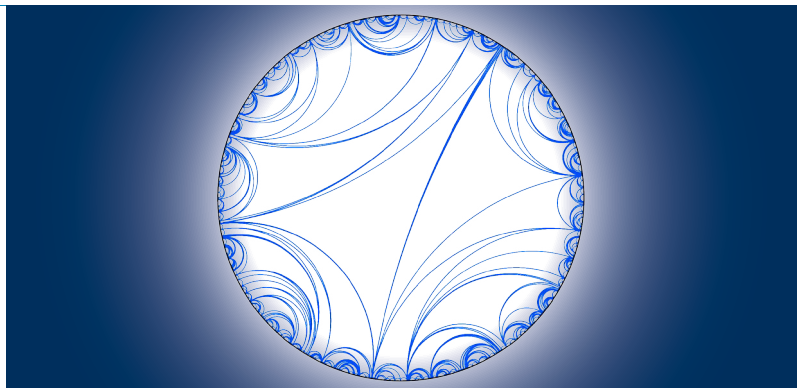
Universität Potsdam



Jan Gairing | Michael Högele | Tetiana Kosenkova | Alexei Kulik

On the calibration of Lévy driven time series with coupling distances

An application in paleoclimate

Preprints des Instituts für Mathematik der Universität Potsdam
3 (2014) 2

Preprints des Instituts für Mathematik der Universität Potsdam

Preprints des Instituts für Mathematik der Universität Potsdam
3 (2014) 2

Jan Gairing | Michael Högele | Tetiana Kosenkova | Alexei Kulik

On the calibration of Lévy driven time series with coupling distances

An application in paleoclimate

Universitätsverlag Potsdam

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.de abrufbar.

Universitätsverlag Potsdam 2014
http://verlag.ub.uni-potsdam.de/

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: 2292
E-Mail: verlag@uni-potsdam.de

Die Schriftenreihe **Preprints des Instituts für Mathematik der Universität Potsdam** wird herausgegeben vom Institut für Mathematik der Universität Potsdam.

ISSN (online) 2193-6943

Kontakt:
Institut für Mathematik
Am Neuen Palais 10
14469 Potsdam
Tel.: +49 (0)331 977 1028
WWW: http://www.math.uni-potsdam.de

Titelabbildungen:
1. Karla Fritze | Institutsgebäude auf dem Campus Neues Palais
2. Nicolas Curien, Wendelin Werner | Random hyperbolic triangulation
Published at: http://arxiv.org/abs/1105.5089

Online veröffentlicht auf dem Publikationsserver der Universität Potsdam
URL http://pub.ub.uni-potsdam.de/volltexte/2014/6978/
URN urn:nbn:de:kobv:517-opus-69781
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-69781

On the calibration of Lévy driven time series with coupling distances – an application in paleoclimate

Jan Gairing [*] Michael Högele[†] Tetiana Kosenkova [‡] Alexei Kulik [§]

February 27, 2014

Abstract

This article aims at the statistical assessment of time series with large fluctuations in short time, which are assumed to stem from a continuous process perturbed by a Lévy process exhibiting a heavy tail behavior. We propose an easily implementable procedure to estimate efficiently the statistical difference between the noisy behavior of the data and a given reference jump measure in terms of so-called *coupling distances*, which were introduced in [10]. After a short introduction to Lévy processes and coupling distances we recall basic statistical approximation results and derive rates of convergence. In the sequel the procedure is elaborated in detail in an abstract setting and eventually applied in a case study to simulated and paleoclimate data. It indicates the dominant presence of a non-stable heavy-tailed jump Lévy component for some tail index $\alpha > 2$.

MSC 2010: 60G51; 60G52; 60J75; 62M10; 62P12;
Keywords: time series with heavy tails; index of stability; goodness-of-fit; empirical Wasserstein distance; limit theorem for the empirical quantile process;

Contents

[*]Institut für Mathematik, Humboldt-Universität zu Berlin, Germany; gairing@math.hu-berlin.de
[†]Institut für Mathematik, Universität Potsdam, Germany; hoegele@math.uni-potsdam.de
[‡]Kyiv National Taras Shevchenko University, Ukraine; tanya.kosenkova@gmail.com
[§]Institute of Mathematics, National Academy of Science, Kiev, Ukraine; kulik@imath.kiev.ua

1 Introduction

In many contexts time series of interest show too large fluctuations in short periods of time in order to justify the assumption of an underlying continuous model. Due to the continuity of Gaussian models it is necessary to go beyond the Gaussian paradigm to model random perturbations and to include the effect of shocks. The natural class including discontinuous perturbations is given by Lévy processes, that is non-Gaussian extensions of Brownian motion, which keep the white noise structure of stationary, δ-correlated increments. Often stochastic modeling consists in the study of deterministic models, which represent preknowledge about the underlying phenomenon, perturbed by (Lévy) noise.

In this contribution we want to follow the ideas of [10] in order to calibrate the jump behavior in our model by means of *coupling distances* based on empirical data. Coupling distances contain a suitably renormalized Wasserstein distance between jump measures of Lévy processes, which metrizes the weak convergence of distributions. In their work they consider a class of Lévy driven dynamical systems and derive a quantitative upper bound for the proximity of their distributions on sample path space. An essential factor in this estimate is the coupling distance between Lévy measures. By construction these distances explore the discrepancy of Lévy measures along decreasing jump sizes. This article wishes to provide the theoretical background as well as an instructive road map to implement the concept of coupling distances in a statistical setting.

The estimation of the tail behavior is also of interest on its own. Many important dynamical features, stability properties or scaling invariance are determined by the mass distribution in the tails. See for instance [17, 18, 21, 13, 16]. In a mathematical paradigm of climate science rapid transitions between stadials and interstadials of the last glaciation period can be described by the impact of unpredictable shocks, that can be interpreted as large discontinuities of a random driving force. In the seminal article [7] Ditlevsen identified an α-stable Lévy component in a climate proxy signal with the help of the statistical analysis of large jumps. Further investigation in that direction has been carried out in [12] and [14] exploiting the selfsimilarity of these processes.

We will concentrate on the calibration of the tails of compound Poisson processes, a crucial subclass of Lévy processes with jumps. We refer to Section 2 for a nutshell and to [1] and [23] for the interested reader. This section also contains the definition of the coupling distance. In Section 3 we provide the statistical tools to assess the Wasserstein distance in the coupling distance. We prove the rate of convergence to a non-trivial limiting random variable. Section 4 details the conceptual and algorithmic approach to the calibration problem. Conceptually we describe the modeling class interpreting large increments as governed by large jumps of the driving noise. This allows to define the statistical quantity of interest, the empirical version of a coupling distance. We present an algorithm how to evaluate this statistic on a data sample. In Section 5 we exemplify the previous strategy for the calibration of polynomial tails. Simulated sample paths confirm the strength of method. Eventually we reassess the climate proxies studied by Ditlevsen et. al with our calibration procedure. Comparing the coupling distance for $\alpha = 1.75$ with values $\alpha > 2$ justifies the assumption of an non-stable jump Lévy component.

2 Lévy processes and coupling distances

Lévy processes: The most prominent representative of a Lévy process is arguably Brownian motion. A standard Brownian motion $B = (B_t)_{t\geqslant 0}$ is defined as a stochastic process starting in $B_0 = 0$, with independent increments satisfying the stationarity condition $B_t - B_s = B_{t-s} \sim N(0, (t-s))$ for $t \geqslant s$. It has almost surely continuous trajectories $t \mapsto B_t$ and moments of all orders.

The concept of a Lévy process drops the assumption of the Gaussianity of increments. A Lévy

process is given as a stochastic process with independent and stationary increments $L_t - L_s \sim L_{t-s}$. Such a process is not necessarily continuous. However, to ensure the separability of the process (trajectories should be determined by any countable dense set of points in time) one imposes stochastic continuity, i.e.

$$\lim_{t\to s} \mathbb{P}(|L_t - L_s| > \eta) = 0 \quad \text{for all } \eta > 0.$$

This property ensures that almost all trajectories are at least right-continuous and with finite left limits (commonly denoted by the French acronym càdlàg). The so-called Lévy-Itô decomposition tells us that a given Lévy process $L = (L_t)_{t\geqslant 0}$ in $\mathbb{R}^d$ can be decomposed almost surely into the sum of four independent components, three of which are stochastic processes. That is there exits a vector $a \in \mathbb{R}^d$, a d-dimensional standard Brownian motion $B = (B_t)_{t\geqslant 0}$, a positive semi-definite (covariance) matrix $A \in \mathbb{R}^{d\otimes d}$ and a measure on the Borel sets $\nu : \mathfrak{B}(\mathbb{R}^d) \to [0,\infty]$ such that for all $\rho > 0$

$$\nu(\{0\}) = 0, \qquad \int_{\mathcal{B}_\rho(0)} |x|^2 \nu(dx) < \infty \qquad \text{and} \qquad \nu(\mathbb{R}^d \setminus \mathcal{B}_\rho(0)) < \infty, \tag{2.1}$$

where $\mathcal{B}_\rho(0) = \{y \in \mathbb{R}^d \mid |y| < \rho\}$. For each fixed $\rho > 0$ the measure ν has two associated processes. First, there is a compound Poisson process $C^\rho = (C^\rho_t)_{t\geqslant 0}$ with intensity $\lambda_\rho = \nu(\mathbb{R}^d \setminus \mathcal{B}_\rho(0))$ and jump distribution ν_ρ

$$\mathfrak{B}(\mathbb{R}^d) \ni E \mapsto \nu_\rho(E) := \frac{\nu(E \cap (\mathbb{R}^d \setminus \mathcal{B}_\rho(0))}{\lambda_\rho}. \tag{2.2}$$

That is a pure jump process with exponentially distributed waiting times between consecutive jumps of intensity λ_ρ, which are independent and distributed according to ν_ρ in (2.2). Note that the jumps of C^ρ are bounded from below by ρ. Second, there is another pure jump process $J^\rho = (J^\rho_t)_{t\geqslant 0}$, whose jumps are bounded from above by ρ to be discussed below. The Lévy process L is decomposed path-wise as

$$L_t = at + A^{1/2}B_t + C^\rho_t + J^\rho_t \qquad \forall t \geqslant 0 \quad \mathbb{P}-\text{a.s.} \tag{2.3}$$

In particular, the marginal laws of L are given by the so-called Lévy-Chinchine representation of the characteristic function

$$\mathbb{E}[\exp(i\langle u, L_t\rangle)] = \exp(t\Psi(u)),$$

with

$$\Psi(u) = iat - \frac{1}{2}\langle a, Aa\rangle + \int_{\mathbb{R}^d\setminus\mathcal{B}_\rho(0)} \big[e^{i\langle u,y\rangle} - 1\big]\nu(dy) + \int_{\mathcal{B}_\rho(0)} \big[e^{i\langle u,y\rangle} - 1 - i\langle u, y\rangle\big]\nu(dy), \qquad u \in \mathbb{R}^d.$$

This representation tells us that $J^\rho = (J^\rho_t)_{t\geqslant 0}$ can be understood as the superposition of independent (recentered) compound Poisson processes, with jumps that take values in rings given by $R_j = \{\rho_j < |y| \leqslant \rho_{j-1}\}$, for a strictly decreasing sequence $\rho = \rho_0 > \rho_1 > \cdots > 0$ of radii $\rho_j \searrow 0$, such that

$$\mathcal{B}_\rho(0) \setminus \{0\} = \bigcup_{j\in\mathbb{N}_0} R_j$$

with joint (possibly infinite) intensity

$$\sum_{j\in\mathbb{N}_0} \nu(R_j) = \nu(\mathcal{B}_\rho(0)) \leqslant \infty$$

and individual jump distribution $E \mapsto \nu_{\rho_{j+1}}(E) - \nu_{\rho_j}(E)$. If $\nu(\mathcal{B}_\rho(0)) < \infty$ we can choose formally $\rho = 0$ and hence $J^\rho \equiv 0$.

Remark 2.1. This construction relies heavily on the fact that the integrals in the exponent Ψ are additive over disjoint supports. Hence the characteristic function decomposes into a product of characteristic functions and thus independent components. As a consequence for every repartition $V_1, \dots, V_n$ of $\mathbb{R}^d$ ($\bigcup_i V_i = \mathbb{R}^d$ and $V_i \bigcap V_j = \emptyset$ unless $i = j$) the Lévy process decomposes into the sum of n independent Lévy processes L^i with jumps taking values in V_i.

Coupling distances between Lévy measures: In the article [10] the authors construct a distance on the set of Lévy measures in $\mathbb{R}^d$, that exploits the approximation of the discontinuous part of L by compound Poisson processes with decreasing lower bound on the jump size.

The idea is that given two jump (probability) distributions μ and μ' on $\mathbb{R}^d$ we want to measure the distance $\mathbb{E}[|X-Y|^2]$ between two random variables $X \sim \mu$ and $Y \sim \mu'$ on a common probability space. Obviously, there is more than one distribution of the random vector (X, Y), that guarantees the marginals $X \sim \mu$ and $Y \sim \mu'$. Any probability measure Π on the product space $\mathbb{R}^d \times \mathbb{R}^d$ with these marginals, is called a coupling of μ and μ'. Hence the set of all couplings is given by

$$\begin{aligned}\mathfrak{C}(\mu, \mu') := \Big\{ &\Pi : \mathfrak{B}(\mathbb{R}^d) \otimes \mathfrak{B}(\mathbb{R}^d) \to [0,1] \text{ probability measure, with} \\ &\Pi(E \times \mathbb{R}^d) = \mu(E), \quad \Pi(\mathbb{R}^d \times E) = \mu'(E) \quad \text{for all } E \in \mathfrak{B}(\mathbb{R}^d) \Big\}. \end{aligned} \tag{2.4}$$

Since the expectation operator is a functional entirely determined by $\Pi \in \mathfrak{C}(\mu, \mu')$, the proximity of μ and μ' can be quantified minimizing the function $\Pi \mapsto \mathbb{E}_\Pi[|X - Y|^2]$ over the set of all couplings in $\mathfrak{C}(\mu, \mu')$. This motivates the following abstract definition.

Definition 2.1. *The Wasserstein metric of order* 2 *between two probability measures μ, μ' on the Borel sets of $\mathbb{R}^d$ is defined by*

$$W_2(\mu, \mu') := \inf_{\substack{\Pi \in \mathfrak{C}(\mu, \mu') \\ \Pi \sim (X,Y)}} \mathbb{E}[|X - Y|^2]^{\frac{1}{2}}.$$

Any minimizer is referred to as optimal coupling between μ and μ'.

It is well-known in the mathematical literature [22] that the convergence $W_2(\mu_n, \mu) \to 0$ is equivalent to the weak convergence $\mu_n \rightharpoonup \mu$ and the convergence of the second moments. In this work we will consider the following example.

Example 2.2. Consider $(\mathbb{R}^d, |\cdot|_1)$ with $|x|_1 = \min(|x|, 1)$ and the Euclidean norm $|\cdot|$. This will be the space where we approximate the laws of the jumps. We introduce the cutoff norm, since we do not require the jumps to have second moments.

Remark 2.2. On $(\mathbb{R}, |\cdot|)$ one can show that for two distribution functions $F(x) = \mu((-\infty, x])$ and $F'(x) = \mu'((-\infty, x])$ the optimal coupling is realized by the random vector

$$(X, Y) = (F^{-1}(U), (F')^{-1}(U)),$$

where U has the uniform distribution in $[0, 1]$. Therefore the Wasserstein metric is easily evaluated by

$$W_2^2(\mu, \mu') = \int_0^1 |(F^{-1}(u) - (F')^{-1}(u)|^2 du.$$

The optimality of the pair (X, Y) relies on the specific metric on $\mathbb{R}$. The law of (X, Y) is obviously a coupling of μ and μ', and the right-hand side provides at least an upper bound for the Wasserstein distance.

Definition 2.3. *For two absolutely continuous Lévy measures $\nu = f dx$ and $\nu' = f' dx$ on $\mathbb{R}^d$ and $0 < \lambda < \min(\nu(\mathbb{R}^d), \nu'(\mathbb{R}^d))$ let*

$$\rho(\lambda) := \inf\{r > 0 \mid \nu(\mathbb{R}^d \setminus \mathcal{B}_r(0)) \geqslant \lambda\}$$

and $\rho'(\lambda)$ analogously. We introduce a family of semimetrics T_λ

$$T_\lambda(\nu, \nu') := \lambda^{\frac{1}{2}} \, W_2(\nu_{\rho(\lambda)}, \nu'_{\rho'(\lambda)}).$$

First note that given ν and $\lambda > 0$ we have the following equality $\lambda_{\rho(\lambda)} = \lambda$, with λ_ρ defined before (2.2). Intuitively, given Lévy measures μ and μ', the semimetric $T_\lambda(\nu, \nu')$ compares the jump distributions $\nu_{\rho(\lambda)}$ and $\nu'_{\rho'(\lambda)}$ of two compound Poisson processes with common rate $\lambda = \lambda_{\rho(\lambda)} = \lambda_{\rho'(\lambda)}$.

Note further that the bivariate function T_λ is symmetric and satisfies the triangle inequality hence it is a semimetric. Clearly $T_\lambda(\nu, \nu) = 0$ does not guarantee that ν the zero measures, since $\nu|_{B_{\rho(\lambda)}}$ is not taken into account. Therefore it is not a proper metric. In order to overcome this shortcoming and provided that $\nu(\mathbb{R}^d) = \nu'(\mathbb{R}^d) = \infty$ we introduce the following.

Definition 2.4. *For two absolutely continuous Lévy measures $\nu = f dx$ and $\nu' = f' dx$ on $\mathbb{R}^d$ with $\nu(\mathbb{R}^d) = \nu'(\mathbb{R}^d) = \infty$ we define*

$$T(\nu, \nu') := \sup_{\lambda > 0} T_\lambda(\nu, \nu').$$

Remark 2.5. Both restrictions in the definitions above can be removed. For details we refer to the original work [10].

- The restriction on absolute continuity is overcome by an interpolation procedure.
- The requirement of infinite mass can be dropped by the ad hoc introduction of an artificial point mass in 0 carrying the missing weight. In this way also finite Lévy measures fall into the setup.

We also refer to [20] for further applications of coupling distances.

3 Statistical considerations

In this section we provide the technical background to compare the jump statistics of a data set to a given reference distribution in terms of the coupling distance. For this purpose we collect the necessary statistical theory. Since our case study is essentially one dimensional we stick to the scalar case. For higher dimensions we refer to Remark 3.1 at the end of this section.

Basic notions: Let us consider a sequence of independent and identically distributed random variables $(X_i)_{i \in \mathbb{N}}$ with common law μ on the real line. Denote by μ_n the *empirical distribution* based on the sample of size n given by

$$\mu_n(E) := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(E) = \frac{\#\{X_i \in E\}}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \in E\} \ , \quad E \in \mathfrak{B}(\mathbb{R}) \ , \tag{3.1}$$

where $\delta_a(\cdot) = \delta_0(\,\cdot - a)$ is the Dirac measure at a. The corresponding *empirical distribution function* F_n is the distribution function of μ_n

$$F_n(x) := \mu_n((-\infty, x]) = \frac{\#\{X_i \leqslant x\}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leqslant x\}\ , \quad x \in \mathbb{R}\ . \tag{3.2}$$

The strong law of large numbers (*Glivenko-Cantelli Theorem*) tells us that if F is the distribution function of the common distribution μ of X, we have for almost all $\omega \in \Omega$

$$\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| \to 0 \quad (n \to \infty)\ . \tag{3.3}$$

A proper scaling of this quantity leads to a non-trivial limit. Indeed if we fix $x \in \mathbb{R}$, the random variables $\mathbf{1}\{X_i \leqslant x\}$, $i \in \mathbb{N}$ are i.i.d. Bernoulli variables with $F(x) = \mathbb{P}(X_i \leqslant x)$. Now the central limit theorem (*de Moivre-Laplace Theorem*) states that

$$\sqrt{n}\,(F_n(x) - F(x)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}\{X_i \leqslant x\} - F(x) \xrightarrow{d} \mathcal{N}\,(0, F(x)(1-F(x)))\,. \tag{3.4}$$

This quantity can be viewed as a stochastic process indexed by $x \in \mathbb{R}$ which leads to the following definition.

Definition 3.1. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables in $\mathbb{R}$ with common distribution function F and let F_n, $n \in \mathbb{N}$ be its empirical distribution function. We define the associated* empirical (error) process *by*

$$G_n(x,\omega) := \sqrt{n}\,(F_n(x,\omega) - F(x))\,, \quad 0 \leqslant x \leqslant 1\ . \tag{3.5}$$

There is a huge literature on empirical processes, for an overview and more details we refer for instance to [2] or [24]. It is easily seen that G_n is a random element of the space $\mathbb{D}[0,1]$, the space of càdlàg functions $\varphi : [0,1] \to \mathbb{R}$, and the one dimensional (marginal) distributions of G_n are determined by (3.5). Moreover it is well known that the random variable $F(X)$ is uniformly distributed on the interval $[0,1]$ whenever X has the continuous strictly increasing distribution function F, since in this case F is one to one and by an easy change of variables

$$\mathbb{P}(F(X) \in [a,b]) = \int_{F^{-1}(a)}^{F^{-1}(b)} f(x)dx = b - a \qquad \text{for all } 0 \leqslant a < b \leqslant 1.$$

However F is usually not known and to be estimated. In any case we have $F(X) \in [0,1]$ whether F corresponds to the actual distribution of X or not. Hence distributions on the interval $[0,1]$ are of particular interest. G_n is then a random element of $\mathbb{D}[0,1]$ and its behavior is well understood.

Theorem 3.2 ([3]Theorem 14.3, p.149)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables on $[0,1]$ with common distribution function F and $(G_n)_{n\in\mathbb{N}}$ be the associated empirical processes. Then there exists a Gaussian random element G with values in $\mathbb{D}[0,1]$ such that*

$$\sup_{x\in\mathbb{R}} |G_n(x,\omega) - G(x,\omega)| \xrightarrow{d} 0 \quad (n \to \infty).$$

Moreover G is the unique Gaussian process determined by $\mathbb{E}[G] = 0$ and covariance

$$\mathbb{E}[G_s \cdot G_t] = F(s)(1 - F(t)) \quad \text{for } 0 \leqslant s \leqslant t.$$

In particular if F is the uniform distribution on $[0,1]$, then $G = B^0$ is a Brownian bridge*, that is a Brownian motion conditioned to end in 0 at time $x = 1$.*

Quantiles: In view of the Wasserstein distance it is also interesting to look at the *empirical quantile function* F_n^{-1}. For $n \in \mathbb{N}$ denote by $X_{i:n}$ the i-th *order statistic* of a sample of size n, i.e. the ordered sample

$$X_{i:1} \leqslant X_{i:2} \leqslant \cdots \leqslant X_{n:n}.$$

As F_n is discontinuous it is certainly not invertible yet the concept of quantiles for non invertible distributions allows to define for $0 < u \leqslant 1$

$$\begin{aligned} F_n^{-1}(u) &= \inf\{x \in \mathbb{R} : \ F_n(x) \geqslant u\} = \inf\{x \in \mathbb{R} : \sharp\{X_i \leqslant x\} \geqslant nu\} \\ &= \min\{X_{i:n} : \sharp\{X_{i:n} \leqslant x\} \geqslant nu\} = X_{\lceil nu \rceil :n} \ . \end{aligned} \tag{3.6}$$

It is the left continuous inverse of the right continuous F_n. In analogy to (3.3) we have by the law of large numbers

$$|F_n^{-1}(u,\omega) - F(u)^{-1}| \to 0 \ , \qquad \text{for all } u \in (0,1) \text{ as } n \to \infty \text{ for almost all } \omega \in \Omega.$$

However in general we cannot expect a uniform convergence in u, since for unbounded support of μ the values will be infinity. For a precise analogue to the Glivenko-Cantelli theorem one must therefore stay away from the endpoints of the open support of μ if they are positive or negative infinity. Following the analogy to the central limit theorem of formula (3.5) we introduce the empirical quantile process.

Definition 3.3. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of real valued i.i.d. random variables with common distribution function F. The* empirical quantile process *or simply* quantile process *is defined as*

$$Q_n(u,\omega) := \sqrt{n}\left(F_n^{-1}(u,\omega) - F^{-1}(u)\right), \quad 0 \leqslant u \leqslant 1, \ n \in \mathbb{N} \ . \tag{3.7}$$

For uniform distributions we obtain an analogue of Theorem 3.2.

Theorem 3.4. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of real valued i.i.d. random variables with common uniform distribution function $F = U$. Then there exists a Brownian bridge $(B_n^0)_{n\in\mathbb{N}}$ such that*

$$\sup_{0\leqslant u\leqslant 1} |Q_n(u,\omega) - B_u^0(\omega)| \xrightarrow{d} 0. \tag{3.8}$$

As already mentioned for general distributions F uniform convergence of the quantile process is out of reach, instead we will consider cutoff L^2 distances later. In the following we link the empirical quantile process to Wasserstein distances.

The empirical Wasserstein distance: We can now calculate the Wasserstein distance of an empirical measure $\mu_n(\omega)$ to some given reference measure μ. We introduce the *Wasserstein statistic*

$$\mathrm{w_n}(\omega) := W_2^2(\mu_n(\omega), \mu) \, . \tag{3.9}$$

We can calculate this distance by

$$\mathrm{w_n} = \int_0^1 |F_n^{-1}(u) - F^{-1}(u)|^2 du = \int_0^1 |X_{\lceil nu \rceil :n} - F^{-1}(u)|^2 du = \sum_{i=1}^n \int_{\frac{i}{n}}^{\frac{i}{n}} (X_{(i-1):n} - F^{-1}(u))^2 du \ , \ \ n \in \mathbb{N} \, . \tag{3.10}$$

The last expression actually turns out to be a quadratic polynomial in the order statistic, that is

$$\mathrm{w_n} = \sum_{i=1}^{n} a_i X_{i:n}^2 + b_i X_{i:n} + c \ , \tag{3.11}$$

where the coefficients are determined by the binomial formula and given by

$$a_i = \frac{1}{n}, \quad b_i = -2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} F^{-1}(u) du, \quad c = \int_0^1 \big(F^{-1}(u)\big)^2 du \ . \tag{3.12}$$

This formula can also be adapted for the trimmed version

$$W_2^2(\mu_n, \mu) \leqslant \int_0^1 \big(|F_n^1(u) - F^{-1}(u)|^2 \wedge 1\big) du \tag{3.13}$$

where the right-hand side can be calculated similarly to (3.11) as follows

$$\mathrm{w_n^*} = \sum_{i=1}^{n} \beta_i^1 X_{i:n}^2 + \beta_i^2 X_{i:n} + \beta_i^3 + \beta^4 \ , \tag{3.14}$$

where for the sequence $0 \leqslant a_1^* \leqslant b_1^* \leqslant a_2^* \leqslant \cdots \leqslant a_n^* \leqslant b_n^* \leqslant 1$ given by

$$a_i^* = \left(\frac{i-1}{n} \vee |F(X_{i:n} - 1)|\right) \wedge \frac{i}{n} \ , \qquad b_i^* = \frac{i-1}{n} \vee \left(|F(X_{i:n} + 1)| \wedge \frac{i}{n}\right) \tag{3.15}$$

the coefficients are calculated by

$$\beta_i^1 = b_i^* - a_i^*, \quad \beta_i^2 = -2 \int_{a_i^*}^{b_i^*} F^{-1}(u) du, \quad \beta_i^3 = \int_{a_i^*}^{b_i^*} \big(F^{-1}(u)\big)^2 du, \quad \beta^4 = 1 - \sum_{i=1}^{n} \beta_i^1 \ . \tag{3.16}$$

Remark 3.1. The concept of coupling distances was introduced in [10] for distributions in $\mathbb{R}^d$. However, in higher dimensions the concept of empirical quantile functions turns out to be much more involved.

Asymptotic distribution and rate of convergence: For rigorous statistical applications it is necessary to determine the rate of convergence of the statistic of interest, in our case $\mathrm{w_n^*}$. By definition $\mathrm{w_n^*}$ tends to zero. Quantifying the rate of convergence amounts to finding the correct renormalization to obtain a non-trivial (random) limit.

For this purpose we need the notion of slow variation. A (measurable) function $\ell : (0,1) \to (0,1)$ is *slowly varying* at zero (at one) if it satisfies for all $x \in (0,1)$

$$\lim_{\substack{u \to 0 \\ (u \to 1)}} \frac{\ell(ux)}{\ell(u)} = 1.$$

The prototype of a slowly varying function at 0 is $\ell(u) = \ln(u)$. An easy calculation shows

$$\frac{\ell(ux)}{\ell(u)} = \frac{\ell(u)}{\ell(u)} + \frac{\ell(x)}{\ell(u)} \to 1 \text{ as } u \to 0+ \, .$$

We summarize the relevant properties. A classical reference is [4].

- Let ℓ be slowly varying. Then also ℓ^γ is slowly varying for all $\gamma \in \mathbb{R}$.
- For $\gamma > 0$, we have $u^\gamma \ell(u) \to 0$ and $u^{-\gamma}\ell(u) \to \infty$ as $u \to 0+$.

In the following we prove a polynomial rate of convergence of w_n^* up to a slowly varying contribution.

Theorem 3.5. *Let F be a distribution function on $\mathbb{R}$ with density f such that $f \circ F^{-1}$ is* regularly varying *at zero and at one, of index $\kappa, \kappa' > 0$ in the sense that*

$$f(F^1(u)) = u^\kappa \ell(u) \quad \textit{near } 0, \qquad \textit{and} \qquad f(F^1(u)) = u^{\kappa'} \ell(u) \quad \textit{near } 1$$

for a common slowly varying function ℓ in zero and one. Assume $|\kappa - \kappa'| < \frac{1}{2}$ and without loss of generality let $\kappa < \kappa'$. Choose now γ

$$0 < \frac{2\kappa' - 3}{2\kappa' - 2} < \gamma < \frac{2\kappa - 2}{2\kappa - 1} < 1.$$

Then we have the limit

$$n^{1-2(1-\gamma)(\kappa-1)}\, \ell^2\Big(n^{\gamma-1}\Big)\, \mathrm{w_n}^* \xrightarrow{d} \big(1 + \mathbb{1}(\kappa = \kappa')\big) \int_0^1 (u^{\kappa-2} B_u)^2 du \quad \textit{as} \quad n \to \infty\ ,$$

where $B = (B_u)_{u\in[0,1]}$ is a standard Brownian motion.

Remark 3.2. It is easy that

$$\frac{2\kappa' - 3}{2\kappa' - 2} < \frac{2\kappa - 2}{2\kappa - 1}$$

is equivalent to $\kappa' - \frac{1}{2} \leqslant \kappa \leqslant \kappa'$.

Proof. The proof follows along the lines of [6]. First consider a sequence of intermediate points $(k_n)_{n\in\mathbb{N}}$ such that

$$1 \leqslant k_n \leqslant n\ , \quad k_n \nearrow \infty \qquad \text{and} \qquad \frac{k_n}{n} \searrow 0 \quad \text{as } n \to \infty\ .$$

We then decompose our integral

$$\int_0^1 \Big(|F_n^{-1}(u) - F^{-1}(u)|^2 \wedge 1\Big) du = \Big(\int_0^{\frac{k_n}{n}} + \int_{\frac{k_n}{n}}^{\frac{1}{2}} + \int_{\frac{1}{2}}^{\frac{n-k_n}{n}} + \int_{\frac{n-k_n}{n}}^1 \Big) \Big(|F_n^{-1}(u) - F^{-1}(u)|^2 \wedge 1\Big) du\ . \quad (3.17)$$

We will first treat the second integral in of (3.17) with the help of Theorem 2.4 in [5]. With a slight adaption to our situation it states that for the quantile process $Q_n = \sqrt{n}(U_n - u)$ of the uniform distribution on $[0,1]$ (cf. Definition 3.3), any $\kappa > 1$ and any slowly varying function $\ell : (0,1) \to (0,1)$ we have the limit

$$\Big(\frac{k_n}{n}\Big)^{2(\kappa-1)}\, \ell^2\Big(\frac{k_n}{n}\Big) \int_{\frac{k_n}{n}}^{\frac{1}{2}} \Big(\frac{Q_n(u)}{u^\kappa \ell(u)}\Big)^2 du \xrightarrow{d} \int_0^1 u^{2\kappa-4} |B_u|^2 du \quad \textit{as} \quad n \to \infty. \quad (3.18)$$

The mean value theorem tells us that for some (random) intermediate value

$$u \wedge U_n(u) \leqslant \vartheta_n(u) \leqslant u \vee U_n(u) \tag{3.19}$$

we have for each $\omega \in \Omega$

$$\begin{aligned} n \int_{\frac{k_n}{n}}^{\frac{1}{2}} \Big(|F_n^{-1}(u) - F^{-1}(u)|^2 \wedge 1 \Big) du &= n \int_{\frac{k_n}{n}}^{\frac{1}{2}} \bigg(\Big(\frac{\partial}{\partial u} F^{-1}(\vartheta_n(u)) \Big)^2 |U_n^{-1}(u) - U^{-1}(u)|^2 \wedge 1 \bigg) du \\ &= \int_{\frac{k_n}{n}}^{\frac{1}{2}} \Big(\frac{|\sqrt{n}(U_n^{-1}(u) - u)|^2}{f(F^{-1}(\vartheta_n(u))^2} \wedge n \Big) du. \end{aligned}$$

By continuity of f, inequality (3.19), the regular variation of $f(F^{-1}(u))$ we have $f(F^{-1}(\theta_n(u)) \to \ell(u)u^\kappa$ as $n \to \infty$. Hence (3.18) yields

$$\begin{aligned} & n \Big(\frac{k_n}{n} \Big)^{2\kappa-2} \ell^2 \Big(\frac{k_n}{n} \Big) \int_{\frac{k_n}{n}}^{\frac{1}{2}} \Big(|F_n^{-1}(u) - F^{-1}(u)|^2 \wedge 1 \Big) du \\ &= \Big(\frac{k_n}{n} \Big)^{2\kappa-2} \ell^2 (\frac{k_n}{n}) \int_{\frac{k_n}{n}}^{\frac{1}{2}} \Big(\frac{Q_n(u)^2}{f(F^{-1}(\vartheta_n(u))^2} \wedge n \Big) du \xrightarrow{d} \int_0^1 u^{2\kappa-4} |B_u|^2 du. \end{aligned}$$

By symmetry of the problem analogous arguments hold for the upper half interval (the third integral of (3.17)) with κ'. For the remainder integrals we have the simple bound

$$n \int_0^{\frac{k_n}{n}} \Big(|F_n^{-1}(u) - F^{-1}(u)|^2 \wedge 1 \Big) du \leqslant n \frac{k_n}{n} = k_n \to 0. \tag{3.20}$$

By the properties of ℓ outlined above only the polynomial terms define the two conditions on k_n

$$n \left(\frac{k_n}{n} \right)^{2\kappa-2} \ell^2 \Big(\frac{k_n}{n} \Big) \nearrow \infty \quad \text{and} \quad k_n \left(\frac{k_n}{n} \right)^{2\kappa-2} \ell^2 \Big(\frac{k_n}{n} \Big) \searrow 0 \ . \tag{3.21}$$

We choose $k_n = n^\gamma$, with $\gamma < 1$ and we get the following requirement

$$\frac{2\kappa - 3}{2\kappa - 2} < \gamma < \frac{2\kappa - 2}{2\kappa - 1} \ . \tag{3.22}$$

Now we have treated all integrals in (3.17). The two tail integrals do not contribute to the limit as long as the upper bound on γ in (3.22) holds for both κ, κ'. Actually only the integral with the smaller exponent matters (therefore two times if $\kappa = \kappa'$). It remains to see that the bounds in (3.22) are monotonically increasing for $\kappa > 1$ to obtain the bounds on γ stated in the theorem. □

Example 3.3. The assumptions on the underlying distribution are not restrictive.

1. Prominent examples of heavy tailed distributions are (symmetric) α-stable distributions. See [8]. They admit a density f and the distribution function F is regularly varying of order $-\alpha, \alpha \in (0,2)$ at $\pm\infty$. L'Hospital's rule yields that f is also regularly varying of order $-(\alpha+1)$. Obviously the quantile function F^{-1} is then regularly varying at 0 and 1 with index $-\frac{1}{\alpha}$ and
$$f(F^{-1}(u)) = u^{\frac{\alpha+1}{\alpha}}\ell(u)$$
for a function ℓ slowly varying at $\pm\infty$. In this case $\kappa = \kappa' = 1 + \frac{1}{\alpha}$ and we obtain a rate of convergence of our statistic of polynomial order
$$n^{1-2(\kappa-1)(1-\gamma)} \quad \text{with } 1 - \frac{\alpha}{2} < \gamma < \frac{1}{1+\frac{\alpha}{2}}.$$
Minimizing γ we can achieve any polynomial rate of convergence slower than $n^{\frac{1}{1+\frac{2}{\alpha}}}$. Since $\alpha \in (0,2)$ this is slower than the order $\sqrt{n}$ of the central limit theorem.

2. The simplest examples of one-sided polynomial tails are Pareto distributions where
$$F(x) = 1 - \frac{c^\alpha}{x^\alpha}, \qquad x \geqslant c,\ \alpha > 1.$$
By analogous reasoning we can still achieve convergence rates arbitrarily close to
$$n^{\frac{1}{1+\frac{2}{\alpha}}},$$
now including values $\alpha > 2$. In this regime we are faster than the central limit theorem but always less than one.

Corollary 3.4. *In the situation of Theorem (3.5) our statistic* w_n^* *renormalized as in (3.21) tends to a limit with expectation* $\frac{1}{2\kappa-2}$ *and for* $\kappa > \frac{5}{4}$ *the variance is bounded by* $\frac{1}{4\kappa-5}$.

Proof. Since $\mathbb{E}[B_u^2] = u$ we have by Fubini's theorem
$$\mathbb{E}[\int_0^1 u^{2\kappa-4}|B_u|^2 du] = \int_0^1 u^{2\kappa-3}du = \frac{1}{2\kappa-2},$$
and since $\frac{B_u}{\sqrt{u}} \sim N(0,1)$ we have
$$\mathbb{E}\Big[\Big(\int_0^1 (u^{\kappa-2}|B_u|)^2 du - \frac{1}{2\kappa-2}\Big)^2\Big] = \mathbb{E}\Big[\Big(\int_0^1 u^{2\kappa-3}\big(\frac{|B_u|^2}{u} - 1\big)du\Big)^2\Big]$$
$$\leqslant \int_0^1 \Big[u^{2(2\kappa-3)}\mathbb{E}\big(\frac{|B_u|^2}{u} - 1\big)^2\Big]du = \int_0^1 u^{2(2\kappa-3)}du = \frac{1}{2(2\kappa-3)+1}.$$
□

4 The procedure in detail

In this section we work through the program laid out in the previous sections. In order to keep calculations and the strategy easily tractable we will restrict ourselves to a simple example, which can be obviously adapted and extended. Many real-world phenomena exhibit strong fluctuations in short time, which are hardly explained by a continuous evolution.

A reasonable modeling approach is hence to consider a process of the type

$$Y_t = G_t + L_t, \qquad t \in [0,T] \text{ for fixed } T > 0, \tag{4.1}$$

where $G = (G_t)_{t\in[0,T]}$ is a continuous process and $L = (L_t)_{t\in[0,T]}$ is a purely discontinuous Lévy process. We have seen in Section 2, that L is determined by a Lévy triplet of the form $(0, 0, \nu)$, where ν is a Lévy measure defined in (2.1). For instance, the solutions of stochastic differential equations

$$Y_t = x + \int_0^t f(Y_s)ds + L_t, \qquad t \in [0,T],$$

for globally Lipschitz continuous functions $f : \mathbb{R} \to \mathbb{R}$ fall into this class. The aim of this procedure is now to determine the nature of L and thus of its Lévy measure ν.

Given a data set $y = (y_i)_{i=0,\dots n}$ we interpret y as a realization of a process Y given in the class of models (4.1) observed at discrete times $t_1 < \cdots < t_i < \cdots < t_n$, that is

$$y_i = Y_{t_i}(\omega) \qquad \text{for some } \omega \in \Omega.$$

We make the following modeling assumptions: Fix a threshold $\rho > 0$.

1. The observation frequency is sufficiently high in comparison to the occurrence of large jumps given as increments beyond our threshold ρ. That means we assume that in each time interval $[t_{i-1}, t_i)$ at most one large jump occurs.

2. The behavior of small jumps is sufficiently benign in comparison to the large jump threshold ρ during our observation. That means in particular, we assume that over one time interval $[t_{i-1}, t_i)$ that small jump and continuous contributions cannot accumulate to this threshold.

These assumptions can be made rigorous by further model assumptions on Y e.g. with the Lipschitz continuity of G. Under the assumptions it is justified to estimate

$$Y_{t_i} - Y_{t_i-1} \approx C^\rho_s - C^\rho_{s-0} \quad \text{for exactly one } s \in [t_{i-1}, t_i),$$

for the compound Poisson process C^ρ given by (2.3) and hence can be considered as the realization of an i.i.d. sequence $X = (X_i)_{i=1,\dots,n}$. We denote by $x = (x_i)_{i=1,\dots n}$ the vector of large increments

$$x_i = (y_i - y_{i-1})\mathbf{1}\{|y_i - y_{i-1}| > \rho\}$$

this means

$$x_i = X_i(\omega) \qquad \text{for some } \omega \in \Omega.$$

Let μ_n be the empirical measure of the data X. The Glivenko-Cantelli theorem (3.2) tells us that for almost all $\omega \in \Omega$

$$\mu_n(\omega, \cdot) \to \mu \qquad n \to \infty, \text{ weakly}$$

for the common distribution μ of X. Since by construction $\mu = \nu_\rho$ we obtain

$$\mu_n(\omega) \approx_n \nu_\rho.$$

Since the Wasserstein distance encoded in the coupling distance metrizes the weak convergence we have for almost all $\omega \in \Omega$

$$T_{\lambda_\rho}(\mu_n, \nu_\rho) \to 0 \qquad \text{as } n \to \infty.$$

In particular we have an estimator for the tail of the Lévy measure ν_ρ. We are now in the position to estimate the distance between the Lévy measure of the compound Poisson approximation C^ρ of L (respectively Y) and the tail ν_ρ^* of a suspected reference Lévy measure ν^*. In particular

$$T_{\lambda_\rho}(\nu_\rho, \nu^*) \leqslant T_{\lambda_\rho}(\nu_\rho, \mu_n) + T_{\lambda_\rho}(\mu_n, \nu_\rho^*) \leqslant (\lambda_\rho \, \mathrm{w_n}^*(\mu_n, \nu^*))^{\frac{1}{2}} + T_{\lambda_\rho}(\nu_\rho, \mu_n), \tag{4.2}$$

where the second to last term tends to 0. The first term can be calculated explicitly due to Section 3. This is carried out in the case of polynomial tails in the next section.

5 Case study

Many observed quantities in nature follow a heavy-tailed distribution, that can be interpreted as the superposition of (large) power law jumps. Small jumps are statistically hard to distinguish from continuous increments and will be neglected in this study. Therefore it is natural to consider jumps away from 0 with polynomial tails. In our simple model we assume

$$\nu(dy) = \frac{c_1 dy}{|y|^{1+\alpha_1}} \mathbf{1}\{y < -\rho_1\} + \frac{c_2 dy}{y^{1+\alpha_2}} \mathbf{1}\{y > \rho_2\}, \tag{5.1}$$

where $c_1, c_2 \geqslant 0$, $\alpha_1, \alpha_2 > 0$ and $\rho_1, \rho_2 > 0$ sufficiently small.

Concrete formulas: To prevent numerical pathologies in the following calculation we exclude α_1 and α_2 from being 1 or 2. Due to Remark (2.1) we can restrict ourselves to one-sided Lévy measures and estimate the right and left tail independently, that is $c = c_1 > 0$, $c_2 = 0$, $\alpha = \alpha_1 > 0$ and $\rho = \rho_1 > 0$. Consider the Lévy measure ν with Pareto density f given by

$$\begin{aligned} f(x) &= \frac{c}{|x|^{\alpha+1}}, && x \leqslant -\rho \\ F(x) := \nu((-\infty, x]) &= \frac{c}{\alpha}|x|^{-\alpha}, && x \leqslant -\rho \\ F^{-1}(u) &= -(\frac{\alpha}{c}u)^{-1/\alpha}, && u \in (0,1). \end{aligned}$$

Since the measure ν is not a probability measure we will introduce the normalized measures ν_λ of precisely mass $\lambda > 0$, that is supported on $(-\infty, -\rho]$, with

$$\rho = F^{-1}(\lambda) = \left(\frac{\alpha\lambda}{c}\right)^{-1/\alpha}. \tag{5.2}$$

We now can look at the normalization

$$\begin{aligned} \tilde{f}_\lambda(x) &= \frac{c}{\lambda|x|^{\alpha+1}}, \quad x \leqslant -\rho \\ \tilde{F}_\lambda(x) &= \frac{c}{\alpha\lambda}|x|^{-\alpha}, \qquad x \leqslant -\rho. \end{aligned} \tag{5.3}$$

and its inverse

$$\tilde{F}_\lambda^{-1}(u) = -(\frac{\alpha\lambda}{c}u)^{-1/\alpha}, \qquad u \in (0,1).$$

In order to calculate the cutoff Wasserstein distance in formula (3.16) we evaluate the primitives

$$\int_z^1 \tilde{F}_\lambda^{-1}(u)du = -\frac{\alpha}{\alpha-1}\left(\frac{\alpha\lambda}{c}\right)^{-1/\alpha}\left(1 - z^{1-1/\alpha}\right), \qquad z \in (0,1) \tag{5.4}$$

$$\int_z^1 \big(\tilde{F}_\lambda^{-1}(u)\big)^2 du = -\frac{\alpha}{\alpha-2}\left(\frac{\alpha\lambda}{c}\right)^{-2/\alpha}\Big(1-z^{1-2/\alpha}\Big), \qquad z\in(0,1). \tag{5.5}$$

Now all necessary functions to implement the empirical Wasserstein distance (3.14) are at our disposal. Recall the modeling assumption of Section 4 and that the time series stems from a process of the form (4.1).

Simulations: In a first test case we simulate $n = 100.000$ data points from a the perfectly symmetric version of jump measure given in (5.1) with minimal jump sizes $\rho = 0.5$ and $\alpha^1 = 1.5$, $\alpha^2 = 1.8$, $\alpha^3 = 2.4$ and $\alpha^4 = 3.0$. We interpret those as the jumps of compound Poisson processes at rate $\lambda = 1$ and denote by μ_n^i the empirical measure of the respective simulation $i = 1,\dots 4$. Furthermore we set $T = 1$ and apply a small linear drift $G(t) = 0.0125 \times t$ to each of them. The paths are shown in Figure 5 (left). We calculate the empirical coupling distance between the simulated data and the jump measures according to the outlined procedure. The right display of Figure 5 shows pronounced and small minima of the coupling distances as a function

$$\alpha \mapsto T_\lambda(\nu_\rho^\alpha, \mu_n^i) \in [0, \lambda^{\frac{1}{2}}]. \tag{5.6}$$

at the original values of $\alpha = \alpha_i$. Due to the cutoff of the Wasserstein distance at height 1 (see Definition 2.3) and the specific choice of the intensity $\lambda = 1$ the values of (5.6) are between 0 and 1. Due to Theorem 3.5, Corollary 3.4 and Example 3.3 we can give a the expected value of $T_\lambda(\nu_\rho^{\alpha_i}, \mu_n^i)$. For large n this expectation should be close to the normalized expectation of the limiting distribution. For the optimal rate for $\gamma = 1 - \frac{\alpha}{2}$ this is given by

$$E^{i,n} := \Big(\frac{\mathbb{E}[\int_0^1 (u^{1+\frac{1}{\alpha}} B_u)^2 du]}{\lambda\, n^{\frac{1}{1+\frac{2}{\alpha}}}}\Big)^{\frac{1}{2}} = \sqrt{\frac{\alpha}{2\lambda}}\; n^{-\frac{1}{2(1+\frac{2}{\alpha})}}. \tag{5.7}$$

The simulation results are compared to the expected value in the following table.

α_i	1.5	1.8	2.4	3.0	
$T_\lambda(\nu_\rho^{\alpha_i}, \mu_n^i)$	0.0575	0.0642	0.0273	0.0183	$\ll 1$
$E^{i,n}$	0.0735	0.0621	0.0474	0.0387	

Apparently the empirical results are better than our prediction $E^{i,n}$. Recall that the optimal rate in Example 3.3 is actually not achieved, hence we renormalize by a too small value in formula (5.7) and consequently slightly overestimate the value of our statistic $T_\lambda(\nu_\rho^{\alpha_i}, \mu_n^i)$.

Paleoclimatic time series: The concentration of calcium ions in ice core data from the Greenland shelf provide a climate proxy for the yearly average temperature distribution during the last glacial period, see [19]. The record shows large fluctuations between (cooler) stadials and (warmer) interstadials, see Figure 5 (left up). In [7] Ditlevsen concentrated on the actual (42) transitions between the different regimes which he interprets as the effect of single large jumps of an underlying noise signal. His analysis indicates a tail index of $\alpha = 1.75$. Moreover he proposed an α-stable Lévy component in the noise signal.

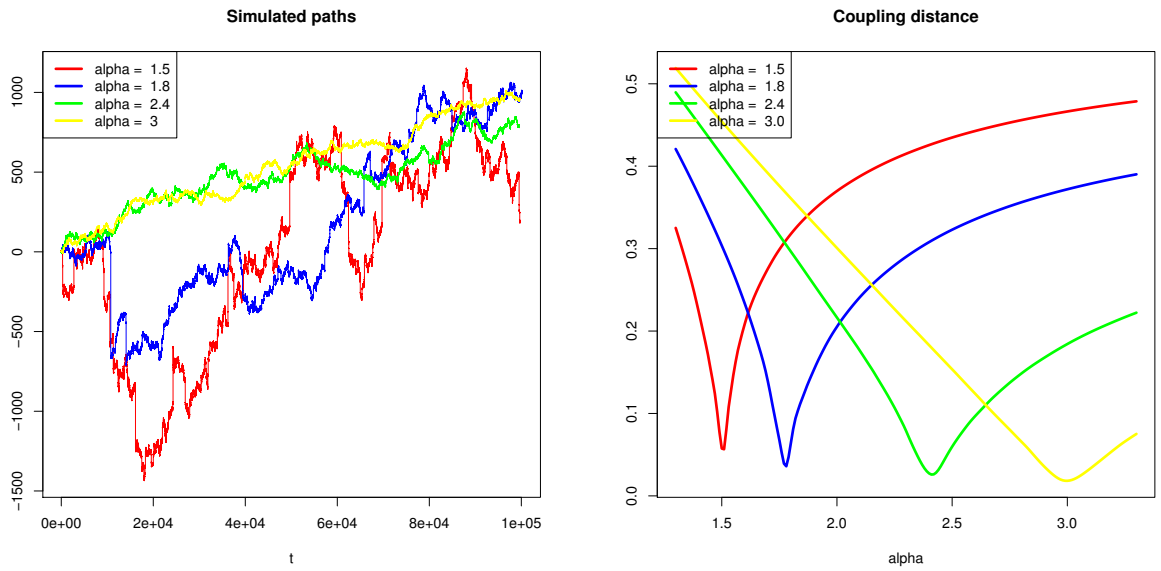


Figure 1: Left: Simulated paths for fixed values of α. Right: Empirical coupling distances between the data of the simulated paths and the jump measures (5.3), as a function of varying α

A series of works [12], [15] and [14] continued the investigation on the mathematical side towards α-stable perturbations and developed an estimation procedure based on the selfsimilarity and characteristic path variations. In the realm of stable diffusions their method confirms the proposed index of stability. In [9] Wasserstein distances were also applied to measure the distance between α-stable distributions and empirical measures, however due to the lack of moments the analysis is restricted to W_p for $p = p(\alpha) < 1$.

These elaborate techniques are hardly applicable beyond this framework. The procedure proposed here does not rely on any features of stable distributions and only requires a monotonic tail behavior of the jump distribution. There is a variety of estimators for the tail index in the literature, our method does not intend to contribute to this list. Instead our method quantifies the fit of a *proposed* tail behavior to a *given* data set. The proposed tails could be derived by standard methods. Certainly coupling distances can also serve in a minimal distance estimation procedure. In order to make such a procedure statistically rigorous it would be necessary to further quantify the separation power by finding uniform lower bound jointly in the parameters (α, λ, ρ).

Our model only describes large jumps, small fluctuations can certainly not be described by the (polynomial) tail behavior. In other words, the modeling assumptions in Section 4 are valid only for increments beyond a certain threshold $\rho > 0$. Small fluctuations are interpreted as contributions of a continuous part G and small jumps as in (4.1). Also note that for $\alpha > 2$ in the limit of ρ to 0 the measure ν from (5.1) is not square integrable in the neighborhood of 0. Hence it is not a Lévy measure (cf. formula (2.1) in Section 2).

Our procedure shows marked minima of the coupling distance for the right and the left side of the one-sided polynomial tails given in 5.1 as a function in α for $\alpha_1 = 3.55$ (left) and $\alpha_2 = 3.6$ (right). The Wasserstein distances, which correspond to the coupling distance to the interval $[0, 1]$ gives the small values $0.089 \ll 1$ (left) and $0.081 \ll 1$ (right). The cutoffs have been chosen $\rho_1 = 0.34$ allowing for $n_1 = 530$ sample points and $\rho_2 = 0.36$ with $n_2 = 894$ sample points. The respective

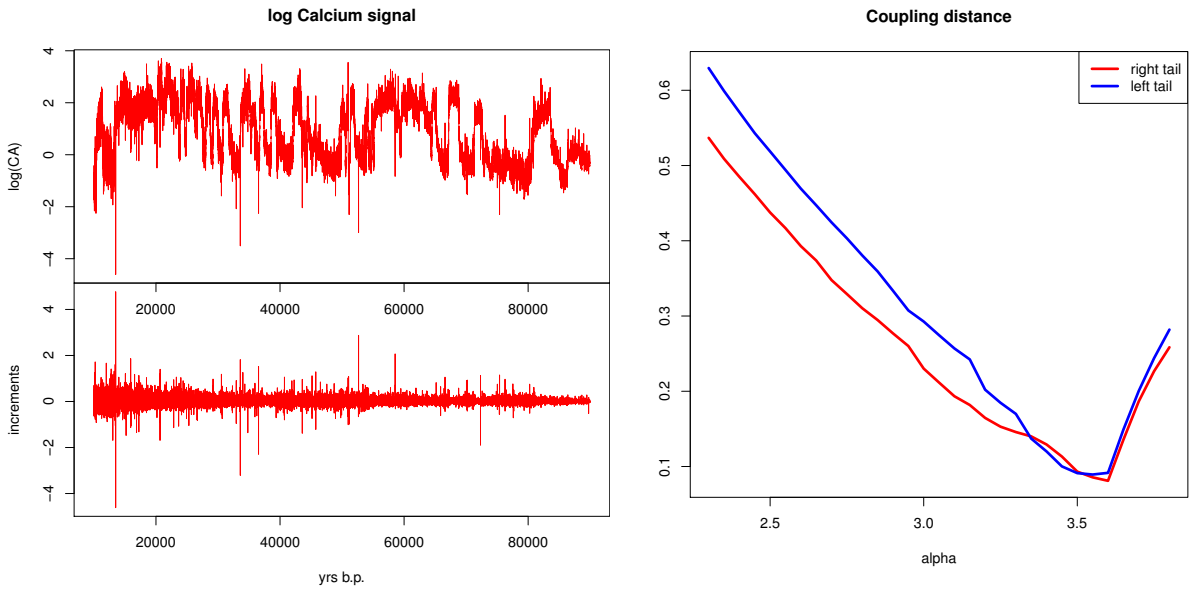


Figure 2: Left (above): Logarithmic calcium signal. Left (below): Increments of the logarithmic calcium signal. Right: Plot of the coupling distance for the left tail (blue) and the right tail (red) of the jump measure.

rates are $\lambda_{\rho_1} = 11$ and $\lambda_{\rho_2} = 8$. The prediction procedure developed above yields $E^{1,n_1} = 0.054$ and $E^{2,n_2} = 0.053$. This under estimation could be explained beyond climatological reasoning by comparably small sample sizes, where the asymptotic regime is not yet fully unfolded. Yet the order of magnitude is caught.

The authors point out that there is a significant degree of freedom in the 3 parameter fit. For instance the cutoff parameter may vary between the floating boundaries of the modeling assumption of a meaningful tail contribution (not corrupted by small scale fluctuations) and a statistically relevant number of data points.

Comparable fits for $\alpha = 1.75$ with a significant number of data points $n > 100$ do not allow for reasonably small Wasserstein distances (all values > 0.5). Consequently we could not confirm the suggestion of a tail index $\alpha = 1.75$ in the literature. The proposed procedure indicates a Lévy jump component with a tail index clearly above 2, that cannot belong to the family of stable diffusions.

References

[1] D. Applebaum. Levy processes and stochastic calculus. *Cambridge university press*, 2009.

[2] E. del Barrio, P. Deheuvels, S. van de Geer. Lectures on Empirical Processes. *Series of Lectures in Mathematics of the EMS*, 2007.

[3] P. Billingsley. Convergence of probability measures. *Wiley series in probability and statistics*, 1999.

[4] N. H. Bingham, C. M. Goldie, J. L. Teugels. Regular variation. *Cambridge University Press*, 1987.

[5] M. Csörgö, C. Horváth, On the distributions of Lp norms and weighted uniform empirical and quantile processes. *The Annals of Probability*, (16) (1) 142–161, 1988.

[6] M. Csörgö, C. Horváth, On the distributions of Lp norms and weighted quantile processes. *Annales de l'I.H.P.*, section B (26) (1) 65–85, 1990.

[7] P.D. Ditlevsen. Observation of a stable noise induced millennial climate changes from an ice-core record. *Geophysical Research Letters*, 26 (10):1441–1444, 1999.

[8] W. Feller. An introduction to probability theory and its applications. *Geophysical Research Letters*, Vol. II, John Wiley & Sons, 1971.

[9] J. Gairing. Speed of convergence of discrete power variations of jump diffusions. *Diplom thesis, Humboldt-Universität zu Berlin, 2011*

[10] J. Gairing, M. Högele, T. Kosenkova, A. Kulik: Coupling distances and applications to noise sensitivity of SDE. *submitted*

[11] J. Gairing, M. Högele, T. Kosenkova, A. Kulik. Coupling distances between Lévy measures and statistical inference for paleoclimate data. *In preparation.*

[12] J. Gairing, P. Imkeller. Stable CLTs and rates for power variation of α-stable Lévy processes. *Methodology and Computing in Applied Probability*, 1-18, 2013.

[13] A. Debussche, M. Högele, P. Imkeller. The dynamics of non-linear reaction-diffusion equations with small Lévy noise. *Springer Lecture Notes in Mathematics, Vol. 2085*, 2013.

[14] C. Hein, P. Imkeller, I. Pavlyukevich. Limit theorems for p-variations of solutions of SDEs driven by additive stable Lévy noise and model selection for paleo-climatic data. *Interdisciplinary Math. Sciences*, 8:137-150, 2009.

[15] C. Hein, P. Imkeller, I. Pavlyukevich. Simple SDE dynamical models interpreting climate data and their meta-stability. Oberwolfach Reports, 2008

[16] M. Högele, I. Pavlyukevich. The exit problem from the neighborhood of a global attractor for heavy-tailed Lévy diffusions. *Stochastic Analysis and Applications*, 32 (1), 163–190 (2013)

[17] P. Imkeller and I. Pavlyukevich. First exit times of SDEs driven by stable Lévy processes. *Stochastic Processes and their Applications*, 116(4):611–642, 2006.

[18] P. Imkeller, I. Pavlyukevich, T. Wetzel. First exit times for Lévy-driven diffusions with exponentially light jumps. *Ann. Probab.*, 37 (2) 530-564, 2009.

[19] Members NGRIP. NGRRIP data. *Nature*, 431:147–151, 2004.

[20] T. Kosenkova, A. Kulik. Explicit bounds for the convergence rate of a sequence of Markov chains to a Lévy-type process. *In preparation.*

[21] I. Pavlyukevich. First exit times of solutions of stochastic differential equations with heavy tails. *Stochastics and Dynamics*, 11, Exp.No.2/3:1–25, 2011.

[22] S. T. Rachev, L. Rüschendorf. Mass Transportation Problems. Vol.I: Theory, Vol.II.: Applications. *Probability and its Applications*. Springer-Verlag, New York 1998.

[23] G. Samoradnitsky, M. Taqqu. Stable non-Gaussian random processes. *Chapman& Hall*. Boca-Raton, London, New York, Washington D.C. 1994.

[24] A.W. van der Vaard, J.A.Wellner, Weak convergence and empirical processes. *Springer Series in Statistics*, 1996.