathbf{x}, \mathbf{y} \in D_i} \|\mathbf{y} - \mathbf{x}\| \right) \|\partial_k f|_{D_i}\|_\infty,$$

where the supremum is clearly finite, since \mathbb{D} is bounded.

Therefore, we conclude

$$|\Delta_\tau f \mathbb{1}_{D_i}| \leq 2^d (1 \vee \sup_{\mathbf{x}, \mathbf{y} \in D_i} \|\mathbf{y} - \mathbf{x}\|) (\|f\|_\infty + \sum_{k=1}^d \|\partial_k f|_{D_i}\|_\infty) \Delta_\tau^*$$

and, consequently,

$$|\Delta_\tau f| \leq 2^d n \left(\sup_{i=1, \dots, n} \left(1 \vee \sup_{\mathbf{x}, \mathbf{y} \in D_i} \|\mathbf{y} - \mathbf{x}\| \right) \right) \left(\|f\|_\infty + \sum_{k=1}^d \sum_{i=1}^n \|\partial_k f|_{D_i}\|_\infty \right) \Delta_\tau^*.$$

□

5.2 Brownian motion with compound Poisson jumps

This section is dedicated to the composition of a univariate Brownian motion with drift and compound Poisson jumps.

Assumption 5.2.1 (Finite Lévy measure) *Let ν be a finite measure.*

Let $\mathbb{D} = [D_{\min}, D_{\max}] \subseteq \mathbb{R}^+$ be a compact interval away from zero. We denote by

$$N_t := J([0, t] \times \mathbb{R}_o) \quad \text{for } t \in \mathbb{R}_0^+$$

the number of jumps up to time t . Then $(N_t)_{t \in \mathbb{R}_0^+}$ is a Poisson process with intensity $\lambda := \nu(\mathbb{R}_o) < \infty$. Under Assumption 5.2.1, X is called *Brownian motion with compound Poisson jumps and drift*. Recalling the Lévy-Khintchine representation (2.5), where the cut-off point for the small jumps was chosen arbitrary, the Lévy-Itô decomposition (2.8) simplifies to

$$X_t = \gamma_0 t + \sigma W_t + \sum_{k=1}^{N_t} Z_k \quad \text{for } t \in \mathbb{R}_0^+, \quad (5.5)$$

where $\{Z_k\}_{k \in \mathbb{N}} \sim \lambda^{-1} \nu$ i. i. d. and $\gamma_0 := \gamma - \int_{\{x: |x| \in]0, 1]\}} \nu(dx)$. Moreover, the Brownian motion W , the Poisson process N and the family $\{Z_k\}_{k \in \mathbb{N}}$ are independent.

Theorem 5.2.2 (Bias condition for (5.5)) *Let X be a Lévy process as in (5.5) such that $p \in \mathcal{C}^1(\mathbb{R}_o)$ with derivative p' that is bounded on \mathbb{D} . Then all \mathbb{R} -valued functions that meet the assumptions of Lemma 5.1.2 satisfy the (inclusive) bias condition $\overline{\text{BC}}1$.*

Proof:

In the light of Lemma 5.1.2, it is sufficient to show that $\Delta_t^* = O(t)$ as $t \rightarrow 0$. Let $y \in \mathbb{D}$ and $\eta \in]0, y[$, then by (5.1)

$$\begin{aligned} |\Delta_t y| &= \left| \frac{1}{t} P \left(\gamma_0 t + \sigma W_t + \sum_{i=1}^{N_t} Z_i \geq y \right) - \lambda P(Z_1 \geq y) \right| \\ &\leq \frac{1}{t} P(\sigma W_t \geq y - \gamma_0 t) e^{-\lambda t} + \lambda \left| P(\sigma W_t + Z_1 \geq y - \gamma_0 t) e^{-\lambda t} - P(Z_1 \geq y) \right| \\ &\quad + \frac{1}{t} \sum_{n=2}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} P \left(\sigma W_t + \sum_{i=1}^n Z_i \geq y - \gamma_0 t \right) =: A_0 + A_1 + A_2. \end{aligned}$$

Firstly, note that $A_2 \leq t^{-1} |e^{\lambda t} - 1 - \lambda t| \leq \lambda^2 t$. Secondly, setting $y'_t := y - \gamma_0 t$, we find

$$A_0 \leq \frac{1}{t} \bar{\Phi} \left(\frac{y'_t}{\sigma \sqrt{t}} \right) \leq \frac{1}{t} \bar{\Phi} \left(\frac{D_{\min} - \gamma_0 t}{\sigma \sqrt{t}} \right) = o(t^q) \text{ as } t \rightarrow 0 \quad (5.6)$$

5 Projection estimation away from zero

for every $q > 0$. Finally,

$$\begin{aligned}
A_1 &\leq \lambda |P(\sigma W_t + Y_1 \geq y'_t) - P(Y_1 \geq y'_t)| + |P(Y_1 \geq y'_t) - P(Y_1 \geq y)| \\
&\leq \int_0^\infty \varphi(u) \left(\int_{y'_t - \sigma\sqrt{t}u}^{y'_t} p(x) dx - \int_{y'_t}^{y'_t + \sigma\sqrt{t}u} p(x) dx \right) du + \left| \int_{y - \gamma_0 t}^y p(x) dx \right| \\
&= \int_0^\infty \varphi(u) I_t^{(1)}(u) du + |I_t^{(2)}(u)|.
\end{aligned}$$

Clearly, $|I_t^{(2)}(u)| \leq |\gamma_0| \|p\|_\infty t$. Since $p \in \mathcal{C}^1$ is assumed to have a bounded derivative on \mathbb{D} , by a Taylor expansion, we conclude

$$\begin{aligned}
I_t^{(1)}(u) &= \int_{y'_t - \sigma\sqrt{t}u}^{y'_t} p(y'_t) + (x - y'_t)p'(\xi_x) dx - \int_{y'_t}^{y'_t + \sigma\sqrt{t}u} p(y'_t) + (x - y'_t)p'(\xi_x) dx \\
&= \int_{y'_t - \sigma\sqrt{t}u}^{y'_t} (x - y'_t)p'(\xi_x) dx - \int_{y'_t}^{y'_t + \sigma\sqrt{t}u} (x - y'_t)p'(\xi_x) dx \\
&\leq 2\|p'\|_{\mathbb{D}, \infty} \sigma^2 u^2 t
\end{aligned}$$

for some $\xi_x \in]x - y'_t \wedge y'_t, x - y'_t \vee y'_t[$ for all $x \in [y'_t - \sigma\sqrt{t}u, y'_t + \sigma\sqrt{t}u]$.

Therefore,

$$A_1 \leq 2\|p'\|_{\mathbb{D}, \infty} \sigma^2 t \int_0^\infty \varphi(u) u^2 du + |\gamma_0| \|p\|_\infty t = O(t) \text{ as } t \rightarrow 0.$$

The bounds given for A_0 , A_1 and A_2 are independent of $y \in \mathbb{D}$. Consequently, for the supremum over all $y \in \mathbb{D}$, we get

$$\sup_{t>0} t^{-1} \Delta_t^* = \sup_{t>0} t^{-1} A_0 + \sup_{t>0} t^{-1} A_1 + \sup_{t>0} t^{-1} A_2 < \infty$$

or, equivalently, $\Delta_t^* = O(t)$ as $t \rightarrow 0$. Finally, Lemma 5.1.2 shows the validity of the inclusive bias condition $\overline{\text{BC}}1$. \square

Remark 5.2.3 *The result holds true if we choose*

$$\mathbb{D} = [-D_{\max}, -D_{\min}] \cup [D_{\min}, D_{\max}],$$

as the proof can be readily adapted for $y < 0$.

5.3 Brownian motion with α -stable jumps

This section is dedicated to the composition of a univariate Brownian motion with drift and an α -stable jump process.

Assumption 5.3.1 (α -stable Lévy density) *We assume there exist $\alpha \in]0, 2[$ and $C_1, C_2 \in \mathbb{R}_0^+$ with $C_1 + C_2 > 0$ such that*

$$\nu(dx) = \frac{C_1}{x^{1+\alpha}} \mathbb{1}_{\mathbb{R}^+}(x)dx + \frac{C_2}{|x|^{1+\alpha}} \mathbb{1}_{\mathbb{R}^-}(x)dx.$$

Let $\mathbb{D} = [D_{\min}, D_{\max}] \subseteq \mathbb{R}^+$ be a compact interval away from zero. Zolotarev [40, Theorem A, p. 4] uses the ‘‘Lévy canonical representation’’ for his parametrisations. Thus instead of the Lévy density p , the ‘‘Lévy spectral function’’ H is considered, where

$$H(x) = \left\{ \begin{array}{ll} -\nu([x, \infty[) & \forall x > 0, \\ \nu(]-\infty, x]) & \forall x < 0. \end{array} \right\} \quad (5.7)$$

Under Assumption 5.3.1, the parameters C_1^Z, C_2^Z used in Zolotarev [40, Theorem C, p. 7] satisfy $C_i^Z = \alpha^{-1}C_i$ for $i \in \{1, 2\}$. (The parameters in [40] are denoted as C_1, C_2 there. The additional ‘‘Z’’ stands for Zolotarev.)

In addition, Zolotarev uses 5 different parametrizations for the characteristic function. In order to clarify which form (namely ‘A’, ‘B’, ‘C’, ‘M’ or ‘E’) a set of parameters is based on, we denote the form together with the parameter set, e. g. $\mathcal{S}(\alpha, \beta, \delta, \varsigma; A)$ with shape α , skewness β , location δ and scale ς .

Under Assumption 5.3.1, X is the composition of a *Brownian motion with drift and an α -stable jump process*. Using Zolotarev [40, Theorem C.2, C.3 and Property 2.3], the Lévy-Itô decomposition (2.8) reduces to

$$X_t = \gamma t + \sigma W_t + S_t \quad \text{for } t \in \mathbb{R}_0^+, \quad (5.8)$$

where $(S_t)_{t \in \mathbb{R}_0^+} \sim \mathcal{S}(\alpha, \beta, \delta, \varsigma t; B)$ is an α -stable process (independent of W) with parameters $(\beta, \delta, \varsigma)$ depending on C_1, C_2 and α only. In the case of $\alpha \neq 1$, we have

$$\beta = \frac{2 \arctan\left(\frac{C_1 - C_2}{C_1 + C_2} \tan\left(\frac{\pi\alpha}{2}\right)\right)}{\pi K(\alpha)}, \quad (5.9)$$

$$\varsigma = \frac{C_1 + C_2}{\alpha} \Gamma(1 - \alpha) \frac{\sin\left(\frac{\pi(1-\alpha)}{2}\right)}{\cos\left(\frac{\pi}{2}\beta K(\alpha)\right)}, \quad (5.10)$$

$$\delta = -\frac{C_1 - C_2}{C_1 + C_2} \tan\left(\frac{\pi}{2}\alpha\right) \cos\left(\frac{\pi}{2}\beta K(\alpha)\right), \quad (5.11)$$

where $K(\alpha) = \alpha - 1 + \text{sgn}(1 - \alpha)$. In the case of $\alpha = 1$, the parameters are given by $\beta = (C_1 - C_2)/(C_1 + C_2)$, $\varsigma = C_1 + C_2$ and $\delta = 0$.

5 Projection estimation away from zero

Theorem 5.3.2 (Bias condition for (5.8)) *Let X be as in (5.8), then all functions $f : \mathbb{D} \rightarrow \mathbb{R}$ that meet the assumptions of Lemma 5.1.2 satisfy*

- (i) *the (exclusive) bias condition BC1 in the case of $\alpha = 1$ and $\beta \neq 0$,*
- (ii) *the (inclusive) bias condition $\overline{\text{BC1}}$ in all other cases.*

Proof:

Depending on the parameters α and β , we have to distinguish four cases regarding different asymptotic behaviour, particularly

- (I) $\alpha < 1$,
- (II) $\alpha > 1$ and $\beta \neq -1$,
- (III) $\alpha = 1$ and $\beta \neq -1$,
- (IV) $\alpha \geq 1$ and $\beta = -1$.

Similar to the proof of Theorem 5.2.2, it is sufficient to show, as $t \rightarrow 0$, that

$$\Delta_t^* = \begin{cases} O(t^\vartheta) \quad \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0, \\ O(t) & \text{in all other cases.} \end{cases}$$

Step 1: “About decomposing the Lévy process”

At first, let $Y \sim \mathcal{S}(\alpha, \beta, 0, 1; B)$. From Zolotarev [40, Property 2.1, p. 60], we know that for all $t \in \mathbb{R}_0^+$

$$S_t \stackrel{d}{=} \begin{cases} (\varsigma t)^{1/\alpha} Y + \varsigma \delta t & \text{for } \alpha \neq 1, \\ (\varsigma t) Y + \varsigma \beta \log((\pi \varsigma t)/2) t & \text{for } \alpha = 1. \end{cases}$$

Subsequently, let $\eta_t > 0$ with $\eta_t \rightarrow \infty$ and $\eta_t^3 \sqrt{t} \rightarrow 0$ as $t \rightarrow 0$. We need the former in Steps 2 and 3 and the latter condition on η_t in Step 5. Conditioning on the value of the Wiener part, we deduce

$$\begin{aligned} P(X_t \geq y) &= P(\gamma t + \sigma W_t + S_t \geq y) \\ &\leq \int_{-\eta_t}^{\eta_t} \varphi(u) P(Y \geq z(t, u)) \, du + 2\bar{\Phi}\left(\frac{\eta_t - \gamma t}{\sigma \sqrt{t}}\right), \end{aligned}$$

where

$$z(t, u) := \begin{cases} (y - u\sigma\sqrt{t} - (\gamma + \varsigma\delta)t) (\varsigma t)^{-1/\alpha} & \text{for } \alpha \neq 1, \\ (y - u\sigma\sqrt{t} - \gamma t + \varsigma\beta t \log(2/(\pi\varsigma t))) (\varsigma t)^{-1} & \text{for } \alpha = 1. \end{cases}$$

5.3 Brownian motion with α -stable jumps

Step 2: “About the asymptotic expansion of the stable tail distribution”

Since $\eta_t^3 \sqrt{t} \rightarrow 0$, we ensure that $z(t, u) \rightarrow +\infty$ as $t \rightarrow 0$ for all $u \in [-\eta_t, \eta_t]$. Clearly, there exists $t' > 0$ such that for all $t < t'$ we have for every $u \in B_{\eta_t}(0)$

$$\begin{aligned} y - u\sigma\sqrt{t} - (\gamma + \varsigma\delta)t &> 0 && \text{for } \alpha \neq 1, \\ y - u\sigma\sqrt{t} - \gamma t + \varsigma\beta t \log(2/(\pi\varsigma t)) &> 0 && \text{for } \alpha = 1. \end{aligned}$$

Even if $\beta < 0$, the latter holds true as $\lim_{t \rightarrow 0} t \log(1/t) = 0$. Hence, z diverges to “ $+\infty$ ” for every combination of parameters $\alpha, \beta, \varsigma, \delta, \gamma$ and σ .

Nevertheless, the tail probability $\bar{F}_Y(z) := P(Y > z)$ shows specific asymptotic behaviour for $z \rightarrow \infty$ depending on shape α and skewness β .

In detail,

$$\bar{F}_Y(z) \sim \begin{cases} \sum_{k=1}^{\infty} c_k z^{-k\alpha} & \text{in Cases I and II,} \\ \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{n!} g_k(\log z) z^{-k} & \text{in Case III,} \\ \frac{1}{\sqrt{2\pi\alpha\varsigma}} e^{-\zeta} \left(1 + \sum_{k=1}^{\infty} g_k(1/\alpha) \left(\frac{\alpha}{\varsigma}\right)^k \right) & \text{in Case IV,} \end{cases}$$

where

$$g_k(x) := \begin{cases} \sum_{l=0}^{k-1} c_{lk} x^l & \text{is a polynomial of degree } k-1 & \text{in Case III,} \\ \sum_{l=0}^{2k} c_{lk} x^l & \text{is a polynomial of degree } 2k & \text{in Case IV,} \end{cases}$$

and

$$\zeta = \zeta(z, \alpha) := \begin{cases} |1 - \alpha|(z/\alpha)^{\alpha/(\alpha-1)} & \text{for } \alpha > 1, \\ \exp(z - 1) & \text{for } \alpha = 1. \end{cases}$$

Firstly, in Case I, we have from Zolotarev [40, p. 89, Theorem 2.4.2] that the coefficients are given by

$$c_k = \frac{(-1)^{k-1} \Gamma(k\alpha + 1)}{\pi\alpha k\Gamma(k+1)} \sin\left(\frac{\pi}{2}k(\alpha + \beta K(\alpha))\right).$$

Secondly, in Case II, we have

$$c_k = \frac{\Gamma(\alpha k)}{\pi\Gamma(k+1)} \sin\left(\frac{\pi}{2}k(2 - \alpha)(1 + \beta)\right)$$

is shown by Zolotarev [40, p. 95]. Finally, in Case III, we have

$$c_{lk} = \sum_{m=l}^k \binom{k}{m} \binom{m}{l} (-1)^{m-l} \Gamma^{(m-l)}(k) \beta^m \left(\frac{\pi}{2}(1 + \beta)\right)^{k-m} \sin\left(\frac{\pi}{2}(k - m)\right)$$

5 Projection estimation away from zero

from Zolotarev [40, Theorem 2.5.4]. The exact formula for c_{lk} in Case IV can be derived from Zolotarev [40, Theorem 2.5.3]. We omit them, as they are dispensable for this proof. In the following, for fixed $\alpha \in [1, 2]$, we deal with $g_k(1/\alpha)$ as if it was a finite constant.

Step 3a: "About the summands of the asymptotic series in Cases I–III"

For every $k \in \mathbb{N}$, we conclude

$$z(t, u)^{-k\alpha} = \frac{\varsigma^k}{y^{k\alpha}} t^k (1 - h_t(u))^{-k\alpha},$$

where

$$h_t(u) := \begin{cases} (u\sigma\sqrt{t} + (\gamma + \varsigma\delta)t)y^{-1} & \text{for } \alpha \neq 1, \\ (u\sigma\sqrt{t} + \gamma t + \varsigma t\beta \log(2/(\pi\varsigma t))y^{-1} & \text{for } \alpha = 1. \end{cases}$$

Furthermore, by Taylor expansion of second order, we have

$$(1 - h_t(u))^{-k\alpha} = 1 + k\alpha h_t(u) + k\alpha(k\alpha + 1) \frac{h_t^2(u)}{2} + R_3(h_t(u); k).$$

Set $\kappa_t := \sup_{u \in B_{\eta_t}(0)} |h_t(u)|$. Since $\eta_t^3 \sqrt{t} \rightarrow 0$, we have $\kappa_t \rightarrow 0$. Hence, we consider the Lagrange form of the remainder term, i. e.

$$R_3(h_t(u); k) = \frac{h_t^3(u)}{6} k\alpha(k\alpha + 1)(k\alpha + 2) (\xi_{h_t(u)})^{-k\alpha-3}$$

for some $\xi_{h_t(u)} \in B_{h_t(u)}(1)$. Further, we can bound the remainder term above by

$$|R_3(h_t(u); k)| \leq \frac{\kappa_t^3}{6} k\alpha(k\alpha + 1)(k\alpha + 2) (1 - \kappa_t)^{-k\alpha-3}$$

uniformly on $B_{\eta_t}(0)$.

Moreover, in all cases, we have $(1 - \kappa_t)^{-k\alpha-3} = 1$. Besides, in Cases I and II, we conclude that

$$\begin{aligned} & \lim_{t \rightarrow 0} t^{-1} \kappa_t^3(k) \\ & \leq \lim_{t \rightarrow 0} \left\{ \sigma^3 \eta_t^3 \sqrt{t} + 3\sigma^2 |\gamma + \varsigma\delta| \eta_t^2 t + 3\sigma |\gamma + \varsigma\delta|^2 \eta_t t^{3/2} + |\gamma + \varsigma\delta|^3 t^2 \right\} = 0 \end{aligned}$$

and, equivalently, in Case III

$$\begin{aligned} & \lim_{t \rightarrow 0} t^{-1} \kappa_t^3 \\ & \leq \lim_{t \rightarrow 0} (\eta_t^3 \varsigma^3 \sqrt{t} + 3\eta_t^2 \varsigma^2 \gamma t + 3\eta_t^2 \varsigma^3 t \beta \log(2/(\pi\varsigma t)) + 3\eta_t \varsigma \gamma^2 t^{3/2}) \\ & \quad + \lim_{t \rightarrow 0} (6\eta_t \varsigma^2 t^{3/2} \gamma \beta \log(2/(\pi\varsigma t)) + 3\eta_t \varsigma^3 t^{3/2} \beta^2 \log^2(2/(\pi\varsigma t)) + \gamma^3 t^2) \\ & \quad + \lim_{t \rightarrow 0} (\gamma^2 t^2 \varsigma \beta \log(2/(\pi\varsigma t)) + 3\gamma \varsigma^2 t^2 \beta^2 \log^2(2/(\pi\varsigma t)) + \varsigma^3 t^2 \beta^3 \log^3(2/(\pi\varsigma t))) \\ & = 0, \end{aligned}$$

5.3 Brownian motion with α -stable jumps

since $\log(1/t) = o(t^{-q})$ as $t \rightarrow 0$ for all $q > 0$. Hence, we have $R_3(h_t(u); k) = o(t)$ as $t \rightarrow 0$ in all three cases.

Additionally, in Case III, we deduce

$$\log z(t, u) = \log \left(\frac{y}{\zeta t} \right) \log(1 - h_t(u)) \leq \kappa_t^* \log \left(\frac{y}{\zeta t} \right)$$

uniformly on $B_{\eta_t}(0)$, where $\kappa_t^* := \sup_{u \in B_{\eta_t}(0)} |\log(1 - h_t(u))| < \infty$. Equivalently, we observe that $g_k(\log z(t, u))$ is uniformly bounded above by

$$g_k^* \left(\frac{y}{\zeta t} \right) := \sum_{l=0}^{k-1} c_{lk} \left(\kappa_t^* \log \left(\frac{y}{\zeta t} \right) \right)^l.$$

Step 3b: "About the asymptotic series in Case IV"

Rewriting ζ as a function of u and t , we get

$$\zeta(u, t) = \begin{cases} |1 - \alpha| (y/\alpha)^{\alpha/(\alpha-1)} (\zeta t)^{-1/(\alpha-1)} (1 - h_t(u))^{\alpha/(\alpha-1)} & \text{for } \alpha > 1, \\ \exp\{y(\zeta t)^{-1} (1 - h_t(u)) - 1\} & \text{for } \alpha = 1, \end{cases}$$

with $h_t(u)$ defined as in Step 3a. Recalling $\kappa_t = \sup_{u \in B_{\eta_t}(0)} |h_t(u)| \rightarrow 0$ as $t \rightarrow 0$, we can uniformly bound ζ above and below on $B_{\eta_t}(0)$. In particular, we have

$$\underline{\zeta}(t) \leq \zeta(u, t) \leq \bar{\zeta}(t),$$

where

$$\underline{\zeta}(t) := \begin{cases} |1 - \alpha| (y/\alpha)^{\alpha/(\alpha-1)} (\zeta t)^{-1/(\alpha-1)} \underline{\kappa}_t & \text{for } \alpha > 1, \\ \exp\{y(\zeta t)^{-1} (1 - \kappa_t) - 1\} & \text{for } \alpha = 1, \end{cases}$$

and

$$\bar{\zeta}(t) := \begin{cases} |1 - \alpha| (y/\alpha)^{\alpha/(\alpha-1)} (\zeta t)^{-1/(\alpha-1)} \bar{\kappa}_t & \text{for } \alpha > 1, \\ \exp\{y(\zeta t)^{-1} (1 + \kappa_t) - 1\} & \text{for } \alpha = 1, \end{cases}$$

with $\bar{\kappa}_t := (1 - \kappa_t)^{\alpha/(\alpha-1)}$ and $\underline{\kappa}_t := (1 + \kappa_t)^{\alpha/(\alpha-1)}$. Consequently, since $\underline{\zeta}$ and $\bar{\zeta}$ share the same asymptotic, for every $\zeta_* \in \{\underline{\zeta}, \zeta, \bar{\zeta}\}$, we see

$$\zeta_*^{-1} = \begin{cases} O(t^{1/(\alpha-1)}) & \text{for } \alpha > 1, \\ O(t^q) \forall q > 0 & \text{for } \alpha = 1, \end{cases}$$

as $t \rightarrow 0$.

Step 4a: "About the convergence of $\Delta_t y$ in Cases I and II"

According to Step 3, $\lim_{t \rightarrow 0} t^{-1} P(X_t \geq y) = \nu([y, \infty])$ is equivalent to $c_1 \zeta = C_1/\alpha$. With the notation given above, in Case I, it holds that

$$c_1 = \frac{\Gamma(\alpha + 1)}{\pi \alpha} \sin \left(\frac{\pi}{2} (1 + \beta) \alpha \right) = +\pi^{-1} \Gamma(\alpha) \sin \left(\frac{\pi}{2} (1 + \beta) K(\alpha) \right), \quad (5.12)$$

5 Projection estimation away from zero

and, similarly, in Case II, we have

$$c_1 = \pi^{-1}\Gamma(\alpha) \sin\left(\frac{\pi}{2}(1+\beta)(2-\alpha)\right) = -\pi^{-1}\Gamma(\alpha) \sin\left(\frac{\pi}{2}(1+\beta)K(\alpha)\right). \quad (5.13)$$

Combining (5.12) and (5.13) with (5.10), we conclude that

$$c_{1\zeta} = \pm \frac{C_1 + C_2}{\alpha} \pi^{-1}\Gamma(\alpha)\Gamma(1-\alpha) \sin\left(\frac{\pi(1-\alpha)}{2}\right) \frac{\sin\left(\frac{\pi}{2}(1+\beta)K(\alpha)\right)}{\cos\left(\frac{\pi}{2}\beta K(\alpha)\right)}.$$

With $\pi^{-1}\Gamma(\alpha)\Gamma(1-\alpha) = (\sin(\pi\alpha))^{-1}$ and $\sin(\pi(1-\alpha)/2) = \cos(\pi\alpha/2)$, we get

$$c_{1\zeta} = \pm \frac{C_1 + C_2}{\alpha} \frac{\cos\left(\frac{\pi\alpha}{2}\right)}{\sin(\pi\alpha)} \left[\sin\left(\frac{\pi}{2}K(\alpha)\right) + \cos\left(\frac{\pi}{2}K(\alpha)\right) \tan\left(\frac{\pi}{2}\beta K(\alpha)\right) \right],$$

where (5.9) implies that

$$\tan\left(\frac{\pi}{2}\beta K(\alpha)\right) = \frac{C_1 - C_2}{C_1 + C_2} \tan\left(\frac{\pi\alpha}{2}\right).$$

As $K(\alpha) = \alpha - 1 + \text{sgn}(1-\alpha) \in \{\alpha, \alpha - 2\}$, we get $\pm \sin\left(\frac{\pi}{2}K(\alpha)\right) = +\sin\left(\frac{\pi}{2}\alpha\right)$ and $\pm \cos\left(\frac{\pi}{2}K(\alpha)\right) = +\cos\left(\frac{\pi}{2}\alpha\right)$. Finally, we the validity of

$$\begin{aligned} c_{1\zeta} &= \frac{C_1 + C_2}{\alpha} \left[\frac{\cos\left(\frac{\pi}{2}\alpha\right) \sin\left(\frac{\pi}{2}\alpha\right)}{\sin(\pi\alpha)} + \frac{C_1 - C_2}{C_1 + C_2} \frac{\cos^2\left(\frac{\pi}{2}\alpha\right) \tan\left(\frac{\pi}{2}\alpha\right)}{\sin(\pi\alpha)} \right] \\ &= \frac{C_1 + C_2}{\alpha} \left[\frac{1}{2} + \frac{C_1 - C_2}{2(C_1 + C_2)} \right] = \frac{C_1}{\alpha}. \end{aligned}$$

Step 4b: “About the convergence of $\Delta_t y$ in Case III”

Similarly to Step 4a, we need to show $c_{01\zeta} = \pi C_1$. By definition, $c_{01} = \pi(1+\beta)/2$. Hence,

$$c_{01\zeta} = \pi \frac{C_1}{C_1 + C_2}$$

by virtue of Zolotarev [40, Theorem C.2 and C.3]. Since $\zeta = C_1 + C_2$, we deduce that $c_{01\zeta} = \pi C_1$ holds.

Step 4c: “About the convergence of $\Delta_t y$ in Case IV”

In Case IV, the Lévy measure has no mass on the positive half-line, as Zolotarev [40, Theorem C.2 and C.3] implies $\beta = -1 \Leftrightarrow C_1 = 0$. Hence, it is sufficient to show $\Delta_t y \rightarrow 0$ as $t \rightarrow 0$. Considering the asymptotic expansion from Step 2 as

5.3 Brownian motion with α -stable jumps

a series in ζ , we observe that the dominating term is $e^{-\zeta}$, where we deduce from Step 3b that $\zeta \rightarrow \infty$ as $t \rightarrow 0$. Therefore, the convergence follows directly from the well-known asymptotic behaviour of the exponential function at “ $-\infty$ ”.

Step 5a: “About the rate of convergence of $\Delta_t y$ in Cases I–III”

To simplify notation, for $k \in \mathbb{N}$ and $t > 0$, we set

$$c'_k(t) := \begin{cases} c_k & \text{in Cases I and II,} \\ g_k^*(y/(\varsigma t))(k\pi)^{-1} & \text{in Case III,} \end{cases}$$

and note that c'_1 is independent of t . Let $y \in \mathbb{D}$, then

$$\begin{aligned} |\Delta_t y| &= \left| \frac{1}{t} P(X_t \geq y) - \nu([y, \infty]) \right| \\ &\leq \left| \frac{1}{t} \int_{-\eta_t}^{\eta_t} \varphi(u) \bar{F}_Y(z(t, u)) du + \frac{2}{t} \bar{\Phi} \left(\frac{\eta_t - \gamma t}{\sigma \sqrt{t}} \right) - \frac{C_1}{\alpha y^\alpha} \right| \\ &\leq \left| \frac{2}{t} \bar{\Phi} \left(\frac{\eta_t - \gamma t}{\sigma \sqrt{t}} \right) \right| + \left| \frac{1}{t} \int_{-\eta_t}^{\eta_t} \varphi(u) \bar{F}_Y(z) du - \frac{c'_1 \varsigma}{\alpha y^\alpha} \right| =: I_t^{(1)} + I_t^{(2)}, \end{aligned}$$

where, $I_t^{(1)} = o(t^q)$ as $t \rightarrow 0$ for all $q > 0$, equivalently to (5.6). Using the results from Step 2, 3 and 4a/b, we get

$$\begin{aligned} I_t^{(2)} &\leq \int_{-\eta_t}^{\eta_t} \varphi(u) \left(\frac{c'_1 \varsigma}{y^\alpha} (\alpha h_t(u) + \alpha(\alpha + 1) \frac{h_t^2(u)}{2} + R_3(h_t(u); 1)) \right) du \\ &\quad + \sum_{k=2}^{\infty} \frac{c'_k(t) \varsigma^k}{y^{k\alpha}} t^{k-1} \int_{-\eta_t}^{\eta_t} \varphi(u) \left(1 + k\alpha h_t(u) + k\alpha(k\alpha + 1) \frac{h_t^2(u)}{2} + R_3(h_t(u); k) \right) du \\ &\quad + \frac{2c'_1 \varsigma}{y^\alpha} \bar{\Phi}(\eta_t) \stackrel{(5.6)}{=} I_t^{(21)} + \sum_{k=2}^{\infty} I_t^{(2k)} + o(t^q). \end{aligned}$$

In Case III, let $\vartheta < 1$ if $\beta \neq 0$ and $\vartheta = 1$ if $\beta = 0$. Note that $\beta = 0 \Leftrightarrow C_1 = C_2$. Then, in Cases I and II,

$$\begin{aligned} \int_{-\eta_t}^{\eta_t} \varphi(u) h_t(u) du &= \sigma \sqrt{t} y^{-1} \int_{-\eta_t}^{\eta_t} u \varphi(u) du + (\gamma + \varsigma \delta) t y^{-1} \int_{-\eta_t}^{\eta_t} \varphi(u) du \\ &\leq y^{-1} (\gamma + \varsigma \delta) t \\ &= y^{-1} O(t) \quad \text{as } t \rightarrow 0, \end{aligned}$$

5 Projection estimation away from zero

and, in Case III,

$$\begin{aligned} \int_{-\eta_t}^{\eta_t} \varphi(u) h_t(u) du &= \sigma \sqrt{t} y^{-1} \int_{-\eta_t}^{\eta_t} u \varphi(u) du + (\gamma + \varsigma \beta \log(2/(\pi \varsigma t))) t y^{-1} \int_{-\eta_t}^{\eta_t} \varphi(u) du \\ &\leq y^{-1} (\gamma + (C_2 - C_1) \log(C_1 + C_2)) t + y^{-1} (C_2 - C_1) t \log(t) \\ &= y^{-1} O(t^\vartheta) \quad \text{as } t \rightarrow 0, \end{aligned}$$

since the normal density is an even function and all moments exist. Equivalently,

$$\begin{aligned} &\int_{-\eta_t}^{\eta_t} \varphi(u) h_t^2(u) du \\ &= \frac{\sigma^2 t}{y^2} \int_{-\eta_t}^{\eta_t} u^2 \varphi(u) du + \left\{ \begin{array}{l} \sigma(\gamma + \varsigma \delta) \\ \sigma(\gamma + \varsigma \beta \log(2/(\pi \varsigma t))) \end{array} \right\} \frac{t^{3/2}}{y^2} \int_{-\eta_t}^{\eta_t} u \varphi(u) du \\ &\quad + \left\{ \begin{array}{l} (\gamma + \varsigma \delta)^2 \\ (\gamma + \varsigma \beta \log(2/(\pi \varsigma t)))^2 \end{array} \right\} \frac{t^2}{y^2} \int_{-\eta_t}^{\eta_t} \varphi(u) du \\ &\leq \left\{ \begin{array}{l} y^{-2} \sigma^2 t + y^{-2} (\gamma + \varsigma \delta)^2 t^2 \\ y^{-2} \sigma^2 t + y^{-2} (\gamma + \varsigma \beta \log(2/(\pi \varsigma t)))^2 t^2 \end{array} \right\} = y^{-2} O(t) \end{aligned}$$

as $t \rightarrow 0$. Thereby,

$$I_t^{(21)} \leq \begin{cases} |y|^{-\alpha-2} O(t^\vartheta) \quad \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0 \\ |y|^{-\alpha-2} O(t) & \text{in all other cases.} \end{cases}$$

Similarly, for all $k > 2$, we have

$$I_t^{(2k)} = \begin{cases} |y|^{-\alpha-2} O(t^{k-2+\vartheta}) \quad \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0, \\ |y|^{-\alpha-2} O(t^{k-1}) & \text{in all other cases.} \end{cases}$$

Consequently,

$$I_t^{(2)} = \begin{cases} |y|^{-\alpha-2} O(t^\vartheta) \quad \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0, \\ |y|^{-\alpha-2} O(t) & \text{in all other cases.} \end{cases}$$

Hence,

$$|\Delta_t y| = \begin{cases} |y|^{-\alpha-2} O(t^\vartheta) \quad \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0. \\ |y|^{-\alpha-2} O(t) & \text{in all other cases,} \end{cases}$$

as $t \rightarrow 0$.

5.3 Brownian motion with α -stable jumps

Step 5b: "About the rate of convergence of $\Delta_t y$ in Case IV"

Let $y \in \mathbb{D}$, then

$$\begin{aligned} |\Delta_t y| &= \left| \frac{1}{t} P(X_t \geq y) - \nu([y, \infty[) \right| \\ &\leq \left| \frac{2}{t} \bar{\Phi} \left(\frac{\eta_t - \gamma t}{\sigma \sqrt{t}} \right) \right| + \left| \frac{1}{t} \int_{-\eta_t}^{\eta_t} \varphi(u) \bar{F}_Y(z(u, t)) du \right| =: I_t^{(1)} + I_t^{(2)}, \end{aligned}$$

where $I_t^{(1)} = o(t^q)$ as $t \rightarrow 0$ for all $q > 0$, equivalently to (5.6). Moreover, from Step 3b, we deduce that

$$|\bar{\Psi}(z)| \leq \frac{1}{\sqrt{2\pi\alpha\underline{\zeta}(t)}} e^{-\underline{\zeta}(t)} \left(1 + \sum_{k=1}^{\infty} g_k(1/\alpha) \alpha^k \underline{\zeta}^{-k}(t) \right).$$

As the right-hand side is independent of u , we integrate $I_t^{(2)}$ independent of $\bar{\Psi}(z)$. Hence,

$$tI_t^{(2)} \leq \frac{1}{\sqrt{2\pi\alpha\underline{\zeta}(t)}} e^{-\underline{\zeta}(t)} \left(1 + \sum_{k=1}^{\infty} g_k(1/\alpha) \alpha^k \underline{\zeta}^{-k}(t) \right).$$

Similarly to $I_t^{(1)}$, we conclude that $tI_t^{(2)} = o(\underline{\zeta}^{-q})$ for every $q > 0$. Recalling from Step 3b that

$$\underline{\zeta}^{-1} = \begin{cases} O(t^{1/(\alpha-1)}) & \text{for } \alpha > 1, \\ O(t^q) \forall q > 0 & \text{for } \alpha = 1, \end{cases}$$

as $t \rightarrow 0$, we conclude, finally, that $I_t^{(2)} = O(t^q)$ for every $q > 0$. In summary, we have $|\Delta_t y| = O(t^q)$ for every $q > 0$ as $t \rightarrow 0$.

Step 6: "Rate of convergence for the supremum Δ_t^ "*

Since the domain of estimation $\mathbb{D} = [D_{\min}, D_{\max}]$ is separated from the origin, we conclude that $\sup_{y \in \mathbb{D}} |y|^{-\alpha-2} = D_{\min}^{-\alpha-2} < \infty$. Finally

$$\Delta_t^* = \sup_{y \in \mathbb{D}} |\Delta_t y| = \begin{cases} O(t^\vartheta) \forall \vartheta < 1 & \text{in Case III and } \beta \neq 0, \\ O(t^q) \forall q > 0 & \text{in Case IV (as } y > 0), \\ O(t) & \text{in all other cases,} \end{cases}$$

as $t \rightarrow 0$. □

Remark 5.3.3 *The result holds true if we choose*

$$\mathbb{D} = [-D_{\max}, -D_{\min}] \cup [D_{\min}, D_{\max}].$$

Clearly, let $Y_1 \sim \mathcal{S}(\alpha, \beta)$ and $Y_2 \sim \mathcal{S}(\alpha, -\beta)$, then Zolotarev [40, Property 2.3] implies $P(Y_1 < y) = P(Y_2 \geq |y|)$ for all $y < 0$.

5.4 General Lévy processes

The two previous sections were dealing with Lévy processes with two specific and well-known jump parts. For general Lévy processes, we derive a bias condition depending on the variation of its sample paths. The following definition is adapted from Blumenthal and Gettoor [6].

Definition 5.4.1 (Blumenthal-Gettoor index) *Let X be a Lévy process with generating triplet (γ, Σ, ν) . Set*

$$\mathcal{I}_X := \left\{ r \in \mathbb{R} : \int_{\mathbb{R}^d} (\|x\|^r \wedge 1) \nu(dx) < \infty \right\}, \quad (5.14)$$

then $\beta_X := \inf \mathcal{I}_X$ is called its Blumenthal-Gettoor index.

The set \mathcal{I}_X is an interval of form $] \beta_X, \infty[$ or $[\beta_X, \infty[$. From Definition 2.2.4, we deduce that $2 \in \mathcal{I}_X$ holds for every Lévy process X .

Example 5.4.2 (Blumenthal-Gettoor index) *The Lévy processes, introduced and analysed in Sections 5.2 and 5.3, are very distinctive. In particular, let X be a compound Poisson process, then $\beta_X = 0$. Conversely, let X be an α -stable process with $\alpha \in]0, 2[$, then $\beta_X = \alpha$.*

Let us remark the following. By definition, a Brownian component has no impact on the Blumenthal-Gettoor index. Nevertheless, the path behaviour of a Brownian motion (seen as a 2-stable process) is similar to the path behaviour of a pure jump Lévy process X with $\beta_X = 2$.

Univariate framework

Theorem 5.4.3 (Bias condition for integral estimation in DT)

Let X be a univariate Lévy process with generating triplet (γ, σ^2, ν) and assume that Assumption 2.3.1 holds. In addition, let $\mathbb{D} \in \mathcal{B}_\circ$ be bounded and the union of finitely many intervals away from zero. Moreover, let $\mathbb{S} \subseteq L^2(\mathbb{D}, \lambda_{\mathbb{D}})$ be a finite dimensional projection space such that all $f \in \mathbb{S}$ meet the assumptions of Lemma 5.1.2. Then all functions $f \in \mathbb{S}$ satisfy

- (i) the bias condition (4.8) BC^{1/2} in general,
- (ii) the bias condition (4.8) $\overline{\text{BC}}1$ if $\sigma = 0$ and $\beta_X < 1$,
- (iii) the bias condition (4.8) BC^{1/β_X} if $\sigma = 0$ and $1 \leq \beta_X \leq 2$.

To prove Theorem 5.4.3, we need the following results that play an important role in the proof afterwards. Firstly, the following Lemma is essentially Rüschemdorf and Woerner [34, Corollary 3.2].

Lemma 5.4.4 (Small-jumps tail estimate) *Let X be a univariate Lévy process with generating triplet (γ, σ^2, ν) . Further, let δ be the smallest non-negative number such that $\{x : |x| \leq \delta\}$ contains the support of its Lévy measure, which should not be identically zero. Then, for all $n \geq 1$ and $\varepsilon > n\delta$, we have*

$$\sup_{y \geq \varepsilon} P(|X_t| \geq y) = O_{\varepsilon, \delta}(t^n) \quad (5.15)$$

as $t \rightarrow 0$.

Secondly, we confer Millar [32, Theorem 2.1] for the following Lemma.

Lemma 5.4.5 (Truncated Lévy r -moment estimate) *Let X be an \mathbb{R}^d -valued Lévy process with generating triplet $(\gamma, 0, \nu)$. The Lévy measure ν is assumed to have bounded support. Further, let $r \in \mathcal{I}_X \cap]0, 2]$, then there is a constant $C(r) < \infty$ depending on the Lévy measure ν and r only such that*

$$\forall t \in [0, 1] : E \|X_t\|^r \leq C(r)t. \quad (5.16)$$

Millar [32] proves in a remark following Theorem 2.1 that (5.16) remains valid if ν has unbounded support as long as $E\|X_t\|^r$ exists, i. e.

$$\int_{\mathbb{R}_0^d} \|x\|^r \nu(dx) < \infty.$$

The condition $t \in [0, 1]$ can be relaxed to $t \in \mathbb{R}_0^+$ as long as either $r \leq 1$ or $EX_1 = 0$. Invoking Lemma 5.4.4 and 5.4.5, we prove Theorem 5.4.3.

Proof of Theorem 5.4.3:

By choice of \mathbb{D} , there is an $\varepsilon > 0$ such that $|y| > \varepsilon$ for all $y \in \mathbb{D}$. W.l.o.g. assume $y > 0$.

Step 1: “About decomposing the Lévy process”

Let $\eta \in]0, \varepsilon/2[$ and, for $t \in \mathbb{R}_0^+$, we set

$$\left. \begin{aligned} Y_t^\eta &:= \iint_{[0, t] \times \mathbb{R}_0 \setminus [-\eta, \eta]} x J(ds, dx) = \sum_{s \leq t} \Delta X_s \mathbb{1}_{|\Delta X_s| > \eta}, \\ X_t^\eta &:= X_t - Y_t^\eta, \end{aligned} \right\} \quad (5.17)$$

then Y^η in (5.17) is referred to as *big-jumps process* of X (w. r. t. the threshold η) and X^η in (5.17) is referred to as *truncated Lévy process* of X . Both, X^η and Y^η ,

5 Projection estimation away from zero

are Lévy processes with Lévy measures $\nu_{X^\eta} := \mathbb{1}_{[-\eta, \eta]}\nu$ and $\nu_{Y^\eta} := \mathbb{1}_{\mathbb{R}_\circ \setminus [-\eta, \eta]}\nu$, respectively. Moreover, let $N_t^\eta := J([0, t] \times \mathbb{R}_\circ \setminus [-\eta, \eta])$ count the number of “big jumps” up to time t . Then there is a family $(Z_k^\eta)_{k \in \mathbb{N}} \sim \frac{\nu_{Y^\eta}}{\nu_{Y^\eta}(\mathbb{R}_\circ)}$ of i. i. d. random variables independent of N^η such that

$$Y_t^\eta = \sum_{k=1}^{N_t^\eta} Z_k \quad \text{for } t \in \mathbb{R}_0^+.$$

We observe that Y^η is a compound Poisson process with rate

$$\lambda_\eta := \nu_{Y^\eta}(\mathbb{R}_\circ) = \nu(\mathbb{R}_\circ \setminus [-\eta, \eta]).$$

Step 2: “Conditioning on the number of big jumps”

Similarly to our approach in Section 5.2, we condition on N^η and conclude that

$$\left. \begin{aligned} |\Delta_\tau y| &= \left| \frac{1}{t} P(X_t^\eta + Y_t^\eta \geq y) - \nu([y, \infty]) \right| \\ &\leq \frac{1}{t} P(X_t^\eta \geq y) e^{-\lambda_\eta t} \\ &\quad + \lambda_\eta \left| P(X_t^\eta + Z_1^\eta \geq y) e^{-\lambda_\eta t} - P(Z_1^\eta \geq y) \right| \\ &\quad + \frac{1}{t} \sum_{n=2}^{\infty} e^{-\lambda_\eta t} \frac{(\lambda_\eta t)^n}{n!} P\left(X_t^\eta + \sum_{k=1}^n Z_k^\eta \geq y\right) \\ &=: A_0 + A_1 + A_2 \end{aligned} \right\} \quad (5.18)$$

Step 3: “Upper bounds for A_0 and A_2 ”

Firstly, note that

$$A_2 \leq t^{-1} |e^{\lambda_\eta t} - 1 - \lambda_\eta t| \leq \lambda_\eta^2 t = O_\eta(t). \quad (5.19)$$

Secondly, since $\varepsilon > 2\eta$, we get

$$A_0 \leq t^{-1} P(X_t^\eta \geq y) + o_\eta(t) = t^{-1} O_{\varepsilon, \eta}(t^2) = O_{\varepsilon, \eta}(t) \quad (5.20)$$

by virtue of Lemma 5.4.4.

Step 4: “General upper bound for A_1 ”

We obviously get

$$\begin{aligned} A_1 &= \lambda_\eta \left| P(X_t^\eta + Z_1^\eta \geq y) e^{-\lambda_\eta t} - P(Z_1^\eta \geq y) \right| \\ &\leq \left| \int P(X_t^\eta \geq y - u) \nu_{Y^\eta}(du) - \nu_{Y^\eta}([y, \infty]) \right| + o_\eta(t) \\ &= \left| \int_{-\infty}^y P(X_t^\eta \geq y - u) \nu_{Y^\eta}(du) - \int_y^\infty P(X_t^\eta < y - u) \nu_{Y^\eta}(du) \right| + o_\eta(t). \end{aligned}$$

Furthermore, we see that

$$\sup_{y \geq \varepsilon} \left| \int_{-\infty}^{-\eta} P(X_t^\eta \geq y - u) \nu_{Y^\eta}(du) \right| \leq \lambda_\eta P(|X_t^\eta| > \eta) = o_\eta(t) \text{ as } t \rightarrow 0$$

by Sato [35, Corollary 8.9]. Likewise, we have

$$\begin{aligned} \sup_{y \geq \varepsilon} \left| \int_{\eta}^{y-\eta} P(X_t^\eta \geq y - u) \nu_{Y^\eta}(du) \right| &\leq \lambda_\eta P(|X_t^\eta| > \eta) = o_\eta(t), \\ \sup_{y \geq \varepsilon} \left| \int_{y+\eta}^{\infty} P(X_t^\eta < y - u) \nu_{Y^\eta}(du) \right| &\leq \lambda_\eta P(|X_t^\eta| > \eta) = o_\eta(t), \end{aligned}$$

since $\nu_{X^\eta}([\eta, \infty]) = 0$. Accordingly, since ν_{Y^η} does not charge $B_\eta(0)$, A_1 is bounded above by

$$A_1 \leq \left| \int_{y-\eta}^y P(X_t^\eta \geq y - u) \nu_{Y^\eta}(du) - \int_y^{y+\eta} P(X_t^\eta < y - u) \nu_{Y^\eta}(du) \right| + o_\eta(t)$$

or, equivalently, by change of variable $z = y - u$ and making use of the finiteness of ν_{Y^η} , we have

$$A_1 \leq \lambda_\eta \left| \int_{-\eta}^0 P(X_t^\eta < z) dz - \int_0^\eta P(X_t^\eta \geq z) dz \right| + o_\eta(t).$$

as $t \rightarrow 0$.

Moreover, since the Lévy measure ν_{X^η} of the truncated Lévy process X^η has no mass outside a (specific) compact set, all its moments are known to exist by virtue of Sato [35, Theorem 25.3]. Let $r > 0$. Then using Markov's inequality, i. e.

$$P(|X_t| \geq z) \leq \frac{\mathbb{E}[g(|X_t|)]}{g(z)}$$

for all monotonically increasing functions $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ with $g(z) > 0$ for all $z > 0$, we conclude with $g(z) = |z|^r$ that A_1 is once more bounded above by

$$A_1 \leq \lambda_\eta \mathbb{E}|X_t^\eta|^r \int_{-\eta}^{\eta} |z|^{-r} dz + o_\eta(t).$$

Clearly, there is a finite constant $C_1(r) < \infty$ such that

$$\int_{-\eta}^{\eta} |z|^{-r} dz < C_1(r),$$

5 Projection estimation away from zero

if and only if $r < 1$. Thereby, we have $A_1 \leq \lambda_\eta C_1(r) \mathbb{E}|X_t^\eta|^r + o_\eta(t)$.

Step 5: “Proof of Theorem 5.4.3”

Let $\vartheta \in [0, 1/\beta_X \wedge 1[$. Then there is an $r \in [0, 1[$ such that $\beta_X < r/\vartheta < 2$. Firstly, since $|\cdot|^r$ is submultiplicative, secondly, by Jensen’s inequality for the concave function $|\cdot|^\vartheta$ and, finally, due to Lemma 5.4.5, we have

$$\begin{aligned} \mathbb{E}|X_t^\eta|^r &\leq \mathbb{E}[(|\sigma W_t| + |X_t^\eta - \sigma W_t|)^r] \leq \mathbb{E}|\sigma W_t|^r + \mathbb{E}|X_t^\eta - \sigma W_t|^r \\ &\leq \left(\mathbb{E}|\sigma W_t|^2 \right)^{r/2} + \left(\mathbb{E}|X_t^\eta - \sigma W_t|^{r/\vartheta} \right)^\vartheta \leq \sigma^r t^{r/2} + C_2 t^\vartheta \end{aligned}$$

for a finite positive constant $C_2 < \infty$ depending only on r and ϑ . Consequently, for all $t \in [0, 1]$, we get

$$\begin{aligned} A_1 &\leq \lambda_\eta C_1(r) \left(\sigma^r t^{r/2} + C_2 t^\vartheta \right) + o_\eta(t) \\ &= \left\{ \begin{array}{ll} O_{\eta, \beta_X}(t^\vartheta) & \text{if } \sigma = 0, \\ O_{\eta, \sigma}(t^{r/2}) \forall r \in [0, 1[& \text{if } \sigma > 0, \end{array} \right\} \end{aligned} \quad (5.21)$$

as $t \rightarrow 0$. The preliminaries of Lemma 5.4.5 in the case of $\beta_X < 1$ and $\sigma = 0$ are also met if $\vartheta = 1$.

As the bounds (5.19), (5.20) and (5.21) for the terms on the right-hand side of (5.18) are uniform in y , we get in case of $\beta_X < 1$ and $\sigma = 0$ that $\Delta_t^* = O_{\varepsilon, \eta}(t)$, in case of $1 \leq \beta_X \leq 2$ and $\sigma = 0$ that $\Delta_t^* = O_{\varepsilon, \eta, \beta_X}(t^\vartheta)$ for all $\vartheta \in [0, 1/\beta_X[$ and in case of $\sigma > 0$ that $\Delta_t^* = O_{\varepsilon, \eta, \sigma}(t^{r/2})$ for all $r \in [0, 1[$. This is sufficient for Theorem 5.4.3 by virtue of Lemma 5.1.2. \square

Multivariate framework

In this subsection, we make use of the notation from Kallsen and Tankov [24] as introduced in Section 5.1. Similarly, we write vectors in bold font again.

Theorem 5.4.6 (BC for integral estimation in DT) *Let \mathbf{X} be an \mathbb{R}^d -valued Lévy process with generating triplet $(\boldsymbol{\gamma}, \Sigma, \nu)$ and assume that Assumption 2.3.1 holds. In addition, let $\mathbb{D} \in \mathcal{B}((\mathbb{R} \setminus \{0\})^d)$ be bounded and the union of finitely many d -dimensional intervals away from the origin. Moreover, let $\mathbb{S} \subseteq L^2(\mathbb{D}, \lambda_{\mathbb{D}})$ be a finite dimensional projection space such that all $f \in \mathbb{S}$ meet the assumptions of Lemma 5.1.10. Then all functions $f \in \mathbb{S}$ satisfy*

- (i) the bias condition (4.8) $\text{BC}^{1/2}$ in general,
- (ii) the bias condition (4.8) $\overline{\text{BC}}1$ if $\Sigma = \mathbf{0}$ and $\beta_{\mathbf{X}} < 1$,
- (iii) the bias condition (4.8) $\text{BC}^{1/\beta_{\mathbf{X}}}$ if $\Sigma = \mathbf{0}$ and $1 \leq \beta_{\mathbf{X}} \leq 2$.

Proof:

Since all norms in a finite dimensional space (like \mathbb{R}^d) are equivalent, we use the supremum norm in the following unless explicitly stated otherwise. For instance, $B_\eta(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_\infty < \eta\}$ is the η -ball w.r.t. the supremum norm. The domain of estimation \mathbb{D} is away from the origin. Thus, there is an $\varepsilon > 0$ such that $\|\mathbf{y}\|_\infty > \varepsilon$ for all $\mathbf{y} \in \mathbb{D}$. W.l.o.g. assume $\mathbf{y} \in (\mathbb{R}^+)^d$.

Step 1: “About decomposing the Lévy process”

Let $\eta \in]0, \varepsilon/2[$ and, for $t \in \mathbb{R}_0^+$, we set

$$\left. \begin{aligned} \mathbf{Y}_t^\eta &:= \iint_{[0,t] \times \mathbb{R}_\circ^d \setminus \overline{B_\eta(\mathbf{0})}} \mathbf{x} J(ds, d\mathbf{x}), \\ \mathbf{X}_t^\eta &:= \mathbf{X}_t - \mathbf{Y}_t^\eta, \end{aligned} \right\} \quad (5.22)$$

equivalently to the univariate case. Both, \mathbf{X}^η and \mathbf{Y}^η , are Lévy processes, with Lévy measures $\nu_{\mathbf{X}^\eta} := \mathbb{1}_{\overline{B_\eta(\mathbf{0})}} \nu$ and $\nu_{\mathbf{Y}^\eta} := \mathbb{1}_{\mathbb{R}_\circ^d \setminus \overline{B_\eta(\mathbf{0})}} \nu$, respectively. Moreover, let $N_t^\eta := J([0, t] \times \mathbb{R}_\circ^d \setminus \overline{B_\eta(\mathbf{0})})$ count the number of jumps with at least one “big” component up to time t . Then there is a family $(\mathbf{Z}_k^\eta)_{k \in \mathbb{N}} \sim \frac{\nu_{\mathbf{Y}^\eta}}{\nu_{\mathbf{Y}^\eta}(\mathbb{R}_\circ^d)}$ of i. i. d. random vectors independent of N^η such that

$$\mathbf{Y}_t^\eta = \sum_{k=1}^{N_t^\eta} \mathbf{Z}_k \quad \text{for } t \in \mathbb{R}_0^+.$$

We observe that \mathbf{Y}^η is a compound Poisson process with rate

$$\lambda_\eta := \nu_{\mathbf{Y}^\eta}(\mathbb{R}_\circ^d) = \nu(\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0})).$$

Step 2: “Conditioning on the number of big jumps”

Similarly to the univariate case, we condition on the number of big jumps N^η and conclude that

$$\left. \begin{aligned} |\Delta_\tau \mathbf{y}| &= \left| \frac{1}{t} P(\mathbf{X}_t^\eta + \mathbf{Y}_t^\eta \geq \mathbf{y}) - \nu([\mathbf{y}, \infty[) \right| \\ &\leq \frac{1}{t} P(\mathbf{X}_t^\eta \geq \mathbf{y}) e^{-\lambda_\eta t} \\ &\quad + \lambda_\eta \left| P(\mathbf{X}_t^\eta + \mathbf{Z}_1^\eta \geq \mathbf{y}) e^{-\lambda_\eta t} - P(\mathbf{Z}_1^\eta \geq \mathbf{y}) \right| \\ &\quad + \frac{1}{t} \sum_{n=2}^{\infty} e^{-\lambda_\eta t} \frac{(\lambda_\eta t)^n}{n!} P\left(\mathbf{X}_t^\eta + \sum_{k=1}^n \mathbf{Z}_k^\eta \geq \mathbf{y}\right) \\ &=: A_0 + A_1 + A_2 \end{aligned} \right\} \quad (5.23)$$

5 Projection estimation away from zero

Step 3: “Upper bounds for A_0 and A_2 ”

Firstly, note that

$$A_2 \leq t^{-1}|e^{\lambda_\eta t} - 1 - \lambda_\eta t| \leq \lambda_\eta^2 t = O_\eta(t). \quad (5.24)$$

Secondly, since $\varepsilon > 2\eta$, there exists an index $k = k(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{D}$ such that $y_k > \eta$. Hence, we get

$$\left. \begin{aligned} A_0 &\leq \frac{1}{t} P\left(X_{1,t}^\eta \geq y_1, \dots, X_{d,t}^\eta \geq y_d\right) + o_\eta(t) \\ &\leq \frac{1}{t} P\left(|X_{k,t}^\eta| \geq \eta\right) + o_\eta(t) = t^{-1} O_{\varepsilon,\eta}(t^2) + o_\eta(t) = O_{\varepsilon,\eta}(t) \end{aligned} \right\} \quad (5.25)$$

by virtue of Lemma 5.4.4.

Step 4: “General upper bound for A_1 ”

Analogously to the univariate case, we get

$$\begin{aligned} A_1 &= \lambda_\eta \left| P(\mathbf{X}_t^\eta + \mathbf{Z}_1^\eta \geq \mathbf{y}) e^{-\lambda_\eta t} - P(\mathbf{Z}_1^\eta \geq \mathbf{y}) \right| \\ &\leq \left| \int P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \nu_{\mathbf{Y}^\eta}(\mathrm{d}\mathbf{u}) - \nu_{\mathbf{Y}^\eta}([\mathbf{y}, \infty[) \right| + o_\eta(t) \\ &= \left| \int_{\mathbb{R}_\circ^d \setminus [\mathbf{y}, \infty[} P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \nu_{\mathbf{Y}^\eta}(\mathrm{d}\mathbf{u}) - \int_{[\mathbf{y}, \infty[} P(\exists k : X_{k,t}^\eta < y_k - u_k) \nu_{\mathbf{Y}^\eta}(\mathrm{d}\mathbf{u}) \right| \\ &\quad + o_\eta(t). \end{aligned}$$

We split $\mathbb{R}_\circ^d \setminus [\mathbf{y}, \infty[$ into $E_0 := [\mathbf{y} - \eta \mathbf{1}, \infty[\setminus [\mathbf{y}, \infty[$ and a partition E_1, \dots, E_d of $\mathbb{R}_\circ^d \setminus [\mathbf{y} - \eta \mathbf{1}, \infty[$, defined through $E_1 := \{\mathbf{u} \in \mathbb{R}_\circ^d : y_1 - u_1 > \eta\}$ and

$$E_k := \left\{ \mathbf{u} \in \mathbb{R}_\circ^d \setminus \bigcup_{l=1}^{k-1} E_l : y_k - u_k > \eta \right\}, \quad k = 2, \dots, d.$$

Then we have $P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \leq P(|X_{k,t}^\eta| > \eta)$ for all $\mathbf{u} \in E_k$. Moreover, since $\nu_{\mathbf{Y}^\eta}$ is a finite measure, we conclude for each $k = 1, \dots, d$ that

$$\sup_{\mathbf{y} \in \mathbb{D}} \left| \int_{E_k(\mathbf{y})} P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \nu_{\mathbf{Y}^\eta}(\mathrm{d}\mathbf{u}) \right| \leq \lambda_\eta P(|X_{k,t}^\eta| > \eta) = o_\eta(t),$$

since $\lim_{t \rightarrow 0} t^{-1} P(|X_{k,t}^\eta| \geq \eta) = \nu_{\mathbf{X}^\eta}(\{\mathbf{x} \in \mathbb{R}^d : |x_k| > \eta\}) = 0$ in coherence with Sato [35, Corollary 8.9].

Furthermore, since $P(\exists k : X_{k,t}^\eta < y_k - u_k) \leq \sum_{k=1}^d P(X_{k,t}^\eta < y_k - u_k)$ is valid in general, we get

$$A_1 \leq \left| \int_{E_0} P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \nu_{\mathbf{Y}^\eta}(\mathbf{d}\mathbf{u}) \right| + \sum_{k=1}^d \left| \int_{[\mathbf{y}, \infty[} P(X_{k,t}^\eta < y_k - u_k) \nu_{\mathbf{Y}^\eta}(\mathbf{d}\mathbf{u}) \right| + o_\eta(t).$$

We split E_0 into E'_1, \dots, E'_d defined through $E'_1 := \{\mathbf{u} \in E_0 : u_1 < y_1\}$ and

$$E'_k := \left\{ \mathbf{u} \in E_0 \setminus \bigcup_{l=1}^{k-1} E'_l : u_k < y_k \right\}, \quad k = 2, \dots, d.$$

Therefore, we conclude that

$$\begin{aligned} & \left| \int_{E'_k} P(\mathbf{X}_t^\eta \geq \mathbf{y} - \mathbf{u}) \nu_{\mathbf{Y}^\eta}(\mathbf{d}\mathbf{u}) \right| \\ & \leq \left| \int_{y_k - \eta}^{y_k} P(X_{k,t}^\eta \geq y_k - u_k) \nu_{\mathbf{Y}^\eta}(\mathbf{d}u_k, \mathcal{T}(\mathbf{y}_{<k} - \eta \mathbf{1}) \times \mathcal{T}(\mathbf{y}_{>k} - \eta \mathbf{1})) \right| \\ & \leq \lambda_\eta \left| \int_0^\eta P(X_{k,t}^\eta \geq z) dz \right| \end{aligned}$$

for all $k = 1, \dots, d$, where we observe

$$\lambda_\eta \int_0^\eta P(X_{k,t}^\eta \geq z) dz \leq \lambda_\eta \mathbb{E}|X_{k,t}^\eta|^r \int_0^\eta |z|^{-r} dz$$

using Markov's inequality, similar to the univariate case. Additionally, we conclude that

$$\left| \int_{[\mathbf{y}, \infty[} P(X_{k,t}^\eta < y_k - u_k) \nu_{\mathbf{Y}^\eta}(\mathbf{d}\mathbf{u}) \right| \leq \lambda_\eta \int_{-\eta}^0 P(X_{k,t}^\eta < z) dz + \lambda_\eta P(X_{k,t}^\eta < -\eta)$$

is valid for all $k = 1, \dots, d$, where equivalently

$$\lambda_\eta \int_{-\eta}^0 P(X_{k,t}^\eta < z) dz + \lambda_\eta P(X_{k,t}^\eta < -\eta) \leq \lambda_\eta \mathbb{E}|X_{k,t}^\eta|^r \int_0^\eta |z|^{-r} dz + o_\eta(t)$$

follows. Finally, since $|X_{k,t}^\eta| \leq \|\mathbf{X}_t^\eta\|_\infty$ by definition, we deduce that A_1 is bounded above by

$$A_1 \leq \lambda_\eta \mathbb{E}\|\mathbf{X}_t^\eta\|_\infty^r \int_{-\eta}^\eta |z|^{-r} dz + o_\eta(t)$$

5 Projection estimation away from zero

where the integral is bounded above by a finite constant $C_1(r) < \infty$ if and only if $r < 1$.

Step 5: “Proof of Theorem 5.4.6”

Let $\vartheta \in [0, 1/\beta_{\mathbf{X}} \wedge 1[$. Then there is an $r \in [0, 1[$ such that $\beta_{\mathbf{X}} < r/\vartheta < 2$. Firstly, since $|\cdot|^r$ is submultiplicative, secondly, by Jensen’s inequality for the concave function $\|\cdot\|^\vartheta$ and, finally, due to Lemma 5.4.5, we have

$$\begin{aligned} \mathbb{E}\|\mathbf{X}_t^\eta\|^r &\leq \mathbb{E}\left[\left(\|\Sigma^{1/2}\mathbf{W}_t\| + \|\mathbf{X}_t^\eta - \Sigma^{1/2}\mathbf{W}_t\|\right)^r\right] \\ &\leq \mathbb{E}\|\Sigma^{1/2}\mathbf{W}_t\|^r + \mathbb{E}\|\mathbf{X}_t^\eta - \Sigma^{1/2}\mathbf{W}_t\|^r \\ &\leq \left(\mathbb{E}\|\Sigma^{1/2}\mathbf{W}_t\|^2\right)^{r/2} + \left(\mathbb{E}\|\mathbf{X}_t^\eta - \Sigma^{r/2}\mathbf{W}_t\|^{r/\vartheta}\right)^\vartheta \\ &\leq \|\Sigma\|^{r/2}t^{r/2} + C_2t^\vartheta \end{aligned}$$

for a finite positive constant $C_2 < \infty$ depending only on r and ϑ . Consequently, for all $t \in [0, 1]$, we get

$$\begin{aligned} A_1 &\leq \lambda_\eta C_1(r) \left(\|\Sigma\|^{r/2}t^{r/2} + C_2t^\vartheta \right) + o_\eta(t) \\ &= \begin{cases} O_{\eta, \beta_{\mathbf{X}}}(t^\vartheta) & \text{if } \Sigma = \mathbf{0}, \\ O_{\eta, \Sigma}(t^{r/2}) \forall r \in [0, 1[& \text{if } \Sigma \neq \mathbf{0}, \end{cases} \end{aligned} \quad (5.26)$$

as $t \rightarrow 0$. The preliminaries of Lemma 5.4.5 in the case of $\beta_{\mathbf{X}} < 1$ and $\Sigma = \mathbf{0}$ are also met if $\vartheta = 1$.

As the bounds (5.24), (5.25) and (5.26) for the terms on the right-hand side of (5.23) are uniform in y , we get in case of $\beta_{\mathbf{X}} < 1$ and $\Sigma = \mathbf{0}$ that $\Delta_t^* = O_{\varepsilon, \eta}(t)$, in case of $1 \leq \beta_{\mathbf{X}} \leq 2$ and $\Sigma = \mathbf{0}$ that $\Delta_t^* = O_{\varepsilon, \eta, \beta_{\mathbf{X}}}(t^\vartheta)$ for all $\vartheta \in [0, 1/\beta_{\mathbf{X}}[$ and in case of $\Sigma \neq \mathbf{0}$ that $\Delta_t^* = O_{\varepsilon, \eta, \sigma}(t^{r/2})$ for all $r \in [0, 1[$. This is sufficient for Theorem 5.4.6 by virtue of Lemma 5.1.10. \square

6 Penalisation in the discrete time framework

This chapter is dedicated to the penalisation method in the discrete time framework. So far, we have proved Theorem 4.2.6 implying an explicit rate of convergence for the projection estimator of a Besov-type smooth Lévy density p if and only if the smoothness of p is known. In Section 6.1, we introduce penalties based on the observation of the discretisation X^τ of a Lévy process X on the regular grid $\tau\mathbb{N}_0$ with mesh size $\tau > 0$. Moreover, we define a penalised projection estimator in the discrete time framework. In Section 6.2, we first adapt Theorem 3.2.1 from the continuous time framework. Secondly, we refine Theorem 3.3.7.

In this chapter, we write vectors in bold font again and let \mathbf{X} be an \mathbb{R}^d -valued Lévy process with Lévy-Khintchine triplet (γ, Σ, ν) for a $d \in \mathbb{N}$. Since all norms in the finite dimensional space \mathbb{R}^d are equivalent, we use the supremum norm in the following again, unless explicitly stated otherwise. We recall that for $\eta > 0$, we denote the open η -ball around $\mathbf{x} \in \mathbb{R}^d$ w. r. t. the supremum norm by

$$B_\eta(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - \mathbf{x}\|_\infty < \eta\}.$$

6.1 Penalised projection estimation in DT

This section is dedicated to the definition of a penalised projection estimator in the discrete time framework. Let $\mathbb{D} \in \mathcal{B}_\circ^d$ such that there exists an $\varepsilon > 0$ with $\mathbb{D} \cap B_\varepsilon(\mathbf{0}) = \emptyset$. Furthermore, let $\{\mathbb{S}_m : m \in M\}$ be a polynomial collection of finite dimensional linear projection spaces. Analogously to the continuous time framework, we restrict ourselves to the subfamily $\{\mathbb{S}_m : m \in M_T\}$ for $T > 0$, where $M_T := \{m \in M : D_m \leq T\}$.

Assumption 6.1.1 (Regularity on finitely many subsets) *For all $m \in M$, assume there exists an $n \in \mathbb{N}$ and a family $E_{m,1}, \dots, E_{m,n}$ of disjoint d -dimensional intervals away from zero such that*

$$i) \quad \bigcup_{i=1}^n E_{m,i} = \mathbb{D},$$

ii) *the restriction $f|_{E_{m,i}}$ of all $f \in \mathbb{S}_m$ to every $E_{m,i}$ belongs to the Sobolev space $\mathcal{W}^{1,q}(E_{m,i})$ for a $q \geq 1$ independent of $i = 1, \dots, n$.*

6 Penalisation in the discrete time framework

To avoid tedious notation, we write

$$\|f'\|_{\mathbb{D},\infty} := \begin{cases} \sup_{i \in \{1, \dots, n\}} \|f'|_{E_{m,i}}\|_{L^\infty(E_{m,i})} & \text{in the case of } d = 1, \\ \sup_{i \in \{1, \dots, n\}} \sup_{j \in \{1, \dots, d\}} \|\partial_j f|_{E_{m,i}}\|_{L^\infty(E_{m,i})} & \text{in the case of } d \geq 2, \end{cases}$$

and for $m \in M$, we introduce as an additional constant

$$D'_m := \sup\{\|f'\|_{\mathbb{D},\infty}^2 : f \in \mathbb{S}_m, \|f\|_\mu^2 = 1\}.$$

In our case, we always have $D'_m < \infty$.

Definition 6.1.2 (Penalised projection estimation) *Let $\tau : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ be a strictly decreasing function and \mathbf{X}^{τ_T} denote the discretisation of \mathbf{X} w.r.t. the mesh size τ_T from (4.2). Furthermore, let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$ such that pen_T is $\sigma(\mathbf{X}_t^{\tau_T} : t \in [0, T])$ -measurable for every $T > 0$. Then $\mathbb{S}_{\text{pen}} := \mathbb{S}_{m_T^{\text{pen}}}$ is called penalised projection space in DT, where*

$$m_T^{\text{pen}} := \arg \min_{m \in M_T} \left\{ -\|\tilde{p}_m\|_\mu^2 + \text{pen}_T(m) \right\}, \quad (6.1)$$

and $\tilde{p}_{\text{pen}} := \tilde{p}(\cdot; \tau, T, \mathbb{S}_{\text{pen}})$ is called penalised projection estimator in DT.

6.2 Oracle inequality for penalised projection estimation

This section is dedicated to the adaption of the oracle inequality (Theorem 3.2.1).
[...]

Theorem 6.2.1 (Asymptotic oracle inequality in DT) *Let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$ such that there exist $c_1 > 1$ and $c_2, \dots, c_7 > 0$ with*

$$\text{pen}_T(m) = \left. \begin{aligned} & \frac{c_1}{T} \sum_{k=1}^{d_m} \tilde{\nu}(f_{m,k}^2) + \frac{c_2 D_m}{T} + \frac{c_3 d_m}{T} \\ & + \frac{c_4 D'_m}{T} + \frac{c_5 D_m^4}{T^3} + \frac{c_6 (D'_m)^4}{T^3} + \frac{c_7 d_m^4}{T^3}. \end{aligned} \right\} \quad (6.2)$$

Assume the mesh size satisfies $\tau = o(T^{-2})$, then there exist finite positive constants $C_1, C_2 < \infty$ with

$$\mathbb{E}\|p - \tilde{p}_{\text{pen}}\|_\mu^2 \leq C_1 \inf_{m \in M_T} \left(\|p - \mathcal{P}_m p\|_\mu^2 + \mathbb{E}[\text{pen}_T(m)] \right) + \frac{C_2}{T}. \quad (6.3)$$

6.2 Oracle inequality for penalised projection estimation

Proof:

We essentially follow Figueroa-López and Houdré [18, Chapter 9].

Step 1: “General notation and fundamental relations”

We denote the *set of observation intervals* on the grid $\tau_T \mathbb{N}_0$ up to time $T > 0$ by

$$G_T = \{g = \Delta_k^\tau : k = 1, \dots, \lfloor T/\tau \rfloor\}.$$

With this notation, we define for every function $f \in L^2(\mathbb{D}, \mu)$

$$\Xi(f) := -\frac{2}{T} \sum_{g \in G_T} f(\mathbf{X}_g) + \|f\|_\mu^2 \quad (6.4)$$

and

$$\Upsilon(f) := \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) - \nu(f) = \tilde{\nu}(f) - \nu(f). \quad (6.5)$$

From classical Hilbert space theory, we deduce that the projection estimator \tilde{p}_m w. r. t. a projection space \mathbb{S}_m is the unique minimiser of Ξ over \mathbb{S}_m and

$$\Xi(\tilde{p}_m) = -\|\tilde{p}_m\|_\mu^2 \quad (6.6)$$

holds for all $m \in M$. Combining (6.1) and (6.6), we conclude that for all $m \in M$

$$\Xi(\tilde{p}_{\text{pen}}) + \text{pen}_T(m_T^{\text{pen}}) \leq \Xi(\tilde{p}_m) + \text{pen}_T(m) \leq \Xi(\mathcal{P}_m p) + \text{pen}_T(m) \quad (6.7)$$

is satisfied. Obviously, we can express Ξ in terms of Υ . In particular, we observe that

$$\Xi(f) = \|f\|_\mu^2 - 2\nu(f) - 2\Upsilon(f) = \|f - p\|_\mu^2 - \|p\|_\mu^2 - 2\Upsilon(f), \quad (6.8)$$

where the second equality follows directly with Lemma 2.3.6.

For the sake of simplicity, for all $f \in L^2(\mathbb{D}, \mu)$, we define $f(\mathbf{x}) = 0$ for all $\mathbf{x} \notin \mathbb{D}$. Let $\eta \in]0, \varepsilon/3 \wedge 1[$, then we decompose the Lévy process \mathbf{X} for $t \in \mathbb{R}_0^+$ as follows.

$$\left. \begin{aligned} \mathbf{V}_t &:= \gamma_\eta t + \Sigma^{1/2} \mathbf{W}_t && \text{with } \gamma_\eta = \gamma - \int_{B_1(\mathbf{0}) \setminus B_\eta(\mathbf{0})} \mathbf{x} \nu(d\mathbf{x}), \\ \mathbf{Y}_t &:= \sum_{k=1}^{N_t} \mathbf{Z}_k := \iint_{[0, t] \times (\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0}))} \mathbf{x} J(ds, d\mathbf{x}), \\ \mathbf{R}_t &:= \mathbf{X}_t - \mathbf{V}_t - \mathbf{Y}_t, \end{aligned} \right\} \quad (6.9)$$

where \mathbf{V} , \mathbf{Y} and \mathbf{R} are independent. In the definition of \mathbf{Y} in (6.9), N is a Poisson process with rate $\lambda_\eta := \nu(\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0})) < \infty$ and the family of jumps $\{\mathbf{Z}_k\}_{k \in \mathbb{N}}$ is an i. i. d. family of random variables (independent of N) with law $P_{\mathbf{Z}_1}$ given by

$$P_{\mathbf{Z}_1}(d\mathbf{x}) = \lambda_\eta^{-1} p(\mathbf{x}) \mathbb{1}_{\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0})}(\mathbf{x}) d\mathbf{x} \quad \text{for } \mathbf{x} \in \mathbb{R}^d.$$

6 Penalisation in the discrete time framework

Step 2: “Excursus: About the number of jumps per observation interval”

If $\tau = o(T^{-1})$, then $\sup_{g \in G_T} N_g$ tends to one in probability as $T \rightarrow \infty$.

Proof: In particular, we observe that

$$P(\forall g \in G_T : N_g \leq 1) = \left(e^{-\lambda_\eta \tau} (1 + \lambda_\eta \tau) \right)^{T/\tau} = \exp \left(-\lambda_\eta T + \frac{T}{\tau} \log(1 + \lambda_\eta \tau) \right)$$

holds for every $T > 0$. Let T be big enough such that $\tau < 1/\lambda_\eta$, then we have

$$\log(1 + \lambda_\eta \tau) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\lambda_\eta \tau)^k}{k}$$

implying that

$$\lim_{T \rightarrow \infty} \exp \left(-\lambda_\eta T + \frac{T}{\tau} \log(1 + \lambda_\eta \tau) \right) = \lim_{T \rightarrow \infty} \exp \left(\sum_{k=2}^{\infty} \frac{\lambda_\eta^k T \tau^{k-1}}{k} \right) = e^0 = 1$$

since $T\tau \rightarrow 0$ by assumption. Finally, since $P(\forall g \in G_T : N_g = 0) = e^{-\lambda_\eta T} \rightarrow 0$, we have proved that

$$\sup_{g \in G_T} N_g \xrightarrow{P} 1 \tag{6.10}$$

as $T \rightarrow \infty$. ◇

Step 3: “Excursus: Implications from Mancini [29, Theorem 3.1]”

The integral estimator $\tilde{\nu}(f)$ tends to $T^{-1} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}}(\cdot)$ in probability.

Proof: Let $r : [0, 1[\rightarrow \mathbb{R}_0^+$ be such that $r(0) = 0$,

$$\lim_{\tau \rightarrow 0} \frac{\tau \log(1/\tau)}{r(\tau)} = 0 \text{ and } \lim_{\tau \rightarrow 0} \frac{\tau}{r(\tau)} = 0.$$

Then we deduce from (6.10) and from Mancini [29, Theorem 3.1] that for P -almost all $\omega \in \Omega$ exists a $T_{\min}(\omega) < \infty$ such that

- (i) $\sqrt{r(\tau_T)} < \varepsilon$,
- (ii) $\sup_{g \in G_T} N_g(\omega) = 1$,
- (iii) for all $g \in G_T : N_g(\omega) = 1 \Leftrightarrow \|\mathbf{V}_g(\omega) + \mathbf{Y}_g(\omega)\|_\infty^2 \geq r(\tau_T)$

hold for all $T > T_{\min}(\omega)$.

6.2 Oracle inequality for penalised projection estimation

For the sake of the argument, let us outline that (iii) is valid in the multivariate case, too. With the abbreviation $C_\Sigma = \|\text{diag}(\Sigma)\|_\infty$, we have to prove that

$$\sup_{g \in G_T} \frac{\|\mathbf{V}_g\|_\infty}{\sqrt{2\tau \log(1/\tau)}} < \|\gamma_\eta\|_\infty + \sqrt{C_\Sigma} < \infty \quad (6.11)$$

is bounded almost surely for large T . For every $j = 1, \dots, d$, we deduce that there is a Wiener process W'_j (possibly correlated to $W'_{j'}$ for $j' \neq j$) such that

$$V_{j, \Delta_h^\tau} = \gamma_{\eta, j} \tau + W'_{j, h\sigma_{jj}\tau} - W'_{j, (h-1)\sigma_{jj}\tau},$$

where σ_{jj} denotes the j -th diagonal element of Σ . We conclude that for large T (hence, small τ)

$$\begin{aligned} \sup_{g \in G_T} \frac{|V_{j, g}|}{\sqrt{2\tau \log(1/\tau)}} &\leq \frac{\gamma_{\eta, j} \tau}{\sqrt{2\tau \log(1/\tau)}} \\ &+ \sup_h \frac{W'_{j, h\sigma_{jj}\tau} - W'_{j, (h-1)\sigma_{jj}\tau}}{\sqrt{2\sigma_{jj}\tau \log(\frac{1}{\sigma_{jj}\tau})}} \sup_h \frac{\sqrt{2\sigma_{jj}\tau \log(\frac{1}{\sigma_{jj}\tau})}}{\sqrt{2C_\Sigma\tau \log(\frac{1}{C_\Sigma\tau})}} \times \\ &\times \sup_h \frac{\sqrt{2C_\Sigma\tau \log(\frac{1}{C_\Sigma\tau})}}{\sqrt{2\tau \log(1/\tau)}} < \|\gamma_\eta\|_\infty + \sqrt{C_\Sigma} \quad \text{a. s. as } T \rightarrow \infty. \end{aligned}$$

The first summand on the right-hand side is seen to be bounded by $\|\gamma_\eta\|_\infty$ for large T , since $\sqrt{\tau/\log(1/\tau)} \rightarrow 0$ as $\tau \rightarrow 0$. The third term in the product on the right-hand side tends to $\sqrt{C_\Sigma}$, and the second term is seen to be bounded by 1, as $\tau \mapsto \tau \log(1/\tau)$ is an increasing function. Finally, the a. s. limit, as $T \rightarrow \infty$ (hence $\tau \rightarrow 0$) of the first factor is almost surely bounded by 1 due to *Lévy's modulus of continuity theorem* (cf. Lévy [28] or Sato [36, p. 10]).

Hence we proved (6.11), as the right-hand side of the inequality is independent of the component j . As we ensured $\min_{k \in \mathbb{N}} \|\mathbf{Z}_k\|_\infty \geq \eta$ by definition of \mathbf{Y} , the remaining of the (univariate) proof of Mancini [29, Theorem 3.1] can be adapted one-to-one to our multivariate setup.

6 Penalisation in the discrete time framework

Therefore,

$$\begin{aligned}
& \tilde{\nu}(f) - \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}}(\cdot) \\
&= \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{\|\mathbf{X}_g\| \geq \varepsilon\}}(\cdot) - \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}}(\cdot) \\
&= \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \left(\mathbb{1}_{\{\|\mathbf{X}_g\| \geq \varepsilon, \|\mathbf{V}_g + \mathbf{Y}_g\|^2 < r(\tau)\}}(\cdot) - \mathbb{1}_{\{\|\mathbf{X}_g\| < \varepsilon, \|\mathbf{V}_g + \mathbf{Y}_g\|^2 \geq r(\tau)\}}(\cdot) \right) \\
&= \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{\|\mathbf{V}_g + \mathbf{R}_g\| \geq \varepsilon\}}(\cdot) \xrightarrow{P} 0
\end{aligned} \tag{6.12}$$

tends to zero in probability, since for a finite constant $K_{\varepsilon, \eta} < \infty$, we conclude that

$$\begin{aligned}
\mathbb{E} \left| \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{\|\mathbf{V}_g + \mathbf{R}_g\| \geq \varepsilon\}} \right| &\leq \|f\|_{\mathbb{D}, \infty} \frac{T}{\tau} P(\|\mathbf{V}_g + \mathbf{R}_g\| \geq \varepsilon) \\
&\leq K_{\varepsilon, \eta} \|f\|_{\mathbb{D}, \infty} T \tau^2 \rightarrow 0
\end{aligned}$$

by virtue of Lemma 5.4.4, the choice of η and the same argument as in Step 3 of the proof of Theorem 5.4.6. \diamond

For the sake of simplicity, let us introduce an abbreviation for the stochastic process appearing last in (6.12). We denote for all $f \in L^2(\mathbb{D}, \mu)$

$$\begin{aligned}
\Psi(f) : \mathbb{R}_0^+ \times \Omega &\rightarrow \mathbb{R}; (t, \omega) \mapsto \sum_{g \in G_t} f(\mathbf{X}_g) \mathbb{1}_{\{\|\mathbf{V}_g + \mathbf{R}_g\| \geq \varepsilon\}}(\omega), \\
\Psi : \mathbb{R}_0^+ \times \Omega &\rightarrow \mathbb{N}; (t, \omega) \mapsto \sum_{g \in G_t} \mathbb{1}_{\{\|\mathbf{V}_g + \mathbf{R}_g\| \geq \varepsilon\}}(\omega).
\end{aligned}$$

Then, for all $T > 0$, we observe

$$\tilde{\nu}(f) - \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} = \frac{1}{T} \Psi(f)_T \leq \frac{\|f\|_{\mathbb{D}, \infty} \Psi_T}{T}. \tag{6.13}$$

Moreover, for every $T > 0$, we conclude that Ψ_T has binomial distribution $B_{\tilde{n}, \rho}$ with parameters $\tilde{n} = \lfloor T/\tau \rfloor$ and $\rho \leq K_{\varepsilon, \eta} \tau^3$.

Step 4: “Decomposing the squared error”

For every $m \in M$, we have

$$\begin{aligned}
\|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 &\leq \|p - \mathcal{P}_m p\|_{\mu}^2 + 2\|\mathcal{P}_{\text{pen}} p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \\
&\quad + 2\Upsilon(\mathcal{P}_{\text{pen}} p - \mathcal{P}_m p) + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}).
\end{aligned} \tag{6.14}$$

6.2 Oracle inequality for penalised projection estimation

Proof: First, we use the fundamental relations from Step 1 to show that for arbitrary $m \in M$

$$\left. \begin{aligned} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 &\stackrel{(6.8)}{=} \Xi(\tilde{p}_{\text{pen}}) + \|p\|_{\mu}^2 + 2\Upsilon(\tilde{p}_{\text{pen}}) \\ &\stackrel{(6.7)}{\leq} \Xi(\mathcal{P}_m p) + \|p\|_{\mu}^2 + 2\Upsilon(\tilde{p}_{\text{pen}}) + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) \\ &\stackrel{(6.8)}{=} \|p - \mathcal{P}_m p\|_{\mu}^2 + 2\Upsilon(\tilde{p}_{\text{pen}} - \mathcal{P}_m p) + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) \end{aligned} \right\} \quad (6.15)$$

holds. Moreover, the representation $\tilde{p}_m - \mathcal{P}_m p = \sum_{k=1}^{d_m} \Upsilon(f_{m,k}) f_{m,k}(\cdot)$ follows directly from the definition of the projection estimator and from (6.5). Let us recall that $\{f_{m,k} : k = 1, \dots, d_m\}$ denotes a μ -ONB of \mathbb{S}_m . Hence, we conclude that

$$\Upsilon(\tilde{p}_m - \mathcal{P}_m p) = \|\tilde{p}_m - \mathcal{P}_m p\|_{\mu}^2$$

holds for every $m \in M$. Thus,

$$\left. \begin{aligned} \Upsilon(\tilde{p}_{\text{pen}} - \mathcal{P}_m p) &= \Upsilon(\tilde{p}_{\text{pen}} - \mathcal{P}_{\text{pen}} p) + \Upsilon(\mathcal{P}_{\text{pen}} p - \mathcal{P}_m p) \\ &= \|\tilde{p}_{\text{pen}} - \mathcal{P}_{\text{pen}} p\|_{\mu}^2 + \Upsilon(\mathcal{P}_{\text{pen}} p - \mathcal{P}_m p). \end{aligned} \right\} \quad (6.16)$$

The validity of (6.14) follows directly from (6.15) and (6.16). \diamond

Step 5: “Decomposing Υ ”

For every T large enough, there exist a Poisson random measure \check{J} with mean measure having μ -density \check{p} and a finite constant $K'_\eta < \infty$ such that

$$\left. \begin{aligned} \Upsilon(f) &\leq \left| \iint_{[0,T] \times \mathbb{D}} \frac{f(\mathbf{x})}{T} (\check{J}(dt, d\mathbf{x}) - \check{p}(\mathbf{x}) dt \mu(d\mathbf{x})) \right| \\ &\quad + \frac{1}{T} K'_\eta \left(\|f\|_{\mathbb{D}, \infty} + \sum_{i=1}^n \sum_{j=1}^d \|\partial_j f|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \right) + \frac{1}{T} \Psi_T \|f\|_{\mathbb{D}, \infty}. \end{aligned} \right\} \quad (6.17)$$

Proof: We deduce from (6.13) that for $T > T_{\min}$

$$\left. \begin{aligned} \Upsilon(f) &= \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) - \nu(f) \\ &\leq \left| \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} - \nu(f) \right| + \frac{1}{T} \Psi_T \|f\|_{\mathbb{D}, \infty} \end{aligned} \right\} \quad (6.18)$$

6 Penalisation in the discrete time framework

holds. Moreover, $\sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}}$ is a compound Poisson process with finite intensity λ_η . Hence, for every $T > T_{\min}$, there exists a Poisson random measure \check{J} on the space $([0, T] \times \mathbb{D}, \mathcal{B}([0, T] \times \mathbb{D}))$ with intensity measure having density $(t, \mathbf{x}) \mapsto \check{p}(\mathbf{x})$ such that

$$\frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} = \iint_{[0, T] \times \mathbb{D}} \frac{f(\mathbf{x})}{T} \check{J}(dt, d\mathbf{x}).$$

Therefore, we conclude that

$$\left| \frac{1}{T} \sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} - \nu(f) \right| \leq \left| \iint_{[0, T] \times \mathbb{D}} \frac{f(\mathbf{x})}{T} (\check{J}(dt, d\mathbf{x}) - \check{p}(\mathbf{x}) dt \mu(d\mathbf{x})) \right| + \left| \frac{1}{T} \mathbb{E} \left[\sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} \right] - \nu(f) \right| \quad (6.19)$$

holds, where we observe that

$$\begin{aligned} \left| \frac{1}{T} \mathbb{E} \left[\sum_{g \in G_T} f(\mathbf{X}_g) \mathbb{1}_{\{N_g=1\}} \right] - \nu(f) \right| &= \left| \frac{1}{\tau} \mathbb{E}[f(\mathbf{X}_\tau); N_\tau = 1] - \nu(f) \right| \\ &= \left| \Delta_\tau (f \mathbb{1}_{\{N_\tau=1\}}) \right|. \end{aligned}$$

Then in the light of Theorem 5.4.3 (Theorem 5.4.6 in the case of $d \geq 2$) and Lemma 5.1.2 (Lemma 5.1.10, respectively), there is a finite constant $K'_\eta < \infty$ such that for every $\vartheta < 1/2$ (or $\vartheta < 1/\beta_{\mathbf{X}} \wedge 1$ in the case of $\sigma = 0$)

$$\sup_{\{\tau: T > T_{\min}\}} \tau^{-\vartheta} \left| \Delta_\tau (f \mathbb{1}_{\{N_\tau=1\}}) \right| \leq \left(\|f\|_{\mathbb{D}, \infty} + \sum_{i=1}^n \sum_{j=1}^d \|\partial_j f|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \right) K'_\eta$$

holds. By the choice of $\tau = o(1/T^2)$, we ensured that

$$\sup_{T > T_{\min}} T \left| \Delta_\tau (f \mathbb{1}_{\{N_\tau=1\}}) \right| \leq \left(\|f\|_{\mathbb{D}, \infty} + \sum_{i=1}^n \sum_{j=1}^d \|\partial_j f|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \right) K'_\eta. \quad (6.20)$$

holds, additionally. Finally, the validity of (6.17) follows directly from (6.18), (6.19) and (6.20). \diamond

In the following, we repeatedly invoke the inequalities below. For all $x_1, x_2 \in \mathbb{R}$ and every arbitrary constant $a > 0$, we have

$$2x_1x_2 \leq ax_1^2 + \frac{1}{a}x_2^2, \quad (6.21)$$

$$(x_1 + x_2)^2 \leq (1 + a)x_1^2 + \left(1 + \frac{1}{a}\right)x_2^2. \quad (6.22)$$

6.2 Oracle inequality for penalised projection estimation

Further, let $B_1, B_2 \in \mathcal{F}$, then

$$(P(B_1) \geq 1 - b_1 \text{ and } P(B_2) \geq 1 - b_2) \Rightarrow P(B_1 \cap B_2) \geq 1 - b_1 - b_2. \quad (6.23)$$

Step 6: "Concentration inequality for Υ "

For arbitrary $a_2, \dots, a_7 > 0$, there exist a positive number k_1 , a quadratic function $f_1 : \xi \mapsto f_1(\xi)$ increasing on $[0, \infty[$ (independent of $\{\mathbb{S}_m : m \in M\}$ and T) and finite constants \bar{K}_1, \bar{K}_2 such that with probability larger than $1 - k_1 e^{-\xi}$

$$\left. \begin{aligned} \|p - \tilde{p}_{\text{pen}}\|_\mu^2 &\leq \|p - \mathcal{P}_m p\|_\mu^2 + 2\|\mathcal{P}_{\text{pen}} p - \tilde{p}_{\text{pen}}\|_\mu^2 + 2a_2\|\mathcal{P}_{\text{pen}} p - \mathcal{P}_m p\|_\mu^2 \\ &\quad + \frac{a_3 D_{m_T^{\text{pen}}}}{T} + \frac{a_4 D_m}{T} + \frac{a_5 D'_{m_T^{\text{pen}}}}{T} + \frac{a_6 D'_m}{T} + \frac{a_7 d_{m_T^{\text{pen}}}}{T} \\ &\quad + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) + \frac{f_1(\xi)}{T} + \frac{\bar{K}_1}{T} + \frac{\bar{K}_2 \Psi_T^2}{T}. \end{aligned} \right\} \quad (6.24)$$

Proof: First, let us recall representation (6.17) for $\Upsilon(\mathcal{P}_{m'} - \mathcal{P}_m)$ and denote the summands therein by A_1, A_2 and A_3 . Then Figueroa-López and Houdré [18, Proposition 9.2] implies that for $m' \in M$ and arbitrary $a_{m'} > 0$

$$A_1 \leq \left\| \frac{\mathcal{P}_{m'} p - \mathcal{P}_m p}{T} \right\|_{L^2(\mathbb{D}, \check{p} \text{d}t \text{d}\mu)} \sqrt{2a_{m'}} + \frac{\|\mathcal{P}_{m'} p - \mathcal{P}_m p\|_{\mathbb{D}, \infty} a_{m'}}{3T} \quad (6.25)$$

holds with probability larger than $1 - e^{-a_{m'}}$. By virtue of (6.23), we deduce that (6.25) holds for all $m' \in M$ with probability greater than $1 - \sum_{m' \in M} e^{-a_{m'}}$.

We recall that \check{p}/λ_η is the density of $P_{\mathbf{V}_\tau + \mathbf{R}_\tau} * P_{\mathbf{Z}_1}$. Hence,

$$|\check{p}(\mathbf{x})| = \left| \int_{\mathbb{R}} \left(p(\mathbf{u}) \mathbb{1}_{\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0})}(\mathbf{u}) \right) F'_{\mathbf{V}_\tau + \mathbf{R}_\tau}(\mathbf{x} - \mathbf{u}) \mu(\text{d}\mathbf{u}) \right| \leq \|p\|_{\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0}), \infty} < \infty$$

holds for every $\mathbf{x} \in \mathbb{D}$. Therefore, we see that

$$\left\| \frac{\mathcal{P}_{m'} p - \mathcal{P}_m p}{T} \right\|_{L^2(\mathbb{D}, \check{p} \text{d}t \text{d}\mu)}^2 \leq \|p\|_{\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0}), \infty} \frac{\|\mathcal{P}_{m'} p - \mathcal{P}_m p\|_\mu^2}{T}$$

is valid. Thus, invoking (6.21) for an arbitrary $a_2 > 0$, we conclude

$$\left\| \frac{\mathcal{P}_{m'} p - \mathcal{P}_m p}{T} \right\|_{L^2(\mathbb{D}, \check{p} \text{d}t \text{d}\mu)} \sqrt{2a_{m'}} \leq a_2 \|\mathcal{P}_{m'} p - \mathcal{P}_m p\|_\mu^2 + \frac{\|p\|_{\mathbb{R}_\circ^d \setminus B_\eta(\mathbf{0}), \infty} a_{m'}}{2a_2 T}$$

for the first term on the right-hand side in (6.25).

6 Penalisation in the discrete time framework

Classical Hilbert space theory reveals that every orthonormal projection operator \mathcal{P} satisfies $\|\mathcal{P}p\|_2 \leq \|p\|_2$. Therefore, we conclude that

$$\|\mathcal{P}_{m'}p - \mathcal{P}_mp\|_{\mathbb{D},\infty} \leq \sqrt{D_{m'}}\|\mathcal{P}_{m'}p\|_\mu + \sqrt{D_m}\|\mathcal{P}_mp\|_\mu \leq \sqrt{D_{m'}}\|p\|_\mu + \sqrt{D_m}\|p\|_\mu$$

holds. Again, invoking (6.21) for arbitrary $a_3 > 0$ and $a_4 > 0$, we have

$$\frac{\|\mathcal{P}_{m'}p - \mathcal{P}_mp\|_{\mathbb{D},\infty} a_{m'}}{3T} \leq \frac{a_3 D_{m'}}{2T} + \frac{a_4 D_m}{2T} + \frac{\|p\|_\mu^2 a_{m'}^2}{36T} \left(\frac{2}{a_3} + \frac{2}{a_4} \right)$$

for the second term on the right-hand side in (6.25).

From (6.17), we observe additionally that

$$A_2 + A_3 = \left. \begin{aligned} & \frac{(K'_\eta + \Psi_T)\|\mathcal{P}_{m'}p - \mathcal{P}_mp\|_{\mathbb{D},\infty}}{T} \\ & + \frac{1}{T} K'_\eta \sum_{i=1}^n \sum_{j=1}^d \|\partial_j(\mathcal{P}_{m'}p - \mathcal{P}_mp)|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \end{aligned} \right\} \quad (6.26)$$

is satisfied. Analogously to the previous argument, we have

$$\frac{(K'_\eta + \Psi_T)\|\mathcal{P}_{m'}p - \mathcal{P}_mp\|_{\mathbb{D},\infty}}{T} \leq \frac{a_3 D_{m'}}{2T} + \frac{a_4 D_m}{2T} + \frac{\|p\|_\mu^2 (2(K'_\eta)^2 + 2\Psi_T^2)}{4T} \left(\frac{2}{a_3} + \frac{2}{a_4} \right)$$

for the first term on the right-hand side in (6.26). By definition, for all $i = 1, \dots, n$ and $j = 1, \dots, d$, we have

$$\|\partial_j(\mathcal{P}_{m'}p - \mathcal{P}_mp)|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \leq \|f'\|_{\mathbb{D},\infty}.$$

Thus, analogously to the previous two arguments, for arbitrary $a_5 > 0$ and $a_6 > 0$, we have

$$\begin{aligned} & \frac{1}{T} K'_\eta \sum_{i=1}^n \sum_{j=1}^d \|\partial_j(\mathcal{P}_{m'}p - \mathcal{P}_mp)|_{E_{m,i}}\|_{L^\infty(E_{m,i})} \\ & \leq \frac{a_5 D_{m'}}{T} + \frac{a_6 D_m}{T} + \frac{\|p\|_\mu^2 (nK'_\eta)^2}{4T} \left(\frac{1}{a_5} + \frac{1}{a_6} \right) \end{aligned}$$

for the second term in (6.26).

With the abbreviation

$$\begin{aligned} \bar{K}_1 & := \frac{\|p\|_\mu^2}{4} \left(2(K'_\eta)^2 \left(\frac{2}{a_3} + \frac{2}{a_4} \right) + (nK'_\eta)^2 \left(\frac{1}{a_5} + \frac{1}{a_6} \right) \right), \\ \bar{K}_2 & := \frac{\|p\|_\mu^2}{2} \left(\frac{2}{a_3} + \frac{2}{a_4} \right), \end{aligned}$$

6.2 Oracle inequality for penalised projection estimation

we proved that

$$\begin{aligned} \Upsilon(\mathcal{P}_{m'} - \mathcal{P}_m) &\leq a_2 \|\mathcal{P}_{m'} p - \mathcal{P}_m p\|_\mu^2 + \frac{a_3 D_{m'}}{T} + \frac{a_4 D_m}{T} \\ &\quad + \frac{a_5 D'_{m'}}{T} + \frac{a_6 D'_m}{T} + \frac{\bar{K}_1}{T} + \frac{\bar{K}_2 \Psi_T^2}{T} \\ &\quad + \frac{\|p\|_{\mathbb{R}_\delta^d \setminus B_\eta(0), \infty} a_{m'}}{2a_2 T} + \frac{\|p\|_\mu^2 a_{m'}^2}{36T} \left(\frac{2}{a_3} + \frac{2}{a_4} \right) \end{aligned}$$

holds with probability greater than $1 - \sum_{m' \in M} e^{-a_{m'}}$. For arbitrary $a'_7 > 0$, we take

$$a_{m'} := a'_7 \sqrt{d_{m'}} \left(\frac{1}{\|p\|_{\mathbb{R}_\delta^d \setminus B_\eta(0), \infty}} \wedge \frac{1}{\|p\|_\mu} \right) + \xi.$$

Then, invoking (6.22) with $a = 1$ and $x_1 + x_2 = a_{m'}^2$, we get

$$\left. \begin{aligned} \Upsilon(\mathcal{P}_{m'} - \mathcal{P}_m) &\leq a_2 \|\mathcal{P}_{m'} p - \mathcal{P}_m p\|_\mu^2 + \frac{a_3 D_{m'}}{T} + \frac{a_4 D_m}{T} \\ &\quad + \frac{a_5 D'_{m'}}{T} + \frac{a_6 D'_m}{T} + \frac{\bar{K}_1}{T} + \frac{\bar{K}_2 \Psi_T^2}{T} \\ &\quad + \left(\frac{(a'_7)^2}{18} \left(\frac{2}{a_3} + \frac{2}{a_4} \right) + \frac{a'_7}{2a_2} \right) \frac{d_{m'}}{T} + \frac{f_1(\xi)}{T}, \end{aligned} \right\} \quad (6.27)$$

where

$$f_1(\xi) = \frac{\|p\|_\mu^2}{18} \left(\frac{2}{a_3} + \frac{2}{a_4} \right) \xi^2 + \frac{\|p\|_{\mathbb{R}_\delta^d \setminus B_\eta(0), \infty} \xi}{2a_2}.$$

We recall that $\{\mathbb{S}_m : m \in M\}$ is assumed to be a polynomial family. Hence there exist $b_1 > 0$ and $b_2 \geq 0$ such that for all $l \in \mathbb{N}$, we have $\#\{m \in M : d_m = l\} \leq b_1 l^{b_2}$. We set

$$k_1 := b_1 \sum_{l=1}^{\infty} l^{b_2} \exp \left(-\sqrt{l} a_7 \left(\frac{1}{\|p\|_{\mathbb{R}_\delta^d \setminus B_\eta(0), \infty}} \wedge \frac{1}{\|p\|_\mu} \right) \right)$$

and deduce that (6.27) holds with probability greater than $1 - k_1 e^{-\xi}$. Renaming the coefficient of $d_{m'}/T$ and combining (6.27) with (6.14), we proved (6.24). \diamond

Step 7: "Reshaping (6.24)"

For an arbitrary $a_8 > 0$, there exist positive constants $C < 1$ and $C' > 1$ (independent of $\{\mathbb{S}_m : m \in M\}$ and T) such that with probability larger than $1 - k_1 e^{-\xi}$

$$\left. \begin{aligned} C \|p - \tilde{p}_{\text{pen}}\|_\mu^2 &\leq C' \|p - \mathcal{P}_m p\|_\mu^2 + (1 + a_8) \|\mathcal{P}_{\text{pen}} p - \tilde{p}_{\text{pen}}\|_\mu^2 \\ &\quad + \frac{a_3 D_{m_T^{\text{pen}}}}{T} + \frac{a_4 D_m}{T} + \frac{a_5 D'_{m_T^{\text{pen}}}}{T} + \frac{a_6 D'_m}{T} + \frac{a_7 d_{m_T^{\text{pen}}}}{T} \\ &\quad + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) + \frac{f_1(\xi)}{T} + \frac{\bar{K}_1}{T} + \frac{\bar{K}_2 \Psi_T^2}{T}. \end{aligned} \right\} \quad (6.28)$$

6 Penalisation in the discrete time framework

Proof: Invoking (6.22), we conclude that

$$\begin{aligned}\|\mathcal{P}_{\text{pen}}p - \mathcal{P}_mp\|_\mu^2 &\leq 2\|p - \mathcal{P}_mp\|_\mu^2 + 2\|p - \mathcal{P}_{\text{pen}}p\|_\mu^2 \\ &\leq 2\|p - \mathcal{P}_mp\|_\mu^2 + 2\|p - \tilde{p}_{\text{pen}}\|_\mu^2 - 2\|\mathcal{P}_{\text{pen}}p - \tilde{p}_{\text{pen}}\|_\mu^2.\end{aligned}$$

Hence,

$$\begin{aligned}\|p - \mathcal{P}_mp\|_\mu^2 + 2\|\mathcal{P}_{\text{pen}}p - \tilde{p}_{\text{pen}}\|_\mu^2 + 2a_2\|\mathcal{P}_{\text{pen}}p - \mathcal{P}_mp\|_\mu^2 - \|p - \tilde{p}_{\text{pen}}\|_\mu^2 \\ \leq (1 + 4a_2)\|p - \mathcal{P}_mp\|_\mu^2 + (2 - 4a_2)\|\mathcal{P}_{\text{pen}}p - \tilde{p}_{\text{pen}}\|_\mu^2 \\ + (4a_2 - 1)\|p - \tilde{p}_{\text{pen}}\|_\mu^2.\end{aligned}$$

For arbitrary $a_8 > 0$, we choose $a_2 = (1 - a_8)/4$ in the case of $a_8 < 1$ and $a_2 \in]0, 1/4[$ in the case of $a_8 \geq 1$. Furthermore, we set

$$C := \begin{cases} a_8, & \text{if } 0 < a_8 < 1, \\ 1 - 4a_2, & \text{if } a_8 \geq 1, \end{cases} \quad C' := \begin{cases} 2 - a_8, & \text{if } 0 < a_8 < 1, \\ 1 + 4a_2, & \text{if } a_8 \geq 1. \end{cases}$$

Finally, we observe

$$\begin{aligned}\|p - \mathcal{P}_mp\|_\mu^2 + 2\|\mathcal{P}_{\text{pen}}p - \tilde{p}_{\text{pen}}\|_\mu^2 + 2a_2\|\mathcal{P}_{\text{pen}}p - \mathcal{P}_mp\|_\mu^2 - \|p - \tilde{p}_{\text{pen}}\|_\mu^2 \\ \leq C'\|p - \mathcal{P}_mp\|_\mu^2 + (1 + a_8)\|\mathcal{P}_{\text{pen}}p - \tilde{p}_{\text{pen}}\|_\mu^2 - C\|p - \tilde{p}_{\text{pen}}\|_\mu^2,\end{aligned}$$

where (6.28) follows directly with (6.24). \diamond

Step 8: "Concentration inequality for the μ -variance"

For arbitrary $a_{10}, \dots, a_{15} > 0$, there exist a finite constant $k_2 > 0$, a quadratic function $f_2 : \xi \mapsto f_2(\xi)$ increasing on $[0, \infty[$ (independent of $\{\mathbb{S}_m : m \in M\}$ and T) and finite constants \bar{K}_3, \bar{K}_4 such that with probability larger than $1 - k_2e^{-\xi}$

$$\left. \begin{aligned}C\|p - \tilde{p}_{\text{pen}}\|_\mu^2 &\leq C'\|p - \mathcal{P}_mp\|_\mu^2 + \frac{(1 + a_{10})}{T} \sum_{k=1}^{d_{m_T}^{\text{pen}}} \langle f_{m,k}^2, \check{p} \rangle_\mu \\ &\quad + \frac{a_4 D_m}{T} + \frac{a_5 D'_{m_T^{\text{pen}}}}{T} + \frac{a_6 D'_m}{T} + \frac{a_{11} D_{m_T^{\text{pen}}}}{T} \\ &\quad + \frac{a_{12} d_{m_T^{\text{pen}}}}{T} + \frac{a_{13} d_{m_T^{\text{pen}}}^4}{T^3} + \frac{a_{14} D_{m_T^{\text{pen}}}^4}{T^3} + \frac{a_{15} (D'_{m_T^{\text{pen}}})^4}{T^3} \\ &\quad + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) \\ &\quad + \frac{f_2(\xi)}{T} + \frac{\bar{K}_3}{T} + \frac{\bar{K}_4(\Psi_T^2 + \Psi_T^4)}{T}.\end{aligned} \right\} \quad (6.29)$$

6.2 Oracle inequality for penalised projection estimation

Proof: Recalling (6.17) and invoking (6.22), we observe that

$$\left. \begin{aligned} & \|\mathcal{P}_{m'}p - \tilde{p}_{m'}\|_\mu^2 = \sum_{k=1}^{d_{m'}} (\Upsilon(f_{m',k}))^2 \\ & \leq 2 \sum_{k=1}^{d_{m'}} \left| \iint_{[0,T] \times \mathbb{D}} f_{m',k}(\mathbf{x}) \frac{\check{J}(dt, d\mathbf{x}) - \check{p}(\mathbf{x})dt\mu(d\mathbf{x})}{T} \right|^2 \\ & \quad + 4 \sum_{k=1}^{d_{m'}} \left(\frac{K'_\eta(\|f_{m',k}\|_{\mathbb{D},\infty} + nd\|f'_{m',k}\|_{\mathbb{D},\infty})}{T} \right)^2 \\ & \quad + 4 \sum_{k=1}^{d_{m'}} \left(\frac{\Psi_T\|f_{m',k}\|_{\mathbb{D},\infty}}{T} \right)^2 \end{aligned} \right\} \quad (6.30)$$

holds for all $m' \in M$. We denote the three sums on the right-hand side of (6.30) by A_4 , A_5 and A_6 . Furthermore, we set

$$\check{D}_{m'} := \sup\{\langle f^2, \check{p} \rangle_\mu : f \in \mathbb{S}_{m'}, \|f\|_\mu^2 = 1\}.$$

Then Figueroa-López and Houdré [18, Proposition 9.3] implies that for $m' \in M$ and arbitrary $a'_{m'} > 0$ and $a_{10} > 0$

$$\sqrt{T}A_4 \leq (1 + a_{10}) \sqrt{\sum_{k=1}^{d_{m'}} \langle f_{m',k}^2, \check{p} \rangle_\mu} + \sqrt{12\check{D}_{m'}a'_{m'}} + \left(1.25 + \frac{32}{a_{10}}\right) \sqrt{\frac{D_{m'}}{T}a'_{m'}} \quad (6.31)$$

holds with probability larger than $1 - e^{-a'_{m'}}$. By virtue of (6.23), we deduce that (6.31) holds for all $m' \in M$ with probability greater than $1 - \sum_{m' \in M} e^{-a'_{m'}}$. We observe that $\check{D}_{m'}$ is bounded above by $\|\check{p}\|_\mu \sqrt{D_{m'}}$. Thus, invoking (6.21), we obtain

$$\sqrt{12\check{D}_{m'}a'_{m'}} \leq a_{11}\sqrt{D_{m'}} + \frac{3\|\check{p}\|_\mu}{a_{11}}a'_{m'}$$

for arbitrary $a_{11} > 0$.

By the choice of

$$a'_{m'} = \frac{a_{12}\sqrt{d_{m'}}}{\frac{3\|\check{p}\|_\mu}{a_{11}} + \left(1.25 + \frac{32}{a_{10}}\right)} + \xi$$

for arbitrary $a_{12} > 0$ and, since $D_{m'} \leq T$ for all $m' \in M_T$, we obtain that

$$\sqrt{T}A_4 \leq (1 + a_{10}) \sqrt{\sum_{k=1}^{d_{m'}} \langle f_{m',k}^2, \check{p} \rangle_\mu} + a_{11}\sqrt{D_{m'}} + a_{12}\sqrt{d_{m'}} + \tilde{f}_2(\xi) \quad (6.32)$$

6 Penalisation in the discrete time framework

holds with probability larger than $1 - k'_2 e^{-\xi}$, where

$$\begin{aligned} \tilde{f}_2(\xi) &= \left(\frac{3\|\check{p}\|_\mu}{a_{11}} + \left(1.25 + \frac{32}{a_{10}} \right) \right) \xi, \\ k'_2 &= b_1 \sum_{l=1}^{\infty} l^{b_2} \exp \left(-\sqrt{l} \frac{a_{12}}{\frac{3\|\check{p}\|_\mu}{a_{11}} + \left(1.25 + \frac{32}{a_{10}} \right)} \right). \end{aligned}$$

We square (6.32) and invoke (6.22) repeatedly. Then, after renaming the constants, we obtain for a quadratic function f_2 increasing on $[0, \infty[$ that

$$(1 + a_8)A_4 + \frac{f_1(\xi)}{T} \leq \frac{(1 + a_{10})}{T} \sum_{k=1}^{d_{m'}} \langle f_{m',k}^2, \check{p} \rangle_\mu + \frac{a_{11}D_{m'}}{T} + \frac{a_{12}d_{m'}}{T} + \frac{f_2(\xi)}{T} \quad (6.33)$$

holds for all $m' \in M_T$ with probability larger than $1 - k'_2 e^{-\xi}$.

Furthermore, invoking (6.22), we observe

$$\begin{aligned} A_5 &\leq \left(\frac{2K'_\eta \sqrt{D_{m'}} + 2ndK'_\eta \sqrt{D'_{m'}}}{T} \right)^2 d_{m'} \\ &= \frac{4(K'_\eta)^2 D_{m'} d_{m'}}{T^2} + \frac{4(ndK'_\eta)^2 D'_{m'} d_{m'}}{T^2}. \end{aligned}$$

Hence, invoking (6.21), we obtain

$$A_5 \leq \frac{2(K'_\eta)^2 D_{m'}^2}{T^2} + \frac{2(K'_\eta)^2 n^2 d^2 (D'_{m'})^2}{T^2} + \frac{2(K'_\eta)^2 (1 + n^2 d^2) d_{m'}^2}{T^2}.$$

Therefore, invoking (6.21), we deduce that

$$\left. \begin{aligned} A_5 &\leq \frac{a'_{13} d_{m'}^4}{T^3} + \frac{a'_{14} D_{m'}^4}{T^3} + \frac{a'_{15} (D'_{m'})^4}{T^3} \\ &\quad + \frac{1}{T} \left(\frac{(K'_\eta)^4 (1 + n^2 d^2)^2}{a'_{13}} + \frac{(K'_\eta)^4}{a'_{14}} + \frac{(K'_\eta)^4 n^4 d^4}{a'_{15}} \right) \end{aligned} \right\} \quad (6.34)$$

holds true for arbitrary $a'_{13}, a'_{14}, a'_{15} > 0$. Analogously, for arbitrary $a''_{13}, a''_{14} > 0$, we observe

$$\left. \begin{aligned} A_6 &\leq \frac{4\Psi_T^2 D_{m'} d_{m'}}{T^2} \leq \frac{2\Psi_T^2 D_{m'}^2}{T^2} + \frac{2\Psi_T^2 d_{m'}^2}{T^2} \\ &\leq \frac{a''_{13} d_{m'}^4}{T^3} + \frac{a''_{14} D_{m'}^4}{T^3} + \frac{\Psi_T^4}{T} \left(\frac{1}{a''_{13}} + \frac{1}{a''_{14}} \right). \end{aligned} \right\} \quad (6.35)$$

6.2 Oracle inequality for penalised projection estimation

We denote

$$\begin{aligned}\bar{K}_3 &:= \bar{K}_1 + \left(\frac{(K'_\eta)^4(1+n^2d^2)^2}{a'_{13}} + \frac{(K'_\eta)^4}{a'_{14}} + \frac{(K'_\eta)^4n^4d^4}{a'_{15}} \right), \\ \bar{K}_4 &:= \bar{K}_2 \vee \left(\frac{1}{a''_{13}} + \frac{1}{a''_{14}} \right), \\ k_2 &:= k_1 + k'_2.\end{aligned}$$

Combining (6.33), (6.34) and (6.35) with (6.28), we observe directly that (6.29) holds with probability larger than $1 - k_2e^{-\xi}$. \diamond

Step 9: “Proof of (6.3)”

As a consequence of Figueroa-López and Houdré [18, Proposition 9.3], we observe that for arbitrary $a'_{16} > 0$ and for every $a''_{m'} > 0$, we have that

$$(1 + a'_{16}) \left(\sum_{k=1}^{d_{m'}} \iint_{[0,T] \times \mathbb{D}} \frac{f_{m',k}^2(\mathbf{x})}{T} \check{J}(dt, d\mathbf{x}) + \left(\frac{1}{2a'_{16}} + \frac{5}{6} \right) \frac{D_{m'}}{T} a''_{m'} \right) \geq \sum_{k=1}^{d_m} \langle f_{m',k}^2, \check{p} \rangle_\mu$$

holds with probability larger than $1 - \sum_{m' \in M_T} e^{-a''_{m'}}$. We derive from (6.12) that

$$\iint_{[0,T] \times \mathbb{D}} \left(\sum_{k=1}^{d_{m'}} \frac{f_{m',k}^2(\mathbf{x})}{T} \right) \check{J}(dt, d\mathbf{x}) = \frac{1}{T} \sum_{g \in G_T} \sum_{k=1}^{d_{m'}} f_{m',k}^2(\mathbf{x}) \mathbb{1}_{\{N_g=1\}} \leq \sum_{k=1}^{d_{m'}} \tilde{\nu}(f_{m',k}^2).$$

Let us recall that $D_{m'} \leq T$ for all $m' \in M_T$. Then by the choice of $a''_{m'} = a'_{17}d_{m'} + \xi$ for an arbitrary $a'_{17} > 0$, we obtain

$$(1 + a'_{16}) \sum_{k=1}^{d_{m'}} \tilde{\nu}(f_{m',k}^2) + \left(\frac{1}{2a'_{16}} + \frac{5}{6} \right) a'_{17}d_{m'} + \left(\frac{1}{2a'_{16}} + \frac{5}{6} \right) \xi \geq \sum_{k=1}^{d_{m'}} \langle f_{m',k}^2, \check{p} \rangle_\mu$$

with probability larger than $1 - k'_3e^{-\xi}$, where $k'_3 = b_1 \sum_{l=1}^{\infty} l^{b_2} \exp(-a'_{17}l)$. Therefore, for arbitrary $a_{16} > 0$ and $a_{17} > 0$, there exist $k_3 < \infty$ and a quadratic polynomial $f_3 : \xi \mapsto f_3(\xi)$ increasing on $[0, \infty[$ such that

$$\left. \begin{aligned} C \|p - \tilde{p}_{\text{pen}}\|_\mu^2 &\leq C' \|p - \mathcal{P}_m p\|_\mu^2 + \frac{(1 + a_{16})}{T} \sum_{k=1}^{d_{m_T}^{\text{pen}}} \tilde{\nu}(f_{m_T}^{\text{pen}}) \\ &\quad + \frac{a_4 D_m}{T} + \frac{a_5 D'_{m_T}^{\text{pen}}}{T} + \frac{a_6 D'_m}{T} + \frac{a_{11} D_{m_T}^{\text{pen}}}{T} \\ &\quad + \frac{a_{12} d_{m_T}^{\text{pen}}}{T} + \frac{a_{13} d_{m_T}^4}{T^3} + \frac{a_{14} D_{m_T}^4}{T^3} + \frac{a_{15} (D'_{m_T}^{\text{pen}})^4}{T^3} \\ &\quad + \text{pen}_T(m) - \text{pen}_T(m_T^{\text{pen}}) \\ &\quad + \frac{f_3(\xi)}{T} + \frac{\bar{K}_3}{T} + \frac{\bar{K}_4 (\Psi_T^2 + \Psi_T^4)}{T} \end{aligned} \right\} \quad (6.36)$$

6 Penalisation in the discrete time framework

holds with probability larger than $1 - k_3 e^{-\xi}$. In particular,

$$f_3(\xi) := f_2(\xi) + \left(\frac{1}{2a'_{16}} + \frac{5}{6} \right) \xi,$$

$$k_3 := k_2 + k'_3.$$

Next, we take $a_{16} = 1 - c_1$, $a_{11} = c_2$, $a_5 = c_4$, $a_{12} = c_3$, $a_{13} = c_7$, $a_{14} = c_5$ and $a_{15} = c_6$ to cancel “ $-\text{pen}_T(m_T^{\text{pen}})$ ” on the right-hand side of (6.36). Hence, we rewrite (6.36) and deduce that

$$A_7 := C \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 - C' \|p - \mathcal{P}_m p\|_{\mu}^2 - \frac{a_4 D_m}{T} - \frac{a_6 D'_m}{T}$$

$$- \text{pen}_T(m) + \frac{\bar{K}_3}{T} - \frac{\bar{K}_4 (\Psi_T^2 + \Psi_T^4)}{T} \leq \frac{f_3(\xi)}{T}$$

holds with probability greater than $1 - k_3 e^{-\xi}$. Using the idea from Figueroa-López and Houdré [18, Lemma 9.4], we observe

$$\mathbb{E}[A_7] \leq \mathbb{E}[A_7^+] \leq \int_0^{\infty} P(A_7 \geq x) dx \leq k_3 \int_0^{\infty} e^{-f_3^{-1}(x)} dx = k_3 \int_0^{\infty} e^{-\xi} f'_3(\xi) d\xi.$$

Since $f_3(0) = 0$, integration by parts yields that

$$C \mathbb{E} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \leq C' \|p - \mathcal{P}_m p\|_{\mu}^2 + \left(1 + \left(\frac{a_4}{c_2} \vee \frac{a_6}{c_4} \right) \right) \mathbb{E}_{P_2}[\text{pen}_T(m)] + \frac{C''}{T} \quad (6.37)$$

holds with

$$C'' = k_3 \int_0^{\infty} e^{-\xi} f_3(\xi) d\xi + \bar{K}_3 + \bar{K}_4 \left(\mathbb{E}[\Psi_T^2] + \mathbb{E}[\Psi_T^4] \right).$$

We recall that Ψ_T has binomial distribution $B_{\tilde{n}, \rho}$ with parameters $\tilde{n} = \lfloor T/\tau \rfloor$ and $\rho \leq K_{\varepsilon, \eta} \tau^3$. Hence, we observe that the second and fourth moment of Ψ_T are finite and converge to zero as $T \rightarrow \infty$ by choice of $\tau = o(T^{-2})$. In particular,

$$\left. \begin{aligned} \mathbb{E}[\Psi_T^2] &= \tilde{n}\rho(1 - \rho) + (\tilde{n}\rho)^2 \leq K_{\varepsilon, \eta} T \tau^2 + K_{\varepsilon, \eta}^2 T^2 \tau^4 \rightarrow 0, \\ \mathbb{E}[\Psi_T^4] &= (\tilde{n} - 3)(\tilde{n} - 2)(\tilde{n} - 1)\tilde{n}\rho^4 + 6(\tilde{n} - 2)(\tilde{n} - 1)\tilde{n}\rho^3 \\ &\quad + 7(\tilde{n} - 1)\tilde{n}\rho^2 + \tilde{n}\rho \\ &= (\tilde{n}\rho)^4 - 6\tilde{n}^3\rho^4 + 11\tilde{n}^2\rho^4 - 6\tilde{n}\rho^4 + 6\tilde{n}^3\rho^3 - 18\tilde{n}^2\rho^3 + 12\tilde{n}\rho^3 \\ &\quad + 7\tilde{n}^2\rho^2 - 7\tilde{n}\rho^2 + \tilde{n}\rho \\ &\leq K_{\varepsilon, \eta}^4 (T^4 \tau^8 + 11T^2 \tau^{10}) + K_{\varepsilon, \eta}^3 (6T^3 \tau^6 + 12T \tau^8) \\ &\quad + 7K_{\varepsilon, \eta}^2 T^2 \tau^4 + K_{\varepsilon, \eta} T \tau^2 \rightarrow 0. \end{aligned} \right\} \quad (6.38)$$

Finally, we set

$$C_1 = \frac{C'}{C} \vee \frac{1 + \left(\frac{a_4}{c_2} \vee \frac{a_6}{c_4}\right)}{C} < \infty \text{ and } C_2 = \frac{C''}{C} < \infty$$

Then we deduce that the oracle inequality (6.3) holds by the combination of (6.37) and (6.38), as $m \in M_T$ is arbitrary. \square

6.3 Estimation of smooth multivariate Lévy densities

In the continuous time framework (cf. Theorem 3.3.7), we gave an explicit rate of convergence for the PPE for univariate Besov-type smooth functions. We can readily adapt this result to the discrete time framework, however, the multivariate case turns out to be more complex. Probably, an equivalent approximation result as Proposition 3.3.3 holds in the case of multivariate Besov-smooth functions. Nevertheless, we restrict ourselves to Sobolev spaces (recall Definition 5.1.7) in the following, in order to use the theory of finite elements as introduced in Brenner and Scott [7]. (It is noteworthy that every Sobolev space is a subspace of a certain Besov space.) Analogously, we obtain bounds for the approximation error of Sobolev-type smooth functions when approximated by piecewise polynomials.

Finite elements

We adapt the concept of finite elements as introduced in Ciarlet [11] and adapted in Brenner and Scott [7, Definition 3.1.1].

Definition 6.3.1 (Finite element) *Let $D \subseteq \mathbb{R}^d$ be a domain with piecewise smooth boundary (the element domain), \mathcal{Q} be a finite-dimensional space of functions on D (the shape functions) and \mathcal{N} be a basis for the dual space $\hat{\mathcal{Q}}$ of \mathcal{Q} (the nodal variables), then $(D, \mathcal{Q}, \mathcal{N})$ is called finite element.*

Example 6.3.2 (Orthotope Lagrange element) *Let $D = [0, 1]^d$ be the d -dimensional unit orthotope and*

$$\mathcal{Q}_k := \left\{ \mathbf{x} \mapsto \sum_j c_j \prod_{i=1}^d f_{j,i}(x_i) : c_j \in \mathbb{R}, f_{j,i} \text{ polynomials of maximum degree } k \right\}.$$

Then $\dim(\mathcal{Q}_k) = \binom{d+k}{k}$, and Brenner and Scott [7, Chapter 3.5 and 3.6] inductively construct a set $\{\mathbf{v}_j : j = 1, \dots, \binom{d+k}{k}\} \subseteq D$ such that

$$\mathcal{N} := \left\{ g_j : \mathcal{Q}_k \rightarrow \mathbb{R}; f \mapsto g_j(f) := f(\mathbf{v}_j) : j = 1, \dots, \binom{d+k}{k} \right\}$$

6 Penalisation in the discrete time framework

is a basis for $\hat{\mathcal{Q}}_k$. Moreover, we have $\mathcal{N} \subseteq \hat{\mathcal{C}}^0$, i. e. every $g_j \in \mathcal{N}$ does not involve any derivative. The finite element $(D, \mathcal{Q}_k, \mathcal{N})$ is called Lagrange element over D .

For a connected set $D \subseteq \mathbb{R}^d$, we denote the *diameter* of D by

$$\text{diam } D := \max\{\|x_1 - x_2\|_\infty : x_1, x_2 \in D\}.$$

Finally, we derive the following Proposition from Brenner and Scott [7, Theorem 4.4.20].

Proposition 6.3.3 (Approximation error for Sobolev-type functions)

Let $\mathbb{D} \in \mathcal{B}_\circ^d$ be a compact d -dimensional interval such that $\mu(\mathbb{D}) < \infty$ and assume that the restriction $p|_{\mathbb{D}}$ of the Lévy density p to \mathbb{D} satisfies $p|_{\mathbb{D}} \in \mathcal{W}^{r,2}(\mathbb{D}, \mu)$ for some $r \in \mathbb{N}$. Further, let \mathcal{D}_m be a partition of \mathbb{D} such that

$$\max\{\text{diam } D : D \in \mathcal{D}_m\} \leq m^{-1} \text{diam } \mathbb{D}$$

and every $D \in \mathcal{D}_m$ is a d -dimensional interval. Moreover, we assume that there is a positive $\rho > 0$ independent of m such that for every $D \in \mathcal{D}_m$, there exists a ball $B_D \subset D$ with $\text{diam } B_D \geq \rho \text{diam } D$. Let \mathbb{S}_m be the projection space consisting of all functions $f : \mathbb{D} \rightarrow \mathbb{R}$ such that the restriction $f|_D$ of f to every $D \in \mathcal{D}_m$ satisfies $f|_D \in \mathcal{Q}_{r-1}$, where \mathcal{Q} is defined as in Example 6.3.2. Then there is a finite positive constant $c(r, \rho) < \infty$ such that

$$\|p - \mathcal{P}_m p\|_\mu \leq c(r, \rho) m^{-r} |p|_{\mathbb{D}}|_{\mathcal{W}^{r,2}}. \quad (6.39)$$

Proof:

We note that every d -dimensional interval is (geometrically spoken) a polyhedron. For every $D \in \mathcal{D}_m$, there exists an affine mapping $\mathbf{x} \mapsto A\mathbf{x} + \mathbf{b}$ for some $A \in \mathbb{R}^{d \times d}$ and $\mathbf{b} \in \mathbb{R}^d$ such that $(D, \mathcal{Q}_k(D), \mathcal{N}(D))$ is a finite element, *affine-equivalent* (w. r. t. the affine mapping) to the Lagrange element from Example 6.3.2 (cf. Brenner and Scott [7, Definition 3.4.1]). As the Lagrange element satisfies the conditions of Brenner and Scott [7, Theorem 4.4.4], we invoke Brenner and Scott [7, Theorem 4.4.20]. Hence, there is an $f \in \mathbb{S}_m$ (in particular the interpolant w. r. t. the finite elements $\{(D, \mathcal{Q}_k(D), \mathcal{N}(D)) : D \in \mathcal{D}\}$) such that for every $0 \leq r' \leq r$, there is a finite constant $c(r, \rho) < \infty$ with

$$\left(\sum_{D \in \mathcal{D}_m} \|p - f|_D\|_{\mathcal{W}^{r',2}(D,\mu)}^2 \right)^{1/2} \leq c(r, \rho) m^{r'-r} |p|_{\mathbb{D}}|_{\mathcal{W}^{r,2}(\mathbb{D},\mu)}.$$

For $r' = 0$, the left-hand side reduces to $\|p - f\|_\mu$ for an $f \in \mathbb{S}_m$. Therefore, (6.39) follows directly by Definition 2.3.5. \square

Rate of convergence and remark

Theorem 6.3.4 (Rate of convergence of the PPE in DT)

Let \mathbf{X} be an \mathbb{R}^d -valued Lévy process with generating triplet (γ, Σ, ν) . Further, let $\mathbb{D} \in \mathcal{B}_\circ^d$ be a compact d -dimensional interval such that $\mu(\mathbb{D}) < \infty$ and assume that the restriction $p|_{\mathbb{D}}$ of the Lévy density p to \mathbb{D} satisfies $p|_{\mathbb{D}} \in \mathcal{W}^{r,2}(\mathbb{D}, \mu)$ for some $r \in \mathbb{N}$ with $r > d/4$. Further, for $m \in M := \mathbb{N}$, let $\mathbb{S}_m \subseteq L^2(\mathbb{D}, \mu)$ be the space of piecewise polynomials of maximum degree $r - 1$ as introduced in Proposition 6.3.3. Moreover, let pen be a penalty on $\{\mathbb{S}_m : m \in M\}$ of form (6.2). Then a mesh size of $\tau = o(T^{-2})$ is sufficient to ensure that the penalised projection estimator in DT from Definition 6.1.2 satisfies

$$\sup_{T>0} T^{\frac{2r}{2r+d}} \mathbb{E} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 < \infty. \quad (6.40)$$

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

$$B(a_1, a_2) := \left\{ f \in \mathcal{W}^{r,2}(\mathbb{D}, \mu) : \|f\|_{\mathbb{R}^d \setminus B_\eta(0), \infty} \leq a_1 \text{ and } |f|_{\mathcal{W}^{r,2}} \leq a_2 \right\}$$

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

$$\sup_{T>0} T^{\frac{2r}{2r+d}} \sup_{p \in B(a_1, a_2)} \mathbb{E} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 < \infty \quad (6.41)$$

holds true for the PPE in DT w. r. t. the penalty pen .

Proof:

From Theorem 6.2.1, with the notation given therein, we infer that there exist finite constants $C_1, C_2 < \infty$ such that

$$\mathbb{E} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \leq C_1 \inf_{m \in M_T} \left(\|p - \mathcal{P}_{\text{pen}} p\|_{\mu}^2 + \mathbb{E}[\text{pen}_T(m)] \right) + \frac{C_2}{T}.$$

Similarly, as in the proof of Theorem 3.3.7, we deduce from Proposition 6.3.3 that there is another finite constant $K_1 = K_1(|p|_{\mathcal{W}^{r,2}}) < \infty$ such that

$$\|p - \mathcal{P}_m p\|_{\mu}^2 \leq K_1 m^{-2r}.$$

Moreover, equivalently to the arguments used in the univariate case, we observe that $d_m = O(m^d)$, $D_m = O(m^d)$ and $D'_m = O(m^d)$. Thus, we derive from (6.2) that

$$\mathbb{E}[\text{pen}_T(m)] \leq K_2 \left(\frac{m^d}{T} + \frac{m^{4d}}{T^3} \right)$$

holds for a finite constant $K_2 = K_2(\|p\|_{\mathbb{R}^d \setminus B_\eta(0), \infty})$.

6 Penalisation in the discrete time framework

The rate of convergence $T^{2r/(2r+d)}$ follows equivalently to the proof of Theorem 3.3.6, since $r > d/4$ makes the term m^{4d}/T^3 negligible compared to m^{-2r} and m/T . Moreover, we note that $B(a_1, a_2)$ is a compact subspace of $\mathcal{W}^{r,2}(\mathbb{D}, \mu)$. Therefore, there exists a uniform bound $K' < \infty$ for $\sup_{f \in B(a_1, a_2)} \{K_1(f), K_2(f)\}$. Hence, the finiteness in (6.41) is clear by virtue of (6.40). \square

Remark 6.3.5 (Critical mesh) *We review the proof of Theorem 6.2.1 and observe that the requirement $\tau = o(T^{-2})$ is crucial for the rate in (6.20) only. In particular, by choosing $\eta > 0$ in Step 1 small enough, we are able to relax the requirement in (6.38) to arbitrary slow convergence of $\tau \rightarrow 0$.*

In the case that Brownian motion is absent, i. e. $\Sigma = \mathbf{0}$, it is sufficient to require $\tau = o(T^{-(1 \vee \beta_{\mathbf{X}})})$ in the light of Theorem 5.4.3, where $\beta_{\mathbf{X}}$ denotes the Blumenthal-Gettoor index of \mathbf{X} , and (6.3) remains valid.

If we relax the mesh size requirement to $\tau = o(T^{-(1+\delta)})$ (or $\tau = o(T^{-((1 \vee \beta_{\mathbf{X}})+\delta)})$ in the case of $\sigma = 0$) for some $0 < \delta \leq 1$, then there is a finite constant $C_3 < \infty$ such that (6.3) rewrites to

$$\mathbb{E} \|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \leq C_1 \inf_{m \in M_T} \left(\|p - \mathcal{P}_{\text{pen}} p\|_{\mu}^2 + \mathbb{E}[\text{pen}_T(m)] \right) + \frac{C_2}{T} + \frac{C_3}{T^{\delta}}.$$

Thus, the rate of convergence in (6.40) decreases to $T^{\frac{2r}{2r+d} \wedge \delta}$.

7 A simulation study

In this chapter, we simulate sample paths of Lévy processes and investigate our estimates in continuous and discrete time framework at work. In Section 7.1, we deal with the case of a univariate Brownian motion with compound Poisson jumps, Section 7.2 is dedicated to the univariate stable cases with finite and infinite variation and the bivariate stable case.

7.1 Brownian motion with compound Poisson jumps

This section is dedicated to the case of the composition of a univariate Brownian motion with compound Poisson jumps as introduced in Section 5.2. Let X be a Lévy process with Lévy-Khintchine triplet (γ, σ^2, ν) satisfying Assumption 5.2.1. We recall that X can be decomposed as

$$X_t = \gamma_0 t + \sigma W_t + \sum_{k=1}^{N_t} Z_k \quad \text{for } t \in \mathbb{R}_0^+, \quad (7.1)$$

where $\gamma_0 := \gamma - \int_{\{x: |x| \in]0,1]\}} \nu(dx)$, W is a standard Brownian motion, N is a Poisson process with intensity $\lambda := \nu(\mathbb{R}_o)$ independent of W and $\{Z_k : k \in \mathbb{N}\}$ is a family of i. i. d. random variables independent of W and N with distribution $\lambda^{-1}\nu$.

Chi-squared jumps

In our first example, we restrict ourselves to positive jumps only. In particular, let $Z_1 \sim \chi_4^2$, N have intensity $\lambda = 10$ and $\sigma = 1$. In contrast to Example 2.3.2, we assume that the compound Poisson jumps are not compensated, i. e. $\gamma_0 = 0$. Since (7.1) allows for separate simulation of W , N and Z , we take the opportunity to investigate the penalised projection estimators in discrete and continuous time framework. We base our analysis on the (discretely observed) sample path on the time horizon $[0, 100]$ shown in Figure 7.1. The magnified plot on the right-hand side reveals the piecewise constant nature of the discretized Lévy process w. r. t. the mesh size $\tau = 0.001$ in coherence with Definition 4.1.1.

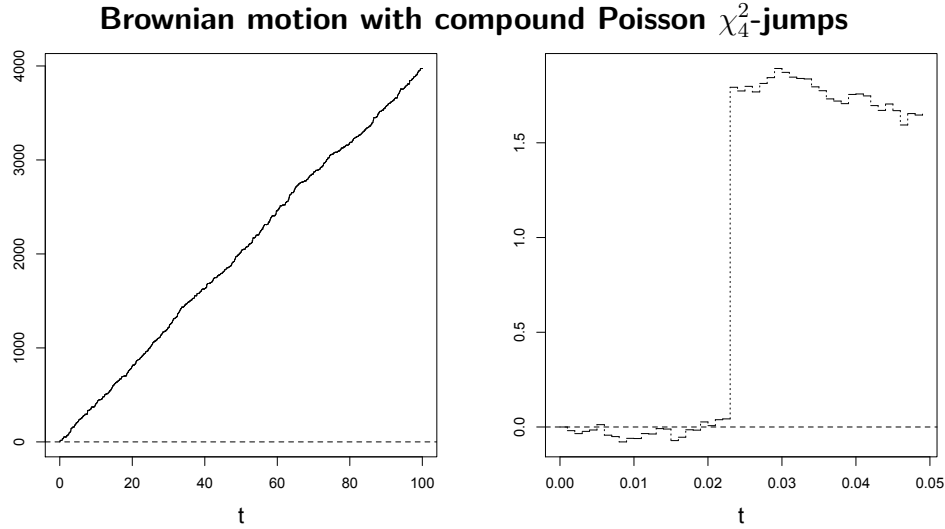


Figure 7.1: We present a sample path $X(\omega)$ of form (7.1) with χ_4^2 distributed jumps on the time horizon $[0, 100]$ (*left*) and a magnified version of $X(\omega)$ on the time horizon $[0, 0.05]$ (*right*).

Projection estimation in the continuous time framework

In order to get a first impression of the influx of discretisation on penalised projection estimation, assume we are able to observe the PRM $J(\cdot, \omega) = \sum_{t \leq T} \delta_{\Delta X_t}$, where δ_{\cdot} is the Dirac measure. We can directly use the methods from Section 2.3 to come up with a PPE for the Lévy measure ν . In our sample path, $N_T(\omega) = 993$ jumps occurred, where $E[N_T] = 1000$ jumps are expected. To illustrate the point, we present an empirical histogram of the occurred jump sizes in Figure 7.2.

We choose $\mathbb{D} =]0, 20]$ as domain of estimation and μ identical to the Lebesgue measure on \mathbb{D} , as there is no additional issue with estimating the Lévy density in a neighbourhood of the origin. This is due to the boundedness of the Lebesgue density of the χ^2 distribution family. In the following, we present penalised projection estimates based on two collections of projection spaces. In particular, let $\{\mathbb{S}_m : m \in M\}$ be the family of spaces consisting of piecewise constant functions on \mathbb{D} based on the regular partition into m classes (see Example (3.1.2)). Similarly, let $\{\mathbb{S}'_m : m \in M\}$ be the family of spaces consisting of piecewise quadratic polynomials based on the same partition. Additionally, we choose the penalty pen to be of form (3.12) with $c_1 = 2$ and $c_2 = c_3 = 1$. We recall that for $T > 0$ and $m \in M_T$

$$\text{pen}_T(m) = \frac{2}{T} \sum_{k=1}^{d_m} \sum_{t \leq T} f_k^2(\Delta X_t) + \frac{D_m + d_m}{T}.$$

We show the results in Figure 7.3 and Table 7.1. We observe that the PPE based

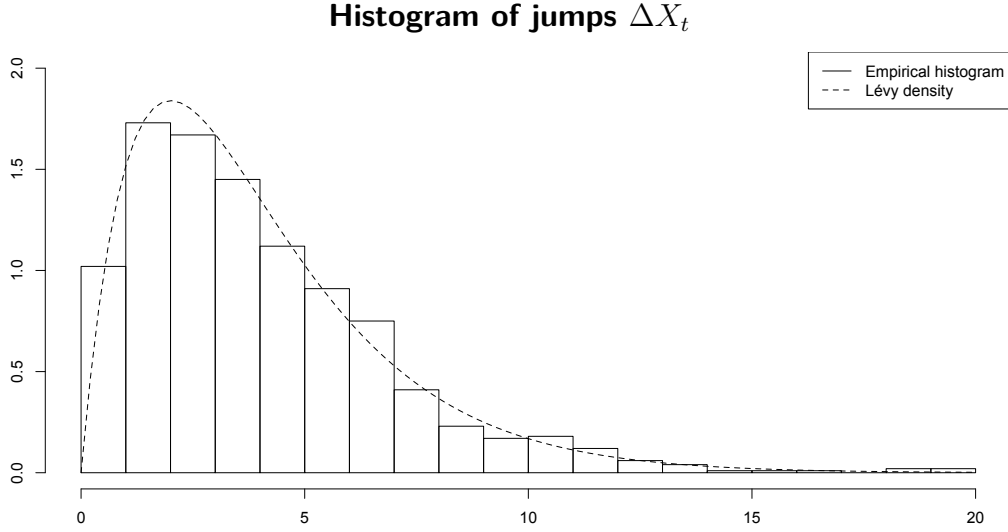


Figure 7.2: An empirical histogram of the jump sizes $\Delta X_t(\omega)$ occurred on the time horizon $[0, 100]$ is presented. The bars are scaled such that the area underneath equals the empirical intensity $N_T(\omega)/T = 9.93$ (*solid*). For comparison, we show, additionally, the Lévy density of $\nu(\cdot) = 10\chi_4^2(\cdot)$ of X (*dashed*).

on piecewise quadratic polynomials appears to have a lower squared error (0.11 compared to 0.28). Moreover, recalling that the dimension of a projection space consisting of piecewise polynomials with maximum degree k on a partition of \mathbb{D} into m classes is equal to $(k+1)m$, we observe that the dimension of the underlying projection space is lower in the case of piecewise quadratics as well ($\dim \mathbb{S}'_7 = 21$ compared to $\dim \mathbb{S}_{31} = 31$).

	m_T^{pen}	SE	AE	SB	MVar
\mathbb{S}	31	0.2811	1.417	0.1036	0.1775
\mathbb{S}'	7	0.1130	0.9625	0.008588	0.1044

Table 7.1: We present the estimated best projection space (m_T^{pen}), the corresponding squared error $\|p - \hat{p}_{\text{pen}}\|_{\mu}^2$ (*SE*) and the absolute error $\|p - \hat{p}_{\text{pen}}\|_{L^1(\mathbb{D}, \mu)}$ (*AE*) for the PPE based on the families $\{\mathbb{S}_m : m \in M\}$ of piecewise constants and $\{\mathbb{S}'_m : m \in M\}$ of piecewise quadratic polynomials. Moreover, we decompose the squared error into the squared μ -bias (*SB*) and the μ -variance (*MVar*).

7 A simulation study

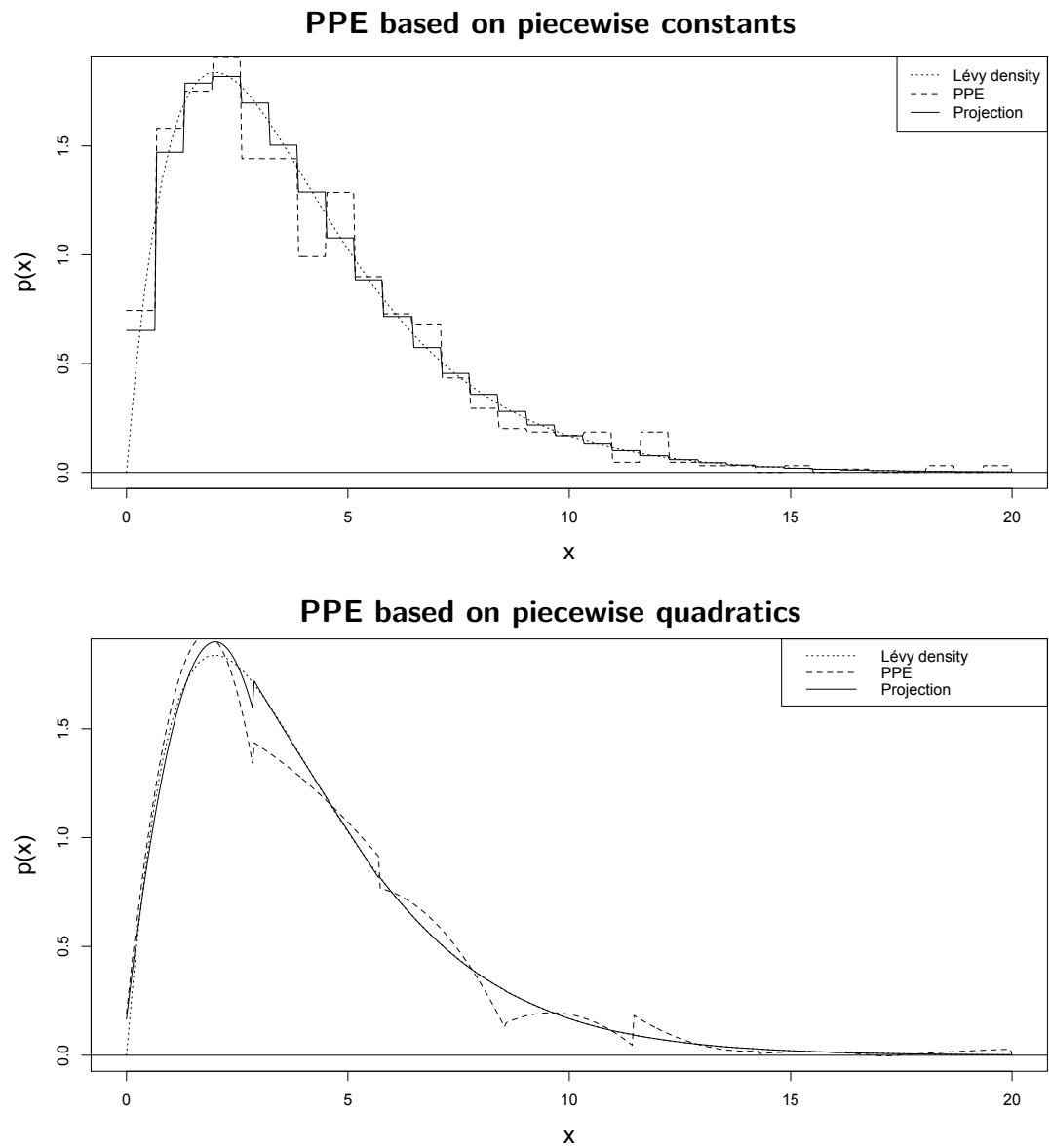


Figure 7.3: We present the μ -orthogonal projection $\mathcal{P}_m p$ (*solid*) and the penalised projection estimator \hat{p}_{pen} (*dashed*) of the restriction of the Lévy density ν' to $\mathbb{D} =]0, 20]$, i.e. the density of $\nu(\cdot) = \lambda \chi_4^2(\cdot)$ with $\lambda = 10$ (*dotted*). The projection and the PPE are presented for the family of piecewise constants based on the partition of \mathbb{D} into $m = 31$ classes (*top*) and for the family of piecewise quadratic polynomials based on the partition of \mathbb{D} into $m = 7$ classes (*bottom*).

Projection estimation in the discrete time framework

Let us focus on penalised projection estimation based on the observed increments as shown in Figure 7.1. We choose $\mathbb{D} = [D_{\min}, 20]$ as domain of estimation and μ identical to the Lebesgue measure on \mathbb{D} . As threshold, we choose

$$D_{\min}(\tau) = \hat{\sigma} \sqrt{2\tau^{0.9} \log\left(\frac{1}{\tau}\right)} \quad \text{for } \tau = \tau(T) > 0$$

derived from Mancini [29, Theorem 3.1], where

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{k=1}^{\lfloor T/\tau \rfloor} X_{\Delta_k^\tau}^2(\omega) \mathbb{1}_{[0, r(\tau)]}(X_{\Delta_k^\tau}^2(\omega)).$$

In our sample, we estimate the volatility to be $\hat{\sigma}^2 = 0.9899$, which is reliable compared to the actual value of $\sigma^2 = 1$. Coherently, we have $D_{\min} = 0.1652$ in this scenario. From the 100 000 observed increments, only 986 exceeding the

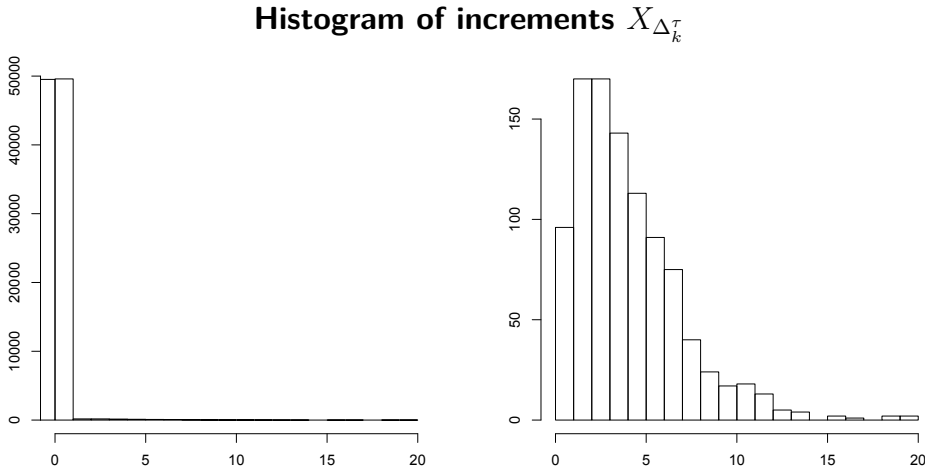


Figure 7.4: We present an empirical histogram of the observed increments $X_{\Delta_k^\tau}$ on the time horizon $[0, 100]$ w.r.t. the mesh size $\tau = 0.001$ (*left*), where we observe that the vast majority of all 100 000 increments is small positive or small negative and caused solely by the Brownian motion. Additionally, we present an empirical histogram of the observed increments $X_{\Delta_k^\tau}$ exceeding the threshold of $D_{\min} = 0.1652$ (*right*), where the principal structure of the χ_4^2 -distribution is observable.

threshold remain. We recall that $N_T(\omega) = 993$ jumps actually occurred. The effect is illustrated in Figure 7.4.

In the following, we present PPEs w.r.t. piecewise constants and piecewise quadratic polynomials, as we have done before in the continuous time framework.

7 A simulation study

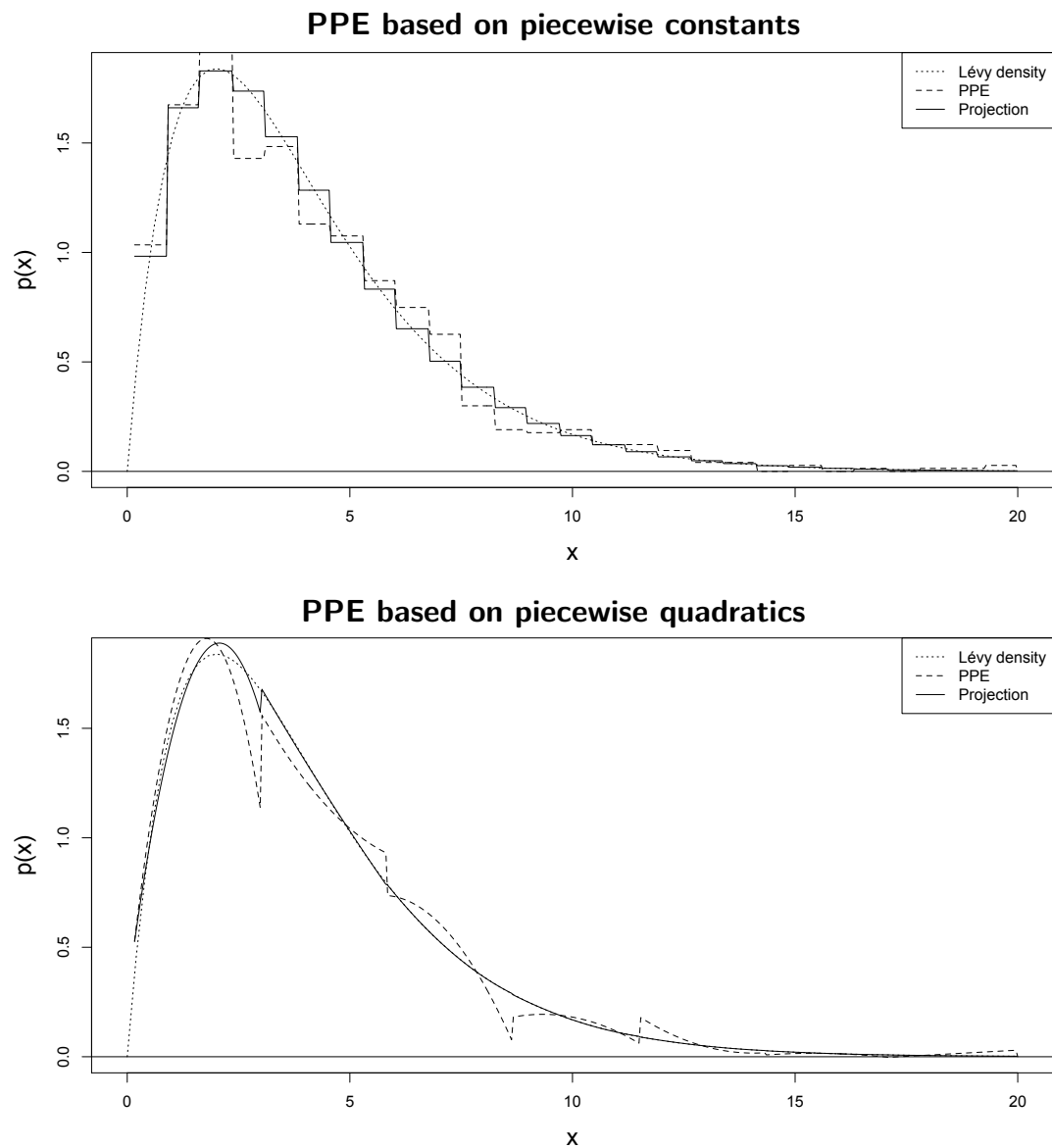


Figure 7.5: We present the μ -orthogonal projection $\mathcal{P}_m p$ (*solid*) and the penalised projection estimator \tilde{p}_{pen} (*dashed*) of the restriction of the Lévy density ν' to $\mathbb{D} = [0.1652, 20]$, i.e. the density of $\nu(\cdot) = \lambda \chi_4^2(\cdot)$ with $\lambda = 10$ (*dotted*). The projection and the PPE are presented for the family of piecewise constants based on the partition of \mathbb{D} into $m = 27$ classes (*top*) and for the family of piecewise quadratic polynomials based on the partition of \mathbb{D} into $m = 7$ classes (*bottom*).

7.1 Brownian motion with compound Poisson jumps

We choose penalty pen of form (6.2) with $c_1 = 2$ and $c_2 = \dots = c_7 = 1$. We show the results in Figure 7.5 and Table 7.2. We observe that the PPE based on piecewise quadratic polynomials appears to have a lower squared error (0.12 compared to 0.24), equivalently to the continuous time framework. Again the complexity of the underlying projection space is lower in the latter case as well ($\dim \mathbb{S}'_7 = 21$ compared to $\dim \mathbb{S}_{27} = 27$). The absolute values for the errors in Tables 7.1 and 7.2 are not directly comparable as the underlying domain of estimation \mathbb{D} differs.

	m_T^{pen}	SE	AE	SB	MVar
\mathbb{S}	27	0.2383	1.327	0.09512	0.1432
\mathbb{S}'	7	0.1248	0.9402	0.006486	0.1183

Table 7.2: We present the estimated best projection space (m_T^{pen}), the corresponding squared error $\|p - \tilde{p}_{\text{pen}}\|_{\mu}^2$ (*SE*) and the absolute error $\|p - \tilde{p}_{\text{pen}}\|_{L^1(\mathbb{D}, \mu)}$ (*AE*) for the PPE based on the families $\{\mathbb{S}_m : m \in M\}$ of piecewise constants and $\{\mathbb{S}'_m : m \in M\}$ of piecewise quadratic polynomials on the domain of estimation $\mathbb{D} = [0.1652, 20]$. Moreover, we decompose the squared error into the squared μ -bias (*SB*) and the μ -variance (*MVar*).

Gaussian jumps

We remain in the framework of univariate Brownian motion with compound Poisson jumps, where the Lévy-Itô decomposition is of form (7.1). In this example, let $Z_1 \sim \mathcal{N}(0, 1)$. We investigate the efficiency of penalised projection estimation in DT at work for different parameters λ and σ and on different time horizons $[0, T]$. Theorem 5.2.2 implies that the critical mesh is given by $\tau = o(1/\sqrt{T})$. In particular, we choose $\tau(T) = 0.1/\sqrt{T}$ for our finitely many time horizons. Further, we choose $\mathbb{D} = [-5, -D_{\min}(\tau)] \cup [D_{\min}(\tau), 5]$ as domain of estimation and μ identical to the Lebesgue measure. As threshold, we choose

$$D_{\min}(\tau) = \hat{\sigma} \sqrt{2\tau^{0.9} \log\left(\frac{1}{\tau}\right)} \quad \text{for } \tau = \tau(T) > 0$$

derived from Mancini [29, Theorem 3.1], where

$$\hat{\sigma}^2 := \frac{1}{T} \sum_{k=1}^{\lfloor T/\tau \rfloor} (X_{\Delta_k^\tau})^2 \mathbb{1}_{((X_{\Delta_k^\tau})^2 \leq 2\tau^{0.9} \log(1/\tau))}.$$

Throughout, we base our estimates on the family $\{\mathbb{S}_m : m \in M\}$ consisting of piecewise quadratic polynomials based on the regular partition of \mathbb{D} into m classes and choose penalty pen to be of form (6.2) with $c_1 = 2$ and $c_2 = \dots = c_7 = 1$.

7 A simulation study

For each parameter set (λ, σ, T) , where we choose $\lambda \in \{10, 100\}$, $\sigma \in \{0.5, 1, 2\}$ and $T \in \{100, 200, 300, 500, 1000\}$, we simulate 100 trajectories. We present the results in Tables 7.3–7.8.

At first, we notice that the estimates $\hat{\sigma}$ for the volatility of the Brownian component are much less biased in the cases of $\sigma = 0.5$ and $\sigma = 1$ compared to $\sigma = 2$. This is due to the fact that the convergence in Mancini [29, Theorem 3.1] takes place in probability and $P((\sigma W_{\Delta_k^\tau})^2 \geq \varepsilon)$ is, obviously, increasing in σ^2 . Consequently, we take relatively low thresholds D_{\min} in the case of $\sigma = 2$. This leads to a higher influx of Brownian increments, causing an overestimation close to the origin. As a refined partition of \mathbb{D} leads to stronger overestimation that is only partially compensated by the penalty, we observe that the complexity of estimated best projection spaces significantly increases in the case of $\sigma = 2$.

Secondly, let us compare the empirical mean squared errors in the different cases. In all cases, we observe that the empirical mean squared errors decrease with a rate close to the theoretical best rate of convergence achievable. As we base our estimates on piecewise quadratic polynomials and the normal density belongs to $\mathcal{B}_\infty^r(L^2)$ for all $r > 0$, we conclude from Theorem 6.3.4 that the PPE converges at most like $T^{-6/7}$. For illustration, we make a least square regression of the logarithmic decrease of the MSE values from Table 7.3 against the log-increase of the time horizon. In particular, we calibrate the linear model

$$\log \left(\frac{\widehat{MSE}_{n+1}}{\widehat{MSE}_n} \right) = a \cdot \log \left(\frac{T_{n+1}}{T_n} \right) + U_n \quad (U_n)_{n \geq 1} \sim \mathcal{N}(0, \sigma_U^2),$$

and observe that the average rate of convergence $\hat{a} = -0.9092$ for the time horizons mentioned above is actually better than $-6/7$. However, the ultimate value, i. e. $\log(\widehat{MSE}_5/\widehat{MSE}_4)/\log(T_5/T_4) = -0.7321$ is worse than the theoretical best.

Additionally, we notice that the empirical MSEs in the high jump activity case (see Tables 7.6–7.8) are higher than the – appropriately scaled – MSEs in the low jump activity case (see Tables 7.3–7.5). In particular, we expect that scaling the jump activity λ and, hence, the Lévy density p by factor 10 would scale the empirical mean squared errors by the squared factor, i. e. 100. Since the observed relative increase of the empirical MSEs is larger, we conclude that scaling the jump activity induces additional effects on the estimator in the discrete time framework.

Naturally, we expect that a smaller mesh size τ is necessary in the higher jump activity case to accurately disentangle the jumps from each other and from the Brownian motion. Figure 7.6 illustrates the point. The PPE for $\lambda = 10$ (*bottom-left*) shows neither significant over- nor underestimation. Contrarily, the PPE for $\lambda = 100$ (*bottom-right*) shows significant underestimation for “medium sized” jumps with absolute values approximately in $]D_{\min}, 2[$. Apparently, $\tau = 0.0031$ is still too large compared to $\lambda = 100$ to detach all jumps from each other.

Brownian motion with Gaussian jumps

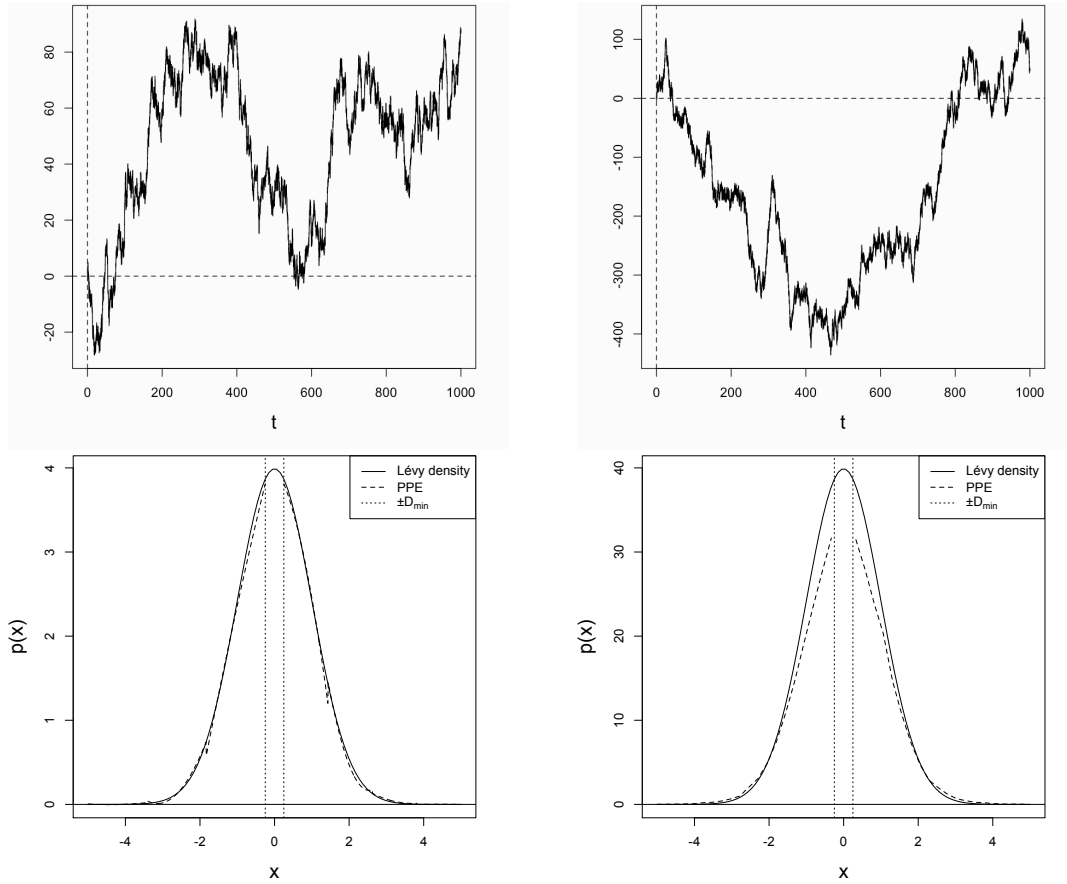


Figure 7.6: We present the sample paths $X(\omega)$ of the composition of a Brownian motion with volatility $\sigma = 1$ and Gaussian jumps with activity $\lambda = 10$ (*top-left*) and of the composition of a Brownian motion with volatility $\sigma = 2$ and Gaussian jumps with activity $\lambda = 100$ (*top-right*). The observation is made on the time horizon $[0, 1000]$ and on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.0031$. Furthermore, we present the corresponding penalised projection estimators (*dashed*) based on the family of piecewise quadratic polynomials on $\mathbb{D} = \{x : |x| \in [0.249, 5]\}$ (*bottom-left*) and $\mathbb{D} = \{x : |x| \in [0.329, 5]\}$ (*bottom-right*) for the restriction of the Lévy density, i. e. the normal density scaled by the factor λ (*solid*). The estimator for the negative part is based on the regular partition of \mathbb{D} into 3 and 4 classes, respectively. The estimator for the positive part is based on the regular partition of \mathbb{D} into 4 and 7 classes, respectively.

7 A simulation study

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	0.5235 (5.946×10^{-3})	3.9	0.2925 (0.1583)	1.077 (0.3662)
$T = 200$ ($\tau = 0.007$)	0.5167 (2.989×10^{-3})	5.9	0.1333 (0.06364)	0.6480 (0.1335)
$T = 300$ ($\tau = 0.0057$)	0.5137 (2.024×10^{-3})	6.3	0.1001 (0.04416)	0.5485 (0.1017)
$T = 500$ ($\tau = 0.0044$)	0.5104 (1.358×10^{-3})	6.4	0.06161 (0.03090)	0.4280 (0.09110)
$T = 1000$ ($\tau = 0.0031$)	0.5070 (7.766×10^{-4})	7.0	0.03709 (0.01396)	0.3311 (0.05223)

Table 7.3: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 10$ and $\sigma = 0.5$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	0.9469 (7.608×10^{-3})	4.1	0.3298 (0.8060)	0.9504 (0.3318)
$T = 200$ ($\tau = 0.007$)	0.9618 (3.670×10^{-3})	5.9	0.1134 (0.06719)	0.5753 (0.1521)
$T = 300$ ($\tau = 0.0057$)	0.9697 (2.8450×10^{-3})	6.2	0.08827 (0.04787)	0.4988 (0.1229)
$T = 500$ ($\tau = 0.0044$)	0.9766 (2.392×10^{-3})	6.4	0.06200 (0.0308)	0.4188 (0.09151)
$T = 1000$ ($\tau = 0.0031$)	0.9841 (1.277×10^{-3})	6.6	0.03259 (0.01420)	0.3034 (0.04996)

Table 7.4: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 10$ and $\sigma = 1$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7.1 Brownian motion with compound Poisson jumps

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	1.207 (8.279×10^{-3})	24.1	12.37 (2.757)	2.9396 (0.3322)
$T = 200$ ($\tau = 0.007$)	1.288 (4.641×10^{-3})	31.3	7.136 (1.327)	2.1161 (0.1540)
$T = 300$ ($\tau = 0.0057$)	1.330 (2.904×10^{-3})	32.6	4.498 (0.6939)	1.7261 (0.2625)
$T = 500$ ($\tau = 0.0044$)	1.382 (2.662×10^{-3})	35.4	2.386 (0.4764)	1.2924 (0.09285)
$T = 1000$ ($\tau = 0.0031$)	1.449 (1.538×10^{-3})	35.1	0.7957 (0.1984)	0.8369 (0.08431)

Table 7.5: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D},\mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 10$ and $\sigma = 2$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	0.6099 (9.022×10^{-3})	7.6	367.5 (13.13)	32.96 (0.5716)
$T = 200$ ($\tau = 0.007$)	0.5967 (5.202×10^{-3})	8.3	232.4 (8.159)	26.06 (0.4690)
$T = 300$ ($\tau = 0.0057$)	0.5855 (3.650×10^{-3})	9.9	172.1 (6.079)	22.42 (0.3936)
$T = 500$ ($\tau = 0.0044$)	0.5716 (2.531×10^{-3})	12.6	116.1 (4.084)	18.46 (0.3136)
$T = 1000$ ($\tau = 0.0031$)	0.5538 (1.388×10^{-3})	15.9	64.92 (2.298)	13.86 (0.2460)

Table 7.6: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D},\mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 100$ and $\sigma = 0.5$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7 A simulation study

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	0.7901 (7.651×10^{-3})	7.6	329.7 (11.94)	30.41 (0.5545)
$T = 200$ ($\tau = 0.007$)	0.8394 (4.903×10^{-3})	8.1	201.7 (8.811)	23.73 (0.4805)
$T = 300$ ($\tau = 0.0057$)	0.8639 (3.738×10^{-3})	8.9	148.2 (5.721)	20.39 (0.3910)
$T = 500$ ($\tau = 0.0044$)	0.8905 (2.284×10^{-3})	11.5	98.91 (4.254)	16.68 (0.3350)
$T = 1000$ ($\tau = 0.0031$)	0.9190 (1.347×10^{-3})	15.1	56.19 (2.076)	12.60 (0.2241)

Table 7.7: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 100$ and $\sigma = 1$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	$\hat{\sigma}$	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}
$T = 100$ ($\tau = 0.01$)	0.9203 (8.278×10^{-3})	19.5	238.5 (10.52)	26.22 (0.5472)
$T = 200$ ($\tau = 0.007$)	1.046 (5.872×10^{-3})	29.3	143.2 (6.431)	19.97 (0.4643)
$T = 300$ ($\tau = 0.0057$)	1.114 (3.890×10^{-3})	34.5	107.0 (4.965)	17.12 (0.4014)
$T = 500$ ($\tau = 0.0044$)	1.199 (2.498×10^{-3})	38.5	73.82 (3.167)	14.16 (0.3052)
$T = 1000$ ($\tau = 0.0031$)	1.304 (1.483×10^{-3})	35.7	42.54 (2.221)	10.77 (0.2520)

Table 7.8: Empirical volatility estimates ($\hat{\sigma}$), model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE, w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE of the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 100$ and $\sigma = 2$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7.1 Brownian motion with compound Poisson jumps

We run another simulation study, to examine the effectiveness of penalisation at work. In particular, for the parameter combinations $\lambda = 10$ and $\sigma = 1$ and, additionally, $\lambda = 100$ and $\sigma = 1$, we simulate 100 trajectories each on the time horizons $[0, 100]$ and $[0, 1000]$. Then we compare the mean squared errors for the projection estimators based on every projection space in $\{\mathbb{S}_m : m = 2, \dots, 10\}$ and $\{\mathbb{S}_m : m = 9, \dots, 17\}$, as appropriate. Still, \mathbb{S}_m denotes the space of piecewise quadratic polynomials based on the partition of \mathbb{D} into m classes.

We illustrate the results in Tables 7.9 – 7.12. Let us state our findings for the low jump activity case, i. e. $\lambda = 10$, first. We observe that penalisation works quite well. In the case of $T = 100$, the empirical MSE is minimal for $m = 5$ (cf. Table 7.9), while the empirical mean of the model, chosen via penalisation, is $\widehat{m}_{\text{pen}} = 4.1$ (cf. Table 7.4). In the case of $T = 1000$, the empirical mean is minimal for $m = 8$ (cf. Table 7.10), while the empirical mean of the model, chosen via penalisation, is $\widehat{m}_{\text{pen}} = 6.6$ (cf. Table 7.4). However, the high jump activity case, i. e. $\lambda = 100$, shows deviant behaviour. For $T = 100$, penalisation works well, as the empirical MSE is minimal for $m = 7$ and $m = 8$ (cf. Table 7.11), while the empirical mean of the model, chosen via penalisation, is $\widehat{m}_{\text{pen}} = 7.6$ (cf. Table 7.7). Nevertheless, efficiency of penalisation decreases in the case of $T = 1000$. The empirical MSE is minimal for $m = 10$ (cf. Table 7.12), while the empirical mean of the model, chosen via penalisation, is $\widehat{m}_{\text{pen}} = 15.1$ (cf. Table 7.7). Apparently, we observe that the MSEs for the model sizes differ by less than 1% and there is a local minimum at $m = 16$. Since our optimisation algorithm searches for local minima only, it appears pleasing to address this efficiency issue to the algorithm rather than the penalisation method.

Finally, let us illustrate the projection estimators for the sample path shown in Figure 7.6 (*top-left*) in comparison to the μ -orthogonal projection. In Figure 7.7, we restrict ourselves to the part of the density on the positive half-line only. The case of two classes corresponds to the case of three or four classes for the two-sided estimators in Table 7.9. Analogously, three classes for the positive tail correspond to five and six classes for the two-sided case and four classes to seven and eight classes. An important observation is made for the discontinuities of the estimates. In particular, we see that increasing the number of classes reduces the impact of discontinuities for the μ -orthogonal projection (at least visually). However, there is no significant enhancement observable for the projection estimator.

7 A simulation study

m	\widehat{MSE}	SB	\widehat{MVar}
2	0.3712 (0.07977)	0.2685	0.1027 (0.07977)
3	0.3487 (0.08392)	0.1442	0.2045 (0.08392)
4	0.2431 (0.08406)	0.01988	0.2232 (0.08406)
5	0.1704 (0.09128)	0.01306	0.1573 (0.09128)
6	0.1852 (0.09163)	6.236×10^{-3}	0.1790 (0.09163)
7	0.2030 (0.09469)	5.683×10^{-3}	0.1974 (0.09469)
8	0.2248 (0.09801)	5.131×10^{-3}	0.2196 (0.09801)
9	0.2448 (0.1002)	2.809×10^{-3}	0.2420 (0.1002)
10	0.2663 (0.1068)	4.875×10^{-4}	0.2658 (0.1068)

Table 7.9: Empirical mean squared errors (MSE), squared μ -biases (SB) and empirical μ -variance (MVar) are presented for the projection estimators for the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 10$ and $\sigma = 1$. The projection spaces \mathbb{S}_m consist of all piecewise quadratic polynomials based on the regular partition of \mathbb{D} into m classes. The process is observed on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.01$ and on the time horizon $[0, 100]$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

m	\widehat{MSE}	SB	\widehat{MVar}
2	0.2828 (0.01147)	0.2694	0.01335 (0.01147)
3	0.1704 (0.01212)	0.1562	0.01412 (0.01212)
4	0.05750 (0.01230)	0.04306	0.01445 (0.01230)
5	0.04465 (0.01323)	0.02655	0.01810 (0.01323)
6	0.03155 (0.01333)	0.01004	0.02151 (0.01333)
7	0.03226 (0.01409)	8.849×10^{-3}	0.02341 (0.01409)
8	0.03141 (0.01444)	4.732×10^{-3}	0.02668 (0.01444)
9	0.03175 (0.01459)	1.180×10^{-3}	0.03057 (0.01459)
10	0.03366 (0.01571)	5.494×10^{-4}	0.03311 (0.01571)

Table 7.10: Empirical mean squared errors (MSE), squared μ -biases (SB) and empirical μ -variance (MVar) are presented for the projection estimators for the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 10$ and $\sigma = 1$. The projection spaces \mathbb{S}_m consist of all piecewise quadratic polynomials based on the regular partition of \mathbb{D} into m classes. The process is observed on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.0031$ and on the time horizon $[0, 1000]$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7.1 Brownian motion with compound Poisson jumps

m	\widehat{MSE}	SB	\widehat{MVar}
2	343.8 (10.52)	26.85	317.9 (10.52)
3	337.5 (10.85)	14.42	323.1 (10.85)
4	331.2 (11.28)	1.988	329.1 (11.28)
5	330.5 (11.26)	1.306	329.1 (11.26)
6	329.8 (11.26)	0.6236	329.2 (11.26)
7	329.7 (11.26)	0.5683	329.2 (11.26)
8	329.7 (11.28)	0.5131	329.2 (11.28)
9	329.9 (11.28)	0.2809	329.6 (11.28)
10	330.1 (11.24)	0.04875	330.0 (11.24)

Table 7.11: Empirical mean squared errors (MSE), squared μ -biases (SB) and empirical μ -variance (MVar) are presented for the projection estimators for the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 100$ and $\sigma = 1$. The projection spaces \mathbb{S}_m consist of all piecewise quadratic polynomials based on the regular partition of \mathbb{D} into m classes. The process is observed on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.01$ and on the time horizon $[0, 100]$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

m	\widehat{MSE}	SB	\widehat{MVar}
9	55.94 (2.066)	0.1180	55.82 (2.066)
10	55.90 (2.072)	0.05494	55.85 (2.073)
11	56.36 (2.214)	0.02755	56.33 (2.214)
12	56.36 (2.214)	0.01335	56.35 (2.214)
13	56.37 (2.218)	9.418×10^{-3}	56.36 (2.218)
14	56.67 (3.405)	5.482×10^{-3}	56.65 (3.405)
15	56.69 (3.401)	3.984×10^{-3}	56.68 (3.401)
16	56.45 (2.209)	2.487×10^{-3}	56.45 (2.209)
17	56.47 (2.218)	1.862×10^{-3}	56.46 (2.218)

Table 7.12: Empirical mean squared errors (MSE), squared μ -biases (SB) and empirical μ -variance (MVar) are presented for the projection estimators for the Lévy density of $X_t = \sigma W_t + \sum_{k=1}^{N_t} Z_k$, where $Z_k \sim \mathcal{N}(0, 1)$, N has intensity $\lambda = 100$ and $\sigma = 1$. The projection spaces \mathbb{S}_m consist of all piecewise quadratic polynomials based on the regular partition of \mathbb{D} into m classes. The process is observed on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.0031$ and on the time horizon $[0, 1000]$. The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

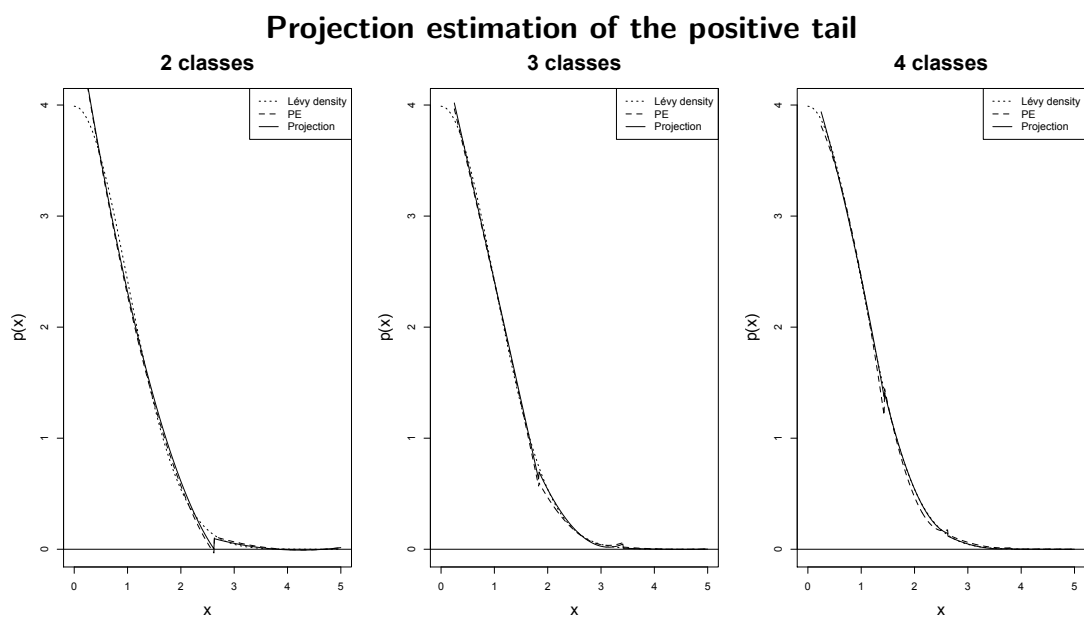


Figure 7.7: We present the μ -orthogonal projection $\mathcal{P}_m p$ (*solid*) and the projection estimator \tilde{p}_m based on the sample path shown top-left in Figure 7.6 (*dashed*) for the restriction of the Lévy density ν' to $\mathbb{D} = [0.249, 5]$, i. e. the density of $\nu(\cdot) = \lambda \mathcal{N}_{0,1}(\cdot)$ with $\lambda = 10$ (*dotted*). The μ -orthogonal projection and the projection estimator are based on the regular partition of $[0.249, 5]$ into $m = 2$ (*left*), $m = 3$ (*middle*) and $m = 4$ (*right*) classes.

7.2 Brownian motion with α -stable jumps

This section is dedicated to the case of the composition of a Brownian motion with α -stable jumps, where $\alpha \in]0, 2[$.

Univariate case

We introduced the univariate case in Section 5.3. Let X be a Lévy process with Lévy-Khintchine triplet (γ, σ^2, ν) satisfying Assumption 5.3.1. We recall that X can be decomposed as

$$X_t = \gamma t + \sigma W_t + S_t \quad \text{for } t \in \mathbb{R}_0^+,$$

where S is a stable process with parameters $(\alpha, \beta, \delta, \varsigma)$ independent of W .

For illustration purposes, we restrict ourselves to the case of a one-sided stable process, i. e. $\beta = 1$. Hence, we strive to estimate the Lévy density

$$\nu(dx) = \frac{C_1}{x^{1+\alpha}} \mathbb{1}_{\mathbb{R}^+}(x)dx + \frac{C_2}{|x|^{1+\alpha}} \mathbb{1}_{\mathbb{R}^-}(x)dx$$

for $C_1 = 1$ and $C_2 = 0$. Let us outline the simulation procedure for the discretised process S^τ on the time horizon $[0, T]$ based on the grid $\tau\mathbb{N}_0$ with mesh size τ . For a fixed set of parameters $(\alpha \neq 1, C_1, C_2)$, we set $(\beta, \delta, \varsigma)$ and $K(\alpha)$ in coherence with Section 5.3. Additionally, let $g_0 = -(\pi\beta K(\alpha))/(2\alpha)$. For each $k = 1, \dots, \lfloor T/\tau \rfloor$, we draw, independently, a random variable $w_k \sim \exp(1)$ and a random variable g_k uniformly distributed on $]-\pi/2, \pi/2[$. Then

$$S_{\Delta_k^\tau} := (\varsigma\tau)^{1/\alpha} \frac{\sin(\alpha(g_k - g_0))}{\cos(g_k)^{1/\alpha}} \left(\frac{\cos(g_k - \alpha(g_k - g_0))}{w_k} \right)^{\frac{1-\alpha}{\alpha}} + \varsigma\delta\tau$$

is $\mathcal{S}(\alpha, \beta, \delta, \varsigma\tau)$ -distributed. Finally, the discretised process is set to be equal to the cumulated sum of the simulated increments.

In our study, we simulate 100 trajectories for each parameter set (α, σ, T) . From Theorems 5.3.2 and 6.3.4, we deduce that the critical mesh is given by $\tau = o(1/T)$ as $T \rightarrow \infty$. We take $\tau = 1/T$ for the finitely many $T \in \{100, 250, 500, 1000\}$ and choose $\alpha \in \{0.75, 1.25\}$ and $\sigma \in \{0.5, 1\}$. We choose $\mathbb{D} = [0.1, 1]$ in the case of $\sigma = 0.5$ and $\mathbb{D} = [0.15, 1]$ in the case of $\sigma = 1$ as domain of estimation. Moreover, we choose the penalty pen to be of form 6.2 with $c_1 = 2$ and $c_2 = \dots = c_7 = 1$.

Estimation based on piecewise quadratic polynomials

In this study, we base our estimates on the family $\{\mathbb{S}_m : m \in M\}$ of piecewise quadratic polynomials on \mathbb{D} (see Example 3.1.2). We present the results in Tables 7.13–7.16. In all four cases, we observe the expected decrease in the empirical

7 A simulation study

mean squared error as a consequence of the decrease in the squared bias (as the projection space complexity increases) and the decrease in the μ -variance. Comparing the empirical mean squared errors from Table 7.13 and Table 7.15 to Table 7.14 to Table 7.16, respectively, we observe that increasing the volatility σ increases the μ -variance and the mean squared error of the PPE, as well. Analogously, we observe that the mean squared errors in the case of $\alpha = 0.75$ are lower by trend than the corresponding MSEs in the case of $\alpha = 1.25$. Naturally, disentangling the stable jumps from the Brownian motion works more efficiently for lower stability indices α .

We show a sample path for each set of parameters (α, σ) on the time horizon $[0, 1000]$ and the corresponding penalised projection estimator for the Lévy density in Figure 7.8. Especially in the cases of $\sigma = 1$, we observe an overestimation of the Lévy density for small $x \in \mathbb{D}$. Additionally, this effect is stronger in the case of $\alpha = 1.25$ compared to $\alpha = 0.75$. A similar explanation as for the MSEs is valid. In fact, the structural overestimation is measured by the MSE, yielding a larger value for the latter as described in the previous paragraph. Additionally, we see that the discontinuities observed on the common boundaries of partition classes are more severe closer to the origin. Again, the effect is stronger in the case of larger volatility σ . Naturally, the (mostly small) Brownian increments have a relative high influence on small (α -stable) increments but a relative small influence on big stable increments. Therefore, the influx of the Brownian motion on the PPE on an interval closer to the origin is higher.

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$ ($\tau = 0.01$)	3.8 (0.65)	4.860 (2.459)	1.161 (0.2003)	0.9348 (0.4859)	3.926 (2.642)
$T = 250$ ($\tau = 0.004$)	6.5 (0.88)	1.159 (0.6589)	0.5897 (0.1047)	0.1552 (0.1137)	1.003 (0.7030)
$T = 500$ ($\tau = 0.002$)	7.2 (0.50)	0.4639 (0.2558)	0.3839 (0.06426)	0.08939 (0.02890)	0.3745 (0.2612)
$T = 1000$ ($\tau = 0.001$)	8.1 (1.8)	0.2636 (0.1258)	0.2924 (0.05201)	0.0677 (0.03261)	0.1959 (0.1419)

Table 7.13: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , *w. r. t.* $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.1, 1]$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 0.75$ and $\sigma = 0.5$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7.2 Brownian motion with α -stable jumps

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$	3.9	159.5	4.341	0.0712	159.4
$(\tau = 0.01)$	(0.64)	(22.62)	(0.2480)	(0.0452)	(22.64)
$T = 250$	7.0	60.62	2.111	4.481×10^{-3}	60.62
$(\tau = 0.004)$	(0.14)	(7.525)	(0.1284)	(2.979×10^{-4})	(7.525)
$T = 500$	7.0	2.789	0.6066	4.606×10^{-3}	2.784
$(\tau = 0.002)$	(0.32)	(0.7602)	(0.06652)	(1.629×10^{-3})	(0.7604)
$T = 1000$	6.6	0.2904	0.2768	0.01177	0.2786
$(\tau = 0.001)$	(1.2)	(0.1480)	(0.05082)	(0.02165)	(0.1530)

Table 7.14: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , *w. r. t.* $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.15, 1]$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 0.75$ and $\sigma = 1$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$	4.2	54.72	2.31	12.41	42.31
$(\tau = 0.01)$	(0.71)	(14.63)	(0.2873)	(6.996)	(11.79)
$T = 250$	7.3	3.09	0.8440	1.6968	1.397
$(\tau = 0.004)$	(0.5478)	(0.8063)	(0.1113)	(0.3679)	(0.8052)
$T = 500$	10.6	1.729	0.6473	0.4581	1.271
$(\tau = 0.002)$	(1.566)	(0.7508)	(0.08831)	(0.3923)	(0.7540)
$T = 1000$	14.6	0.9593	0.5000	0.0963	0.8630
$(\tau = 0.001)$	(1.386)	(0.4690)	(0.07067)	(0.09008)	(0.4806)

Table 7.15: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , *w. r. t.* $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.1, 1]$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 1.25$ and $\sigma = 0.5$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7 A simulation study

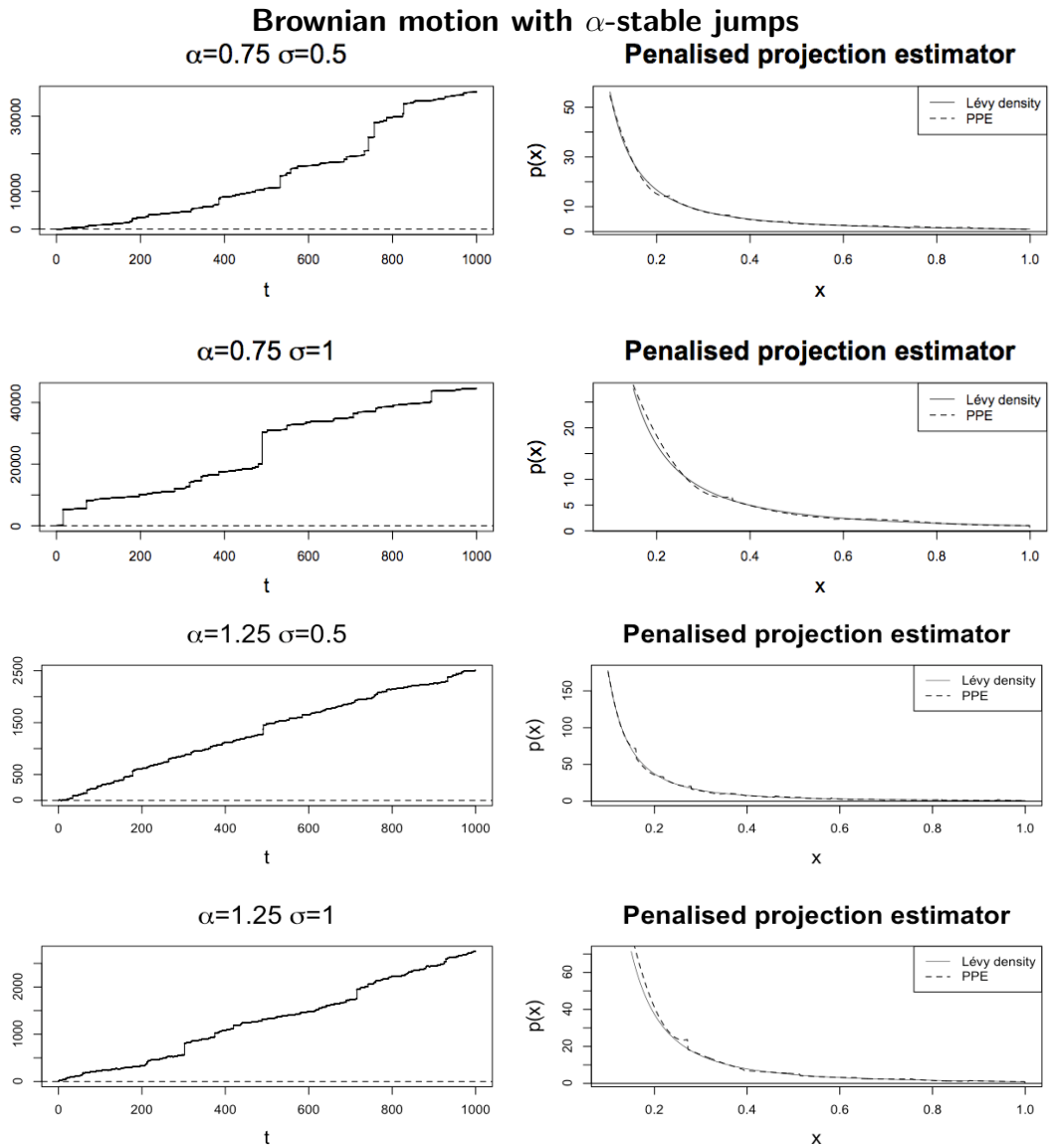


Figure 7.8: We present the sample paths $X(\omega)$ of the composition of a Brownian motion with volatility σ and α -stable jumps (*left column*). The observation is made on the time horizon $[0, 1000]$ and on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.001$. Moreover, we present the corresponding Lévy density on the interval $[0.1, 1]$ (*solid*) and the corresponding penalised projection estimator \tilde{p}_{pen} based on the family of piecewise quadratic polynomials on $\mathbb{D} = [0.1, 1]$ (for $\sigma = 0.5$) and $\mathbb{D} = [0.15, 1]$ (for $\sigma = 1$) (*dashed*). The estimators are based on the partition of \mathbb{D} into 7, 4, 15 and 7 classes, respectively (*right column*).

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$ ($\tau = 0.01$)	3.2 (0.43)	74.00 (14.93)	3.584 (0.3055)	1.527 (0.4364)	72.47 (15.04)
$T = 250$ ($\tau = 0.004$)	7.0 (0.17)	96.41 (10.73)	3.062 (0.1627)	0.06652 (5.278×10^{-3})	96.35 (10.73)
$T = 500$ ($\tau = 0.002$)	7.6 (0.97)	18.54 (3.258)	1.383 (0.1095)	0.05262 (0.01998)	18.49 (3.262)
$T = 1000$ ($\tau = 0.001$)	8.0 (1.3)	2.734 (0.8784)	0.6150 (0.0812)	0.04586 (0.02312)	2.688 (0.8830)

Table 7.16: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.15, 1]$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 1.25$ and $\sigma = 1$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

Estimation based on a non-Lebesgue reference measure

In this study, we choose a non-Lebesgue reference measure μ . In particular, we let μ be defined by $\mu(dx) = |x|^{-2\alpha-2}dx$ in coherence with Example 2.3.4. Moreover, we base our estimate on the family $\{\mathbb{S}'_m : m \in M\}$ of piecewise constant functions on \mathbb{D} , where the partition \mathcal{D}_m of \mathbb{D} is chosen such that $\mu(D) = \mu(\mathbb{D})/m$ for all $D \in \mathcal{D}_m$. From an estimate \tilde{p}_{pen} for the μ -density p , we use the relationship $\nu' = p\mu'$ to come up with an estimate $\tilde{\nu}' := \tilde{p}_{\text{pen}}\mu'$ for the Lévy density.

The empirical mean squared errors, absolute errors and the corresponding decomposition of the MSEs into squared μ -bias and μ -variance for the same parameter sets (α, σ, T) as in the previous simulation study are shown in Tables 7.17 – 7.20. At first glance, the most obvious observation, is that the mean squared errors in the case of $\sigma = 1$ (Tables 7.18 and 7.20) are lower than the corresponding MSEs in the case of $\sigma = 0.5$ (Tables 7.17 and 7.19). However, we have to be careful with the interpretation. The decomposition into squared μ -bias and μ -variance reveals that the latter increases as σ increases, while the decrease is totally due to the decrease in the squared μ -bias. This decrease is due to the different domains of estimations \mathbb{D} we work with in the cases of $\sigma = 1$ and $\sigma = 0.5$. The penalised projection estimators $\tilde{\nu}'$ for the Lévy densities (based on the observation of the sample paths shown in Figure 7.8) are shown in Figure 7.9. The high variability of the estimates close to the lower boundary of \mathbb{D} is due to the use of piecewise constants and the change of measure. The latter induces the partition to be refined closer to the origin. We still note the general overestimation, especially observed in the case of $\alpha = 1.25$ and $\sigma = 1$. The explanation given above remains valid.

7 A simulation study

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$ ($\tau = 0.01$)	4.6 (1.5)	0.4008 (0.04664)	2.504 (0.2523)	0.3959 (0.04693)	4.910×10^{-3} (2.719×10^{-3})
$T = 250$ ($\tau = 0.004$)	13.1 (2.9)	0.2438 (0.03429)	1.5338 (0.1372)	0.2408 (0.03477)	2.981×10^{-3} (1.550×10^{-3})
$T = 500$ ($\tau = 0.002$)	22.4 (4.7)	0.1714 (0.02591)	1.150 (0.08530)	0.1691 (0.02645)	2.271×10^{-3} (1.126×10^{-3})
$T = 1000$ ($\tau = 0.001$)	37.0 (5.4)	0.1125 (0.01412)	0.8683 (0.05339)	0.1106 (0.01449)	1.917×10^{-3} (8.904×10^{-4})

Table 7.17: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.1, 1]$ w. r. t. the reference measure $\mu(dx) = |x|^{-2\alpha-2}dx$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 0.75$ and $\sigma = 0.5$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$ ($\tau = 0.01$)	1.8 (1.0)	0.7529 (0.07264)	5.190 (0.2670)	0.3970 (0.06823)	0.3559 (0.04752)
$T = 250$ ($\tau = 0.004$)	9.9 (2.4)	0.2606 (0.03154)	2.498 (0.1691)	0.1458 (0.02742)	0.1149 (0.01512)
$T = 500$ ($\tau = 0.002$)	16.0 (2.7)	0.1020 (0.01537)	0.9175 (0.08277)	0.09413 (0.01593)	7.913×10^{-3} (2.089×10^{-3})
$T = 1000$ ($\tau = 0.001$)	27.4 (5.5)	0.05677 (0.01447)	0.5548 (0.05216)	0.05369 (0.01494)	3.086×10^{-3} (1.197×10^{-4})

Table 7.18: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.15, 1]$ w. r. t. the reference measure $\mu(dx) = |x|^{-2\alpha-2}dx$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 0.75$ and $\sigma = 1$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7.2 Brownian motion with α -stable jumps

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$	3.2	0.6040	6.503	0.6000	3.992×10^{-3}
$(\tau = 0.01)$	(0.5)	(0.01676)	(0.2449)	(0.01634)	(1.614×10^{-3})
$T = 250$	10.1	0.4794	3.971	0.4788	6.282×10^{-4}
$(\tau = 0.004)$	(1.9)	(0.02125)	(0.2413)	(0.02142)	(5.060×10^{-4})
$T = 500$	21.9	0.3864	2.948	0.3858	5.912×10^{-4}
$(\tau = 0.002)$	(3.6)	(0.02067)	(0.1647)	(0.02077)	(3.372×10^{-4})
$T = 1000$	47.1	0.2914	2.176	0.2906	7.216×10^{-4}
$(\tau = 0.001)$	(6.9)	(0.01741)	(0.1017)	(0.01757)	(2.974×10^{-4})

Table 7.19: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.1, 1]$ w. r. t. the reference measure $\mu(dx) = |x|^{-2\alpha-2}dx$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 1.25$ and $\sigma = 0.5$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

	\widehat{m}_{pen}	\widehat{MSE}	\widehat{MAE}	\widehat{SB}	\widehat{MVar}
$T = 100$	5.2	0.4564	5.245	0.3902	0.0663
$(\tau = 0.01)$	(1.4)	(0.02873)	(0.3098)	(0.03199)	(0.01265)
$T = 250$	11.5	0.3313	4.231	0.2915	0.0398
$(\tau = 0.004)$	(2.6)	(0.02777)	(0.2179)	(0.02889)	(4.594×10^{-3})
$T = 500$	21.9	0.2221	2.250	0.2134	8.646×10^{-3}
$(\tau = 0.002)$	(4.9)	(0.02447)	(0.1658)	(0.02523)	(1.788×10^{-3})
$T = 1000$	37.2	0.1561	1.314	0.1536	2.511×10^{-3}
$(\tau = 0.001)$	(5.6)	(0.01448)	(0.07206)	(0.01481)	(7.359×10^{-4})

Table 7.20: Model complexities (\widehat{m}_{pen}), mean squared errors (MSE) and mean absolute errors (MAE , w. r. t. $\|\cdot\|_{L^1(\mathbb{D}, \mu)}$) are presented for the PPE (on the domain $\mathbb{D} = [0.15, 1]$ w. r. t. the reference measure $\mu(dx) = |x|^{-2\alpha-2}dx$) of the Lévy density of $X = \sigma W + S$, where $S \sim \mathcal{S}(\alpha, 1, \delta, \varsigma; B)$ with $\alpha = 1.25$ and $\sigma = 1$. The estimates are based on the observation of the discretisation X^τ on the time horizon $[0, T]$. Additionally, the MSE is decomposed into the squared μ -bias (SB) and the μ -variance ($MVar$). The values in brackets show the standard deviation of the estimates based on 100 trajectories each.

7 A simulation study

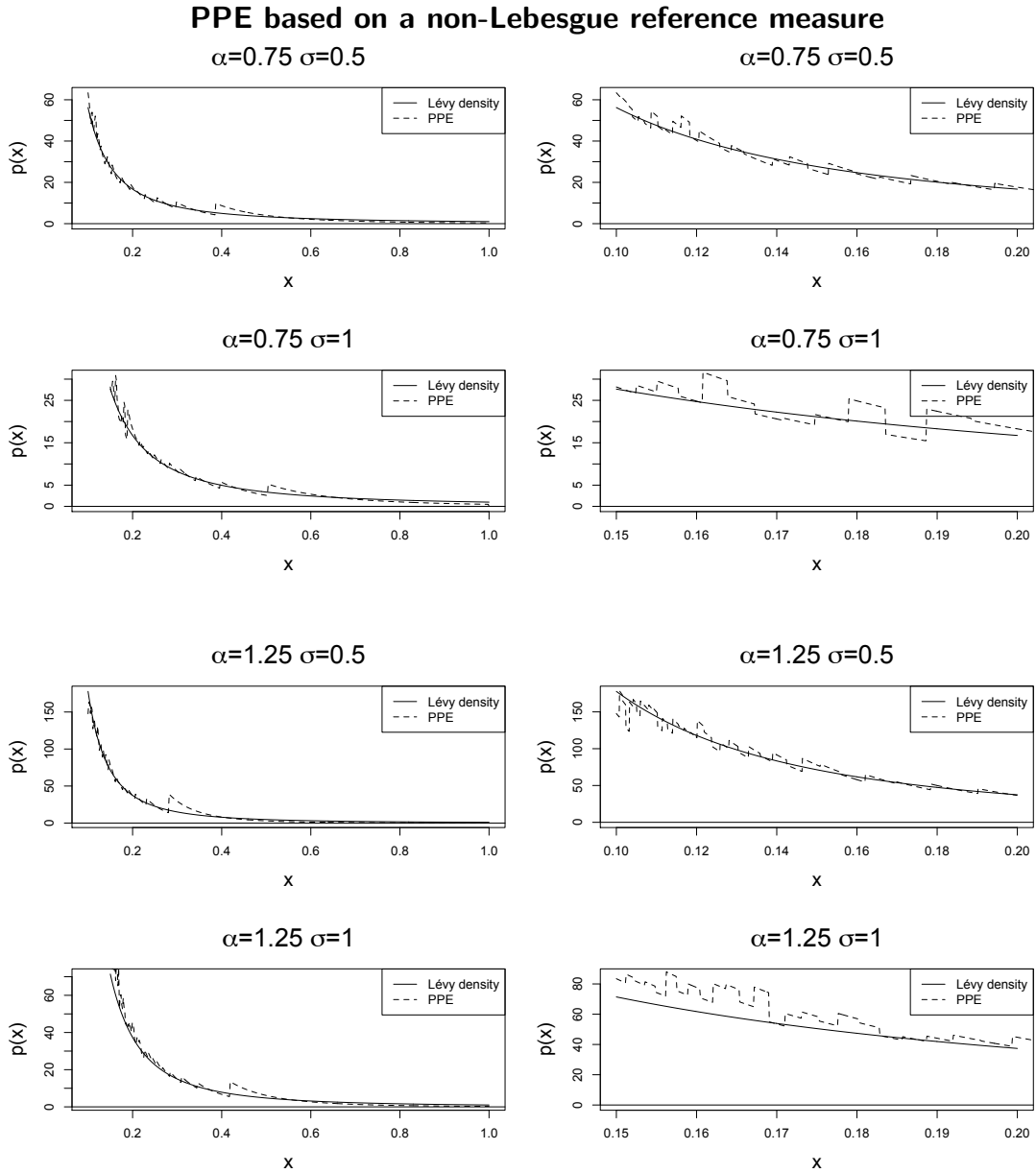


Figure 7.9: We present the penalised projection estimator \tilde{p}_{pen} w.r.t. the reference measure μ defined by $\mu(dx) = |x|^{-2\alpha-2}dx$ and based on the family of piecewise constants on $\mathbb{D} = [0.1, 1]$ and $\mathbb{D} = [0.15, 1]$, respectively (*dashed*), for the Lévy density on the interval $[0.1, 1]$ (*solid*) of the composition of a Brownian motion with volatility σ and α -stable jumps (*left column*). The estimates are based on the sample paths shown in Figure 7.8 on the time horizon $[0, 1000]$ and on the grid $\tau\mathbb{N}_0$ with mesh size $\tau = 0.001$. Moreover, we present a magnified plot of the PPE (*dashed*) and the Lévy density (*solid*) on the subinterval $[0.1, 0.2]$ and $[0.15, 0.2]$, respectively (*right column*).

Bivariate case

Last but not least, we investigate at work the behaviour of the penalised projection estimator in the case of a bivariate α -stable process. We will use the simulation algorithm introduced in Tankov [39, Theorem 4.3], which is based on the notion of *Lévy copulas* as introduced in Kallsen and Tankov [24].

A prominent family of Lévy copulas is the two-parametric family of *Clayton Lévy copulas*. In particular, for $\theta > 0$ and $\eta \in [0, 1]$, Kallsen and Tankov [24, Chapter 5] shows that (in the bivariate case $d = 2$)

$$F(x_1, x_2) := (|x_1|^{-\theta} + |x_2|^{-\theta})^{-1/\theta} (\eta \mathbb{1}_{\{x_1 x_2 \geq 0\}} + (1 - \eta) \mathbb{1}_{\{x_1 x_2 < 0\}}) \quad (7.2)$$

defines a Lévy copula F . Since F is a homogenous function of order 1, i. e.

$$F(cx_1, cx_2) = cF(x_1, x_2)$$

for all $c > 0$ and $(x_1, x_2) \in \mathbb{R}^2$, we deduce from Kallsen and Tankov [24, Theorem 4.8] that a Lévy process $X = (X_1, X_2)$ with α -stable margins X_1 and X_2 having Lévy copula F (in the sense of Kallsen and Tankov [24, Theorem 3.6]) is a bivariate α -stable Lévy process. For the simulation procedure, we refer to Tankov [39, Example 4.1] for further details. As usual in these contexts, we approximate the increments $X_{\Delta\bar{k}}$ on the grid $\tau\mathbb{N}_0$ with mesh size $\tau > 0$ (theoretically composed of infinitely many jumps) by a deterministic finite number of jumps. In particular, we simulate 2000 jumps per unit time interval.

A closed formula for the tail integrals $\nu(\mathcal{T}(x_1, x_2))$ of the Lévy measure is provided in Eder and Klüppelberg [14, Example 4.3]. We take the second partial derivative w. r. t. x_1 and x_2 , to come up with a closed form of the Lévy density.

Lemma 7.2.1 (Bivariate Clayton α -stable Lévy density) *Let X be a Lévy process with α -stable margins X_1 and X_2 such that its marginal Lévy densities satisfy Assumption 5.3.1 with parameters $c_1^+, c_1^- \geq 0$ and $c_2^+, c_2^- \geq 0$, respectively. Further, let X have Lévy copula F of form (7.2) for $\theta > 0$ and $\eta \in [0, 1]$. Then for every $(x_1, x_2) \in (\mathbb{R}_\circ)^2$, we have*

$$p(x_1, x_2) = \alpha^{1+\theta} (1 + \theta) \left(c_1^{\text{sgn}(x_1)} c_2^{\text{sgn}(x_2)} \right)^{-\theta} |x_1 x_2|^{\alpha\theta-1} \times \left. \begin{aligned} & \times \left((c_1^{\text{sgn}(x_1)})^{-\theta} |x_1|^{\alpha\theta} + (c_2^{\text{sgn}(x_1)})^{-\theta} |x_2|^{\alpha\theta} \right)^{-\frac{1}{\theta}-2} \times \\ & \times (\eta \mathbb{1}_{\{x_1 x_2 \geq 0\}} + (1 - \eta) \mathbb{1}_{\{x_1 x_2 < 0\}}), \end{aligned} \right\} \quad (7.3)$$

where the constants are interpreted figuratively as $c_i^{\text{sgn}(x_i)} := c_i^+ \mathbb{1}_{\{x_i \geq 0\}} + c_i^- \mathbb{1}_{\{x_i < 0\}}$.

For illustration purposes, we restrict ourselves to the case of pure α -stable processes without Brownian motion or drift and present a weak and a strong dependent case.

7 A simulation study

Let $\alpha = 0.75$, $\theta = 0.5$ and $\eta = 0.5$ and $c_1^+ = c_1^- = c_2^+ = c_2^- = 1$. These parameters correspond to very weak dependence and a symmetric Lévy process. We aim for estimating the Lévy density on the domain

$$\mathbb{D} = \{(x_1, x_2) = (r \cos(\phi), r \sin(\phi)) : r \in [0.1, 1], \phi \in]-\pi, \pi[)\}.$$

We present a perspective, a contour and a grey-scale image plot of the Lévy density in Figure 7.11. We observe the symmetry and concentration of the density close to the coordinate axes. Nevertheless, we remark that the axes themselves are not charged by the Clayton Lévy measure. Furthermore, we base our estimates on the family $\{\mathbb{S}_m : m \in M\}$ of piecewise constant functions based on a partition of \mathbb{D} , where we separately part the radial and the angular coordinate into m classes. In the light of Theorem 6.3.3, we interpret the *annulus* \mathbb{D} (from the Latin word for “little ring”) and the elements of its partitions as polyhedrons via the polar coordinate transform. As the Lévy measure does not charge the coordinate axes, we do not have any further issues with \mathbb{D} . Finally, we choose the penalty to be of form (6.2) with $c_1 = 2$ and $c_2, \dots, c_7 = 1$. In our example, we take $T = 1000$ and $\tau = 0.001$. A sample path is shown in Figure 7.10.

From this sample, the penalised projection space is estimated to $\hat{m}_T^{\text{pen}} = 16$. The corresponding PPE for the Lévy measure restricted to \mathbb{D} is illustrated next to the Lévy density in Figure 7.11. Let us outline our observations step by step. First, we observe a strong overestimation close to the inner boundary of the domain \mathbb{D} in the perspective plot. The highest tick mark on the z -axis of the plot of the true Lévy density is at the value 60, whereas the highest tick mark of the plot of the PPE is at the value 300. Secondly, the contour plots reveal a strong structural difference of the Lévy density and the PPE. However, we have to look carefully. Let us note that the shown structure of the PPE is fully imposed by the family $\{\mathbb{S}_m : m \in M\}$ of projection spaces we use. Moreover, the level marks in the contour plot reveals more of the true nature of the PPE. The grey-scale plot for the sub-domain $\{x \in \mathbb{D} : \|x\| \leq 0.5\}$ illustrates this point more obvious. As a matter of fact, the grey-scale image of the PPE does show a similar structure as the grey-scale image of the Lévy density. Certainly, the PPE is coarser. Nevertheless, the symmetry and the concentration of the density close to the coordinate axes is clearly disclosed.

To come up with a more comprehensive study similar to the univariate cases, bivariate numerical integration of the squared difference of the PPE and the Lévy density is necessary. Standard algorithms do not provide a result in reasonable computation time. Nevertheless, we use the *Monte-Carlo integration method* to come up with an estimate of the squared error. In particular, we choose a reasonable high number (2 000 000 in our case) of grid points, i. i. d. and uniform in \mathbb{D} . Then we use the law of large numbers for the sample mean of the function values scaled by the Lebesgue measure of \mathbb{D} . In our case, we get $\|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \approx 7 \times 10^4$.

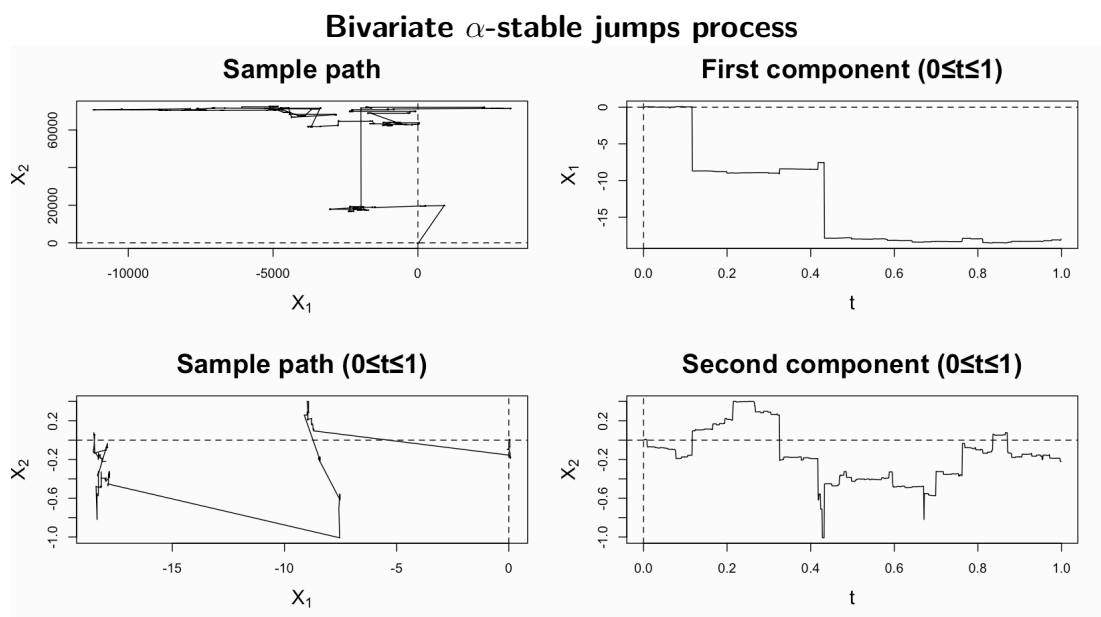


Figure 7.10: We present the sample path $X(\omega)$ of a bivariate α -stable Lévy process with Lévy density of form (7.3) with parameters $\alpha = 0.75$, $c_1^+ = c_1^- = c_2^+ = c_2^- = 1$, $\theta = 0.5$ and $\eta = 0.5$ on the time horizon $[0, 1000]$ (*top-left*) and a magnified version restricted to the time horizon $[0, 1]$ (*bottom-left*). Corresponding to the path shown bottom-left, we present the components $X_1(\omega)$ and $X_2(\omega)$ separately on the time horizon $[0, 1]$ (*right column*). The increments of the path are simulated on the regular grid with mesh size $\tau = 0.001$.

7 A simulation study

Bivariate Clayton Lévy density

Penalised projection estimator

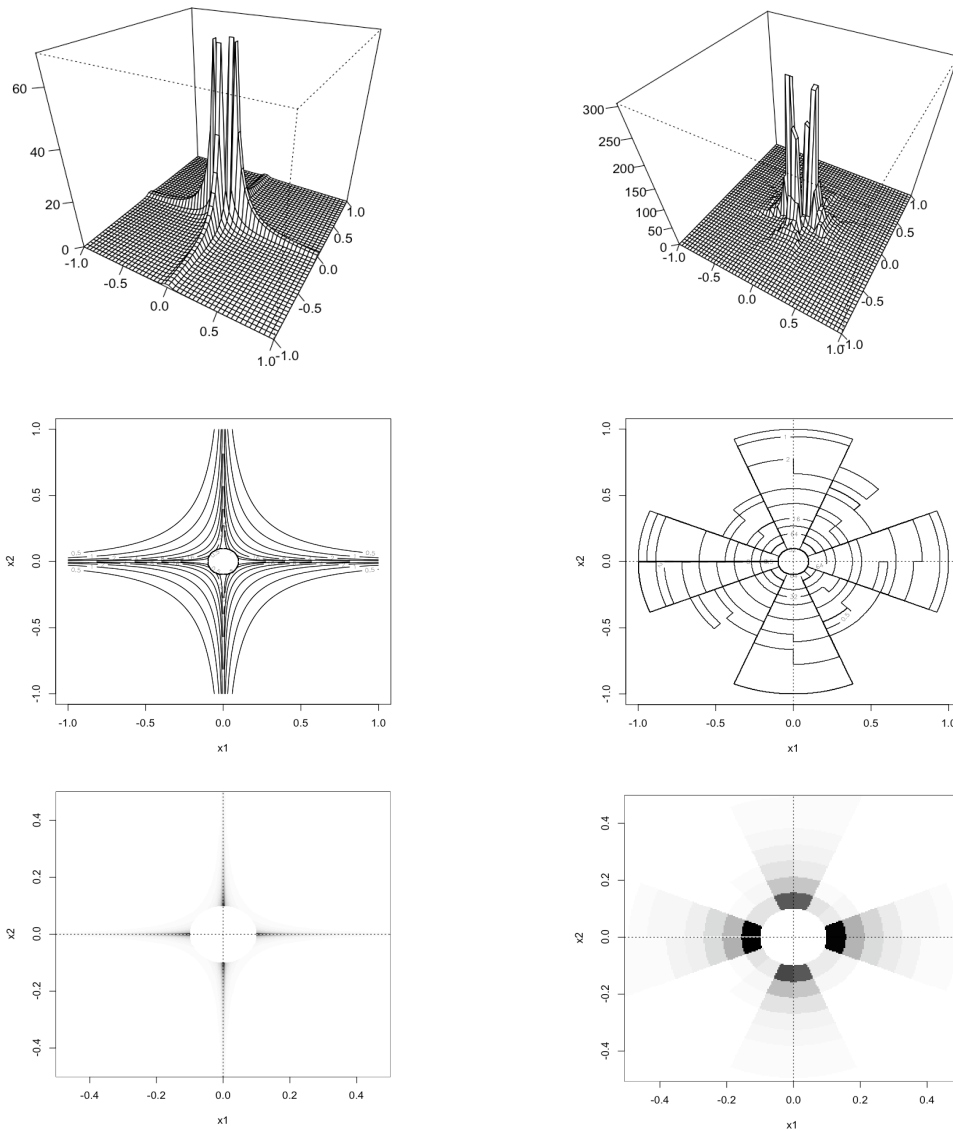


Figure 7.11: We present a perspective, a contour and a grey-scale image plot for the bivariate Clayton Lévy density of form (7.3) with parameters $\alpha = 0.75$, $c_1^+ = c_1^- = c_2^+ = c_2^- = 1$, $\theta = 0.5$ and $\eta = 0.5$ (*left column*) and the penalised projection estimator \tilde{p}_{16} based on the observation of the sample path shown in Figure 7.10 (*right column*).

7.2 Brownian motion with α -stable jumps

Let $\alpha = 0.75$, $\theta = 3$, $\eta = 0.75$, $c_1^+ = c_2^- = 2$ and $c_1^- = c_2^+ = 1$. These parameters correspond to strong dependence and an asymmetric Lévy process. We aim for estimating the Lévy density on the domain

$$\mathbb{D} = \{(x_1, x_2) = (r \cos(\phi), r \sin(\phi)) : r \in [0.1, 1], \phi \in]-\pi, \pi[)\}.$$

We present a perspective, a contour and a grey-scale image plot of the restriction of the Lévy density to \mathbb{D} in Figure 7.13. We observe the asymmetry and concentration of the density close to the negative x_2 and the positive x_1 -axis in the first and third quadrant. Again, we remark that the axes themselves are not charged by the Clayton Lévy measure. As before, we base our estimates on the family $\{\mathbb{S}_m : m \in M\}$ of piecewise constant functions based on a partition of \mathbb{D} , where we separately part the radial and the angular coordinate into m classes. We choose the penalty pen to be of form (6.2) with $c_1 = 2$ and $c_2, \dots, c_7 = 1$. In our example, we take $T = 1000$ and $\tau = 0.001$. A sample path is shown in Figure 7.12.

From this sample, the penalised projection space is estimated to $\hat{m}_T^{\text{pen}} = 12$. The corresponding PPE for the Lévy measure restricted to \mathbb{D} is illustrated next to the Lévy density in Figure 7.13. Let us outline our observations step by step. First, in the perspective plot we observe weaker overestimation close to the inner boundary of the domain \mathbb{D} in comparison to the previous example. The highest tick mark on the z -axis of the plot of the true Lévy density is at the value 300, whereas the highest tick mark on the z -axis of the plot of the PPE is at the value 400. Secondly, the contour plots reveal weaker structural differences of the Lévy density and the PPE as well. Again, we have to look carefully. The shown structure of the PPE is fully imposed by the family $\{\mathbb{S}_m : m \in M\}$ of projection spaces, as it was in the previous example. Nevertheless, the connection to the true Lévy density is more obvious, especially when taking the level marks in the contour plot into account. The grey-scale image plot for the sub-domain $\{x \in \mathbb{D} : \|x\| \leq 0.5\}$ illustrates this point more obvious. As a matter of fact, the grey-scale image of the PPE is certainly coarser once more. Nevertheless, the strong asymmetry is satisfyingly disclosed.

As in the previous example, standard adaptive algorithms for the bivariate numerical integration of the squared difference of PPE and Lévy density do not provide a reliable result in reasonable computation time. Nevertheless, using the *Monte-Carlo integration method* to come up with an estimate, again with 2 000 000 grid points, we get $\|p - \tilde{p}_{\text{pen}}\|_{\mu}^2 \approx 3 \times 10^3$. Comparing this to the approximate error of $\approx 7 \times 10^4$ in the weak dependence case before, we find our visual observations supported.

7 A simulation study

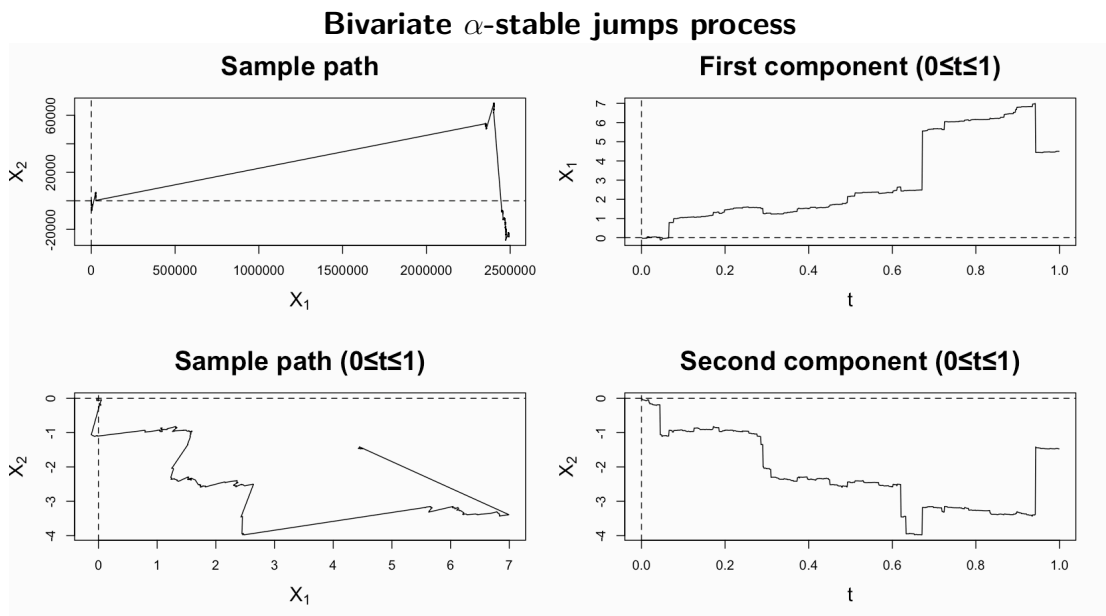


Figure 7.12: We present the sample path $X(\omega)$ of a bivariate α -stable Lévy process with Lévy density of form (7.3) with parameters $\alpha = 0.75$, $c_1^+ = c_2^- = 2$, $c_1^- = c_2^+ = 1$, $\theta = 3$ and $\eta = 0.75$ on the time horizon $[0, 1000]$ (*top-left*) and a magnified version restricted to the time horizon $[0, 1]$ (*bottom-left*). Corresponding to the path shown bottom-left, we present the components $X_1(\omega)$ and $X_2(\omega)$ separately on the time horizon $[0, 1]$ (*right column*). The increments of the path are simulated on the regular grid with mesh size $\tau = 0.001$.

Bivariate Clayton Lévy density

Penalised projection estimator

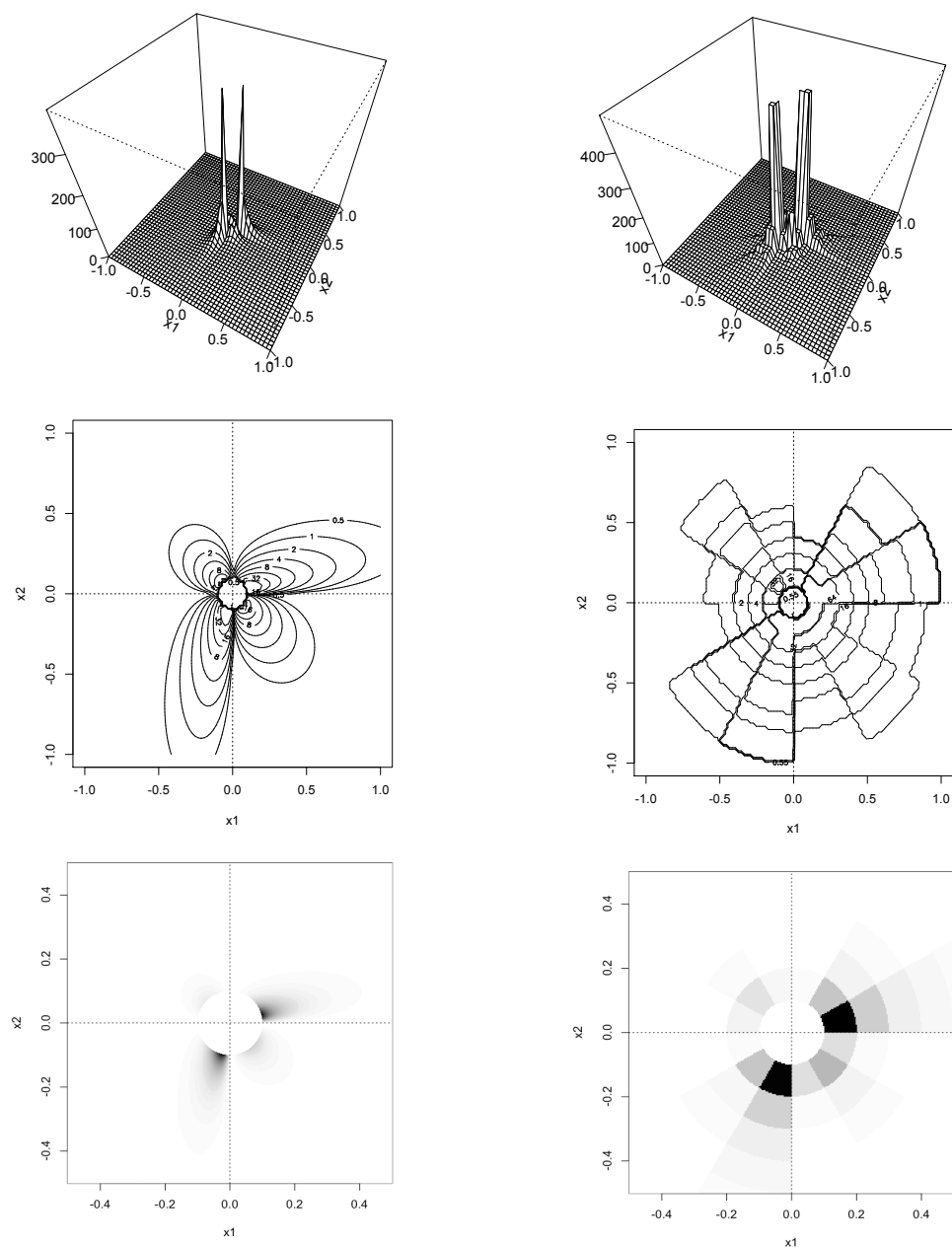


Figure 7.13: We present a perspective, a contour and a grey-scale image plot for the bivariate Clayton Lévy density of form (7.3) with parameters $\alpha = 0.75$, $c_1^+ = c_2^- = 2$, $c_1^- = c_2^+ = 1$, $\theta = 3$ and $\eta = 0.75$ (left column) and the penalised projection estimator \tilde{p}_{12} based on the observation of the sample path shown in Figure 7.12 (right column).

7 *A simulation study*

List of Notation

The page numbers show the first occurrence of a notation.

Number sets

\mathbb{N}	The set of positive integers, i. e. $\{1, 2, \dots\}$	5
\mathbb{N}_0	The set of non-negative integers, i. e. $\mathbb{N} \cup \{0\}$	5
\mathbb{R}_0^+	The set of non-negative real numbers, i. e. $[0, \infty[$	5
\mathbb{R}^+	The set of strictly positive real numbers, i. e. $]0, \infty[$	5
\mathbb{R}^-	The set of strictly negative real numbers, i. e. $] - \infty, 0[$	5
\mathbb{R}_\circ^d	The set $\mathbb{R}^d \setminus \{0\}$	7

Limits

$\lim_{s \nearrow t}$	Left-hand limit or limit from below, i. e. $\lim_{s \rightarrow t; s < t}$	5
$\lim_{s \searrow t}$	Right-hand limit or limit from above, i. e. $\lim_{s \rightarrow t; s > t}$	5

Random variables and stochastic processes

$\Delta_\tau f$	Discretization bias, i. e. $\tau^{-1} \mathbb{E}[f(X_\tau)] - \nu(f)$	29
Δ_τ^*	Discrete transition deviation, i. e. $\sup_{y \in \mathbb{D}} \Delta_\tau y $	33
ΔX_t	The jump size of X at time t , i. e. $X_t - X_{t-}$	6
F_X	(Cumulative) distribution function of X	5
\bar{F}_X	Tail distribution function of X	5
\hat{F}_X	Characteristic function of X	5
F'_X	Probability density of X	5
P_X	Law of a random variable X under P	5
X_{t-}	The point-wise left-sided limit of X , i. e. $\forall \omega \in \Omega : \lim_{s \nearrow t} X_s(\omega)$	6
$X_{\Delta_k^\tau}$	The k -th observed increment (on $\tau\mathbb{N}_0$), i. e. $X_{k\tau} - X_{(k-1)\tau}$	27
X^τ	The discretization of X w. r. t. $\tau > 0$, i. e. $X_t^\tau = X_{\tau \lfloor t/\tau \rfloor}$	27

σ -fields and measures

$\mathcal{B}(\mathbb{X})$	The Borel σ -field on \mathbb{X}	5
\mathcal{B}^d	The Borel σ -field on \mathbb{R}^d , i. e. $\mathcal{B}(\mathbb{R}^d)$	5
\mathcal{B}_\circ^d	The Borel σ -field on \mathbb{R}_\circ^d , i. e. $\mathcal{B}(\mathbb{R}_\circ^d)$	7
$\lambda_{\mathbb{X}}$	The Lebesgue measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$	5

Function spaces and norms

$\mathcal{B}_\infty^{r,q}(\mathbb{D}, \mu)$	Besov space over $L^q(\mathbb{D}, \mu)$ with degree of smoothness r	22
\mathcal{C}^n	The set of all \mathbb{R} -valued, n -times cont. differentiable functions.	39
$L^q(\mathbb{D}, \mu)$	Lebesgue space of all q -th power integrable functions	9
$\mathcal{W}^{r,q}(\mathbb{D}, \mu)$	Sobolev pace over $L^q(\mathbb{D}, \mu)$ with degree of smoothness r	35
$\ \cdot\ _\infty$	The maximum (or essential supremum) norm in \mathbb{R}^d (or $L^\infty(\mathbb{D})$)	5
$\ \cdot\ _\mu$	The norm on $L^2(\mathbb{D}, \mu)$ defined by $\ f\ _\mu^2 := \int_{\mathbb{D}} f(x) ^2 \mu(dx)$	9
$\langle \cdot, \cdot \rangle_\mu$	The scalar product on $L^2(\mathbb{D}, \mu)$, i. e. $\langle f, g \rangle_\mu = \int_{\mathbb{D}} f(x)g(x)\mu(dx)$	11

Estimators

\hat{p}_m	Projection estimator (in CT) w. r. t. the projection space \mathbb{S}_m	16
\tilde{p}_m	Projection estimator (in DT) w. r. t. the projection space \mathbb{S}_m	61
\hat{p}_\star	Oracle (in CT) w. r. t. the collection $\{\hat{p}_m : m \in M_T\}$	17
\tilde{p}_\star	Oracle (in DT) w. r. t. the collection $\{\tilde{p}_m : m \in M_T\}$	32
\hat{p}_{pen}	Penalized projection estimator (in CT) w. r. t. the penalty pen.	19
\tilde{p}_{pen}	Penalized projection estimator (in DT) w. r. t. the penalty pen.	60
$\hat{\nu}(f)$	Integral estimator (in CT)	12
$\tilde{\nu}(f)$	Integral estimator (in DT)	28

List of Abbreviations

The page numbers show the first occurrence of an abbreviation.

a. e.	almost everywhere	30
a. s.	almost surely	6
cf.	confer	1
CT	continuous time framework	2
DT	discrete time framework	2
e. g.	exempli gratia	2
et al.	et alii	18
i. e.	id est	5
i. i. d.	independently and indentially distributed	8
MSE	mean squared error	16
μ -ONB	μ -orthonormal basis	12
p.	page	29
PPE	penalised projecton estimator / estimate	19
PRM	Poisson random measure	7
w. l. o. g.	without loss of generality	24
w. r. t.	with respect to	6

Bibliography

- [1] Bargel, M. and Wenzel, J. (2009) Ist die Mathematik Schuld an der Finanzkrise? Press release of Deutsche Mathematiker-Vereinigung. Available at https://www.dmv.mathematik.de/component/docman/doc_download/68-pm-010409-ist-die-mathematik-schuld-an-der-finanzkrise.html
- [2] Barndorff-Nielsen, O. E. (1998) Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2:41–68
- [3] Barron, A. R., Birgé, L. and Massart, P. (1995) Model selection via penalization. Tech. Rep. 54, Université Paris-Sud
- [4] Bertoin, J. (1996) *Lévy Processes*, Cambridge Tracts in Mathematics, vol. 121. Cambridge University Press, Cambridge
- [5] Birgé, L. and Massart, P. (1994) Minimum contrast estimation on sieves. Tech. Rep. 34, Université Paris-Sud
- [6] Blumenthal, R. M. and Gettoor, R. K. (1961) Sample functions of stochastic processes with stationary independent increments. *Journal of Mathematics and Mechanics*, 10:493–516
- [7] Brenner, S. C. and Scott, L. R. (1994) *The Mathematical Theory of Finite Element Methods*, Texts in Applied Mathematics, vol. 15. Springer, New York
- [8] Cariboni, J. and Schoutens, W. (2009) *Lévy Processes in Credit Risk*. Wiley and Sons, Chichester
- [9] Carr, P., Geman, H., Madan, D. and Yor, M. (2003) Stochastic volatility for Lévy processes. *Mathematical Finance*, 13:345–382
- [10] Carr, P., Madan, D. and Chang, E. (1998) The variance gamma process and option pricing. *European Finance Review*, 2:79–105
- [11] Ciarlet, P. G. (1978) *The Finite Element Method for Elliptic Problems*, Studies in Mathematics and its Applications, vol. 4. North-Holland, Amsterdam, New York, Oxford

- [12] DeVore, R. A. and Lorentz, G. G. (1993) *Constructive Approximation*. Springer Verlag, Berlin
- [13] Eberlein, E. and Keller, U. (1995) Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299
- [14] Eder, I. and Klüppelberg, C. (2009) Pareto Lévy measures and multivariate regular variation. Available at <http://www-m4.ma.tum.de/Papers/Eder/LCMRVSubmitted.pdf>
- [15] Figueroa-López, J. E. (2004) *Non-parametric Estimation of Lévy Processes with a View Towards Mathematical Finance*. Ph.D. thesis, Georgia Institute of Technology. No. etd-04072004-122020
- [16] Figueroa-López, J. E. (2008) Small-time moment asymptotics for Lévy processes. *Statistics and Probability Letters*, 78:3355–3365. Doi:10.1016/j.spl.2008.07.012
- [17] Figueroa-López, J. E. (2009) Non-parametric estimation for Lévy models based on discrete sampling. In Rojo, J. (ed.), *Optimality: The Third Erich L. Lehmann Symposium*, IMS Lecture Notes – Monograph Series, vol. 57, 117–146. Institute of Mathematical Statistics, Beachwood, OH. Doi:10.1214/09-LNMS5709
- [18] Figueroa-López, J. E. and Houdré, C. (2004) Non-parametric estimation of Lévy processes with a view towards mathematical finance. Available at arXiv:0412351v1 [math.ST]
- [19] Figueroa-López, J. E. and Houdré, C. (2006) Risk bounds for the non-parametric estimation of Lévy processes. In Gine, E., Kolchinskii, V., Li, W. and Zinn, J. (eds.), *High Dimensional Probability: Proceedings of the Fourth International Conference*, IMS Lecture Notes – Monograph Series, vol. 51, 96–116. Institute of Mathematical Statistics, Beachwood, OH. Available at arXiv:0612697 [math.ST]
- [20] Figueroa-López, J. E. and Houdré, C. (2008) Small-time expansions for the transition distributions of Lévy processes. Available at arXiv:0809.0849v2 [math.PR]
- [21] Föllmer, H. (2009) Alles richtig und trotzdem falsch? Anmerkungen zur Finanzkrise und zur Finanzmathematik
- [22] Ibragimov, I. A. and Hasminskii, R. Z. (1981) *Statistical Estimation: Asymptotic Theory*, Applications of mathematics, vol. 16. Springer Verlag, New York

- [23] Jacod, J. (2007) Asymptotic properties of power variations of Lévy processes. *European Series in Applied and Industrial Mathematics: Probability and Statistics*, 11:173–196. Doi:10.1051/ps:2007013
- [24] Kallsen, J. and Tankov, P. (2006) Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572. ISSN 0047-259X. Doi:10.1016/j.jmva.2005.11.001
- [25] Klüppelberg, C., Kuhn, G. and Peng, L. (2008) Semi-parametric models for the multivariate tail dependence function – the asymptotically dependent case. *Scandinavian Journal of Statistics*, 35(4):701–718
- [26] Kutoyants, Y. A. (1998) *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics, vol. 134. Springer Verlag, New York
- [27] Li, D. X. (2000) On default correlation: A copula function approach. *Journal of Fixed Income*. Available at <http://ssrn.com/abstract=189908>
- [28] Lévy, P. P. (1937) *Théorie de l'Addition des Variables Aléatoires*. No. 1 in Monographies des Probabilités, publiés sous la direction de E. Borel. Gauthier-Villars, Paris
- [29] 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> with corresponding figures available at <http://www.cmap.polytechnique.fr/SEM/finance/2004-2005/mancini2.pdf>
- [30] Mandelbrot, B. B. (1963) The variation of certain speculative prices. *The Journal of Business of the University of Chicago*, 36:394–419
- [31] Mazja, W. (1979) *Einbettungssätze für Sobolewsche Räume (Teil 1)*. Teubner-Texte zur Mathematik. Teubner Verlagsgesellschaft, Leipzig
- [32] Millar, P. W. (1971) Path behavior of processes with stationary independent increments. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 17(1):53–73. Doi:10.1007/BF00538475
- [33] Reynaud-Bouret, P. (2002) Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. Tech. Rep. 1102-001, School of mathematics, Georgia Institute of Technology
- [34] Rüschenendorf, L. and Woerner, J. H. C. (2002) Expansion of transition distributions of Lévy processes in small time. *Bernoulli*, 8(1):81–96

- [35] Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge Studies in Advanced Mathematics, vol. 68. Cambridge University Press, Cambridge
- [36] Sato, K. (2001) Basic results on Lévy processes. In Barndorff-Nielsen, O. E., Mikosch, T. and Resnick, S. I. (eds.), *Lévy Processes Theory and Applications*, 3–37. Birkhauser, Boston
- [37] Schoutens, W. (2003) *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley series in Probability and Statistics. Wiley and Sons, Chichester
- [38] Szpiro, G. (2009) Eine falsch angewendete Formel und ihre Folgen. *Neue Zürcher Zeitung*. Issue from March 18, 2009
- [39] Tankov, P. (2006) Simulation and option pricing in Lévy copula models. Available at http://www.math.jussieu.fr/~tankov/levycopulas_ima.pdf
- [40] Zolotarev, V. M. (1986) *One-dimensional Stable Distributions*, Translations of Mathematical Monographs, vol. 65. American Mathematical Society, Providence (RI)