SFB 823

Discussion Paper

# Nonparametric inference on Lévy measures and copulas

Axel Bücher, Mathias Vetter

Nr. 16/2012

SFB 823

# Nonparametric inference on Lévy measures and copulas

Axel Bücher Mathias Vetter *

May 2, 2012

**Abstract**

In this paper nonparametric methods to assess the multivariate Lévy measure are introduced. Starting from high-frequency observations of a Lévy process $\mathbf{X}$, we construct estimators for its tail integrals and the Pareto Lévy copula and prove weak convergence of these estimators in certain function spaces. Given $n$ observations of increments over intervals of length $\Delta_n$, the rate of convergence is $k_n^{-1/2}$ for $k_n = n\Delta_n$ which is natural concerning inference on the Lévy measure. Analytic properties of the Pareto Lévy copula which, to the best of our knowledge, have not been mentioned before in the literature are provided as well. We conclude with a short simulation study on the performance of our estimators.

*Keywords*: Copula, Lévy copula, Lévy measure, Lévy process, nonparametric statistics, Pareto Lévy copula, weak convergence.

*AMS 2000 subject classifications*: primary, 60F05, 60G51, 62H20; secondary, 62G32, 62M09.

*Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany. E-mail: axel.buecher@rub.de, mathias.vetter@rub.de. The authors are thankful for financial support through the collaborative research center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

# 1 Introduction

The modeling and estimation of dependencies is attracting an increasing attention over the last decades in various fields of science like mathematical finance, actuarial science or hydrology, among others.

In discrete time models, one of the most popular approaches is the concept of copulas which allows to separate the effects of dependence of a random vector from its univariate marginal behavior. In the bivariate case, the copula of a continuous random vector $(X, Y)$ is the unique function $C : [0,1]^2 \to [0,1]$ for which

$$\mathbb{P}[X \le x, Y \le y] = C\big(\mathbb{P}[X \le x], \mathbb{P}[Y \le y]\big).$$

This formula, known as Sklar's Theorem, is usually interpreted in the way that the copula completely characterizes the stochastic dependence between $X$ and $Y$ and hence represents the primary object of interest for investigating dependencies. For introductions to the concept of copulas in the aforementioned fields of science see McNeil et al. (2005), Frees and Valdez (1998), Genest and Favre (2007) and references therein. The books of Joe (1997) and Nelsen (2006) provide compendiums on the mathematical background and on various parametric models. The huge amount of applications gave rise to a great demand for statistical methods, of which semi- and nonparametric estimation in discrete time i.i.d. models has been investigated in Genest et al. (1995), Fermanian et al. (2004) and Segers (2011). Nonparametric generalizations to the case of serially dependent stationary time series have recently been considered in Bücher and Volgushev (2011).

On the other hand, a huge amount of models in applied stochastics relies on an underlying process which is defined in continuous time. A basic tool in this framework is the class of (multidimensional) Lévy processes which provides a flexible way to model empirically observed behaviour and includes prime examples such as Brownian motion and the (compound) Poisson process. Statistical methods in this context depend on the nature of the observation schemes which are usually classified as high frequency and low frequency setups. In both areas the literature on nonparametrics has grown considerably over the last decade. To mention only a few approaches we refer to Jacod (2007) and Figueroa-López (2009) for the case of high frequency observations, whereas seminal papers in the low frequency setting are due to Neumann and Reiß (2009) and recently to Nickl and Reiß (2012).

Our aim in this work is to combine both strands of the literature and to provide nonparametric methods to estimate the dependence structure of a multivariate Lévy process. For the sake of brevity we will concentrate on the bivariate case solely, but extensions to the general $d$-dimensional setting are straightforward to obtain as well. Thus, let $\mathbf{X} = (X^{(1)}, X^{(2)})$ be a two-dimensional Lévy process with Lévy-Itô decomposition

$$\mathbf{X}_t = \mathbf{a}t + \mathbf{B}_t + \int_0^t \int_{\|\mathbf{u}\| \le 1} \mathbf{u} \star (\mu - \bar{\mu})(ds, d\mathbf{u}) + \int_0^t \int_{\|\mathbf{u}\| > 1} \mathbf{u} \star \mu(ds, d\mathbf{u}), \tag{1.1}$$

where $\mathbf{a} \in \mathbb{R}^2$ is a drift vector, $\mathbf{B}$ is a bivariate Brownian motion with some covariance matrix $\Sigma$, and $\mu$ and $\bar{\mu}$ are the jump measure of the Lévy process and its compensator, respectively. It is well-known that the compensator takes the form $\bar{\mu}(ds, d\mathbf{u}) = ds\ \nu(d\mathbf{u})$, where $\nu$ is the so called Lévy measure of $\mathbf{X}$. Given the choice of the truncation function $h(\mathbf{u}) = 1_{\{\|\mathbf{u}\|>1\}}$, the law of $\mathbf{X}$ is uniquely determined by the Lévy triplet $(\mathbf{a}, \Sigma, \nu)$.

As noted above, in the framework of statistics for stochastic processes it is inevitable to lose some words on the underlying observation scheme. We decide to work in a high frequency setting which means in the simplest case that at stage $n$ one is able to observe one realization of the process $\mathbf{X}$ at the equidistant times $i\Delta_n$, $i = 0, \ldots, n$, for a mesh $\Delta_n \to 0$. An outlook on extensions to a more general setup including irregularly spaced data and asynchronous observations will be provided in a concluding section at the end of the paper. Within the class of high frequency settings a further distinction regards the nature of the covered time horizon. Usually we have either $n\Delta_n = T$, corresponding to a finite time horizon (a trading day, say), whereas $n\Delta_n \to \infty$ means that the process is eventually observed on the entire time span $[0, \infty)$.

Due to the independence of the continuous part and the jump part of a Lévy process, the analysis of the stochastic nature of $\mathbf{X}$ canonically splits into inference on the covariance matrix $\Sigma$ and inference on the jump measure $\nu$, since no joint contribution of the two components is involved. However, estimation of the characteristics of the Brownian part of $\mathbf{X}$ with our without additional jumps is well understood in the high frequency setup (among others, see Jacod (2008) for a thorough theory on the behaviour of more general Itô semimartingales), so our focus in this paper will be on the jump dependence of the two components. In analogy to standard copulas for random vectors we will employ a concept of a Lévy copula to capture the dependence structure within $\nu$ which dates back to Cont and Tankov (2004) and Kallsen and Tankov (2006). We will follow a slightly different approach due to Klüppelberg and Resnick (2008) and Eder and Klüppelberg (2012), however, and focus on nonparametric methods to assess the closely related Pareto Lévy copula.

Besides parametric approaches to infer the (Pareto) Lévy copula such as Esmaeili and Klüppelberg (2011), nonparametric methods in this area are hardly available. To the best of our knowledge, the only concept is due to the unpublished work of Laeven (2011) who constructs an estimator for the Lévy copula based on a representation in the limit involving ordinary copulas and provides some asymptotic properties, but for which no explicit proof is available. On the other hand, since the (Pareto) Lévy copula captures the tendency of the process to have joint (largely negative) jumps, the need for reliable nonparametric estimators is evident from practice, particularly with a view on finance. This convinces us that there is a clear gap in the literature which we aim to fill in this work.

In contrast to Laeven's method, our approach will be based directly on the defining relation of the Pareto Lévy copula $\Gamma$ which involves tail integrals of both the Lévy measure and its marginals. For simplicity, we will focus on the spectrally positive case only, that is we assume that $\mathbf{X}$ has only positive jumps in both directions, or equivalently that the Lévy measure $\nu$ has support on $[0, \infty)^2 \backslash \{(0, 0)\}$. $\Gamma$ will then naturally be a function on the same space. In the case where all tail integrals are continuous, we obtain a representation of $\Gamma$ as a functional of those, and we propose to estimate $\Gamma$ by using appropriate estimators for the tail integrals. It turns out that in order to do so, we are forced to work in the high frequency setting with infinite time horizon, that is $n\Delta_n \to \infty$. Under some rather mild assumptions we are then able to prove weak convergence of a suitably standardized version of $\hat{\Gamma} - \Gamma$ in a certain function space, which will be our main result. As a by-product we obtain a Donsker theorem for the bivariate Lévy measure as well, a result which is similar in spirit to the recent work of Nickl and Reiß (2012), but in a high-frequency setting rather than a low-frequency world.

The paper is organized as follows: Section 2 is devoted to a brief discussion on jump dependence of bivariate Lévy processes. We summarize the concept of Pareto Lévy copulas and derive some of their analytical properties. In Section 3 we define estimators for bivariate tail integrals, as well as for their associated Pareto Lévy copulas. Weak convergence of these estimators is discussed in Section 4. A brief discussion of our results and a small simulation study are provided in Section 5, whereas some conclusions are given in Section 6. Finally, some technical results are postponed to Section 7.

## 2 Jump dependence and the Pareto Lévy copula

Suppose that we are given a bivariate Lévy process $\mathbf{X}$ of the form (1.1) where $\nu$ denotes its Lévy measure. As already stated in the introduction, one assumption will be that $\nu$ has support on $[0,\infty)^2 \backslash \{(0,0)\}$, which means that both components of $\mathbf{X}$ only have positive jumps. This condition is for notational convenience in first place, as we will see later that one can follow a similar approach in order to estimate the jump dependence in the other three quadrants as well.

Let us review some recent concepts on jump dependence. The basic quantity in this framework is the bivariate tail integral $U$ of $\nu$, which for the moment will be defined as a function from $[0,\infty]^2 \backslash \{(0,0)\}$ to $\mathbb{R}$ given by

$$U(\mathbf{x}) = \nu([x_1,\infty] \times [x_2,\infty]), \qquad \mathbf{x} = (x_1,x_2). \tag{2.1}$$

From the theory of Lévy processes it is well-known that this quantity gives the average amount of jumps of $\mathbf{X}$ which fall into the interval $[x_1,\infty] \times [x_2,\infty]$ during a time period of length one. Since $\mathbf{X}$ has càdlàg paths, $U(\mathbf{x})$ is necessarily finite. In the same way, we are able to introduce marginal tail integrals. Precisely, let $U_i : [0,\infty] \to [0,\infty]$, $i = 1,2$, be defined via

$$U_1(x_1) = \nu([x_1,\infty] \times \mathbb{R}) \quad \text{and} \quad U_2(x_2) = \nu(\mathbb{R} \times [x_2,\infty]). \tag{2.2}$$

Again, $U_i(x_i)$ is finite for $x_i > 0$, but in the infinite activity case we have $U_i(0) = \infty$ and since this is typically satisfied for Lévy processes, we will assume such a property for $i = 1,2$ as well.

It is obvious that the entire information about $\nu$ is contained in the tail integral $U$. Therefore, just as for regular copulas, one might be interested in splitting $U$ into several functions which are related to the jump behaviour of $\mathbf{X}$ in the marginals (naturally given by the univariate tail integrals $U_i$) and a Lévy copula $C$ which captures the specific tendency of $\mathbf{X}$ to have joint jumps. Having this intuition in mind, Cont and Tankov provided the following definition.

**Definition 2.1** *A bivariate Lévy copula for Lévy processes with positive jumps is a function* $C : [0,\infty]^2 \setminus \{(\infty,\infty)\} \to [0,\infty)$ *which*

*(i) is grounded, that is* $C(x,0) = C(0,x) = 0$ *for all* $x \in [0,\infty]$*;*

*(ii) has uniform margins, so* $C(x,\infty) = C(\infty,x) = x$ *for all* $x \in [0,\infty)$*;*

*(iii) is 2-increasing, that is* $C(x_1,x_2) - C(x_1,y_2) - C(y_1,x_2) + C(y_1,y_2) \geq 0$ *for all* $x_1 \leq y_1$ *and* $x_2 \leq y_2$*.*

The main result on Lévy copulas is a version of the famous Sklar's theorem which states that for each tail integral $U$ with marginals $U_1$ and $U_2$ there exists a Lévy copula $C$ such that

$$U(\mathbf{x}) = C(U_1(x_1), U_2(x_2)), \qquad \mathbf{x} = (x_1,x_2) \in [0,\infty]^2 \setminus \{(0,0)\},$$

holds. Similarly to the usual copula, $C$ is uniquely defined if $U_1$ and $U_2$ are continuous. Therefore continuity of $U_i$ is a natural condition in order to secure that the concept of copulas is appropriate, and it becomes our third main assumption. Also, if both marginal tail integrals are strictly decreasing, we obtain a representation of $C$ via

$$C(\mathbf{u}) = U(U_1^{-1}(u_1), U_2^{-1}(u_2)), \qquad \mathbf{u} = (u_1,u_2) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}. \tag{2.3}$$

We will see later that some smoothness assumptions on $\nu$ are necessary for estimation purposes from which strict monotonicity of the marginal tail integrals follows.

The inverse statement of Sklar's theorem is true as well, which states that knowledge of the marginals $U_i$ and the Lévy copula $C$ determines $U$ completely and thus in turn $\nu$. A drawback of the approach of Cont and Tankov (2004) is, however, that $C$ is not a tail integral – in contrast to the regular copula of a random vector which couples marginal distribution functions and is a bivariate distribution function itself. This circumstance makes the interpretation of a Lévy copula quite difficult, and for that reason it appears to be natural to focus on an alternative notion of copula in this setting.

**Definition 2.2** *A bivariate Pareto Lévy copula for Lévy processes with positive jumps is a function* $\Gamma : [0,\infty]^2 \setminus \{(0,0)\} \to [0,\infty)$ *which*

*(i) is grounded, that is* $\Gamma(u,\infty) = \Gamma(\infty,u) = 0$ *for all* $u \in (0,\infty]$*;*

*(ii) has Pareto margins, so* $\Gamma(u,0) = \Gamma(0,u) = 1/u$ *for all* $u \in (0,\infty]$*;*

*(iii) is 2-increasing.*

As usual, we set $1/\infty = 0$ and vice versa. Following Eder and Klüppelberg (2012), Sklar's theorem now reads as follows: Given $U$ and its marginals, we have

$$U(\mathbf{x}) = \Gamma\left(1/U_1(x_1), 1/U_2(x_2)\right), \qquad \mathbf{x} = (x_1,x_2) \in [0,\infty]^2 \setminus \{(0,0)\} \tag{2.4}$$

for some unique Pareto Lévy copula $\Gamma$, and we obtain the relation

$$\Gamma(\mathbf{u}) = U\left(U_1^{-1}(1/u_1), U_2^{-1}(1/u_2)\right), \qquad \mathbf{u} = (u_1, u_2) \in [0, \infty]^2 \setminus \{(0,0)\}. \tag{2.5}$$

The difference to the approach of Cont and Tankov (2004) is that the marginals of $\Gamma$ correspond to Pareto tails, which are the tail integrals of a 1-stable Lévy process on the positive half line. Since $\Gamma$ is 2-increasing as well, it is a simple task to deduce that it satisfies the properties of a tail integral of a spectrally positive Lévy process as claimed. Thus the Pareto Lévy copula allows for the interpretation that the marginals of $\nu$ are standardized to the Lévy measures of a 1-stable Lévy process, which is similar in spirit to the ordinary copula concept where marginals are standardized to uniform distributions.

Finally, we collect some basic properties of Pareto Lévy copulas, some of which already have been stated in Cont and Tankov (2004) and Kallsen and Tankov (2006) in the context of Lévy copulas.

**Proposition 2.3** *Every Pareto Lévy copula $\Gamma$ has the following properties.*

*(i)* **('Lipschitz continuity')** $|\Gamma(\mathbf{u}) - \Gamma(\mathbf{v})| \leq \left|\frac{1}{u_1} - \frac{1}{v_1}\right| + \left|\frac{1}{u_2} - \frac{1}{v_2}\right|$.

*(ii)* **(Monotonicity)** $\Gamma$ *is 2-increasing and the functions* $\Gamma(u, \cdot)$ *and* $\Gamma(\cdot, u)$ *are non-increasing for each fixed* $u \geq 0$.

*(iii)* **('Fréchet-Hoeffding bounds')** $\Gamma_\perp \leq \Gamma \leq \Gamma_\parallel$, *where* $\Gamma_\perp(\mathbf{u}) = u_1^{-1} 1_{\{u_2=0\}} + u_2^{-1} 1_{\{u_1=0\}}$ *and* $\Gamma_\parallel(\mathbf{u}) = (u_1 \vee u_2)^{-1}$ *denote the Pareto Lévy copulas corresponding to independence and to perfect positive dependence, respectively.*

*(iv)* **(Partial derivatives)** $\dot\Gamma_1(u_1, 0) = -u_1^{-2}$ *and* $\dot\Gamma_1(u_1, \infty) = 0$. *For fixed* $u_2 \in (0, \infty)$, *the partial derivative* $\dot\Gamma_1(u_1, u_2)$ *exists for almost all* $u_1 \in (0, \infty)$ *and for such* $u_1$ *and* $u_2$

$$0 \geq \dot\Gamma_1(u_1, u_2) \geq -u_1^{-2}.$$

*Similarly,* $\dot\Gamma_2(0, u_2) = -u_2^{-2}$, $\dot\Gamma(\infty, u_2) = 0$ *and for each* $u_1 \in (0, \infty)$ *the partial derivative* $\dot\Gamma_2(u_1, u_2)$ *exists for almost all* $u_2 \in (0, \infty)$ *with*

$$0 \geq \dot\Gamma_2(u_1, u_2) \geq -u_2^{-2}.$$

*Furthermore, the mappings* $u_2 \mapsto \dot\Gamma_1(u_1, u_2)$ *and* $u_1 \mapsto \dot\Gamma_1(u_1, u_2)$ *are defined and non-decreasing almost everywhere.*

**Proof.** Observing $\Gamma(\mathbf{u}) = C(1/u_1, 1/u_2)$ with the Lévy copula $C$ assertion (i) follows from Lemma 3.2 in Kallsen and Tankov (2006). Assertion (ii) follows from the fact that $\Gamma$ is 2-increasing and grounded by Definition 2.2. The lower bound in (iii) is obvious. By Theorem 5.1 in Kallsen and Tankov (2006) we have $\Gamma(\mathbf{u}) = \lim_{t\to 0} t^{-1} C_t(t/u_1, t/u_2)$ for some (ordinary) copulas $C_t : [0,1]^2 \to [0,1]$. It is well-known that every copula is bounded above by the Fréchet-Hoeffding bound $M(\mathbf{u}) = u_1 \wedge u_2$, whence setting $C_t = M$ for all $t$ yields assertion (iii). Regarding (iv) we only consider $\dot\Gamma_1$. The assertion is obvious for

$u_2 \in \{0, \infty\}$. Monotonicity of $u_1 \mapsto \Gamma(u_1, u_2)$ for each $u_2$ proves existence of $\dot{\Gamma}_1(u_1, u_2) \leq 0$ for almost all $u_1 \in (0, \infty)$ and all $u_2 \in (0, \infty)$. Moreover, for each such $u_1, u_2$ by Lipschitz continuity,

$$|\dot{\Gamma}_1(u_1, u_2)| = \lim_{t \to 0} \left| \frac{\Gamma(u_1 + t, u_2) - \Gamma(u_1, u_2)}{t} \right| \leq \lim_{t \to 0} \left| \frac{1/(u_1 + t) - 1/u_1}{t} \right| = \frac{1}{u_1^2}.$$

Finally, fix $v_2 \leq u_2$ and consider $u_1 \mapsto \Gamma(u_1, v_2) - \Gamma(u_1, u_2)$. This mapping is non-increasing according to part (ii), and hence its first derivative $\dot{\Gamma}_1(u_1, v_2) - \dot{\Gamma}_1(u_1, u_2)$ exists almost everywhere and is non-positive. This proves the final assertion. □

# 3 Estimation of bivariate tail integrals and Pareto Lévy copulas

In the following we are interested in the construction of an estimator $\hat{\Gamma}$ for $\Gamma$ which is based on relation (2.5) and empirical versions of the tail integrals $U$, $U_1$ and $U_2$. Such estimators have for instance been discussed in Figueroa-López (2008) in the univariate setting, and we will transfer them naturally to the bivariate case.

Before we introduce these empirical versions, is turns out to be convenient to change the domain of $U$ slightly. Since by assumption no negative jumps are involved, we have

$$\nu([x_1, \infty] \times [0, \infty]) = \nu([x_1, \infty] \times [-\infty, \infty])$$

for each $x_1 > 0$, and similarly for the second component. Therefore it is equally well possible to define $U$ in the same way as before, but as a function $U : \mathbb{H} \to \mathbb{R}$, where

$$\mathbb{H} = (0, \infty]^2 \cup (\{-\infty\} \times (0, \infty]) \cup ((0, \infty] \times \{-\infty\}).$$

Note that $U$ corresponds on the stripes through $-\infty$ to the marginal tail integrals $U_1$ and $U_2$, respectively.

Our estimator for the function $U$ will be defined on $\mathbb{H}$ as well, and precisely we set

$$U_n(\mathbf{x}) = \frac{1}{k_n} \sum_{j=1}^{n} 1_{\{\Delta_j^n X^{(1)} \geq x_1, \Delta_j^n X^{(2)} \geq x_2\}}, \qquad \mathbf{x} = (x_1, x_2), \tag{3.1}$$

where $k_n = n\Delta_n$ and $\Delta_j^n X^{(i)} = X^{(i)}_{j\Delta_n} - X^{(i)}_{(j-1)\Delta_n}$ denotes the $j$-th increment of $X^{(i)}$, $i = 1, 2$. Having the role of the stripes through $-\infty$ in mind, we obtain empirical versions of the univariate tail integrals through

$$U_{n,1}(x_1) = U_n(x_1, -\infty) = \frac{1}{k_n} \sum_{j=1}^{n} 1_{\{\Delta_j^n X^{(i)} \geq x_1\}}, \quad x_1 \in (0, \infty], \tag{3.2}$$

and analogously for $U_{n,2}$. Weak convergence of $U_n$ in an appropriate function space is established in Proposition 4.2 below.

The underlying idea behind $U_n$ is rather natural, given the interpretation of $U$ as the average number of jumps of a certain size during the unit interval. Stationarity and indepedence of increments of a Lévy process ensure that the same behaviour is to be expected over intervals of arbitrary size, as long as $U$ is standardized accordingly. Therefore a canonical idea is to count joint large increments of $X^{(1)}$ and $X^{(2)}$, as they indicate joint large jumps over the corresponding time interval, and this is precisely what $U_n$ does. Note that in order for $U_n$ to be consistent, it is necessary to be in the high-frequency setting with infinite time horizon, that is $k_n \to \infty$. On each fixed time interval $[0, T]$ there are only finitely many jumps larger than a given size, which is clearly not sufficient to draw inference on the entire distribution of the jumps.

In order to construct an empirical version of (2.5) a notion of a generalized inverse function is of importance. For any $f : (0, \infty] \to [0, \infty)$ which is monotonically decreasing, left-continuous and satisfies $f(\infty) = 0$ we define $f^- : (0, \infty] \to [0, \infty)$ via

$$f^-(z) = \inf\{x > 0 \mid f(x) \leq z\}. \tag{3.3}$$

**Definition 3.1** *Let $U$ be the tail integral of a bivariate Lévy process with positive jumps and $U_1$, $U_2$ be its marginal tail integrals. Using their empirical versions (3.1) and (3.2) we define the empirical Pareto Lévy copula as*

$$\hat{\Gamma}_n(\mathbf{u}) = U_n\left(\overline{U_{n,1}^-(1/u_1)}, \overline{U_{n,2}^-(1/u_2)}\right), \qquad \mathbf{u} = (u_1, u_2) \in [0, \infty]^2 \setminus \{(0, 0)\}, \tag{3.4}$$

*where $U_{n,i}^-$ is the generalized inverse function of $U_{n,i}$ as defined in (3.3), with the convention that $U_{n,i}^-(1/\infty) = U_{n,i}^-(0) = \infty$ and where $\overline{a} = a1_{\{a>0\}} - \infty 1_{\{a=0\}}$ for some $a \in [0, \infty]$. Finally, we set $U_n(-\infty, -\infty) = n/k_n$.*

**Remark 3.2** In order to understand why $\overline{a}$ has to be introduced, suppose that we are intested in estimating $\Gamma(u_1, 0)$ (even though it is known to take the value $1/u_1$). Our estimator becomes $U_n(U_{n,1}^-(1/u_1), -\infty)$ then, which is in general close to $1/u_1$ due to the definition of $U_{n,1}$. On the other hand, if we forget about $\overline{a}$, we obtain $U_n(U_{n,1}^-(1/u_1), 0)$ which only counts those increments of $\mathbf{X}$ where the first component exceeds $U_{n,1}^-(1/u_1)$ and the second one is non-negative. Due to the existence of a Brownian part in $\mathbf{X}$, however, we cannot expect these two estimators to be close, since a number of increments in the second component is indeed negative and thus this estimator is considerably small than $\hat{\Gamma}(u_1, 0)$. □

**Remark 3.3** In the general case of arbitrary jumps a similar construction allows the estimation of $\Gamma$ in the interior of each of the four quadrants separately. Indeed, Eder and Klüppelberg (2012) give a general notion of tail integrals and Pareto Lévy copulas in their Definition 4, and from Sklar's theorem in this context (which is their Theorem 1) we know that the same relation as (2.5) holds for $\mathbf{u} \in (\mathbb{R}\backslash\{0\})^2$ and determines $\Gamma$ uniquely. For the sake of brevity we dispense with the entire theory in this setting. □

# 4 Results on weak convergence

Our aim in this section is to prove a result on weak convergence of the estimator $\hat{\Gamma}_n$, but as a by-product we obtain such a claim for $U_n$ as well. Before we come to the main theorem, let us briefly resume our assumptions on $\nu$ which mostly have already been given in the previous paragraphs.

**Assumption 4.1** *Let* $\mathbf{X}$ *be a bivariate Lévy process with the representation (1.1). The following assumptions on* $\nu$ *are in order:*

*(i)* $\nu$ *has support on* $[0,\infty)^2 \backslash \{(0,0)\}$.

*(ii) On this set it takes the form* $\nu(d\mathbf{u}) = s(\mathbf{u})d\mathbf{u}$ *for a positive Lévy density* $s$ *which satisfies*

$$\sup_{u \in M_\eta} \left(|s(\mathbf{u})| + \|\nabla s(\mathbf{u})\|\right) < \infty$$

*for any* $\eta \in (0,\infty)^2$, *where*

$$M_\eta = (\eta,\infty)^2 \cup (\{0\} \times (\eta,\infty)) \cup ((\eta,\infty) \times \{0\})$$

*and* $\nabla s$ *denotes the gradient of* $s$ *on* $(\eta,\infty)^2$ *and the univariate derivative on the stripes through 0, respectively.*

*(iii)* $\nu$ *has infinite activity, that is* $\nu([0,\infty) \times [0,\infty)) = \infty$.

Assumption 4.1 (ii) had not been stated previously. It is used to prove a second order condition regarding the difference between $U$ and the expectation of $U_n$ for which we generalize a result due to Figueroa-López and Houdré (2009) from the univariate setting to the multidimensional case. Continuity and (strict) monotonicity of the marginal tail integrals as claimed before are obvious consequences of it.

We begin with a result on weak convergence of $U_n$, and to this end we have to define the function space on which the asymptotics take place. Let $\mathcal{B}_\infty(\mathbb{H})$ be the space of all functions $f : \mathbb{H} \to \mathbb{R}$ which are bounded on any subset of $\mathbb{H}$ that is bounded away from the origin and from the points $(-\infty, 0)$ and $(0,-\infty)$. We consider the metric inducing the topology of uniform convergence on those subsets, defined by

$$d(f,g) = \sum_{k=1}^{\infty} 2^{-k} \left(\|f-g\|_{T_k} \wedge 1\right),$$

where $T_k = [1/k,\infty]^2 \cup (\{-\infty\} \times [1/k,\infty]) \cup ([1/k,\infty] \times \{-\infty\})$ and $\|f\|_{T_k} = \sup_{\mathbf{u} \in T_k} |f(\mathbf{u})|$. This space is a complete metric space, and a sequence converges in $\mathcal{B}_\infty(\mathbb{H})$, if and only if it converges uniformly on each $T_k$.

**Proposition 4.2** *Assume that* $\mathbf{X}$ *is a Lévy process satisfying Assumption 4.1. If the observation scheme meets the conditions*

$$\Delta_n \to 0, \quad k_n \to \infty, \quad \sqrt{k}_n \Delta_n \to 0, \tag{4.1}$$

*then we have*

$$\gamma_n(\mathbf{x}) = \sqrt{k_n}\left\{U_n(\mathbf{x}) - U(\mathbf{x})\right\} \xrightarrow{w} \mathbb{B}(\mathbf{x})$$

*in* $(\mathcal{B}_\infty(\mathbb{H}), d)$*, where* $\mathbb{B}$ *is a tight, centered Gaussian process with covariance*

$$\mathbb{E}\left[\mathbb{B}(\mathbf{x})\mathbb{B}(\mathbf{y})\right] = U(\mathbf{x} \vee \mathbf{y}) = U(x_1 \vee y_1, x_1 \vee y_2).$$

*The sample paths of* $\mathbb{B}$ *are uniformly continuous on each* $T_k$ *with respect to the pseudo distance*

$$\rho(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[(\mathbb{B}(\mathbf{x}) - \mathbb{B}(\mathbf{y}))^2\right]^{1/2} = \{U(\mathbf{x}) + U(\mathbf{y}) - 2U(\mathbf{x} \vee \mathbf{y})\}^{1/2} = |U(\mathbf{x}) - U(\mathbf{y})|^{1/2}.$$

For the proof of Proposition 4.2 the following lemma is extremely useful. Its univariate version is a special case of a more general result in Figueroa-López and Houdré (2009).

**Lemma 4.3** *Suppose that Assumption 4.1 holds and let* $\delta > 0$ *be fixed. Then there exist constants* $K = K(\delta)$ *and* $t_0 = t_0(\delta)$ *such that the uniform bound*

$$\left|\mathbb{P}(X_t^{(1)} \geq x_1, X_t^{(2)} \geq x_2) - t\nu([x_1, \infty) \times [x_2, \infty))\right| < Kt^2$$

*holds for all* $\mathbf{x} = (x_1, x_2) \in [\delta, \infty]^2 \cup (\{-\infty\} \times [\delta, \infty]) \cup ([\delta, \infty] \times \{-\infty\})$ *and* $0 < t < t_0$*.*

Before we come to the result on $\hat{\Gamma}$, let us introduce an oracle estimator for $\Gamma$. We set

$$\widetilde{\Gamma}_n(\mathbf{u}) = U_n\left(U_1^{-1}(1/u_1), U_2^{-1}(1/u_2)\right), \qquad \mathbf{u} = (u_1, u_2) \in [0, \infty]^2 \setminus \{(0,0)\}, \tag{4.2}$$

which means that we replace the inverses of the empirical marginal tail integrals by the unobservable true ones. Thanks to Proposition 4.2 we obtain weak convergence of a restricted version of this intermediate estimator in the space $\mathcal{B}_\infty((0,\infty]^2)$ of all real functions on $(0,\infty]^2$ that are bounded on sets which are bounded away from the origin. In a similar sprit as before, we equip this space with the metric $d(f,g) = \sum_{k=1}^\infty 2^{-k}\left(\|f-g\|_{T_k} \wedge 1\right)$, where $T_k = [1/k, \infty]^2$. Setting $\mathbf{x} = (U_1^{-1}(1/u_1), U_2^{-1}(1/u_2))$ and observing that $U_i^{-1}(k) \geq k' > 0$, the continuous mapping theorem immediately yields the following result.

**Corollary 4.4** *Under the conditions of Proposition 4.2 we have*

$$\widetilde{\alpha}_n(\mathbf{u}) = \sqrt{k_n}\left(\widetilde{\Gamma}_n(\mathbf{u}) - \Gamma(\mathbf{u})\right) \xrightarrow{w} \mathbb{B}\left(U_1^{-1}(1/u_1), U_2^{-1}(1/u_2)\right)$$

*in* $(\mathcal{B}_\infty((0,\infty]^2), d)$ *with* $\mathbb{B}$ *as defined in Proposition 4.2.*

From a statistical point of view there is no loss in information when estimating $\Gamma(\mathbf{u})$ on $(0,\infty]^2$ instead of the entire domain $[0,\infty]^2 \setminus \{(0,0)\}$, since a Pareto Lévy copula is grounded by definition and thus known on stripes through 0. This remark remains valid for the final result of this section as well, which is on weak convergence of the estimator $\hat{\Gamma}_n(\mathbf{u})$.

**Theorem 4.5** *Assume that* $\mathbf{X}$ *is a Lévy process satisfying Assumption 4.1. If (4.1) holds, then we have*

$$\alpha_n(\mathbf{u}) = \sqrt{k_n}\left(\hat{\Gamma}_n(\mathbf{u}) - \Gamma(\mathbf{u})\right) \xrightarrow{w} \mathbb{G}(\mathbf{u})$$

*in* $(\mathcal{B}_\infty((0,\infty]^2), d)$*. Here the process* $\mathbb{G}$ *is defined as*

$$\mathbb{G}(\mathbf{u}) = \widetilde{\mathbb{G}}(\mathbf{u}) + u_1^2\, \dot{\Gamma}_1(\mathbf{u})\, \widetilde{\mathbb{G}}(u_1, -\infty) + u_2^2\, \dot{\Gamma}_2(\mathbf{u})\, \widetilde{\mathbb{G}}(-\infty, u_2), \tag{4.3}$$

*where* $\widetilde{\mathbb{G}}$ *denotes a tight centered Gaussian field on* $\mathbb{H}$ *with covariance structure*

$$\mathbb{E}\left[\widetilde{\mathbb{G}}(\mathbf{u})\widetilde{\mathbb{G}}(\mathbf{v})\right] = \Gamma(\mathbf{u} \vee \mathbf{v}) = \Gamma(u_1 \vee v_1, u_2 \vee v_2)$$

*using the convention* $\Gamma(u, -\infty) = \Gamma(-\infty, u) = 1/u$*. The sample paths of* $\widetilde{\mathbb{G}}$ *are uniformly continuous on each* $T_k$ *with respect to the pseudo distance*

$$\rho(\mathbf{u}, \mathbf{v}) = \mathbb{E}\left[\left(\widetilde{\mathbb{G}}(\mathbf{u}) - \widetilde{\mathbb{G}}(\mathbf{v})\right)^2\right]^{1/2} = |\Gamma(\mathbf{u}) - \Gamma(\mathbf{v})|^{1/2}.$$

If both coordinates of $\mathbf{u}$ are distinct from $\infty$, then $\dot{\Gamma}_i(\mathbf{u})$ exists as a consequence of (2.5) and Assumption 4.1, and $\mathbb{G}(\mathbf{u})$ is well-defined. On the other hand, if one of the components equals $\infty$, we have $\widetilde{\mathbb{G}}(\mathbf{u}) = 0$ almost surely; and also $\dot{\Gamma}_1(u_1, \infty) = 0$ and $\dot{\Gamma}_2(\infty, u_2) = 0$ from Proposition 2.3. Hence, the right hand side of (4.3) is well-defined as well, and we have $\mathbb{G}(\mathbf{u}) = 0$ almost surely in this case.

**Proof of Lemma 4.3.** For main parts the proof is almost similar to the one of the result in Figueroa-López and Houdré (2009) which is why we will only give the main steps and restrict ourselves to the genuine bivariate case of $x_1, x_2 \neq -\infty$. First, let $\varepsilon < (\delta/2 \wedge 1)$ and pick a smooth function $c_\varepsilon : \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$1_{[-\varepsilon/2, \varepsilon/2]}(\|\mathbf{u}\|) \leq c_\varepsilon(\mathbf{u}) \leq 1_{[-\varepsilon, \varepsilon]}(\|\mathbf{u}\|).$$

Here and throughout the proof, $\|\cdot\|$ denotes the Euklidian norm on $\mathbb{R}^2$. We also define the function $\bar{c}_\varepsilon$ via $\bar{c}_\varepsilon(\mathbf{u}) = 1 - c_\varepsilon(\mathbf{u})$. It is straightforward to see that there exist independent processes $\mathbf{X}^\varepsilon$ and $\widetilde{\mathbf{X}}^\varepsilon$ such that $\mathbf{X} \sim \mathbf{X}^\varepsilon + \widetilde{\mathbf{X}}^\varepsilon$ and where $\widetilde{\mathbf{X}}^\varepsilon$ is a compound Poisson process with intensity $\lambda_\varepsilon = \int \bar{c}_\varepsilon(\mathbf{u})\nu(d\mathbf{u})$ and jump distribution $f_\varepsilon(d\mathbf{u}) = \bar{c}_\varepsilon(\mathbf{u})\nu(d\mathbf{u})/\lambda_\varepsilon$ and $\mathbf{X}^\varepsilon$ is a Lévy process with triplet $(\mathbf{b}_\varepsilon, \Sigma, c_\varepsilon(\mathbf{u})\nu(d\mathbf{u}))$, where we set $\mathbf{b}_\varepsilon = \mathbf{b} - \int 1_{\{\|\mathbf{u}\| \leq 1\}} \mathbf{u}\bar{c}_\varepsilon(\mathbf{u})\nu(d\mathbf{u})$.

Since our result is a distributional one only, it is possible to work with this particular representation of $\mathbf{X}$ in the following. Call $N_t^\varepsilon$ the number of jumps of $\widetilde{\mathbf{X}}^\varepsilon$ up to time $t$.

Define $f(\mathbf{u}) = 1_{\{\mathbf{u} \geq \mathbf{x}\}}$ in a componentwise sense. Using the law of total expectation we then have

$$\mathbb{E}[f(\mathbf{X}_t)] = \sum_{k=0}^{\infty} e^{-\lambda_\varepsilon t} \frac{(\lambda_\varepsilon t)^k}{k!} \mathbb{E}[f(\mathbf{X}_t) | N_t^\varepsilon = k] = e^{-\lambda_\varepsilon t} \mathbb{E}[f(\mathbf{X}_t^\varepsilon)] + e^{-\lambda_\varepsilon t} \lambda_\varepsilon t \mathbb{E}[f(\mathbf{X}_t^\varepsilon + \xi_1)]$$
$$+ \sum_{k=2}^{\infty} e^{-\lambda_\varepsilon t} \frac{(\lambda_\varepsilon t)^k}{k!} \mathbb{E}\left[ f\left( \mathbf{X}_t^\varepsilon + \sum_{l=1}^{k} \xi_l \right) \right],$$

where the $\xi_l$ are i.i.d. $\sim f_\varepsilon$. As noted before, we may proceed similarly to Figueroa-López and Houdré (2009) now: Using their equation (3.3), the condition $\varepsilon < \delta/2$ ensures the existence of $K$ and $t_0$, both depending on $\delta$ only, such that

$$e^{-\lambda_\varepsilon t} \mathbb{E}[f(\mathbf{X}_t^\varepsilon)] \leq \mathbb{P}(X_t^{(\varepsilon,1)} \geq \delta) < K t^2$$

for all $0 < t < t_0$, where $X^{(\varepsilon,1)}$ denotes the first component of $\mathbf{X}^\varepsilon$. Also,

$$e^{-\lambda_\varepsilon t} \sum_{k=2}^{\infty} \frac{(\lambda_\varepsilon t)^k}{k!} < K t^2.$$

It therefore remains to focus on $\mathbb{E}[f(\mathbf{X}_t^\varepsilon + \xi_1)]$. The distribution of $\xi_1$ is $s(\mathbf{u}) \bar{c}_\varepsilon(\mathbf{u}) d\mathbf{u} / \lambda_\varepsilon$, and as a consequence of Assumption 4.1 (ii) it follows that

$$g(\mathbf{u}) = \mathbb{E}[f(\mathbf{u} + \xi_1)] = \mathbb{P}(u_1 + \xi_1^{(1)} \geq x_1, u_2 + \xi_1^{(2)} \geq x_2)$$

is twice continuously differentiable with bounded derivatives. Using independence of $\mathbf{X}^\varepsilon$ and $\xi_1$, it is sufficient to discuss $\mathbb{E}[g(\mathbf{X}_t^\varepsilon)]$, for which we can use Itô formula now: For arbitrary $\mathbf{Y}$ we have

$$g(\mathbf{Y}_t) = g(\mathbf{Y}_0) + \int_0^t \nabla g(\mathbf{Y}_{s-}) d\mathbf{Y}_s + \frac{1}{2} \sum_{1 \leq i,j \leq 2} g_{ij}(\mathbf{Y}_{s-}) d[Y^i, Y^j]_s^c$$
$$+ \sum_{0 < s \leq t} \left( g(\mathbf{Y}_s) - g(\mathbf{Y}_{s-}) - \nabla g(\mathbf{Y}_{s-}) \Delta \mathbf{Y}_s \right), \quad (4.4)$$

where the quadratic covariation $[Y^i, Y^j]_s^c$ becomes $\Sigma_{ij} s$ in case of a Lévy process and $\Delta \mathbf{Y}_s$ is the jump size at time $s$. Also, $\nabla g$ and $g_{ij}$ denote the gradient and the corresponding partial derivatives of $g$. Plugging in $\mathbf{X}^\varepsilon$ for $\mathbf{Y}$ we discuss each of the four summands above separately: first, $\mathbf{u} \geq \mathbf{x}$ implies $\|\mathbf{u}\| \geq \|\mathbf{x}\| \geq \delta > \varepsilon$, and thus

$$g(\mathbf{X}_0^\varepsilon) = g(0) = \mathbb{P}(\xi_1^{(1)} \geq x_1, \xi_1^{(2)} \geq x_2) = \frac{1}{\lambda_\varepsilon} \int 1_{\{\mathbf{u} \geq \mathbf{x}\}} s(\mathbf{u}) \bar{c}_\varepsilon(\mathbf{u}) d\mathbf{u}$$
$$= \frac{1}{\lambda_\varepsilon} \int 1_{\{\mathbf{u} \geq \mathbf{x}\}} s(\mathbf{u}) d\mathbf{u} = \frac{1}{\lambda_\varepsilon} \nu([x_1, \infty) \times [x_2, \infty)).$$

Second, the Lévy triplet of $\mathbf{X}^\varepsilon$ is $(\mathbf{b}_\varepsilon, \Sigma, c_\varepsilon(\mathbf{u}) \nu(d\mathbf{u}))$. From $\varepsilon < 1$ we conclude that $X^\varepsilon$ does not admit jumps larger than 1, and therefore $d\mathbf{X}_s^\varepsilon$ consists of three summands, of which two correspond to martingales. Therefore

$$\left| \mathbb{E}\left[ \int_0^t \nabla g(\mathbf{X}_{s-}^\varepsilon) d X_s^\varepsilon \right] \right| \leq \int_0^t \left| \mathbb{E}[\nabla g(\mathbf{X}_{s-}^\varepsilon)] \mathbf{b}_\varepsilon \right| ds < K t$$

due to boundedness of the first derivatives of $g$. We may proceed similarly for the third term in (4.4), whereas conditioning on $\mathbf{X}^\varepsilon_{s-}$ gives

$$\sum_{0<s\le t} \mathbb{E}\left[g(\mathbf{X}^\varepsilon_s) - g(\mathbf{X}^\varepsilon_{s-}) - \nabla g(\mathbf{X}^\varepsilon_{s-})\Delta\mathbf{X}^\varepsilon_s\right] =$$
$$\int_0^t \int \mathbb{E}[g(\mathbf{X}^\varepsilon_{s-} + \mathbf{u}) - g(\mathbf{X}^\varepsilon_{s-}) - \nabla g(\mathbf{X}^\varepsilon_{s-})\mathbf{u}]c_\varepsilon(\mathbf{u})\nu(d\mathbf{u})ds$$

for the final quantity. Multidimensional Taylor formula proves that the inner integrand above may be bounded by $K\|\mathbf{u}\|^2$. Since $\nu$ is a Lévy measure, we obtain

$$\left|\mathbb{E}[g(\mathbf{X}^\varepsilon_t)] - \frac{1}{\lambda_\varepsilon}\nu([x_1,\infty)\times[x_2,\infty))\right| < Kt.$$

From $|1-\exp(-\lambda_\varepsilon t)| < Kt$ for $0<t<t_0$ the conclusion follows. □

**Proof of Proposition 4.2.** Before we begin with the proof, note that due to Theorem 1.6.1 in van der Vaart and Wellner (2007) weak convergence in $\mathcal{B}_\infty((0,\infty]^2)$ is equivalent to weak convergence on each $\ell^\infty(T_k)$, which is the space of all bounded functions on $T_k$ endowed with the uniform norm. Therefore it is possible to fix one such $T_k$ throughout the rest of the proof.

Let us introduce some additional notation. We define a class of functions $\mathcal{F}_n = \{f_{n,\mathbf{x}} : \mathbf{x}\in T_k\}$ via

$$f_{n,\mathbf{x}}(\mathbf{p}) = \sqrt{n/k_n}\left(1_{\{\mathbf{p}\ge\mathbf{x}\ge(0,0)\}} + 1_{\{p_1\ge x_1, x_2=-\infty\}} + 1_{\{p_1\ge x_2, x_1=-\infty\}}\right).$$

Furthermore, we set

$$\overline{\gamma}_n(\mathbf{x}) = \sqrt{k_n}\left(U_n(\mathbf{x}) - \mathbb{E}[U_n(\mathbf{x})]\right) = n^{-1/2}\sum_{j=1}^n \left(f_{n,\mathbf{x}}(\Delta^n_j\mathbf{X}) - \mathbb{E}[f_{n,\mathbf{x}}(\Delta^n_j\mathbf{X})]\right).$$

A consequence of Lemma 4.3 is that it is sufficient to discuss weak convergence of $\overline{\gamma}_n(\mathbf{x})$ only. Indeed, let $\mathbf{x}\in T_k$. Then by stationarity of increments of $\mathbf{X}$ and using $k_n = n\Delta_n$ we have

$$\mathbb{E}[U_n(\mathbf{x})] - U(\mathbf{x}) = \Delta_n^{-1}\mathbb{P}\left(\Delta^n_1 X^{(1)} \ge x_1, \Delta^n_1 X^{(2)} \ge x_2\right) - \nu([x_1,\infty)\times[x_2,\infty)).$$

This quantity is bounded by $K\Delta_n$ due to Lemma 4.3, so the growth condition $\sqrt{k}_n\Delta_n \to 0$ ensures that $\sqrt{k_n}(\gamma_n(\mathbf{x}) - \overline{\gamma}_n(\mathbf{x}))$ is uniformly small on each fixed $T_k$.

In order to prove $\overline{\gamma}_n(\mathbf{x}) \xrightarrow{w} \mathbb{B}(\mathbf{x})$ on $\ell^\infty(T_k)$ we will employ Theorem 11.20 in Kosorok (2008) for which several intermediate results have to be shown. To begin with, set

$$F_n(\mathbf{p}) = \sqrt{n/k_n}1_{\{\mathbf{p}\in T_k\}},$$

which is a sequence of integrable (with respect to any probability measure) envelopes. The first two steps are related to the class of functions $\mathcal{F}_n$. We start with the proof of an entropy condition, namely

$$\limsup_{n\to\infty}\sup_Q \int_0^1 \sqrt{\log N(\varepsilon\|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q))}d\varepsilon < \infty, \tag{4.5}$$

where $N$ denotes the covering number of the set $\mathcal{F}_n$ and the supremum runs over all probability measures $Q$ with finite support such that $\|F_n\|_{Q,2} = \big(\int F_n^2(\mathbf{p})dQ(\mathbf{p})\big)^{1/2} > 0$. Thanks to the special form of $\mathcal{F}_n$, this result is a simple consequence of Lemma 11.21 in Kosorok (2008): it suffices to check that each $\mathcal{F}_n$ is a VC-class with VC-index 5. This follows from the fact that each finite subset of $\mathbb{H}$ of size 5 has either a subset of 3 elements in $[0,\infty]^2\backslash\{(0,0)\}$, or a subset of two elements in one of the stripes through $-\infty$. In neither of the cases theses subsets can be shattered by the sets deduced from the indicators in the definition of $f_{n,\mathbf{x}}$.

The second condition to check is that $\mathcal{F}_n$ is almost measurable Suslin, and it follows from Lemma 11.15 and the discussion on page 224 in Kosorok (2008) that it is sufficient to prove separability of $\mathcal{F}_n$, that is the existence of a countable subset $T_{n,k}$ of $T_k$ such that

$$\mathbb{P}^*\big(\sup_{\mathbf{x}\in T_k}\inf_{\mathbf{y}\in T_{n,k}}\big|f_{n,\mathbf{x}}(\Delta_j^n\mathbf{X}) - f_{n,\mathbf{y}}(\Delta_j^n\mathbf{X})\big| > 0\big) = 0.$$

Here, $\mathbb{P}^*$ denotes the outer expectation, since measurability of the event within the brackets is not ensured. Set $T_{n,k} = T_k \cap \overline{\mathbb{Q}}^2$. Then for each $\omega$ and each $\mathbf{x} \in T_k$, there exists a $\mathbf{y} \in T_{n,k}$ such that $f_{n,\mathbf{x}}(\Delta_j^n\mathbf{X}(\omega)) = f_{n,\mathbf{y}}(\Delta_j^n\mathbf{X}(\omega))$, since the $f_{n,\mathbf{x}}$ are indicator functions. This proves separability of $\mathcal{F}_n$.

The remaining steps regard the behaviour of the variances and covariances of the $f_{n,\mathbf{x}}$ and their envelopes. We have

$$\lim_{n\to\infty}\mathbb{E}[\overline{\gamma}_n(\mathbf{x})\overline{\gamma}_n(\mathbf{y})] = \lim_{n\to\infty}\mathbb{E}[f_{n,\mathbf{x}}(\Delta_j^n\mathbf{X})f_{n,\mathbf{y}}(\Delta_j^n\mathbf{X})] = U(\mathbf{x}\vee\mathbf{y}) \tag{4.6}$$

as well as

$$\lim_{n\to\infty}\mathbb{E}[F_n^2(\Delta_j^n\mathbf{X})] \le U(1/k,-\infty) + U(-\infty,1/k)$$

and

$$\lim_{n\to\infty}\mathbb{E}[F_n^2(\Delta_j^n\mathbf{X})1_{\{F_n(\Delta_j^n\mathbf{X})>\varepsilon\sqrt{n}\}}] \le \lim_{n\to\infty}\mathbb{E}[F_n^2(\Delta_j^n\mathbf{X})]\big(\varepsilon\sqrt{k_n}\big)^{-1} \to 0.$$

Finally, as in (4.6) we have for $\mathbf{x},\mathbf{y}\in T_k$ that

$$\begin{aligned}\rho_n(\mathbf{x},\mathbf{y}) &= \Big(\mathbb{E}\Big[\big(f_{n,\mathbf{x}}(\Delta_j^n\mathbf{X}) - f_{n,\mathbf{y}}(\Delta_j^n\mathbf{X})\big)^2\Big]\Big)^{1/2}\\ &\to (U(\mathbf{x}) + U(\mathbf{y}) - 2U(\mathbf{x}\vee\mathbf{y}))^{1/2} = \rho(\mathbf{x},\mathbf{y}),\end{aligned}$$

and due to Lemma 4.3 the convergence holds uniformly as well. This completes the proof. □

**Proof of Theorem 4.5.** Let $\mathcal{B}_\infty^0((0,\infty]^2) \subset \mathcal{B}_\infty((0,\infty]^2)$ and $\mathcal{B}_\infty^0((0,\infty]) \subset \mathcal{B}_\infty((0,\infty])$ denote the space of all tail integrals of bivariate Lévy measures concentrated on the first quadrant or of univariate Lévy measures concentrated on $(0,\infty]$, respectively. Consider the mapping $\Phi : \mathcal{B}_\infty^0((0,\infty]^2) \times (\mathcal{B}_\infty^0((0,\infty]))^2 \to \mathcal{B}_\infty((0,\infty]^2)$, defined by

$\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$ with

$$\begin{aligned}
\Phi_1 :& \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^0_\infty((0,\infty]))^2 \to \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^-_\infty((0,\infty]))^2 \\
& (U, U_1, U_2) \mapsto (U, U_1^-, U_2^-) \\
\Phi_2 :& \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^-_\infty((0,\infty]))^2 \to \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^p_\infty([0,\infty)))^2 \\
& (U, V_1, V_2) \mapsto (U, V_1 \circ P, V_2 \circ P) \\
\Phi_3 :& \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^p_\infty([0,\infty)))^2 \to \mathcal{B}_\infty((0,\infty]^2) \\
& (U, G_1, G_2) \mapsto U(G_1, G_2),
\end{aligned}$$

where $P(x) = 1/x$ and where, in the last step, $G_i(\infty) = \infty$. Moreover, $\mathcal{B}^-_\infty((0,\infty]) \subset \mathcal{B}_\infty((0,\infty])$ and $\mathcal{B}^p_\infty([0,\infty)) \subset \mathcal{B}_\infty([0,\infty))$ are defined as the images of the associated function spaces under the respective mappings. Set also $\widetilde{\Gamma}_{n,1}(x) = U_n(U_1(1/x), -\infty)$ and $\widetilde{\Gamma}_{n,2}(x) = U_n(-\infty, U_2(1/x))$. The proof will now basically consist of two steps. We start with discussing weak convergence of

$$\sqrt{k_n}\left(\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - \Phi(\Gamma, P, P)\right) \xrightarrow{w} \mathbb{G}, \tag{4.7}$$

whereas this result is transferred to the original claim later on.

Let us begin with the proof of (4.7). This assertion follows from the functional delta method in topological vector spaces, see van der Vaart and Wellner (1996), if we prove first that

$$\sqrt{k_n}\left\{(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - (\Gamma, P, P)\right\} \xrightarrow{w} (\widetilde{\mathbb{G}}, \widetilde{\mathbb{G}}(\cdot, -\infty), \widetilde{\mathbb{G}}(-\infty, \cdot))$$

in $\mathcal{B}_\infty((0,\infty]^2) \times (\mathcal{B}_\infty((0,\infty]))^2$ and second that $\Phi$ is Hadamard-differentiable at $(\Gamma, P, P)$ tangentially to suitable subspaces with derivative

$$\left(\Phi'_{(\Gamma,P,P)}(U, U_1, U_2)\right)(\mathbf{u}) = U(\mathbf{u}) + u_1^2\, \dot{\Gamma}_1(\mathbf{u})\, U_1(u_1) + u_2^2\, \dot{\Gamma}_2(\mathbf{u})\, U_2(u_2), \tag{4.8}$$

where the summands involving the partial derivatives on the right-hand side are defined as 0 if one of the coordinates of $\mathbf{u}$ equals $\infty$. The first claim follows easily from Proposition 4.2 and the continuous mapping theorem. Regarding the second assertion we need to clarify the metrics on the corresponding spaces. The canonical definitions are

$$d(f,g) = \sum_{k=1}^\infty 2^{-k} \left(\|f-g\|_{T_k} \wedge 1\right),$$

where $T_k = [1/k, \infty]^2$ in case of $\mathcal{B}_\infty((0,\infty]^2)$, while $T_k = [1/k, \infty]$ and $T_k = [0,k]$ for $\mathcal{B}_\infty((0,\infty])$ and $\mathcal{B}_\infty([0,\infty))$, respectively. Unfortunately, the mapping $\Phi_1$ is not Hadamard-differentiable with respect to these metrics (see the proof of Lemma 7.2 below for details), whence we need to consider the weaker modifications

$$d_2(f,g) = \sum_{k=1}^\infty 2^{-k} \left(\|f-g\|_{S_k} \wedge 1\right),$$

where $S_k = ([1/k, k] \cup \{\infty\})^2$ in case of $\mathcal{B}_\infty((0,\infty]^2)$, while $S_k = [1/k, k] \cup \{\infty\}$ and $S_k = \{0\} \cup [1/k, k]$ for $\mathcal{B}_\infty((0,\infty])$ and $\mathcal{B}_\infty([0,\infty))$, respectively. With these modifications, it follows from Lemma 7.1 and the chain rule that

$$\Phi : (\mathcal{B}^0_\infty((0,\infty]^2, d) \times (\mathcal{B}^0_\infty((0,\infty]), d)^2 \to (\mathcal{B}^0_\infty((0,\infty]^2, d_2)$$

is Hadamard-differentiable at $(\Gamma, P, P)$ with derivative as specified in (4.8) tangentially to

$$\mathbb{D}_0 = \{(U, U_1, U_2) \in \mathcal{C}((0,\infty]^2) \times (\mathcal{C}((0,\infty]))^2 \mid U_j(\infty) = 0, \lim_{x\to 0} x^2 U_j(x) = 0\}. \tag{4.9}$$

Here, $\mathcal{C}((0,\infty]^2)$ and $C((0,\infty])$ denote the set of all functions on $(0,\infty]^2$ and $(0,\infty]$ that are continuous with respect to the pseudo metrics $\rho(\mathbf{u}, \mathbf{v}) = |\Gamma(\mathbf{u}) - \Gamma(\mathbf{v})|^{1/2}$ and $\rho(u, v) = |1/u - 1/v|^{1/2}$, respectively. Hence, observing $(\widetilde{\mathbb{G}}, \widetilde{\mathbb{G}}(\cdot, -\infty), \widetilde{\mathbb{G}}(-\infty, \cdot)) \in \mathbb{D}_0$, the functional delta method yields

$$\sqrt{k_n}\left(\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - \Phi(\Gamma, P, P)\right) \xrightarrow{w} \mathbb{G}$$

in $(\mathcal{B}_\infty((0,\infty]^2), d_2)$.

We will use the approximation Theorem 4.2 in Billingsley (1968), adapted to the concept of weak convergence in the sense of Hoffmann-Jørgensen, to transfer this result to weak convergence in $(\ell^\infty([\eta,\infty]^2), \|\cdot\|_\infty)$ for all $\eta > 0$ and hence in $(\mathcal{B}_\infty((0,\infty]^2), d)$. To this end, define

$$W_n(\mathbf{u}) = \sqrt{k_n}(\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - \Phi(\Gamma, P, P))(\mathbf{u})$$

and

$$W_{n,M}(\mathbf{u}) = \sqrt{k_n}(\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - \Phi(\Gamma, P, P))(\mathbf{u}) 1_{\{\mathbf{u} \in [\eta, M]^2\}}.$$

Then $W_{n,M}(\mathbf{u}) \xrightarrow{w} \mathbb{G}_M(\mathbf{u}) := \mathbb{G}(\mathbf{u}) 1_{\{\mathbf{u}\in[\eta,M]^2\}}$ for $n \to \infty$ and $\mathbb{G}_M(\mathbf{u}) \xrightarrow{w} \mathbb{G}(\mathbf{u})$ for $M \to \infty$ in $(\ell^\infty([\eta,\infty])^2, \|\cdot\|_\infty)$, and it remains to prove that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\left(\sup_{u_1 > M \text{ or } u_2 > M} |\sqrt{k_n}(\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2}) - \Phi(\Gamma, P, P))(\mathbf{u})| > \varepsilon\right) = 0.$$

Noting that $\Phi(\Gamma, P, P) = \Gamma$ the probability can be bounded by

$$\mathbb{P}^*\left(\sup_{u_1 \geq M/2 \text{ or } u_2 \geq M/2} |\sqrt{k_n}(\widetilde{\Gamma}_n - \Gamma)(\mathbf{u})| > \varepsilon\right)$$
$$+ \mathbb{P}^*\left(\exists \mathbf{u} \text{ with } u_1 > M \text{ or } u_2 > M : \widetilde{\Gamma}^-_{ni} \circ P(u_i) < M/2, i = 1, 2\right).$$

The Portmanteau Theorem implies that the lim sup of the first probability converges to 0 for $M \to \infty$ using Proposition 4.2. Furthermore, some thoughts reveal that $\widetilde{\Gamma}^-_{n,i}(z) = 1/(U_i(U^-_{n,i}(z)))$ for all $z > 0$. Due to monotonicity of $\widetilde{\Gamma}^-_{ni} \circ P$, the second probability is bounded by $\mathbb{P}(\widetilde{\Gamma}^-_{ni} \circ P(M) \leq M/2, i = 1, 2)$, which thus converges to 0 for $n \to \infty$ observing that $\widetilde{\Gamma}^-_{ni} \circ P(M) = M + o_P(1)$.

In the final step we will prove $\sqrt{k_n}(\hat{\Gamma}_n - \Gamma) \xrightarrow{w} \mathbb{G}$ in each $(\ell^\infty([\eta,\infty]^2), \|\cdot\|_\infty)$, for which we heavily rely on the fact that the same result holds for the statistic discussed above. A consequence of the identity $\widetilde{\Gamma}^{-}_{n,i}(z) = 1/(U_i(U^{-}_{n,i}(z)))$ is that $\Phi(\widetilde{\Gamma}_n, \widetilde{\Gamma}_{n,1}, \widetilde{\Gamma}_{n,2})(\mathbf{u})$ and $\hat{\Gamma}_n(\mathbf{u})$ coincide as long as $U^{-}_{n,i}(1/u_i) \neq 0$ for $i = 1, 2$. By monotonicity it is therefore sufficient to prove that the probability of $U^{-}_{n,i}(1/\eta) = 0$ becomes small, which is precisely

$$\lim_{n\to\infty} \mathbb{P}\left(U^{-}_{n,i}(1/\eta) = 0\right) = 0.$$

To this end, let $N_i(n)$ denote the number of positive increments of $X^{(i)}$. By definition of the generalized inverse function in (3.3) we have that $U^{-}_{n,i}(1/\eta) = 0$ is equivalent to $1/\eta \geq N_i(n)/k_n$ or $N_i(n) \leq k_n/\eta$. Furthermore, letting $M_i(n)$ be the number of positive increments of the process $Z^{(i)}_t = a_i t + B^{(i)}_t$, we see that it is sufficient to prove

$$\lim_{n\to\infty} \mathbb{P}\left(M_i(n) \leq k_n/\eta\right) = 0,$$

since $\mathbf{X}$ does not admit negative jumps. Note that we have

$$\mathbb{P}\left(\Delta^n_j Z^{(i)} > 0\right) = \mathbb{P}\left(\Delta^n_j B^{(i)} > -a_i \Delta_n\right) = \mathbb{P}\left(N > -a_i \Delta_n^{1/2}\right) = \frac{1}{2} + o(1),$$

where $N$ is a standard Gaussian variable. Let $n$ be large enough in order for the probability above to be larger than 1/3. For such $n$ we conclude easily that

$$\mathbb{P}\left(M_i(n) \leq k_n/\eta\right) \leq \mathbb{P}\left(\mathrm{Bin}(n, 1/3) \leq k_n/\eta\right) \to 0,$$

e.g., from Markov inequality and (4.1). This finishes the proof. □

# 5 Discussion and simulations

## 5.1 An asymptotic comparison

Suppose a statistician has knowledge of the marginal tail integrals. In this case, the results in Section 4 provide two competitive asymptotically unbiased estimators for the Pareto Lévy copula, namely the oracle estimator $\widetilde{\Gamma}_n$ exploiting knowledge of the marginals and the empirical Pareto Lévy copula $\hat{\Gamma}_n$ ignoring this additional information. The following proposition gives a partial answer to the question of which estimator is (asymptotically) preferable. Perhaps surprisingly, ignoring the additional knowledge decreases the asymptotic variance under certain growth conditions on $\Gamma$. A similar observation has recently been made in the context of copula estimation, see Genest and Segers (2010).

**Proposition 5.1** *Suppose that the Pareto Lévy copula $\Gamma$ has continuous first order partial derivatives and that the functions*

$$u_1 \mapsto u_1\,\Gamma(u_1,u_2) = \frac{\Gamma(u_1,u_2)}{\Gamma(u_1,0)}, \qquad u_2 \mapsto u_2\,\Gamma(u_1,u_2) = \frac{\Gamma(u_1,u_2)}{\Gamma(0,u_2)} \tag{5.1}$$

*are non-decreasing for fixed* $u_2 \in (0, \infty]$ *and* $u_1 \in (0, \infty]$*, respectively. Then the Gaussian fields* $\mathbb{G}$ *and* $\widetilde{\mathbb{G}}$ *satisfy the inequality*

$$\mathrm{Cov}\{\mathbb{G}(\mathbf{u}), \mathbb{G}(\mathbf{v})\} \leq \mathrm{Cov}\{\widetilde{\mathbb{G}}(\mathbf{u}), \widetilde{\mathbb{G}}(\mathbf{v})\}$$

*for all* $\mathbf{u}, \mathbf{v} \in (0, \infty]^2$*. Particularly,* $\mathrm{Var}\{\mathbb{G}(\mathbf{u})\} \leq \mathrm{Var}\{\widetilde{\mathbb{G}}(\mathbf{u})\}$.

**Proof.** The proof is rather straightforward whence we restrict ourselves to the main idea. We have $\mathrm{Cov}(\mathbb{G}(\mathbf{u}), \mathbb{G}(\mathbf{v})) - \mathrm{Cov}(\widetilde{\mathbb{G}}(\mathbf{u}), \widetilde{\mathbb{G}}(\mathbf{v})) = \sum_{i=1}^{8} A_i$ where $A_i = A_i(\mathbf{u}, \mathbf{v})$ is defined as

$$\begin{aligned}
A_1 &= u_1^2 \dot{\Gamma}_1(\mathbf{u})\, v_1^2 \dot{\Gamma}_1(\mathbf{v})\, 1/(u_1 \vee v_1) & \qquad A_5 &= u_1^2 \dot{\Gamma}_1(\mathbf{u})\, \Gamma(u_1 \vee v_1, v_2) \\
A_2 &= u_1^2 \dot{\Gamma}_1(\mathbf{u})\, v_2^2 \dot{\Gamma}_2(\mathbf{v})\, \Gamma(u_1, v_2) & A_6 &= u_2^2 \dot{\Gamma}_2(\mathbf{u})\, \Gamma(v_1, u_2 \vee v_2) \\
A_3 &= u_2^2 \dot{\Gamma}_2(\mathbf{u})\, v_1^2 \dot{\Gamma}_1(\mathbf{v})\, \Gamma(v_1, u_2) & A_7 &= v_1^2 \dot{\Gamma}_1(\mathbf{v})\, \Gamma(u_1 \vee v_1, u_2) \\
A_4 &= u_2^2 \dot{\Gamma}_2(\mathbf{u})\, v_2^2 \dot{\Gamma}_2(\mathbf{v})\, 1/(u_2 \vee v_2) & A_8 &= v_2^2 \dot{\Gamma}_2(\mathbf{v})\, \Gamma(u_1, u_2 \vee v_2).
\end{aligned}$$

The four summands on the left-hand side are non-negative, whereas the other four ones are non-positive. For symmetry reasons we may suppose $u_1 \leq v_1$. Distinguishing the two cases $u_2 \leq v_2$ and $u_2 > v_2$ some easy calculations (which frequently exploit condition (5.1)) show that $A_5 + A_1, A_6 + A_4, A_7 + A_3, A_8 + A_2 \leq 0$ in the first case, while $A_5 + A_2, A_6 + A_3, A_7 + A_1, A_8 + A_4 \leq 0$ in the second case. $\square$

Under the assumptions of Proposition 5.1 the condition in (5.1) is equivalent to

$$u_1\, \dot{\Gamma}_1(\mathbf{u}) + \Gamma(\mathbf{u}) \geq 0, \quad u_2\, \dot{\Gamma}_2(\mathbf{u}) + \Gamma(\mathbf{u}) \geq 0$$

for each $\mathbf{u} = (u_1, u_2) \in (0, \infty]^2$, which is easily accessible for most parametric classes of Pareto Lévy copulas. For instance, for the Clayton Pareto Lévy copula given by

$$\Gamma(\mathbf{u}) = (u_1^\theta + u_2^\theta)^{-1/\theta}$$

we have

$$u_1\, \dot{\Gamma}_1(\mathbf{u}) + \Gamma(\mathbf{u}) = (u_1^\theta + u_2^\theta)^{-1/\theta - 1} u_2^\theta, \qquad u_2\, \dot{\Gamma}_2(\mathbf{u}) + \Gamma(\mathbf{u}) = (u_1^\theta + u_2^\theta)^{-1/\theta - 1} u_1^\theta$$

which is readily seen to be non-negative. In Figure 1 we depict the graph of the asymptotic relative efficiency

$$[0, 2]^2 \to [0, \infty), \mathbf{u} \mapsto \frac{\mathrm{Var}\{\mathbb{G}(\mathbf{u})\}}{\mathrm{Var}\{\widetilde{\mathbb{G}}(\mathbf{u})\}}$$

of the oracle estimator $\widetilde{\Gamma}_n$ to the empirical Pareto Lévy copula $\hat{\Gamma}_n$ for $\mathbf{u} \in [0, 2]^2$. The Clayton parameter is chosen as $\theta = 0.5$. Close to the axis the relative efficiency decreases to 0, while the maximal relative efficiency is attained on the diagonal with a value of $21/32 \approx 0.656$. Even in this best case, the difference is seen to be substantial.

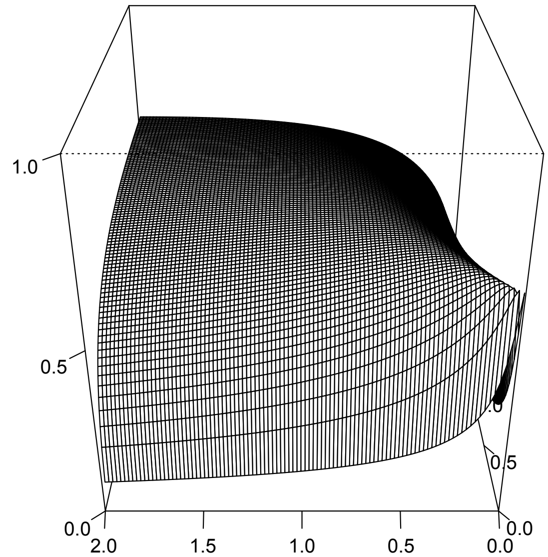


Figure 1: *The graph of the asymptotic relative efficiency of* $\widetilde{\Gamma}_n$ *to* $\hat{\Gamma}_n$ *for the Clayton Pareto Lévy copula with* $\theta = 0.5$.

## 5.2 Simulation study

In order to obtain an impression on the performance of the asymptotic results stated in the previous section we will discuss some finite sample properties concerning Proposition 4.2 and Theorem 4.5. In both cases, the setting is as follows: We simulate (essentially) two 1/2 stable subordinators, i.e., both tail integrals are given by $U_i(x) = (\pi x)^{-1/2}$, which are coupled by a Clayton Pareto Lévy copula with $\theta = 1/2$. Sometimes we add two independent Brownian motions with variance 1/2 each, sometimes we assume to observe the pure jump processes only. Throughout the study we use $n = 22,500$ observations and run the simulation 500 times each.

What differs from setting to setting is the choice of $k_n$, or, equivalently, of $\Delta_n$. Recall that the rate of convergence is $k_n^{-1/2}$ (which in light of the results in Figueroa-López and Houdré (2009) appears to be a natural one in the context of estimating the Lévy measure). Hence, a larger $k_n$ suggests a better performance of the normal approximation, whereas Lemma 4.3 indicates that the magnitude of the bias grows with $k_n$ as well. Both intuitive properties are visible from the simulation study provided in the following and from additional results which we do not show for the sake of brevity.

Despite the fact that we have proven weak convergence of our estimators in certain function spaces we restrict ourselves to an analysis of the finite dimensional properties of our estimators. Let us begin with the asympotics in Proposition 4.2 for which we estimate $U(x, x)$ for $x = 2, 1, 0.5$. Table 1 gives estimated bias and (co)variance for different choices of $k_n$. Note that we have $\mathrm{Cov}(\mathbb{B}(\mathbf{x}), \mathbb{B}(\mathbf{y})) = (32\pi)^{-1/2} \approx 0.0997$ whenever $\mathbf{x}$ or $\mathbf{y}$ equals $(2, 2)$, whereas $\mathrm{Cov}(\mathbb{B}(\mathbf{x}), \mathbb{B}(\mathbf{y})) = (16\pi)^{-1/2} \approx 0.1410$ if the "larger" vector is $(1, 1)$ and

| $x, y$ | 2,2 | | 1,1 | | 0.5,0.5 | | 2,0.5 | 2,1 | 1,0.5 |
|---|---|---|---|---|---|---|---|---|---|
| $k_n$ | bias | var | bias | var | bias | var | cov | cov | cov |
| 50 | -0.0106 | 0.1007 | -0.0077 | 0.1400 | 0.0023 | 0.1915 | 0.0988 | 0.0978 | 0.1376 |
| 75 | -0.0330 | 0.0972 | -0.0229 | 0.1453 | -0.0395 | 0.1956 | 0.1015 | 0.1001 | 0.1435 |
| 100 | 0.0168 | 0.1021 | 0.0223 | 0.1375 | 0.0341 | 0.1893 | 0.0996 | 0.0927 | 0.1300 |
| 150 | 0.0037 | 0.1061 | 0.0154 | 0.1480 | 0.0470 | 0.2180 | 0.1073 | 0.1106 | 0.1531 |
| 50 | -0.0281 | 0.0893 | -0.0120 | 0.1208 | -0.0042 | 0.1863 | 0.0840 | 0.0854 | 0.1233 |
| 75 | 0.0252 | 0.0949 | 0.0115 | 0.1187 | 0.0226 | 0.1861 | 0.0861 | 0.0894 | 0.1216 |
| 100 | 0.0126 | 0.0922 | 0.0043 | 0.1320 | 0.0401 | 0.1940 | 0.0933 | 0.0932 | 0.1323 |
| 150 | -0.0085 | 0.0929 | -0.0127 | 0.1337 | 0.0277 | 0.1991 | 0.0931 | 0.0962 | 0.1371 |

Table 1: *Empirical bias and (co)variances of* $\sqrt{k_n}(U_n(\mathbf{x}) - U(\mathbf{x}))$ *for various choices of* $k_n$. *Upper four lines: Pure subordinator; lower four lines: Subordinator + Brownian Motion.*

finally $\mathrm{Var}(\mathbb{B}(\mathbf{x})) = (8\pi)^{-1/2} \approx 0.1995$ for $\mathbf{x} = (0.5, 0.5)$.

Generally, the theoretical (co)variances are well reproduced in both situations, even though the results look probably a bit better in the first four lines. This is of course no surprise, since additional Brownian increments make it harder to infer on the jump measure. In order to assess how well the normal approximation works apart from bias and variance, Figure 2 gives QQ-plots for the medium choice of $k_n = 75$. These plots confirm that the finite sample properties are indeed satisfying, despite the discrete nature of the test statistic which simply counts exceedances of certain levels and is rescaled afterwards.

Let us come to the estimation of the Pareto Lévy copula. We proceed in the same way as before and discuss convergence of the finite dimensional distributions only. For simplicity, we estimate $\Gamma(x, x)$ for $x = 2, 1, 0.5$ again, but these are of course different quantities now. In this case, the variances compute to $\mathrm{Var}(\mathbb{G}(\mathbf{x})) = 21/(128x)$, which becomes approximately 0.0820 for $x = 2$, 0.1641 for $x = 1$, and 0.3281 for $x = 0.5$. Also, for $x > y$ we have $\mathrm{Cov}(\mathbb{G}(\mathbf{x}), \mathbb{G}(\mathbf{y})) = 7/32(1/x - \Gamma(x, y))$. Therefore $\mathrm{Cov}(\mathbb{G}(\mathbf{2}), \mathbb{G}(\mathbf{0.5})) \approx 0.0608$, $\mathrm{Cov}(\mathbb{G}(\mathbf{2}), \mathbb{G}(\mathbf{1})) \approx 0.0718$, and $\mathrm{Cov}(\mathbb{G}(\mathbf{1}), \mathbb{G}(\mathbf{0.5})) \approx 0.1437$. We state their empirical versions in Table 2.

In this case the growth in bias for larger $k_n$ is clearly visible, and we also have a larger bias when estimating $\Gamma(0.5, 0.5)$. Overall, however, the results are satisfying again, and we see from the QQ-plot in Figure 3 that the normal approximation works very well for $k_n = 75$, no matter if a Brownian motion is added or not.

## 6 Conclusions

In this paper we have investigated the problem of estimating both the bivariate Lévy measure and the (Pareto) Lévy copula in a nonparametric way. Our estimators are based on counting joint large increments of a bivariate Lévy process, and in both cases we were able to prove weak convergence in appropriate function spaces. At least two natural extensions of our work are of interest for future research.

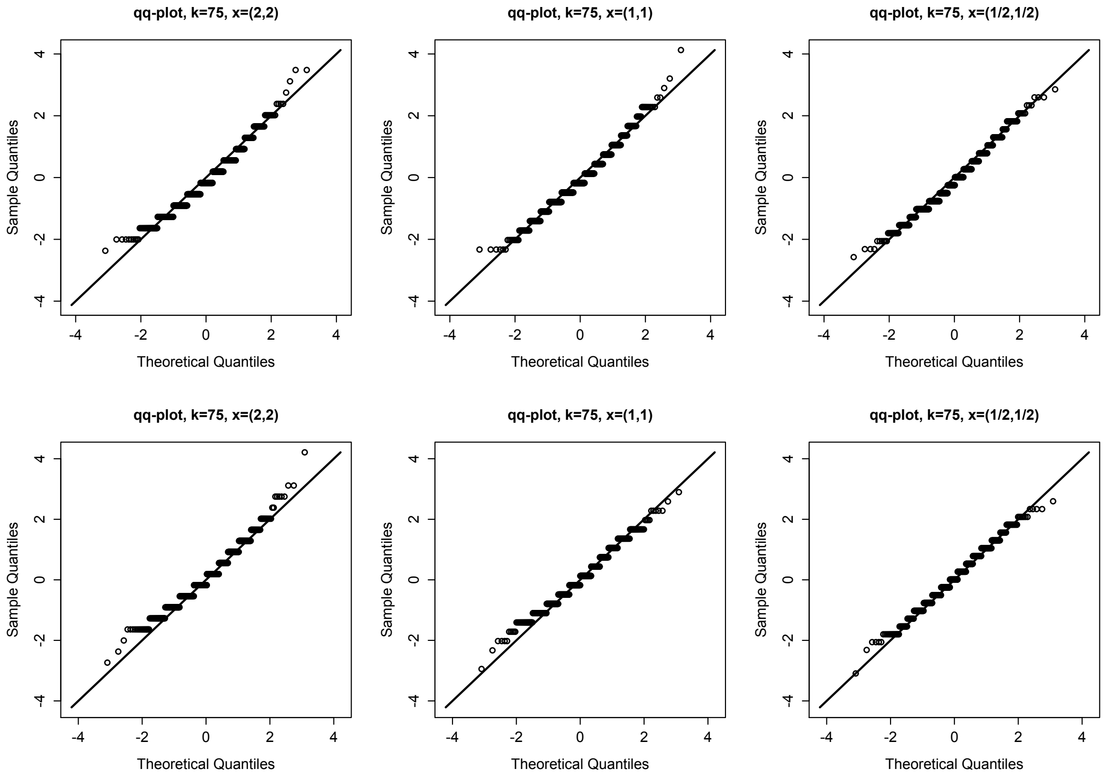


Figure 2: *QQ-plots of the empirical quantiles of* $\sqrt{k_n}(U_n(\mathbf{x})-U(\mathbf{x}))$ *divided by their sample standard deviation vs. the theoretical quantiles of the standard normal distribution. Upper three pictures: Pure subordinator; lower three pictures: Subordinator + Brownian Motion.*

First, on the observational side several robustness issues could be discussed: The prime question in this context is: How realistic are observations of a bivariate Lévy process at synchronous times and equally spaced, if we are faced with real data? The simplest extension probably is to introduce estimators in case where the observation intervals are not of equal size. Then we stay in the context of independent increments, for which theory of weak convergence is established as well; see e.g. Kosorok (2008). If both univariate processes are observed at different times, the situation is less clear. It might be promising to follow the approach due to Hayashi and Yoshida (2005) for diffusion processes then, but mathematics appear to be tough. Finally, one could move to bivariate Itô semimartingales for which both the Brownian part and the Lévy measure depend on a time index and estimate local versions of $\nu$ and related quantities.

From a statistical point of view it might be interesting to construct several nonparametric tests concerning the dependence structure of a multivariate Lévy process. This could include estimation of certain functionals of $\Gamma$ or $U$ as well as tests for independence or tests for a parametric form of these functions. For this reason, it would be important to establish a thorough theory concerning (Pareto) Lévy copulas which relates functionals of

| $x, y$ | 2,2 | | 1,1 | | 0.5,0.5 | | 2,0.5 | 2,1 | 1,0.5 |
|---|---|---|---|---|---|---|---|---|---|
| $k_n$ | bias | var | bias | var | bias | var | cov | cov | cov |
| 50 | 0.0141 | 0.0827 | 0.0455 | 0.1740 | 0.0863 | 0.3520 | 0.0777 | 0.0668 | 0.1599 |
| 75 | -0.0082 | 0.0874 | 0.0173 | 0.1653 | 0.1252 | 0.3428 | 0.0740 | 0.0690 | 0.1459 |
| 100 | 0.0502 | 0.0783 | 0.0894 | 0.1708 | 0.1748 | 0.3400 | 0.0685 | 0.0508 | 0.1547 |
| 150 | 0.0356 | 0.0862 | 0.1182 | 0.1646 | 0.3176 | 0.3324 | 0.0744 | 0.0698 | 0.1421 |
| 50 | 0.0345 | 0.0790 | 0.0560 | 0.1637 | 0.1021 | 0.3263 | 0.0699 | 0.0639 | 0.1389 |
| 75 | 0.0091 | 0.0886 | 0.0753 | 0.1760 | 0.1522 | 0.3508 | 0.0832 | 0.0729 | 0.1522 |
| 100 | 0.0312 | 0.0745 | 0.0776 | 0.1530 | 0.1480 | 0.3033 | 0.0610 | 0.0558 | 0.1305 |
| 150 | 0.0284 | 0.0866 | 0.0988 | 0.1694 | 0.2074 | 0.3337 | 0.0746 | 0.0725 | 0.1486 |

Table 2: *Empirical bias and (co)variances of* $\sqrt{k}(\hat{\Gamma}_n(\mathbf{x}) - \Gamma(\mathbf{x}))$ *for various choices of* $k_n$. *Upper four lines: Pure subordinator; lower four lines: Subordinator + Brownian Motion.*

$\Gamma$ to certain dependence properties, as in the case of ordinary copulas for which standard measures such as Kendall's $\tau$ or Spearman's $\rho$ can be written as integrals over $C$ and are thus accessible through nonparametric estimation of the copula.

## 7 Auxiliary results

**Lemma 7.1** *Let* $P : (0, \infty] \to [0, \infty)$ *denote the function* $P(x) = 1/x$.

*a) The mapping*

$$\Phi_1 : (\mathcal{B}^0_\infty((0,\infty]^2), d) \times (\mathcal{B}^0_\infty((0,\infty]), d)^2 \to (\mathcal{B}^0_\infty((0,\infty]^2), d) \times (\mathcal{B}^-_\infty((0,\infty]), d_2)^2$$

*defined by* $\Phi_1(U, U_1, U_2) = (U, U_1^-, U_2^-)$ *is Hadamard-differentiable at* $(\Gamma, P, P)$ *tangentially to* $\mathbb{D}_{0,1} = \mathbb{D}_0$ *as defined in* (4.9) *with derivative*

$$\Phi'_{1,(\Gamma,P,P)}(U, U_1, U_2) = (U, x_1^{-2} U_1(1/x_1), x_2^{-2} U_2(1/x_2)).$$

*b) The mapping*

$$\Phi_2 : (\mathcal{B}^0_\infty((0,\infty]^2), d) \times (\mathcal{B}^-_\infty((0,\infty]), d_2)^2 \to (\mathcal{B}^0_\infty((0,\infty]^2), d) \times (\mathcal{B}^p_\infty([0,\infty)), d_2)^2$$

*defined by* $\Phi_2(U, V_1, V_2) = (U, V_1 \circ P, V_2 \circ P)$ *is Hadamard-differentiable at* $(\Gamma, P, P)$ *tangentially to* $\mathbb{D}_{0,2} = \Phi'_{1,(\Gamma,P,P)}(\mathbb{D}_{0,1})$ *with derivative*

$$\Phi'_{2,(\Gamma,P,P)}(U, V_1, V_2) = (U, V_1 \circ P, V_2 \circ P).$$

*c) Suppose* $\Gamma$ *has continuous first order partial derivatives on* $(0,\infty)^2$. *The mapping*

$$\Phi_3 : (\mathcal{B}^0_\infty((0,\infty]^2), d) \times (\mathcal{B}^p_\infty([0,\infty)), d_2)^2 \to \mathcal{B}_\infty((0,\infty]^2, d_2)$$

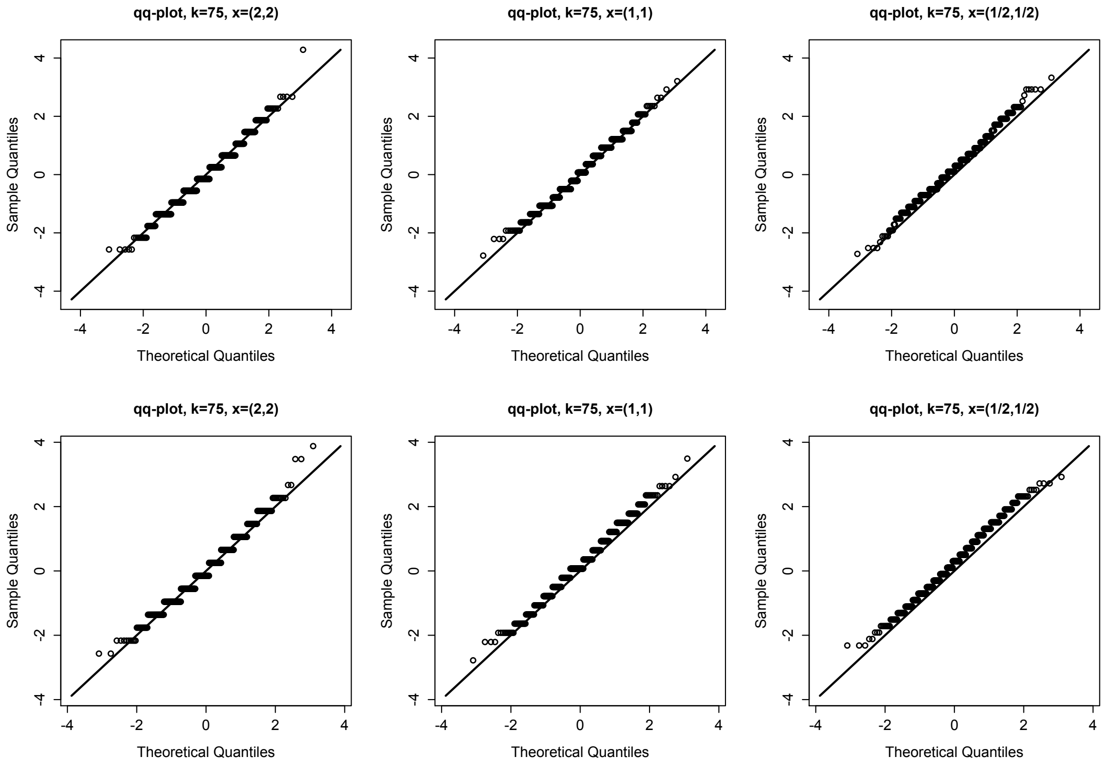


Figure 3: *QQ-plots of the empirical quantiles of* $\sqrt{k_n}(\hat{\Gamma}_n(\mathbf{x}) - \Gamma(\mathbf{x}))$ *divided by their sample standard deviation vs. the theoretical quantiles of the standard normal distribution. Upper three pictures: Pure subordinator; lower three pictures: Subordinator + Brownian Motion.*

*defined by* $\Phi_3(U, G_1, G_2) = U(G_1, G_2)$ *is Hadamard-differentiable at* $(\Gamma, \mathrm{id}, \mathrm{id})$ *tangentially to* $\mathbb{D}_{0,3} = \Phi'_{2,(\Gamma,P,P)}(\mathbb{D}_{0,2})$ *with derivative*

$$\Phi'_{3,(\Gamma,\mathrm{id},\mathrm{id})}(U, G_1, G_2)(\mathbf{u}) = U(\mathbf{u}) + \sum_{j=1}^{2} \dot{\Gamma}_j(\mathbf{u}) G_j(u_j),$$

*where the sum on the right-hand side is defined as* $0$ *if one of the coordinates of* $\mathbf{u}$ *equals* $\infty$.

**Proof.** The assertion in a) is a consequence of Lemma 7.2 below, whereas the assertion in b) follows from linearity of $\Phi_2$. Regarding c) let $t_n \to 0$, $(U_n, G_{n1}, G_{n2}) \to (U, G_1, G_2) \in \mathbb{D}_{0,3}$ such that $(\Gamma + t_n U_n, \mathrm{id}_{[0,\infty)} + t_n G_{n1}, \mathrm{id}_{[0,\infty)} + t_n G_{n2}) \in \mathcal{B}^0_\infty((0,\infty]^2) \times (\mathcal{B}^p_\infty([0,\infty)))^2$. First consider $\mathbf{u} \in S_{k,1} = [1/k, k]^2$, which allows to decompose

$$t_n^{-1}\{\Phi_3(\Gamma + t_n U_n, \mathrm{id}_{[0,\infty)} + t_n G_{n1}, \mathrm{id}_{[0,\infty)} + t_n G_{n1}) - \Phi_3(\Gamma, \mathrm{id}, \mathrm{id})\}(\mathbf{u}) = L_{n1}(\mathbf{u}) + L_{n2}(\mathbf{u}), \tag{7.1}$$

where

$$L_{n1}(\mathbf{u}) = t_n^{-1}\{\Gamma(u_1 + t_n G_{n1}(u_1), u_2 + t_n G_{n1}(u_2)) - \Gamma(\mathbf{u})\},$$
$$L_{n2}(\mathbf{u}) = U_n(u_1 + t_n G_{n1}(u_1), u_2 + t_n G_{n1}(u_2)).$$

$U_n$ converges uniformly on $S_{2k,1} = [1/(2k), 2k]^2$ to $U$, which is uniformly continuous on $S_{2k,1}$. Hence, since $\sup_{u_j \in [1/(2k),2k]} |t_n G_{nj}(u_j)| \to 0$, we obtain $\sup_{\mathbf{u} \in S_{k,1}} |L_{n2}(\mathbf{u}) - U(\mathbf{u})| \to 0$. It remains to consider the summand $L_{n1}$. A Taylor expansion yields

$$L_{n1}(\mathbf{u}) = \sum_{j=1}^{2} \dot\Gamma_j(\mathbf{u}) G_{nj}(u_j) + r_n(\mathbf{u}),$$

where the remainder term is given by

$$r_n(\mathbf{u}) = \sum_{j=1}^{2} \{\dot\Gamma_j(\mathbf{v}_n) - \dot\Gamma_j(\mathbf{u})\} G_{nj}(u_j)$$

with some intermediate point $\mathbf{v}_n$. Observing that $|\dot\Gamma_j(\mathbf{u})| \le u_j^{-2} \le k^2$, uniform convergence of $G_{nj}$ to $G_j$ on $[1/k, k]$ implies that the dominating term in the expansion of $L_{n1}$ converges to $\sum_{j=1}^2 \dot\Gamma_j(\mathbf{u}) G_j(u_j)$, uniformly on $S_{k,1}$. By uniform continuity of $\dot\Gamma_j$ on $S_{k,1}$ and boundedness of $G_{nj}$ we obtain $r_n(\mathbf{u}) = o(1)$, uniformly. The other cases, i.e., $\mathbf{u} \in ([1/k, k] \times \{\infty\}) \cup (\{\infty\} \times [1/k, k]) \cup \{(\infty, \infty)\}$ are treated similarly, the details are omitted for the sake of brevity. □

**Lemma 7.2** *Let $\mathbb{D}_\Psi \subset \mathcal{B}_\infty((0,\infty])$ consist of all functions $f : (0,\infty] \to [0,\infty)$ that are non-increasing and left-continuous with $f(\infty) = 0$. Recall (3.3) for the definition of the generalized inverse function. Then the mapping*

$$\Psi : (\mathbb{D}_\Psi, d) \to (\mathcal{B}_\infty((0,\infty]), d_2), \quad f \mapsto f^-$$

*is Hadamard-differentiable at $P(x) = x^{-1}$ tangentially to the space*

$$\mathbb{D}_0 = \left\{ h \in \mathcal{C}((0,\infty]) \mid h(\infty) = 0, \lim_{x\to 0} x^2 h(x) = 0 \right\}.$$

*with derivative $(\Psi'_P(h))(x) = x^{-2} h(x^{-1})$.*

Proof. Let $t_n \in \mathbb{R} \setminus \{0\}, h_n \in \mathcal{B}_\infty((0,\infty])$ and $h \in \mathbb{D}_0$ such that $t_n \to 0, d(h_n, h) \to 0$ and $P + t_n h_n \in \mathbb{D}_\Psi$. It suffices to show that for each $\varepsilon \in (0,1), M \in (1,\infty)$

$$\sup_{z \in [\varepsilon, M]} \left| \frac{(P + t_n h_n)^-(z) - z^{-1}}{t_n} - z^{-2} h(z^{-1}) \right| \to 0.$$

For $z \in [\varepsilon, M]$ set $\xi_n(z) = (P + t_n h_n)^-(z)$. Choose $n_0 \in \mathbb{N}$ such that $\sup_{x \ge (2M)^{-1}} |t_n h_n(x)| \le \varepsilon/2$ for all $n \ge n_0$. We begin the proof by showing that $\xi_n(z) \in [1/(2M), 2/\varepsilon]$ for all $n \ge n_0$. By monotonicity of $\xi_n$ we obtain

$$\begin{aligned} \xi_n(z) \le \xi_n(\varepsilon) &\le \inf\{x \ge 1/(2M) \mid 1/x + t_n h_n(x) \le \varepsilon\} \\ &\le \inf\{x \ge 1/(2M) \mid 1/x \le \varepsilon/2\} = 2/\varepsilon. \end{aligned}$$

For the lower bound note that $\xi_n(z) \geq \xi_n(M)$ and set $x_0 = 1/(2M)$. Then

$$(P + t_n h_n)(x_0) = 2M + t_n h_n(1/(2M)) \geq 3M/2.$$

By monotonicity of $P + t_n h_n$ we obtain $(P + t_n h_n)(x) \geq 3M/2$ for all $x \leq x_0$ and hence $\xi_n(M) \geq x_0 = 1/(2M)$ as asserted.

By definition of the inverse, for $\varepsilon_n(z) = t_n^2 \wedge \xi_n(z)$,

$$(P + t_n h_n)(\xi_n(z)) \geq z \geq \varepsilon \vee \{(P + t_n h_n)(\xi_n(z) + \varepsilon_n(z))\} > 0.$$

Some careful calculations convert the latter estimate into

$$\frac{t_n \xi_n(\xi_n + \varepsilon_n) h_n(\xi_n + \varepsilon_n) - \varepsilon_n}{\{\varepsilon(\xi_n + \varepsilon_n)\} \vee \{1 + t_n(\xi_n + \varepsilon_n) h_n(\xi_n + \varepsilon_n)\}} \leq \xi_n - \frac{1}{z} \leq \frac{t_n \xi_n^2 h_n(\xi_n)}{1 + t_n \xi_n h_n(\xi_n)}, \tag{7.2}$$

where we used the abbreviations $\xi_n = \xi_n(z)$ and $\varepsilon_n = \varepsilon_n(z)$. Since $\xi_n(z) \in [1/(2M), 2/\varepsilon]$, boundedness of $h_n$ on $[1/M, 4/\varepsilon]$ implies $\sup_{z \in [\varepsilon, M]} |\xi_n(z) - z^{-1}| \to 0$. Dividing equation (7.2) by $t_n$ and exploiting the facts that $\varepsilon_n \leq t_n^2$ and that $x^2 h(x)$ is uniformly continuous on $[1/(2M), 4/\varepsilon]$ the assertion follows. □

## References


Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons Inc., New York.

Bücher, A. and Volgushev, S. (2011). Empirical and sequential empirical copula processes under serial dependence. *arXiv:1111.2778.*

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

Eder, I. and Klüppelberg, C. (2012). Pareto Lévy copulas and multivariate regular variation. *Advances in Applied Probability*, 44:117–138.

Esmaeili, H. and Klüppelberg, C. (2011). Parametric estimation of a bivariate stable Lévy process. *J. Multivariate Anal.*, 102(5):918–930.

Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Figueroa-López, J. E. (2008). Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.*, 78(18):3355–3365.

Figueroa-López, J. E. (2009). Nonparametric estimation of time-changed Lévy models under high-frequency data. *Adv. in Appl. Probab.*, 41(4):1161–1188.

Figueroa-López, J. E. and Houdré, C. (2009). Small-time expansions for the transition distributions of Lévy processes. *Stochastic Process. Appl.*, 119(11):3862–3889.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *N. Am. Actuar. J.*, 2(1):1–25.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Engineering*, 12:347–368.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Genest, C. and Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.*, 101(8):1837–1845.

Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379.

Jacod, J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM Probab. Stat.*, 11:173–196.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.*, 118(4):517–559.

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

Kallsen, J. and Tankov, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivariate Anal.*, 97(7):1551–1572.

Klüppelberg, C. and Resnick, S. I. (2008). The Pareto copula, aggregation of risks, and the emperor's socks. *J. Appl. Probab.*, 45(1):67–84.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Laeven, R. J. A. (2011). Non-parametric estimation for multivariate Lévy processes. Technical report. unpublished.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management.* Princeton Series in Finance. Princeton University Press, Princeton, NJ. Concepts, techniques and tools.

Nelsen, R. B. (2006). *An introduction to copulas.* Springer Series in Statistics. Springer, New York, second edition.

Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248.

Nickl, R. and Reiß, M. (2012). A Donsker theorem for Lévy measures. Technical report. unpublished.

Segers, J. (2011). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *arXiv:1012.2133v2, to appear in Bernoulli.*

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes - Springer Series in Statistics.* Springer, New York.

van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: particles, processes and inverse problems*, volume 55 of *IMS Lecture Notes Monogr. Ser.*, pages 234–252. Inst. Math. Statist., Beachwood, OH.