Estimating Bivariate Tail: a copula based approach

Elena Di Bernardino[1], Véronique Maume-Deschamps[2], Clémentine Prieur[3]

Abstract

This paper deals with the problem of estimating the tail of a bivariate distribution function. To this end we develop a general extension of the POT (Peaks-Over-Threshold) method, mainly based on a two-dimensional version of the Pickands-Balkema-de Haan Theorem. We introduce a new parameter that describes the nature of the tail dependence, and we provide a way to estimate it. We construct a two-dimensional tail estimator and study its asymptotic properties. We also present real data examples which illustrate our theoretical results.

Keywords: Extreme Value Theory, Peaks Over Threshold method, Pickands-Balkema-de Haan Theorem, Tail dependence.
2000 MSC: 62H12, 62H05, 60G70.

1. Introduction

The univariate POT (Peaks-Over-Threshold) method is common for estimating extreme quantiles or tail distributions (see e.g. McNeil 1997, 1999 and references therein). A key idea of this method is that a distribution is in the domain of attraction of an extreme value distribution if and only if the distribution of excesses over high thresholds is asymptotically generalized Pareto (GPD) (e.g. Balkema and de Haan, 1974; Pickands, 1975):

$$V_{\xi,\sigma}(x) := \begin{cases} 1-\left(1-\frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}}, & \text{if } \xi \neq 0,\ \sigma>0, \\ 1-\mathrm{e}^{\frac{-x}{\sigma}}, & \text{if } \xi = 0,\ \sigma>0, \end{cases} \tag{1}$$

and $x \geq 0$ for $\xi \leq 0$ or $0 \leq x < \frac{\sigma}{\xi}$ for $\xi > 0$. This univariate modeling is well understood, and has been discussed by Davison (1984), Davison and Smith (1990) and other papers of these authors.
In this paper, we are interested in the problem of fitting the joint distribution of bivariate observations exceeding high thresholds. To this end we develop

[1]Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 avenue Tony Garnier, 69366 Lyon, France, Tel.: +33 4 37 28 74 46, elena.di-bernardino@univ-lyon1.fr
[2]Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 avenue Tony Garnier, 69366 Lyon, France, Tel.: +33 4 37 28 74 28, veronique.maume@univ-lyon1.fr
[3]Université Joseph Fourier, Tour IRMA, MOISE-LJK B.P. 53 38041 Grenoble, France, Tel.: +33 4 76 63 59 63, clementine.prieur@imag.fr

a bivariate estimation procedure, mainly based on a version of the Pickands-Balkema-de Haan Theorem in dimension 2 (Theorem 2.1). This extension allows us to consider a two-dimensional structure of dependence between both continuous random components X and Y. This dependence is modeled via a copula C, which is supposed to be unknown.

We recall here some classical bivariate threshold models, based on a characterization of the joint tail by Resnick (1987). Letting F denote the joint distribution of (Y_1, Y_2) with marginals F_j, $j = 1, 2$. Define $Z_j = -1/\log(F_j(Y_j))$, $j = 1, 2$, i.e. each Y_j is transformed to a unit Fréchet variable and $\mathbb{P}(Z_j \leq z) = \exp^{-1/z}$, for $0 < z < \infty$. Let F_* denote the joint distribution of (Z_1, Z_2), we have $F(y_1, y_2) = F_*(z_1, z_2)$. The assumption that F is in the maximum domain of attraction (MDA) of a bivariate extreme value distribution G is equivalent assuming F_* to be in the domain of attraction of a bivariate extreme value distribution G_*, where the marginals of G_* are unit Fréchet. The characterization of Resnick (1987) can be written as

$$\lim_{t\to\infty} \frac{\log(F_*(tz_1, tz_2))}{\log(F_*(t,t))} = \lim_{t\to\infty} \frac{1 - F_*(tz_1, tz_2)}{1 - F_*(t,t)} = \frac{\log(G_*(z_1, z_2))}{\log(G_*(1,1))}. \tag{2}$$

Equating the left and the right-hand terms for large t leads to the following model for the joint tail of F (see Ledford and Tawn, 1996):

$$\mathfrak{F}_1(y_1, y_2) = \exp\{-l\left(-\log(F_{Y_1}(y_1)), -\log(F_{Y_2}(y_2))\right)\}, \tag{3}$$

for $y_j > u_j$, where u_j are high thresholds for the marginal distributions and l is the stable tail dependence function of the limiting extreme value distribution G_*. Then approximation (3) can be estimated by

$$\widehat{\mathfrak{F}}_1^*(y_1, y_2) = \exp\{-\widehat{l}(-\log(\widehat{F}^*_{Y_1}(y_1)), -\log(\widehat{F}^*_{Y_2}(y_2)))\}, \tag{4}$$

for high values of y_1 and y_2, where $\widehat{F}^*_{Y_1}(y_1)$ (resp. $\widehat{F}^*_{Y_2}(y_2)$) is an estimator for the marginal tail of Y_1 (resp. Y_2). For instance $\widehat{F}^*_{Y_1}(y_1)$ (resp. $\widehat{F}^*_{Y_2}(y_2)$) comes from the univariate POT method described in Section 4.1. In (4) $\widehat{l}$ is an estimator of the stable tail dependence function (see Drees and Huang, 1998; Draisma *et al.*, 2004; Einmahl *et al.*, 2008). For another approach, based on the estimation of the so-called univariate dependence function of Pickands (Pickands, 1981), see for instance Capéraà and Fougères (2000). Problems arise with both these bivariate techniques when (Y_1, Y_2) are asymptotically independent i.e.,

$$\lambda := \lim_{t\to 0} \mathbb{P}[F^{-1}_{Y_1}(Y_1) > 1 - t \,|\, F^{-1}_{Y_2}(Y_2) > 1 - t] = 0. \tag{5}$$

When the data exhibit positive or negative association that only gradually disappears at more and more extreme levels, these methods produce a significant bias. In order to overcome this problem, Ledford and Tawn (1996, 1997, 1998) introduced a model in which the tail dependence is characterized by a coefficient $\eta \in (0, 1]$. In these works the joint survival distribution function of a bivariate random vector (Z_1, Z_2) with unit Fréchet marginals is assumed to satisfy $\mathbb{P}[Z_1 > z, Z_2 > z] \sim L(z)\mathbb{P}[Z_1 > z]^{\frac{1}{\eta}}$, where L is a slowly varying function at infinity. Various methods to estimate this coefficient η are proposed in Peng (1999), Draisma *et al.* (2004), Beirlant *et al.* (2011). For some counter-examples of the Ledford and Tawn's model see Schlather (2001).

Contrarily to this approach, we propose a model based on regularity conditions of the copula and on the explicit description of the dependence structure in the joint tail (see condition in (8) in Proposition 2.1). The study of tail dependence from a distributional point of view by means of appropriate copulae has received attention in the past decade. The interested reader is referred to Juri and Wüthrich (2002, 2004), Wüthrich (2004), Charpentier and Juri (2006), Charpentier and Segers (2006), Javid (2009).

The general idea of our model is to decompose the estimation of $\mathbb{P}(X \leq x, Y \leq y)$, for x, y above some marginal thresholds u_X, u_Y, in the estimation of different bivariate regions. For the joint upper tail in $[u_X, x] \times [u_Y, y]$ we use the non parametric estimators coming from Theorem 2.1 (see Section 2). For the *lateral regions* $[-\infty, x] \times [-\infty, u_Y]$ and $[-\infty, u_X] \times [-\infty, y]$ we approximate the distribution function F using (3). The stability of our estimation compared to the one of $\widehat{\mathcal{F}}_1^*$ is analyzed on some real cases (Section 7) which have been studied in other papers (e.g. Beirlant *et al.*, 2011; Frees and Valdez 1998; Lescourret and Robert, 2006). Therefore our estimator, in a different way from the Ledford and Tawn's method, covers situations less restrictive than dependence or perfect independence above thresholds. Note also that our method is free from the pre-treatment of data because we can work directly with the original general samples without the transformation in Fréchet marginal distributions.

Finally, we recall that, in the past decade, bivariate extensions of the POT method via generalized Pareto distribution have been developed in a series of papers by Falk and Reiss (2005 and references therein) or in Reiss and Thomas (2007; Chapter 13). Recently a multivariate generalization is treated in Beirlant *et al.* (2004), Rootzén and Tajvidi (2006) and Michel (2008). The role of multivariate generalized Pareto distributions in the framework of extreme value theory is still under scrutiny. In contrast to the univariate case it is not intuitively clear, how exceedances over high thresholds are to be defined. Our paper makes a contribution to this part of recent literature. To the best of our knowledge the POT procedure we propose in this paper can not be directly deduced from POT methods proposed in works cited above. Moreover we provide an estimation of bivariate tails such this type of estimation is not obtained in the papers cited above. However, some ingredients for a comparison are investigated in Theorem 4.2 in Juri and Wüthrich (2004).

The paper is organized as follows. In Section 2 we state an extension of the Pickands- Balkema-de Haan Theorem in the case of bivariate distributions with different marginals (Theorem 2.1). In Section 3 we provide a new non parametric estimator for the dependence structure of a bivariate random sample in the upper tail. In Section 4.2 we recall the POT procedure for univariate distributions and we use Theorem 2.1 in order to build a new estimator for the tail of the bivariate distribution. The study of the asymptotic properties of our estimator makes use of a convergence result in univariate case (Theorem 5.1) dealing with asymptotic behavior of the absolute error between the theoretical distribution function and its tail estimator. In Section 6 we present the consistency result of our estimator with its convergence rate both in the asymptotic dependent case (Theorem 6.1) and in the asymptotic independent one (Theorem 6.2). Examples with real data are presented in Section 7. Some auxiliary results and more technical proofs are postponed to the Appendix.

Remark 1 Assume we observe $X_1, \ldots, X_n$ i.i.d. with common distribution function F. If we fix some high threshold u, let N denote the number of excesses above u. In the following, two approaches will be considered. In the first one, we work conditionally on N. If n is the sample size and u_n the associated threshold, the number of excesses is m_n, with $\lim_{n\to\infty} m_n = \infty$ and $\lim_{n\to\infty} m_n/n = 0$. The second approach considers the number of excesses N_n as a binomial random variable (which is the case in the simulations), $N_n \sim \operatorname{Bi}(n, 1-F(u_n))$ with $\lim_{n\to\infty} 1-F(u_n) = 0$ and $\lim_{n\to\infty} n(1-F(u_n)) = \infty$. Keeping in mind these considerations will be useful in the following (in particular in Section 5).

2. On the two-dimensional Pickands-Balkema-de Haan Theorem

A central one dimensional result in univariate tail estimation is the so-called Pickand-Balkema-de Haan Theorem. As our aim is the estimation of bivariate tails, we are interested in two-dimensional extensions of this theorem. Such a two dimensional generalization can be found in the literature (e.g. see Juri and Wüthrich, 2004; Wüthrich, 2004) with the assumption $F_X = F_Y$. Starting from Theorem 4.1 in Juri and Wüthrich (2004) and Theorem 3.1 in Charpentier and Juri (2006), we provide here a precise formulation and proof of a general bivariate Pickands-Balkema-de Haan Theorem (Theorem 2.1 below). We first introduce some notation and recall results from Juri and Wüthrich (2004) and Nelsen (1999), which we will need later.
We consider a 2-dimensional copula $C(u,v)$ and the associated survival copula $C^*(u,v)$. In a first time we assume that X and Y are uniformly distributed on $[0,1]$. Let us fix a threshold $u \in [0,1)$ such that $\mathbb{P}[X > u, Y > u] > 0$, i.e. such that $C^*(1-u, 1-u) > 0$. We consider the distribution of X and Y conditioned on $\{X > u, Y > u\}$:

$$\forall\, x \in [0,1], \quad \overline{F}_{X,\,u}(x) := \mathbb{P}[\, X \leq x \,|\, X > u, Y > u\,] = 1 - \frac{C^*(1 - x \vee u, 1-u)}{C^*(1-u, 1-u)}, \tag{6}$$

$$\forall\, y \in [0,1], \quad \overline{F}_{Y,\,u}(y) := \mathbb{P}[\, Y \leq y \,|\, X > u, Y > u\,] = 1 - \frac{C^*(1-u, 1 - y \vee u)}{C^*(1-u, 1-u)}. \tag{7}$$

Note that the continuity of the copula C implies that $\overline{F}_{X,\,u}$ and $\overline{F}_{Y,\,u}$ are also continuous.

Definition 2.1 Let X and Y be uniformly distributed on $[0,1]$. Assume that for a threshold $u \in [0,1)$, $C^*(1-u, 1-u) > 0$. We define the *Upper-tail dependence copula* at level $u \in [0,1)$ relative to the copula C by

$$C_u^{up}(x,y) := \mathbb{P}[\, X \leq \overline{F}_{X,\,u}^{-1}(x), Y \leq \overline{F}_{Y,\,u}^{-1}(y) \,|\, X > u, Y > u\,],$$

$\forall\,(x,y) \in [0,1]^2$, where $\overline{F}_{X,\,u}$, $\overline{F}_{Y,\,u}$ are given by (6)-(7).

Note that $\mathbb{P}[X \leq x, Y \leq y \,|\, X > u, Y > u]$ obviously defines a two-dimensional distribution function whose marginals are given by $\overline{F}_{X,\,u}$ and $\overline{F}_{Y,\,u}$. We remark that $C_u^{up}(x,y)$ is a copula and from the continuity of $\overline{F}_{X,\,u}$ and $\overline{F}_{Y,\,u}$ we obtain the uniqueness of C_u^{up}. Moreover, the asymptotic behavior of C_u^{up} for u around 1 describes the dependence structure of X, Y in their upper tails.
In order to provide an explicit form for $\lim_{u\to 1} C_u^{up}(x,y)$, we state Proposition 2.1 below, which is a modification of Theorem 3.1 in Charpentier and Juri (2006). More precisely we adapt Theorem 3.1 in Charpentier and Juri (2006) in the case of Upper-tail dependence copula, assuming that C satisfies suitable regularity condition under the direction $(1-u,\, 1-u)$ (see the limit in (8)). For comparisons we refer to Section 3 in Charpentier and Juri (2006).

Proposition 2.1 *Assume that* $\partial C^*(1-u, 1-v)/\partial u < 0$ *and* $\partial C^*(1-u, 1-v)/\partial v < 0$ *for all* $u, v \in [0,1)$. *Furthermore, assume that there is a positive function* G *such that*

$$\lim_{u\to 1} \frac{C^*(x\,(1-u),\, y\,(1-u))}{C^*(1-u,\, 1-u)} = G(x,y), \quad \textit{for all } x,\, y\, > \, 0. \tag{8}$$

Then for all $(x,y) \in [0,1]^2$

$$\lim_{u\to 1} C_u^{up}(x,y) = x + y - 1 + G(g_X^{-1}(1-x), g_Y^{-1}(1-y)) := C^{*\,G}(x,y), \tag{9}$$

where $g_X(x) := G(x,1)$, $g_Y(y) := G(1,y)$. *Moreover there is a constant* $\theta > 0$ *such that, for* $x > 0$

$$G(x,y) = \begin{cases} x^\theta g_Y(\frac{y}{x}) & \textit{for } \frac{y}{x} \in [0,1], \\ y^\theta g_X(\frac{x}{y}) & \textit{for } \frac{y}{x} \in (1,\infty). \end{cases} \tag{10}$$

The proof of Proposition 2.1 is postponed to the Appendix. We adapt in our setting the proof of Theorem 3.1 by Charpentier and Juri (2006). Since $\partial C^*(1-u, 1-v)/\partial u < 0$ and $\partial C^*(1-u, 1-v)/\partial v < 0$ for all $u, v \in [0,1)$, we have $C^*(1-u, 1-u) > 0$, for all $u \geq 0$, i.e. C_u^{up} is well defined for all $u \geq 0$. Then we ask that the joint survival distribution function of X and Y, uniformly distributed on $[0,1]$, is strictly decreasing in each coordinate. As in Remark 3.2 in Charpentier and Juri (2006) one can prove that the convergence in (9) is uniform in $[0,1]^2$. From Proposition 2.1, functions G, g_X, and g_Y characterize the asymptotic behavior of the dependence structure for extremal events.

Remark 2

. We note that ${C^*}^G(x,y)$ defined in (9) is the survival copula of the copula $C^G(x,y) := G(g_X^{-1}(x), g_Y^{-1}(y))$ and thus, in particular, is a copula (for more details see Section 3 in Charpentier and Juri, 2006).

. In the case of symmetric copula, i.e. $C(u,v) = C(v,u)$ for all u and v, the limit G in (8) is continuous, symmetric, with marginals $G(x,1) = G(1,x) = g(x)$, where $g : [0,\infty) \to [0,\infty)$ is a strictly increasing function and $g(x) = x^\theta g(1/x)$ for all $x \in (0,\infty)$ (for more details about properties of G in the symmetric case see Section 2 in Juri and Wüthrich, 2004).

In the univariate setting de Haan (1970) proves that $F \in MDA(H_\xi)$ is equivalent to the existence of a positive measurable function $a(\cdot)$ such that, for $1 - \xi\, x > 0$ and $\xi \in \mathbb{R}$,

$$\lim_{u \to x_F} \frac{1 - F(u + x\, a(u))}{1 - F(u)} = \begin{cases} (1 - \xi\, x)^{\frac{1}{\xi}}, & \text{if } \xi \neq 0, \\ \mathrm{e}^{-x}, & \text{if } \xi = 0, \end{cases} \tag{11}$$

where $x_F := \sup\{x \in \mathbb{R} \,|\, F(x) < 1\}$. It allows stating below a rigorous formulation of the two-dimensional Pickands-Balkema-de Haan Theorem in a general case.

Theorem 2.1 *Let X and Y be two continuous real valued random variables, with different marginal distributions, respectively F_X, F_Y, and copula C. Suppose that $F_X \in MDA(H_{\xi_1})$, $F_Y \in MDA(H_{\xi_2})$ and that C satisfies assumptions of Proposition 2.1. Then*

$$\sup_{\mathcal{A}} \Big| \mathbb{P}\big[X - u \le x, Y - F_Y^{-1}(F_X(u)) \le y \big| X > u, Y > F_Y^{-1}(F_X(u))\big]$$
$$- {C^*}^G\big(1 - g_X(1 - V_{\xi_1, a_1(u)}(x)), 1 - g_Y(1 - V_{\xi_2, a_2(F_Y^{-1}(F_X(u)))}(y))\big)\Big| \xrightarrow[u \to x_{F_X}]{} 0, \tag{12}$$

where $V_{\xi_i, a_i(\cdot)}$ is the GPD with parameters $\xi_i, a_i(\cdot)$ defined in (1)*, $a_i(\cdot)$ is as in* (11)*, for $i = 1, 2$, $\mathcal{A} := \{(x,y) : 0 < x \le x_{F_X} - u, 0 < y \le x_{F_Y} - F_Y^{-1}(F_X(u))\}$, with $x_{F_X} := \sup\{x \in \mathbb{R} \,|\, F_X(x) < 1\}$, $x_{F_Y} := \sup\{y \in \mathbb{R} \,|\, F_Y(y) < 1\}$.*

The proof of Theorem 2.1 is postponed to the Appendix.

3. Estimating dependence structure in the bivariate framework

It is well known that a bivariate distribution function F with continuous marginal distribution functions F_X, F_Y is said to have a *stable tail dependence function* l if for $x \ge 0$ and $y \ge 0$ the following limit exists:

$$\lim_{t \to 0} \frac{1}{t}\, \mathbb{P}[1 - F_X(X) \le tx \text{ or } 1 - F_Y(Y) \le ty] := l(x,y) \tag{13}$$

or similarly

$$\lim_{t\to 0}\frac{1}{t}\,\mathbb{P}[1-F_X(X)\leq tx, 1-F_Y(Y)\leq ty] := R(x,y) = x+y-l(x,y), \quad (14)$$

see e.g. Huang (1992). If F_X, F_Y are in the maximum domain of attraction of two extreme value distributions H_X, H_Y and if (13) holds then F is in the domain of attraction of an extreme value distribution H with marginals H_X, H_Y and with copula determined by l. Furthermore (13) is equivalent to

$$\lim_{t\to 0}\frac{1}{t}\,(1-C(1-tx,1-ty)) = l(x,y), \text{ for } x\geq 0,\, y\geq 0. \quad (15)$$

Note that the upper tail dependence coefficient defined in (5) is such that $\lambda = R(1,1)$. We introduce the non parametric estimators for l and R (see Einmahl *et al.*, 2006):

$$\widehat{l}(x,y) = \frac{1}{k_n}\sum_{i=1}^{n} 1_{\{R(X_i)>n-k_nx+1 \text{ or } R(Y_i)>n-k_ny+1\}}, \quad (16)$$

$$\widehat{R}(x,y) = \frac{1}{k_n}\sum_{i=1}^{n} 1_{\{R(X_i)>n-k_nx+1,\, R(Y_i)>n-k_ny+1\}}, \quad (17)$$

where $k_n\to\infty$, $k_n/n \to 0$ and $R(X_i)$ is the rank of X_i among $(X_1,\ldots,X_n)$, $R(Y_i)$ is the rank of Y_i among $(Y_1,\ldots,Y_n)$, for $i=1,\ldots,n$.

3.1. Asymptotic dependent case

If X and Y are asymptotically dependent ($\lambda>0$) we introduce an estimator for G, g_X and g_Y which will be used later to estimate the tail of the bivariate distribution function. Using (13)-(15), we write

$$g_X(x) = \frac{x+1-l(x,1)}{2-l(1,1)} = \frac{R(x,1)}{R(1,1)}, \quad g_Y(y) = \frac{y+1-l(1,y)}{2-l(1,1)} = \frac{R(1,y)}{R(1,1)},$$

$$G(x,y) = \frac{x+y-l(x,y)}{2-l(1,1)} = \frac{R(x,y)}{R(1,1)}.$$

Using (10), as R is homogeneous of order one then $\theta=1$. As $\eta\in(0,1]$ in the Ledford and Tawn's model (see Ledford and Tawn 1996, 1997, 1998), θ describes the nature of the tail dependence, it does not depend on the marginal distribution functions.
In order to estimate g_X, g_Y and G, we use the non parametric estimator for R in (17) and we obtain

$$\widehat{g}_X(x) = \frac{\widehat{R}(x,1)}{\widehat{R}(1,1)}, \quad \widehat{g}_Y(x) = \frac{\widehat{R}(1,y)}{\widehat{R}(1,1)}, \quad \text{and} \quad \widehat{G}(x,y) = \frac{\widehat{R}(x,y)}{\widehat{R}(1,1)}. \quad (18)$$

Using (18) we get a non parametric estimator for θ, for $x > 0$

$$\widehat{\theta}_{\frac{y}{x}} = \begin{cases} \frac{\log(\widehat{G}(x,y)) - \log(\widehat{g}_Y(\frac{y}{x}))}{\log(x)} & \text{if } \frac{y}{x} \in [0,1], \\ \frac{\log(\widehat{G}(x,y)) - \log(\widehat{g}_X(\frac{x}{y}))}{\log(y)} & \text{if } \frac{y}{x} \in (1,\infty). \end{cases} \tag{19}$$

Following Remark 2, in the case of a symmetric copula, using $g_X(x) = g_Y(x) = g(x) = x^{\theta} g(1/x)$ for $x > 0$, we get

$$\widehat{\theta}_x = \frac{\log(\widehat{g}(x)) - \log(\widehat{g}(\frac{1}{x}))}{\log(x)}. \tag{20}$$

Using Theorem 2.2 in Einmahl *et al.* (2006) (see Theorem C in Appendix) we state the following consistency result for $\widehat{G}$, $\widehat{g}_X$ and $\widehat{g}_Y$:

Corollary 3.1 *Under assumptions of Theorem C if we have v_n such that $v_n/\sqrt{k_n} \to 0$, for $n \to \infty$, and $\lambda > 0$ we obtain*

$$v_n \sup_{0<x,y\leq 1} \left| \widehat{G}(x,y) - G(x,y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

$$v_n \sup_{0<x\leq 1} \left| \widehat{g}_X(x) - g_X(x) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \; v_n \sup_{0<y\leq 1} \left| \widehat{g}_Y(y) - g_Y(y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

with $\widehat{g}_X(x)$, $\widehat{g}_Y(y)$ and $\widehat{G}(x,y)$ as in (18), $k_n \to \infty$, $k_n/n \to 0$ and $k_n = o\big(n^{\frac{2\alpha}{1+2\alpha}}\big)$.

We now provide an illustration for two different copulae: survival Clayton and Logistic copulae. We remark that they are two symmetric copulae with $\lambda > 0$. In particular we observe the sensitivity of $\widehat{\theta}_x$ in (20) to the sequence k_n (Figure 1). We draw the mean curve on 100 samples of size $n = 1000$ (full line) and the empirical standard deviation (dashed lines).

On simulations it seemed to us that for each value of x one could exhibit a range of values of k_n under which our estimate well behaved. In the following we fixe x for each simulation and may vary k_n. The choice of k_n does not appear to be crucial for $\widehat{\theta}_x$. In Figure 2 the mean squared error for $\widehat{\theta}_x$ is calculated on 100 samples of size $n = 1000$.

3.2. Asymptotic independent case

We say that X and Y are asymptotically independent if $\lambda = R(1,1) = 0$. In terms of copula this means that $C(u,u) = 1 - 2(1-u) + o(1-u)$, for $u \to 1$. The problem, with respect to Section 3.1, is that $g_X(x) = \frac{R(x,1)}{R(1,1)}$ and $g_Y(y) = \frac{R(1,y)}{R(1,1)}$ have no sense as $\lambda = 0$ and $R(x,y) = x + y - l(x,y) = 0, \;\; \forall\, x,\, y$.

We thus need to introduce a second-order refinement of condition in (8). More precisely, as in Draisma *et al.* (2004), we shall assume that:

$$\lim_{t\to 0} \frac{\frac{C^*(tx,ty)}{C^*(t,t)} - G(x,y)}{q_1(t)} := Q(x,y), \tag{21}$$

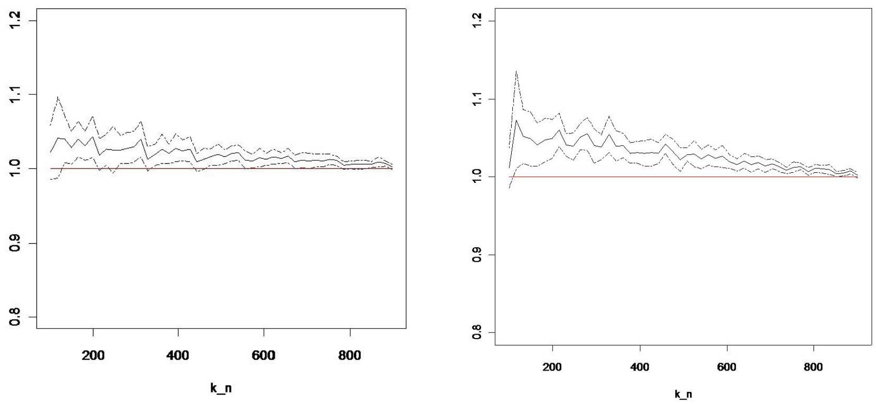

Figure 1: Estimator for θ, $(k, \widehat{\theta}_x)$ (left) $x = 0.07$, survival Clayton copula with parameter 1 (right) $x = 5$, Logistic copula with parameter 0.5

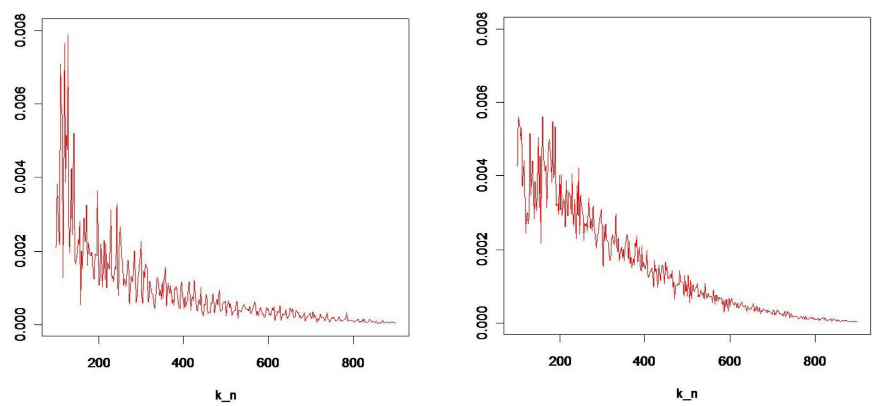

Figure 2: Mean squared error for $\widehat{\theta}_x$ (left) $x = 0.07$, survival Clayton copula with parameter 1 (right) $x = 5$, Logistic copula with parameter 0.5

for all $x, y \geq 0$, $x + y > 0$, where q_1 is some positive function and Q is neither a constant nor a multiple of G. Moreover we assume that convergence in (21) is uniform on $\{x^2 + y^2 = 1\}$. Let $q(t) := \mathbb{P}[1 - F_X(X) < t, 1 - F_Y(Y) < t]$ and $q^{\leftarrow}$ its inverse function. Then, using (21), we obtain the following consistency result for $\widehat{G}$, $\widehat{g}_X$ and $\widehat{g}_Y$:

Proposition 3.1 *Suppose (8) and (21) hold. We assume* $\lim_{t \to 0} q(t)/t = \lambda = 0$. *Then, for a sequence k_n such that $a_n := n\, q(k_n/n) \to \infty$ (this implies $k_n \to \infty$), $k_n/n \to 0$, $\sqrt{a_n}\, q_1(q^{\leftarrow}(a_n/n)) \to 0$, it holds that*

$$\psi_n \sup_{0 < x, y \leq 1} \left| \widehat{G}(x, y) \; - \; G(x, y) \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0,$$

$$\psi_n \sup_{0<x\leq 1} \left|\widehat{g}_X(x) - g_X(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \psi_n \sup_{0<y\leq 1} \left|\widehat{g}_Y(y) - g_Y(y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\psi_n << \sqrt{a_n}$, $\widehat{g}_X(x)$, $\widehat{g}_Y(y)$ *and* $\widehat{G}(x,y)$ *as in* (18).

Details of the proof are postponed to the Appendix. It is mainly based on Lemma 6.1 in Draisma *et al.* (2004).
In Proposition 3.2 below, by assuming some regularity properties on C, we deduce a specific form for G, g_X, g_Y and θ.

Proposition 3.2 *If* $\lambda = 0$ *and* C *is a twice continuously differentiable copula with the determinant of the Hessian matrix of* C *at* $(1,1)$ *different to zero, then*

$$\lim_{u\to 1} \frac{C^*(x(1-u),\, y(1-u))}{C^*(1-u, 1-u)} = x\, y, \qquad \forall\; x,\, y\; > 0,$$

$g_X(x) = g_Y(x) = x$ *and* $\theta = 2$.

Details of the proof will be omitted here. The main ingredient is the second-order development of copula C.
The assumptions of Proposition 3.2 are satisfied for a large class of asymptotic independent copulae: Ali Mikhail-Haq, Frank, Clayton with $a \geq 0$, Independent and Fairlie-Gumbel-Morgenstern copulae. An example of a non symmetric copula that satisfies the assumptions of Proposition 3.2 is $C(u,v) = x\, y + \frac{1}{9}(1-|2\,x-1|)\,(1-(2\,y-1)^2)$. This type of asymmetric copula is proposed in Benth and Kettler (2011) to model the evolution of price spread between electricity and gas prices.

We introduce some examples of asymptotic independent copulae that do not satisfy the assumptions of Proposition 3.2.
We consider the Ledford and Tawn's model (e.g. see Ledford and Tawn, 1996): $2\,u - 1 + C(1-u, 1-u) = (1-u)^{\frac{1}{\eta}} L(1-u)$, with L a slowly varying function at zero and $\eta \in (0,1]$. Then, for $\eta > 1/2$, $\lim_{u\to 1} \left(C(1,1) - C(1-u,1) - C(1,1-u) - C(1-u,1-u)\right)/(1-u)^2 = \infty$. Thus $\frac{\partial^2 C}{\partial u \partial v}$ does not exist at the point $(1,1)$. In particular this is the case of the Gaussian Copula with correlation parameter $\rho > 0$. However, from Theorem 5.3 in Juri and Wüthrich (2004), for a Gaussian Copula with $\mid \rho \mid < 1$ it holds that $\lim_{u\to 1} C_u^{up}(x,y) = x\, y$, for $(x,y) \in [0,1]^2$.
Let $C(u,v) = x\, y - \frac{1}{8}(1-|2\,x-1|)\,(1-(2\,y-1)^2)$, (for furthers details see Benth and Kettler, 2011). In this case $\frac{\partial^2 C}{\partial u \partial v}(1,1) = 0$. However we can calculate the limit in (8), and using (10) we obtain

$$G(x,y) = x\, y^2, \quad g_X(x) = x, \quad g_Y(y) = y^2, \quad \theta = 3.$$

We now provide an illustration for a Clayton copula. In particular we observe the sensitivity of $\widehat{\theta}_x$ in (20) to the sequence k_n (Figure 3). We draw the mean curve on 100 samples of size $n = 1000$ (full line) and the empirical standard deviation (dashed lines). Furthermore the mean squared error for $\widehat{\theta}_x$ is calculated on 100 samples of size $n = 1000$.

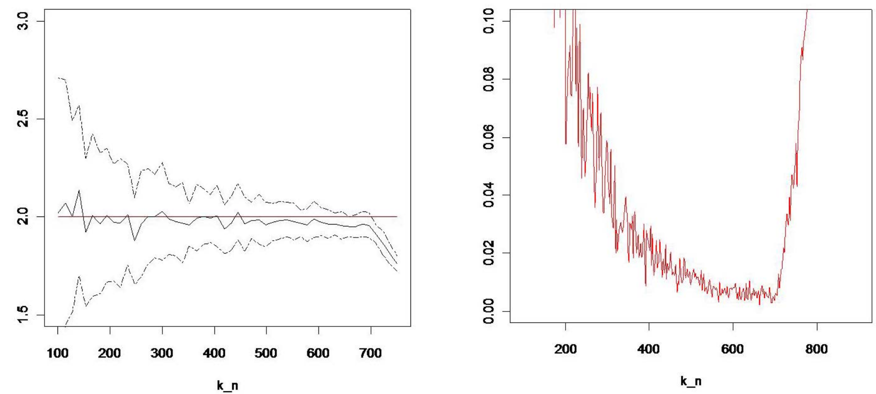

Figure 3: Clayton copula with parameter 0.05: (left) estimator for θ, $(k, \widehat{\theta}_x)$ with $x = 0.7$; (right) mean squared error for $\widehat{\theta}_x$ with $x = 0.7$.

4. Estimating tail distributions

4.1. Estimating the tail of univariate distributions

The estimation of the tail of bivariate distributions requires first the estimation of one-dimensional tail (McNeil, 1999; El-Aroui and Diebolt, 2002). Fix a threshold u and define $F_u(x) = \mathbb{P}[X \leq x \,|\, X > u]$. Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with unknown distribution function F and $\widehat{F}_X(u)$ the empirical distribution function evaluated at the threshold u. Recall that the univariate tail may be estimated by

$$\widehat{F}^*(x) = (1 - \widehat{F}_X(u))V_{\widehat{\xi},\widehat{\sigma}}(x-u) + \widehat{F}_X(u), \quad \text{for } x > u, \tag{22}$$

where $\hat{\xi}$, $\hat{\sigma}$ are the maximum likelihood estimators (MLE) based on the excesses above u. Using (22) we get the estimator, proposed by Smith (1987)

$$1 - \widehat{F}^*(y) = \begin{cases} \frac{N}{n}\left(1 - \widehat{\xi}\,\frac{(y-u)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{\xi}}}, & \text{if } \widehat{\xi} \neq 0, \\ \frac{N}{n}\left(\mathrm{e}^{\frac{-(y-u)}{\widehat{\sigma}}}\right), & \text{if } \widehat{\xi} = 0, \end{cases} \tag{23}$$

with $u < y < \infty$ (if $\widehat{\xi} \leq 0$) or $u < y < \frac{\widehat{\sigma}}{\widehat{\xi}}$ (if $\widehat{\xi} > 0$) and N the random number of excesses above the threshold.

4.2. Estimating the tail of bivariate distributions

In this section we present the main construction of this paper. We propose indeed a POT procedure in order to estimate the two-dimensional distribution function $F(x, y)$. Asymptotic properties for this estimator are stated and proved in Section 6.

This construction generalizes the one-dimensional POT construction stated in Section 4.1. Let X and Y be two real valued random variables with different continuous marginal distributions F_X and F_Y. The structure of dependence between X and Y is represented by copula C.

Construction of the tail estimator:
Given a high threshold u and $u_Y := F_Y^{-1}(F_X(u))$, we introduce the distribution of excesses: $F_u(x,y) := \mathbb{P}[\, X-u \le x, Y-u_Y \le y \,|\, X > u, Y > u_Y\,]$. Using (3) for large value of u and $x > u$, $y > u_Y$, we can approximate $F(u,y)$ and $F(x,u_Y)$ as

$$F_1^*(u,\, y) = \mathrm{e}^{\{-l(-\log(F_X(u)),-\log(F_Y(y)))\}}, \tag{24}$$

$$F_2^*(x,u_Y) = \mathrm{e}^{\{-l(-\log(F_X(x)),-\log(F_Y(u_Y)))\}}, \tag{25}$$

where l is the stable tail dependence function defined by (13). We recall that behind approximations (24)-(25), in order to avoid significant bias, we suppose that the data structure is characterized by dependence (or perfect independence) in the lateral regions $[-\infty,x]\times[-\infty,u_Y]$ and $[-\infty,u_X]\times[-\infty,y]$.
From Theorem 2.1 we now know that, for u around x_F, we can approximate the distribution of excesses with $C^{*\,G}$. So we obtain, for $x > u$, $y > u_Y$,

$$F^*(x,y) = (\overline{F}(u,u_Y))\cdot C^{*\,G}\big(1-g_X(1-V_{\xi_X,\sigma_X}(x-u)), 1-g_Y(1-V_{\xi_Y,\sigma_Y}(y-u_Y))\big) + F_1^*(u,y) + F_2^*(x,u_Y) - F(u,u_Y). \tag{26}$$

Then, we estimate $F(u,u_Y)$ and $\overline{F}(u,u_Y)$ in (26) from the data $\{X_i,Y_i\}_{i=1,\ldots,n}$, using the empirical distribution estimates

$$\widehat{F}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le u_Y\}}, \quad \widehat{\overline{F}}(u,u_Y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i> u,\, Y_i> u_Y\}}. \tag{27}$$

From (24)-(25) and using the non parametric estimator (16) we obtain

$$\widehat{F}_1^*(u,y) = \exp\{-\widehat{l}(-\log(\widehat{F}_X(u)), -\log(\widehat{F}_Y^*(y)))\}, \tag{28}$$

$$\widehat{F}_2^*(x,u_Y) = \exp\{-\widehat{l}(-\log(\widehat{F}_X^*(x)), -\log(\widehat{F}_Y(u_Y)))\}, \tag{29}$$

where $\widehat{F}_X(u)$ and $\widehat{F}_Y(u_Y)$ are the empirical univariate estimators evaluated at respective thresholds and $\widehat{F}_X^*(x)$ and $\widehat{F}_Y^*(y)$ are one-dimensional POT tail estimators of the marginal distribution functions, defined by (22). Now, using (27), (28) and (29), we can approximate $F^*(x,y)$, for $x > u$, $y > u_Y = F_Y^{-1}(F_X(u))$ and u large, by

$$\begin{aligned}\widetilde{F}^*(x,y) = {}& \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i>u,\, Y_i>u_Y\}}\right)\big(1-\widehat{g}_X(1-V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u)) \\ & -\widehat{g}_Y(1-V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-u_Y)) + \widehat{G}\big(1-V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u), 1-V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-u_Y)\big)\big) \\ & + \widehat{F}_1^*(u,y) + \widehat{F}_2^*(x,u_Y) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le u,\, Y_i\le u_Y\}},\end{aligned} \tag{30}$$

where $\widehat{\xi}_X$, $\widehat{\sigma}_X$ (resp. $\widehat{\xi}_Y$, $\widehat{\sigma}_Y$) are MLE based on the excesses of X (resp. Y). Finally we remark that the second threshold in (30) depends on the unknown marginal distribution functions F_X and F_Y. Then, in order to compute in practice $\widehat{F}^*(x,y)$, we propose to estimate u_Y by $\widehat{u}_Y = \widehat{F}_Y^{-1}(\widehat{F}_X(u))$, with $\widehat{F}_X(u) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leq u\}}$ and $\widehat{F}_Y^{-1}$ the empirical quantile function of Y.
So we obtain, from (30), the tail estimator for the two-dimensional distribution function for $x > u$ and $y > \widehat{u}_Y$:

$$\boxed{\begin{aligned}\widehat{F}^*(x,y) = &\left(\frac{1}{n}\sum_{i=1}^n 1_{\{X_i>u,\, Y_i>\widehat{u}_Y\}}\right)\big(1-\widehat{g}_X(1-V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u))\\ &-\widehat{g}_Y(1-V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y))+\widehat{G}\big(1-V_{\widehat{\xi}_X,\widehat{\sigma}_X}(x-u),1-V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(y-\widehat{u}_Y)\big)\big)\\ &+\widehat{F}_1^*(u,y)+\widehat{F}_2^*(x,\widehat{u}_Y)-\frac{1}{n}\sum_{i=1}^n 1_{\{X_i\leq u,\, Y_i\leq \widehat{u}_Y\}}, \quad (31)\end{aligned}}$$

In the case with same marginal distributions we have a particular case of (30), with the same threshold u for X and Y, and we do not need to estimate the second threshold.

Remark 3 Note that $\widehat{F}^*(x,y)$ in (31), is only valid for $x > u$ and $y > \widehat{u}_Y$, when u is *large enough*. The expression large enough is a fundamental problem of the POT method. The choice of the threshold u is indeed a compromise: u has to be large for the GPD approximation to be valid, but if it is too large, the estimation of the parameters ξ_X, ξ_Y and σ_X, σ_Y will suffer from a lack of observations over the thresholds. The compromise will be explained in Sections 5 and 6.

5. Convergence results in the univariate case

In order to study asymptotic properties of our bivariate tail estimator we present in this section some slight modifications of one-dimensional convergence results in Smith (1987; Theorems 3.2 and 8.1). Incidentally we get asymptotic confidence intervals for the unknown theoretical univariate function $F(x)$, using Theorem 5.1. From now on we assume that the tail of F decays like a power function, i.e. is in the domain of attraction of Fréchet i.e. $\overline{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$, with $\alpha > 0$. As in Smith (1987), Section 3, we shall assume that L satisfies the following condition

- SR2: $\frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x)), \quad$ as $x \to \infty, \quad \forall\, t > 0$,

where $\phi(x) > 0$ and $\phi(x) \to 0$ as $x \to \infty$. Let R_ρ be the set of $\rho-$regularly varying functions. Condition SR2 implies, excluding trivial cases, $\phi \in R_\rho$, for some $\rho \leq 0$, and $k(t) = c\, h_\rho(t)$, with $h_\rho(t) = \int_1^t u^{\rho-1}\mathrm{d}u$; (for more details see Section 3 in Smith, 1987 or Goldie and Smith, 1987).

The study of the asymptotic properties of the maximum likelihood estimators of the scale and shape parameters of the generalized Pareto distribution in the POT method has received attention in the literature. For instance asymptotic normality of $\widehat{\xi}$ and $\widehat{\sigma}$, in the case of random threshold in the POT procedure is studied in depth in Drees *et al.* (2004). Smith (1987) examines a slightly different version of the MLE's that is based on the excesses over a deterministic threshold and on the second-order Condition SR2. For details about the difference between these two approaches see, for instance, Remark 2.3 in Drees *et al.* (2004). In this paper we follow the approach proposed in Smith (1987). In particular Theorems 3.2. and 8.1. in Smith (1987) are written conditionally on $N = m_n$, with N denoting the number of excesses above the threshold. In practice we work with some deterministic threshold u and N is considered as random (see Remark 1 in Section 1). Therefore we give the version of Theorem 3.2 in Smith (1987) (resp. Theorem 8.1), Corollary 5.1 (resp. Corollary 5.2), unconditionally on N.

Corollary 5.1 *Suppose L satisfies SR2. Let n be the sample size and $u_n := \overline{f}(n)$ the threshold, such that $\overline{f}(n) \to \infty$, for $n \to \infty$. Let $N = N_n$ denote the random number of excesses of u_n. We define $\xi = -\alpha^{-1}$ and $\sigma_n = \overline{f}(n)\,\alpha^{-1}$. If*

$$n(1 - F(u_n)) \xrightarrow[n\to\infty]{} \infty, \tag{32}$$

$$\sqrt{n(1 - F(u_n))}c\,\phi(u_n) \xrightarrow[n\to\infty]{} \mu(\alpha - \rho), \tag{33}$$

then there exists, with probability 1, a local maximum $(\widehat{\sigma}_n, \widehat{\xi}_n)$ of the GPD log likelihood function, such that

$$\sqrt{N}\begin{pmatrix} \frac{\widehat{\sigma}_n}{\sigma_n} - 1 \\ \widehat{\xi}_n - \xi \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\begin{pmatrix} \frac{\mu(1-\xi)(1+2\xi\rho)}{1-\xi+\xi\rho} \\ \frac{\mu(1-\xi)\xi(1+\rho)}{1-\xi+\xi\rho} \end{pmatrix}; \begin{pmatrix} 2(1-\xi) & (1-\xi) \\ (1-\xi) & (1-\xi)^2 \end{pmatrix}\right).$$

Proof: If (32) and (33) hold then $N(n\,(1 - F(u_n)))^{-1} \xrightarrow[n\to\infty]{\mathbb{P}} 1$, and (3.2) in Smith (1987) holds in probability, i.e.

$$\frac{\sqrt{N}\,c\,\phi(u_n)}{\alpha - \rho} = \frac{\sqrt{N}\,c\,\phi(\overline{f}(n))}{\alpha - \rho} \xrightarrow[n\to\infty]{\mathbb{P}} \mu \in (-\infty, \infty).$$

Hence we conclude with a Skorohod-type construction of probability spaces on which (3.2) in Smith (1987) holds almost surely. □

Corollary 5.2 *Suppose all the assumptions of Corollary 5.1 are satisfied. Let n be the sample size, $u_n := \overline{f}(n) \to \infty$ and $z_n := f(n) \to \infty$, for $n \to \infty$, such that $(z_n)^{-s\,\rho}\frac{\phi(u_n(z_n)^s)}{\phi(u_n)} \to 1$, for $n \to \infty$ and $s \in [0,1]$. Let $N = N_n$ denote the random number of excesses above u_n. If*

$$\frac{\log(z_n)}{\sqrt{n(1 - F(u_n))}} \xrightarrow[n\to\infty]{} 0, \tag{34}$$

then

$$\frac{\sqrt{N}}{\log(f(n))}\left[\frac{1-\widehat{F}^*(\overline{f}(n)\,f(n))}{1-F(\overline{f}(n)\,f(n))}-1\right]\xrightarrow[n\to\infty]{d}\mathcal{N}(\nu,\tau^2),$$

where $\widehat{F}^*$ *is as in* (23), $\nu=0$ *if* $\rho=0$, $\nu=\frac{\mu\alpha(\alpha+1)(1+\rho)}{1+\alpha-\rho}$ *for* $\rho<0$ *and* $\tau^2=\alpha^2(1+\alpha)^2$.

Proof: If (32), (33) and (34) hold, then (8.7), (8.8) and (8.11) in Smith (1987) hold in probability, i.e

$$\frac{\log(z_n)}{\sqrt{N}}\xrightarrow[n\to\infty]{\mathbb{P}}0,\qquad \frac{\sqrt{N}}{\log(z_n)}\left[\frac{N}{n\,(1-F(u_n))}-1\right]\xrightarrow[n\to\infty]{\mathbb{P}}0.$$

We conclude as for Corollary 5.1. □
Note that, in simple cases, we often have $\phi(x)=x^\rho$; in which case $\frac{(z_n)^{-s\,\rho}\,\phi(u_n(z_n)^s)}{\phi(u_n)}\to 1$, for $n\to\infty$, is automatic satisfied. From Corollary 5.2 the following result can be obtained.

Theorem 5.1 *Assume that all the assumptions of Corollary 5.2 are satisfied. We use the same notation. If*

$$(z_n)^\alpha(n(1-F(u_n)))^{-1/2}\xrightarrow[n\to\infty]{}0, \tag{35}$$

then

$$\frac{\sqrt{N}}{\log(f(n))\,\widehat{\overline{F}}(\overline{f}(n)\,f(n))}\left[F(\overline{f}(n)\,f(n))-\widehat{F}^*(\overline{f}(n)\,f(n))\right]\xrightarrow[n\to\infty]{d}\mathcal{N}(\nu,\tau^2), \tag{36}$$

where $\widehat{\overline{F}}$ *is the univariate empirical survival function,* $\widehat{F}^*$ *is as in* (23), $\nu=0$ *if* $\rho=0$, $\nu=\frac{\mu\alpha(\alpha+1)(1+\rho)}{1+\alpha-\rho}$ *for* $\rho<0$ *and* $\tau^2=\alpha^2(1+\alpha)^2$.

The proof of Theorem 5.1 is postponed to the Appendix. As a byproduct, from (36) it is possible to construct in practice asymptotic confidence intervals for $F(\overline{f}(n)\,f(n))$.

6. Convergence results in the bivariate case

In this section we provide our main result: the consistency property of our bivariate tail estimator (31) with convergence rate. We consider:

Remark 4 Let n be the sample size. We choose, from Theorem 2.1,

$$u_{1\,n}:=\overline{f}_1(n)\xrightarrow[n\to\infty]{}\infty\quad\text{threshold for }X,$$

$$u_{2\,n}:=\overline{f}_2(n)=F_Y^{-1}(F_X(\overline{f}_1(n))),\xrightarrow[n\to\infty]{}\infty\quad\text{threshold for }Y.$$

Remark 5 As in Section 4.2 in the following we propose to estimate the second threshold $\overline{f}_2(n)$ by $\widehat{\overline{f}}_2(n) := \widehat{F}_Y^{-1}(\widehat{F}_X(\overline{f}_1(n)))$, with $\widehat{F}_X(\overline{f}_1(n)) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leq \overline{f}_1(n)\}}$ and $\widehat{F}_Y^{-1}$ the empirical quantile function of Y.

In the following we state and prove separately our consistency result in the asymptotic dependent case (Theorem 6.1) and in the asymptotic independent one (Theorem 6.2).

6.1. Asymptotic dependent case

The proof of Theorem 6.1 below, makes use of a result by Einmahl *et al.* (2006) which specifies the asymptotic behavior of $\hat{l}(x,y)$ uniformly in $0 \leq x, y \leq 1$ and provides a convergence rate (see Theorem C in Appendix). More precisely in the asymptotic dependent case, using (18) and applying Corollary 3.1, we obtain the following main result:

Theorem 6.1 *Suppose F_X and F_Y belong to the maximum domain of attraction of Fréchet, L_X, L_Y satisfy Condition SR2. Assume that $\lambda > 0$ and that the assumptions of Theorem 2.1 and Corollary 3.1 are satisfied. If sequences $f_1(n)$, $f_2(n)$, $\overline{f}_1(n)$, $\overline{f}_2(n)$, defined by Remark 4, satisfy conditions of Theorem 5.1 then*

$$\left|\sqrt{k_n}(F^*(x_n,y_n) - \widetilde{F}^*(x_n,y_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \qquad (37)$$

with $x_n = \overline{f}_1(n)f_1(n)$, $y_n = \overline{f}_2(n)f_2(n)$. Moreover if $\widehat{\overline{f}}_2(n)$ satisfies conditions of Theorem 5.1 in probability then

$$\left|\sqrt{k_n}(F^*(x_n,\widehat{y}_n) - \widehat{F}^*(x_n,\widehat{y}_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \qquad (38)$$

with $\widehat{y}_n = \widehat{\overline{f}}_2(n)f_2(n)$. In (37)-(38) we have $k_n \to \infty$, $k_n/n \to 0$, $k_n/N_X \xrightarrow{\mathbb{P}} 0$, $k_n/N_Y \xrightarrow{\mathbb{P}} 0$, $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$, $\alpha > 0$.

The proof of Theorem 6.1 is postponed to the Appendix.

Remark 6 Let us study, on a class of examples, the assumption of Theorem 6.1. First if we suppose that the function $\phi(x)$ in Condition SR2 (Section 5) has the general form $\phi(x) = x^\rho$, with $\rho \leq 0$, then

$$(z_n)^{-s\,\rho}\,\phi(\widehat{\overline{f}}_2(n)(z_n)^s)\,/\,\phi(\widehat{\overline{f}}_2(n)) = 1, \quad \forall\ s\ \in [0,1].$$

For instance this is the case of Burr or Fréchet univariate distributions. Furthermore if we assume that F_Y belongs to the maximum domain of attraction of Fréchet (i.e. $\overline{F}_Y(y) = y^{-\alpha}L(y)$), F_Y has positive density $f_Y \in R_{-1-\alpha}$ and $\overline{f}_2(n)$ satisfies conditions in (32)-(35) then also the estimated second threshold $\widehat{\overline{f}}_2(n)$ satisfies, in probability, conditions in (32)-(35).

We remark indeed that $\widehat{F}_X(\overline{f}_1(n))$ is a high quantile within the sample (see Embrechts *et al.*, 1997), i.e. $\widehat{F}_X(\overline{f}_1(n)) \xrightarrow[n\to\infty]{\mathbb{P}} 1$ and $n(1-\widehat{F}_X(\overline{f}_1(n))) \xrightarrow[n\to\infty]{\mathbb{P}} \infty$. Then, using Theorem 6.4.14 in Embrechts *et al.* (1997) and a Skorohod-type construction of probability spaces we obtain $\widehat{\overline{f}}_2(n)\ \left(\overline{f}_2(n)\right)^{-1} \xrightarrow[n\to\infty]{\mathbb{P}} 1$.
Furthermore, using Condition SR2,
$\frac{\overline{F}_Y(\widehat{\overline{f}}_2(n))}{\overline{F}_Y(\overline{f}_2(n))} = \frac{\widehat{\overline{f}}_2(n)^{-\alpha}}{\overline{f}_2(n)^{-\alpha}} \frac{L(\widehat{\overline{f}}_2(n))}{L(\overline{f}_2(n))} = \frac{\widehat{\overline{f}}_2(n)^{-\alpha}}{\overline{f}_2(n)^{-\alpha}} \left[1 + k\left(\frac{\widehat{\overline{f}}_2(n)}{\overline{f}_2(n)}\right)\phi(\overline{f}_2(n)) + o(\phi(\overline{f}_2(n)))\right].$
Using properties of k and ϕ (see Section 5) we obtain $\frac{\overline{F}_Y(\widehat{\overline{f}}_2(n))}{\overline{F}_Y(\overline{f}_2(n))} \xrightarrow[n\to\infty]{\mathbb{P}} 1$.
Then $\widehat{\overline{f}}_2(n)$ satisfies, in probability, condition in (32):

$$n(1 - F_Y(\widehat{\overline{f}}_2(n))) = \frac{\overline{F}_Y(\widehat{\overline{f}}_2(n))}{\overline{F}_Y(\overline{f}_2(n))}\, n(1 - F_Y(\overline{f}_2(n))) \xrightarrow[n\to\infty]{\mathbb{P}} \infty.$$

The proof for conditions in (33)-(35) is completely analogue to that of condition in (32).

6.2. Asymptotic independent case

As noticed in Section 3.2 in the asymptotic independent case we need to introduce a second-order refinement of condition in (8). Under condition in (21) we obtain the following main result:

Theorem 6.2 *Suppose F_X and F_Y belong to the maximum domain of attraction of Fréchet, L_X, L_Y satisfy Condition SR2. Assume that the assumptions of Theorem 2.1, Proposition 3.1 and Corollary 7.1 are satisfied. If sequences $f_1(n)$, $f_2(n)$, $\overline{f}_1(n)$, $\overline{f}_2(n)$, defined by Remark 4, satisfy conditions of Theorem 5.1 then*

$$\left|\sqrt{a_n}\,(F^*(x_n, y_n) - \widetilde{F}^*(x_n, y_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \tag{39}$$

where $x_n = \overline{f}_1(n) f_1(n)$, $y_n = \overline{f}_2(n) f_2(n)$. Moreover if $\widehat{\overline{f}}_2(n)$ satisfies conditions of Theorem 5.1 in probability then

$$\left|\sqrt{a_n}\,(F^*(x_n, \widehat{y}_n) - \widehat{F}^*(x_n, \widehat{y}_n))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \tag{40}$$

with $\widehat{y}_n = \widehat{\overline{f}}_2(n) f_2(n)$. In (39)-(40) we have $a_n = n\, q(k_n/n) \to \infty$ (this implies $k_n \to \infty$), $k_n/n \to 0$, $\sqrt{a_n}\, q_1(q^{\leftarrow}(a_n/n)) \to 0$, $k_n/N_X \xrightarrow{\mathbb{P}} 0$, $k_n/N_Y \xrightarrow{\mathbb{P}} 0$, and $k_n = o(n^{\frac{2\alpha}{1+2\alpha}})$, for some $\alpha > 0$.

The proof of Theorem 6.2 is postponed to the Appendix.

7. Illustrations with real data

In this section we present four real cases (see Figures 4-5) for which we estimate bivariate tail probabilities to illustrate the finite sample properties of our estimator. We analyze the stability of our estimation compared to the one of $\widehat{\mathfrak{F}}_1^*$, as well the estimation of parameter θ of these real cases.

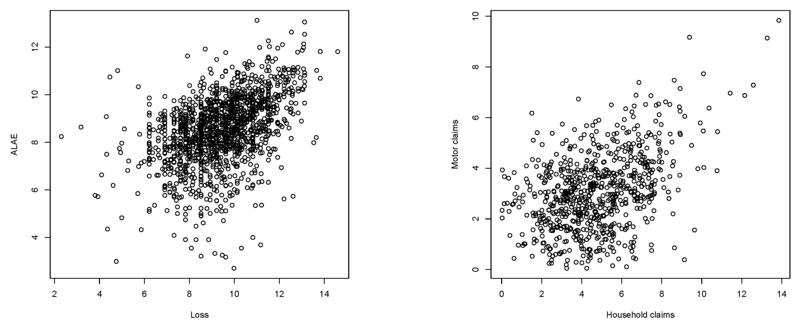

Figure 4: Logarithmic scale (left) ALAE versus Loss; (right) Storm damages.

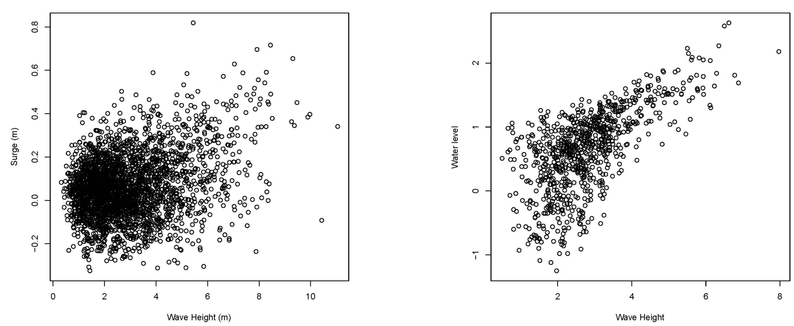

Figure 5: (left) Wave Height (m) versus Surge (m); (right) Wave heights versus Water level.

We consider the **Loss-ALAE data** (for details see Frees and Valdez, 1998). Each claim consists of an indemnity payment (the loss, X) and an allocated loss adjustment expense (ALAE, Y). We estimate $F(2.10^5, 10^5)$. The empirical probability, defined by (27), is 0.9506667 and the survival empirical probability is 0.006 (for a comparison using the Ledford and Tawn's model see Beirlant *et al.*, 2011). Figure 6 shows the sensitivity of $\widehat{\theta}$ and $\widehat{F}^*$ to the sequence k_n and provides a comparison with the estimator $\widehat{\mathfrak{F}}_1^*$.

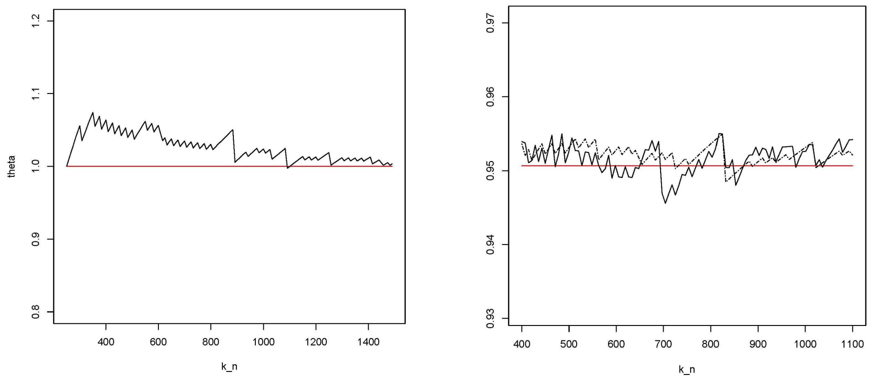

Figure 6: (left) $\widehat{\theta}_{0.04}$; (right) $\widehat{F}^*(2.10^5, 10^5)$ (full line), $\widehat{\mathfrak{F}}_1^*(2.10^5, 10^5)$ (dashed line), with the empirical probability indicated with a horizontal line; Loss-ALAE data.

We now consider an **example from storm insurance**: aggregate claims of motor policies (Y) and aggregate claims of household policies (X) from a French insurance portfolio for 736 storm events (for a detailed description see Lescourret and Robert, 2006). We estimate $F(8000, 950)$. The empirical probability is 0.96875 and the survival empirical probability is 0.014. The stability of our estimation compared to the one of $\widehat{\mathfrak{F}}_1^*$, as well the estimation of parameter θ are presented in Figure 7.

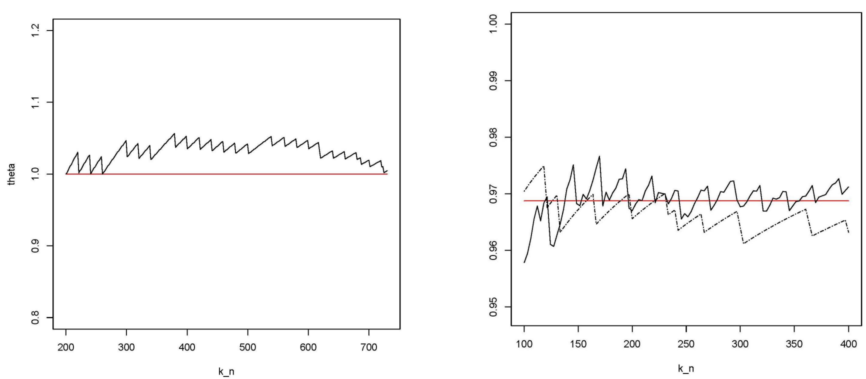

Figure 7: (left) $\widehat{\theta}_{0.05}$; (right) $\widehat{F}^*(8.10^3, 950)$ (full line), $\widehat{\mathfrak{F}}_1^*(8.10^3, 950)$ (dashed line), with the empirical probability indicated with a horizontal line; Storm insurance data.

We study the **wave surge data** comprising 2894 bivariate events that occurred during $1971 - 1977$ in Cornwall (England) (for details see Coles and Tawn, 1994 or Ramos and Ledford, 2009). We estimate $F(8.32, 0.51)$. The empirical probability is 0.9903 and the survival empirical probability is 0.00069. The sensitivity of $\widehat{\theta}$ and $\widehat{F}^*$ to the sequence k_n and the estimation of θ are presented in Figure 8.

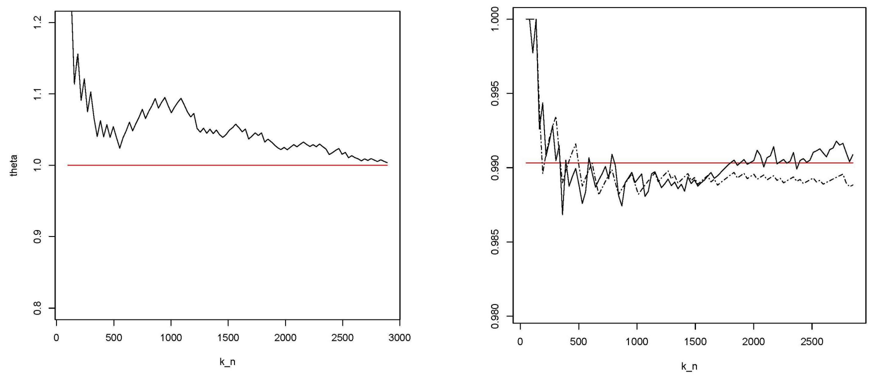

Figure 8: (left) $\widehat{\theta}_{0.02}$; (right) $\widehat{F}^*(8.32, 0.51)$ (full line), $\widehat{\mathfrak{F}}_1^*(8.32, 0.51)$ (dashed line), with the empirical probability indicated with a horizontal line; Wave-Surge data.

Finally we analyze the **Wave height versus Water level data**, recorded during 828 storm events spread over 13 years in front of the Dutch coast near the town of Petten (for details see Draisma *et al.*, 2004). We estimate $F(5.93, 1.87)$. The empirical probability is 0.97584 and the survival empirical probability is 0.00604. The sensitivity of $\widehat{\theta}$ and $\widehat{F}^*$ to the sequence k_n and the estimation of θ are presented in Figure 9. From Draisma *et al.* (2004) it seems that the coefficient η of Ledford and Tawn's model is smaller than 1, then it is plausible to assume asymptotic independence between the wave heights and the water level. Analogously, in our model the estimated parameter $\widehat{\theta}$ is greater than one (see Figure 9).

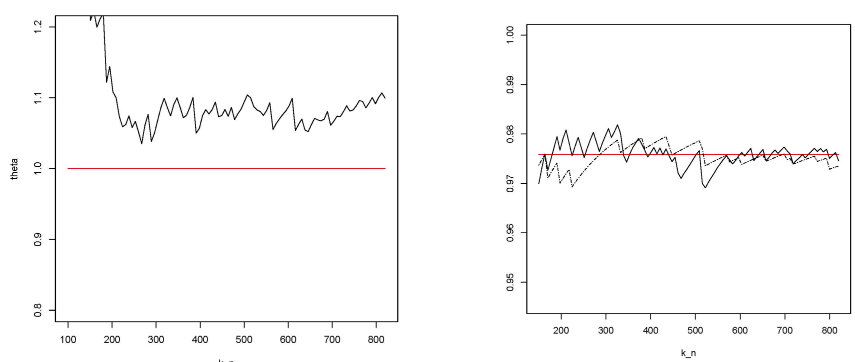

Figure 9: (left) $\widehat{\theta}_{\frac{0.1}{0.11}} = \widehat{\theta}_{0.91}$ as in (19); (right) $\widehat{F}^*(5.93, 1.87)$ (full line), $\widehat{\mathfrak{F}}_1^*(5.93, 1.87)$ (dashed line), with the empirical probability indicated with a horizontal line; Wave height-Water level data.

Appendix: proofs and auxiliary results

Proof [**Proposition 2.1**] :
We know that
$C_u^{up}(x,y) = 1 - \frac{C^*(1-\overline{F}_{X,u}^{-1}(x), 1-u)}{C^*(1-u,1-u)} - \frac{C^*(1-u, 1-\overline{F}_{Y,u}^{-1}(y))}{C^*(1-u,1-u)} + \frac{C^*(1-\overline{F}_{X,u}^{-1}(x), 1-\overline{F}_{Y,u}^{-1}(y))}{C^*(1-u,1-u)}$.

Then $$\lim_{u\to 1} C_u^{up}(x,y) = \lim_{u\to 1}\left[x + y - 1 + \frac{C^*(1-\overline{F}_{X,u}^{-1}(x),\, 1-\overline{F}_{Y,u}^{-1}(y))}{C^*(1-u,\,1-u)}\right]. \quad (41)$$

We introduce the following lemma.

Lemma A *(Charpentier and Juri, 2006; Lemma 6.1) Suppose that the random vectors (X_n, Y_n) have continuous, strictly increasing marginals and are such that $\lim_{n\to\infty}(X_n, Y_n) = (X, Y)$ in distribution for some (X, Y). Then*

$$\lim_{n\to\infty} ||C_n - C||_\infty = 0,$$

where C_n and C denote the copulas of (X_n, Y_n) and (X, Y), respectively.

Let (X, Y) have distribution function C. Note that

$$\mathbb{P}\left[X > x(1-u)\,|\,X > u, Y > u\right] = \frac{C^*(1-x\,(1-u),\,1-u)}{C^*(1-u,\,1-u)}, \quad (42)$$

$$\mathbb{P}\left[Y > y(1-u)\,|\,X > u, Y > u\right] = \frac{C^*(1-u,\,1-y\,(1-u))}{C^*(1-u,\,1-u)}, \quad (43)$$

$$\mathbb{P}\left[X > x(1-u),\, Y > y(1-u)\,|\,X > u, Y > u\right] = \frac{C^*(1-x\,(1-u),\,1-y\,(1-u))}{C^*(1-u,\,1-u)}. \quad (44)$$

The distributions in (42)-(44) are respectively the survival conditional distributions of $\frac{X}{1-u}$, $\frac{Y}{1-u}$ and $\left(\frac{X}{1-u}, \frac{Y}{1-u}\right)$, given that $X > u$ and $Y > u$. Since $\partial C^*(1-u, 1-v)/\partial u < 0$ and $\partial C^*(1-u, 1-v)/\partial v < 0$, for all $u, v \in [0, 1)$, it follows that the distributions in (42)-(43) are continuous and strictly increasing. By hypothesis, we have

$$\lim_{u\to 1} \frac{C^*(x\,(1-u),\, y\,(1-u))}{C^*(1-u,\,1-u)} = G(x,y), \quad \text{for all } x,\, y > 0, \quad (45)$$

implying that the expressions in (42)-(43) respectively converge to $g_X(1-x)$ and $g_Y(1-y)$ as $u \to 1$, with $g_X(x) := G(x, 1)$, $g_Y(y) := G(1, y)$.
Since copulas are invariant under strictly increasing transformations of the underlying variables, it follows that we can use the conditional distributions in

(42)-(43), instead of $\overline{F}_{X,\,u}$ and $\overline{F}_{Y,\,u}$, to construct $C_u^{up}(x,y)$. Then, from (41) and using Lemma A, we have

$$\lim_{u\to 1} C_u^{up}(x,y) = \lim_{u\to 1}\left[x+y-1+\frac{C^*(g_X^{-1}(1-x)\,(1-u), g_Y^{-1}(1-y)\,(1-u))}{C^*(1-u,\,1-u)}\right]$$
$$= x\,+\,y\,-\,1\,+\,G(g_X^{-1}(1-x),\,g_Y^{-1}(1-y)).$$

As in the proof of Theorem 3.1 in Charpentier and Juri (2006), the limit in (45) implies that there is a $\theta>0$ such that G is homogeneous of order θ, i.e. for all $t,\,x,\,y\,>\,0$,

$$G(t\,x,\,t\,y) = t^\theta\,G(x,\,y). \tag{46}$$

By a discussion of the general solution of functional (46) we obtain the explicit form of G:

$$G(x,y) = \begin{cases} x^\theta g_Y(\frac{y}{x}) & \text{for } \frac{y}{x}\in[0,1],\\ y^\theta g_X(\frac{x}{y}) & \text{for } \frac{x}{y}\in(1,\infty).\end{cases}$$

For this part of the proof we refer the interested reader to Theorem 3.1 in Charpentier and Juri, (2006). □

Proof [**Theorem 2.1**] :
From (11) we obtain the existence of $a_1(\cdot)$ and $a_2(\cdot)$ such that, for $p := u+x\,a_1(u)$ and $q := u_Y + y\,a_2(u_Y)$

$$V_{\xi_1,1}(x) = \lim_{u\to x_{F_X}} 1-\frac{1-F_X(p)}{1-F_X(u)} = \lim_{u\to x_{F_X}} \mathbb{P}[X\le p|X>u], \tag{47}$$

$$V_{\xi_2,1}(y) = \lim_{u_Y\to x_{F_Y}} 1-\frac{1-F_Y(q)}{1-F_Y(u_Y)} = \lim_{u_Y\to x_{F_Y}} \mathbb{P}[Y\le q|Y>u_Y]. \tag{48}$$

From $Y \stackrel{d}{=} F_Y^{-1}(F_X(X))$, we take $u_Y = F_Y^{-1}(F_X(u))$ and from (47)-(48), as $u\to x_{F_X}$, we have

$$1-(1-V_{\xi_1,1}(x))(1-F_X(u)) \sim F_X(u+x\,a_1(u)),$$

$$1-(1-V_{\xi_2,1}(y))(1-F_Y(F_Y^{-1}(F_X(u)))) \sim F_Y(F_Y^{-1}(F_X(u))+y\,a_2(F_Y^{-1}(F_X(u)))).$$

Then

$$\lim_{u\to x_{F_X}} \mathbb{P}\left[\frac{X-u}{a_1(u)}>x,\,\frac{Y-F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))}>y\,\middle|\,X>u, Y>F_Y^{-1}(F_X(u))\right]$$
$$= \lim_{u\to x_{F_X}} \frac{C^*\big(1-F_X(u+x\,a_1(u)), 1-F_Y(F_Y^{-1}(F_X(u))+y\,a_2(F_Y^{-1}(F_X(u))))\big)}{C^*\big(1-F_X(u), 1-F_Y(F_Y^{-1}(F_X(u)))\big)}$$
$$= \lim_{u\to x_{F_X}} \frac{C^*\big((1-V_{\xi_1,1}(x))(1-F_X(u)), (1-V_{\xi_2,1}(y))(1-F_Y(F_Y^{-1}(F_X(u))))\big)}{C^*\big(1-F_X(u), 1-F_Y(F_Y^{-1}(F_X(u)))\big)}$$
$$= \lim_{\nu\to 1} \frac{C^*\big((1-V_{\xi_1,1}(x))(1-\nu), (1-V_{\xi_2,1}(y))(1-\nu)\big)}{C^*\big(1-\nu, 1-\nu\big)}. \tag{49}$$

Now, if $h := (1-\xi_1\, x)^{\frac{1}{\xi_1}}$, $\xi_1 \neq 0$ or if $h := \mathrm{e}^{-x}$, $\xi_1 = 0$ then $1-V_{\xi_1,1}(x) = V_{1,1}(h)$. So (49) becomes $\lim_{\nu\to 1} C^*\big(V_{1,1}(h)(1-\nu), V_{1,1}(w)(1-\nu)\big)/C^*\big(1-\nu, 1-\nu\big)$. As C satisfies hypotheses of Proposition 2.1, the above limit is equal to $G(V_{1,1}(h), V_{1,1}(w)) = G(1-V_{\xi_1,1}(x), 1-V_{\xi_2,1}(y))$. Then

$$\lim_{u\to x_{F_X}} \mathbb{P}\bigg[\frac{X-u}{a_1(u)} \leq x,\ \frac{Y-F_Y^{-1}(F_X(u))}{a_2(F_Y^{-1}(F_X(u)))} \leq y \,\bigg|\, X > u, Y > F_Y^{-1}(F_X(u))\bigg]$$
$$= C^{*\,G}\big(1-g_X(1-V_{\xi_1,1}(x)), 1-g_Y(1-V_{\xi_2,1}(y))\big). \quad (50)$$

Since the limit is a continuous distribution function (as $C^{*\,G}$, g and the GPD are), (50) can be strengthened to uniform convergence (see e.g. Embrechts *et al.* 1997, p. 552). Then (12) follows. $\square$

Proof [**Theorem 5.1**]**:**
To begin with, we work conditionally on $N_n = m_n$. First we have to prove that

$$\widetilde{r}_{m_n}\left[F(u_{m_n}\, z_{m_n}) - \widehat{F}^*(u_{m_n}\, z_{m_n})\right] \xrightarrow[n\to\infty]{d} \mathcal{N}(\nu, \tau^2), \quad (51)$$

with $\widetilde{r}_{m_n} = \frac{\sqrt{m_n}}{\log(z_{m_n})}\left(\frac{1}{1-\frac{1}{n}\sum_{i=1}^n 1_{(X_i \leq u_{m_n} z_{m_n})}}\right) = \frac{\sqrt{m_n}}{\log(z_{m_n})\,\widehat{\overline{F}}(u_{m_n}\, z_{m_n})}$.

To this end we need to prove that

$$\frac{\overline{F}(u_{m_n}\, z_{m_n})}{\widehat{\overline{F}}(u_{m_n}\, z_{m_n})} \xrightarrow[n\to\infty]{\mathbb{P}} 1, \quad (52)$$

then, using Theorem 8.1 in Smith (1987) and the Slutsky theorem, we obtain (51). To prove (52) we use the following result:

Proposition B *(Einmahl, 1990; Corollary 1) Let a sequence of i.i.d random variables* $X_1, X_2, \dots$ *from a distribution function* F*. We denote with* $\{k_n\}_{n=1}^\infty$ *an arbitrary sequence of positive numbers, such that* $k_n \leq n$ *and* $k_n \to \infty$, $\lim_{n\to\infty} \frac{k_n}{n} = 0$*. Let* $\{\gamma_n\}_{n=1}^\infty$ *be a sequence of positive numbers, such that* $\lim_{n\to\infty} \frac{\gamma_n}{\sqrt{k_n}} = \infty$*, then* $\sup_{t\geq F^{-1}(1-\frac{k_n}{n})}\left(\frac{n}{\gamma_n}\right)\left|\widehat{\overline{F}}(t) - \overline{F}(t)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

We choose an arbitrary sequence $\{k_n\}_{n=1}^\infty := \{m_n\}_{n=1}^\infty$ (number of excesses on a sample of size n) such that $m_n \leq n$, $\lim_{n\to\infty} m_n = \infty$ and $\lim_{n\to\infty} \frac{m_n}{n} = 0$ (see Remark 1 in Section 1). We take $\{\gamma_n\}_{n=1}^\infty := \{\sqrt{m_n}\,\alpha_n\}_{n=1}^\infty$, where α_n is an arbitrary sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = \infty$.
Then, using Proposition B, we have for $u_{m_n}\, z_{m_n} \geq F^{-1}(1-\frac{m_n}{n})$

$$\left(\frac{n}{\sqrt{m_n}\,\alpha_n}\,\overline{F}(u_{m_n}\, z_{m_n})\right)\left|\frac{\widehat{\overline{F}}(u_{m_n}\, z_{m_n}) - \overline{F}(u_{m_n}\, z_{m_n})}{\overline{F}(u_{m_n}\, z_{m_n})}\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

We choose α_n such that for large n

$$\exists\, c > 0 : \quad 0 < \frac{\sqrt{m_n}\,\alpha_n}{n\,\overline{F}(u_{m_n}\, z_{m_n})} \leq c. \quad (53)$$

In the Fréchet case we have $L(x) = x^{\alpha}\,\overline{F}(x)$, for $\alpha > 0$ and $\forall\ t > 0$, $\frac{L(tx)}{L(x)} = 1 + k(t)\phi(x) + o(\phi(x))$ for $x \to \infty$. Then, using Assumptions (8.7) and (8.8) of Theorem 8.1. in Smith (1987), we obtain

$$\frac{\overline{F}(u_{m_n}\, z_{m_n})}{\overline{F}(u_{m_n})} = z_{m_n}^{-\alpha}\left[1 + k(z_{m_n})\phi(u_{m_n}) + o(\phi(u_{m_n}))\right].$$

Hence $\frac{n\,\overline{F}(u_{m_n}\, z_{m_n})}{\sqrt{m_n}}$ is equal to $\frac{n}{\sqrt{m_n}}\,\overline{F}\,(u_{m_n})\big[z_{m_n}^{-\alpha}\big(1{+}k(z_{m_n})\phi(u_{m_n}){+}o(\phi(u_{m_n}))\big)\big]$ which, for n large, can be approximated by

$$\sqrt{m_n} z_{m_n}^{-\alpha}\left(1 + k(z_{m_n})\phi(u_{m_n}) + o(\phi(u_{m_n}))\right). \tag{54}$$

Assume now $\frac{(z_{m_n})^{\alpha}}{\sqrt{m_n}} \xrightarrow[n\to\infty]{} 0$, that is the analogue of condition in (35) conditionally on $N_n = m_n$. Then the properties of k and ϕ insure that the right hand side of (54) increases to infinity hence one can choose α_n satisfying (53). To conclude the proof, we use assumption (35) and a Skorohod type argument. $\square$

Proof [**Theorem 6.1**]**:**
To prove (37) we first observe, using Corollary 3.1, Proposition 7.1 and the analogue of Kolmogorov-Smirnov Theorem in dimension 2 (e.g. see Dudley, 1966), that

$$\sqrt{k_n}\,\Bigg|\,C^{*\,G}\big(1{-}g_X(1{-}V_{\xi_X,\sigma_X}(\overline{f}_1(n)f_1(n){-}\overline{f}_1(n))),1{-}g_Y(1{-}V_{\xi_Y,\sigma_Y}(\overline{f}_2(n)f_2(n){-}\overline{f}_2(n))))\,\cdot\overline{F}(\overline{f}_1(n),\overline{f}_2(n)){-}\widehat{\overline{F}}(\overline{f}_1(n),\overline{f}_2(n)){\cdot}\,C^{*\,G}\Bigg(1{-}\widehat{g}_X(1{-}V_{\widehat{\xi}_X,\widehat{\sigma}_X}(\overline{f}_1(n)f_1(n){-}\overline{f}_1(n))),\ 1-\widehat{g}_Y(1-V_{\widehat{\xi}_Y,\widehat{\sigma}_Y}(\overline{f}_2(n)f_2(n)-\overline{f}_2(n)))\Bigg)\Bigg|\xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Furthermore $r_n\left|\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\le\overline{f}_1(n),\,Y_i\le\overline{f}_2(n)\}} - F(\overline{f}_1(n),\overline{f}_2(n))\right|\xrightarrow[n\to\infty]{\mathbb{P}} 0$, with $r_n << \sqrt{n}$. At last using Corollary 3.1, Theorem 5.1, we obtain convergence (37). If $\widehat{\overline{f}}_2(n)$ satisfies conditions of Theorem 5.1 in probability then with the same proof structure we obtain (38). $\square$

Proof [**Proposition 3.1**]**:** Under assumptions of Proposition 3.1, as in the proof of Lemma 6.1 in Draisma *et al.* (2004) we obtain

$$\sup_{0<x,y\le 1}\left|\sqrt{a_n}\left(\frac{1}{a_n}\sum_{i=1}^{n} 1_{\{R(X_i)>n-k_nx+1;\,R(Y_i)>n-k_ny+1\}} - G(x,y)\right) - W(x,y)\right|\xrightarrow[n\to\infty]{a.s.} 0,$$

where $a_n = n\,q(k_n/n)$ and $W(x,y)$ is a zero-mean gaussian process with $\mathbb{E}(W(x_1,y_1)W(x_2,y_2)) = G(x_1\wedge x_2, y_1\wedge y_2)$. Then, in particular

$$\psi_n \sup_{0<x,y\le 1}\left|\frac{\sum_{i=1}^{n}\frac{1}{k_n}1_{\{R(X_i)>n-k_n\,x+1;\,R(Y_i)>n-k_n\,y+1\}}}{\sum_{i=1}^{n}\frac{1}{k_n}1_{\{R(X_i)>n-k_n+1;\,R(Y_i)>n-k_n+1\}}} - G(x,y)\right| = \psi_n \sup_{0<x,y\le 1}\left|\,\widehat{G}(x,y)\ -\ G(x,y)\,\right|\xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\psi_n << \sqrt{a_n} = \sqrt{n\, q(k_n/n)}$ and $\widehat{G}$ as in (18). Finally for the marginals g_X and g_Y we have

$$\psi_n \sup_{0<x\leq 1} \left|\widehat{g}_X(x) - g_X(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \psi_n \sup_{0<y\leq 1} \left|\widehat{g}_Y(y) - g_Y(y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\widehat{g}_X$ and $\widehat{g}_Y$ as in (18). $\square$

Proof [**Theorem 6.2**]**:**
Under assumptions of Theorem 6.2 and Proposition 3.1 we obtain asymptotic convergence results for $\widehat{G}(x,y)$, $\widehat{g}_X(x)$ and $\widehat{g}_Y(y)$, with convergence rate $\psi_n << \sqrt{n\, q(k_n/n)}$ and $\widehat{g}_X,\ \widehat{g}_Y,\ \widehat{G}$ as in (18).
With the same proof structure of Theorem 6.1, using Corollary 7.1 and Proposition 7.1 we obtain convergence (39). Moreover if $\widehat{\widehat{f}}_2(n)$ satisfies conditions of Theorem 5.1 in probability then we obtain (40). $\square$

Auxiliary results

Theorem C *(Einmahl et al. 2006; Theorem 2.2) Assume that exists the limit $R(x,y)$ in* (14) *such that, for some $\alpha > 0$*

$$\frac{1}{t}\mathbb{P}(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty) - R(x,y) = O(t^\alpha), \ \text{as } t \to 0, \qquad (55)$$

uniformly for $\max(x,y) \leq 1,\ x,y \geq 0$. *Let $k_n \to \infty$, $k_n/n \to 0$ and $k_n = o\big(n^{\frac{2\alpha}{1+2\alpha}}\big)$. If $R_1(x,y) := \frac{\partial R(x,y)}{\partial x}$ and $R_2(x,y) := \frac{\partial R(x,y)}{\partial y}$ are continuous then*

$$\sup_{0<x,y\leq 1} \left|\sqrt{k_n}(\widehat{l}(x,y) - l(x,y)) + B(x,y)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

where $B(x,y) := W(x,y) - R_1(x,y)W_1(x) - R_2(x,y)W_2(y)$, with W a continuous mean zero Gaussian process on $[0,x]\times[0,y]$ with covariance structure $\mathbb{E}(W(x_1,y_1)\,W(x_2,y_2)) = R(x_1\wedge x_2, y_1\wedge y_2)$ and with marginal processes defined by $W_1(x) = W([0,x]\times[0,\infty])$, $W_2(y) = W([0,\infty]\times[0,y])$.

Note that (55) is a second-order condition quantifying the speed of convergence in (14) and condition $k_n = o\big(n^{\frac{2\alpha}{1+2\alpha}}\big)$ gives an upper bound on the speed with which k_n can grow to infinity. This upper bound is related to the speed of convergence in (55) by α. If C is a twice continuously differentiable copula on $[0,\,1]^2$ then (55) holds for any $\alpha \geq 1$. Furthermore, it is easily seen that $\widehat{l}(x,y) + \widehat{R}(x,y) = \frac{\lceil k_n\, x\rceil + \lceil k_n\, y\rceil - 2}{k_n} \leq \frac{\lceil k_n\, x\rceil + \lceil k_n\, y\rceil}{k_n}$, almost surely, for each $0 < x,y \leq 1$, where $\lceil z\rceil$ is the smallest integer $\geq z$. Then under assumption of Theorem C we can easily obtain a gaussian approximation for $\widehat{R}(x,y) - R(x,y)$. Note that the asymptotic variance of $\sqrt{k_n}(\widehat{l}(x,y) - l(x,y))$, in Theorem C, vanishes in the asymptotic independent case. Then, with $\lambda = 0$, we obtain:

Corollary 7.1 *Assume that, for some $\alpha > 0$*

$$\frac{1}{t}\mathbb{P}(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty) = O(t^\alpha), \ \text{as } t \to 0,$$

uniformly for $\max(x,y) \leq 1,\ x,y \geq 0$. *Let* $k_n \to \infty$, $k_n/n \to 0$ *and* $k_n = o\big(n^{\frac{2\alpha}{1+2\alpha}}\big)$. *Then it holds*

$$\sup_{0<x,y\leq 1} \left|\sqrt{k_n}(\widehat{l}(x,y) - l(x,y))\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proposition 7.1 *Let* $V_{\xi,\sigma}(x)$ *the Generalized Pareto Distribution (GPD) and* $\widehat{\xi}_n$, $\widehat{\sigma}_n$, *the maximum likelihood estimators of the parameters* $\xi = -\alpha^{-1} < 0$ *and* $\sigma = u_n\alpha^{-1}$, *in the case unconditionally on* N. *If all the conditions of Corollary 5.1 hold then*

$$p_{nx} \sup_{x\in[0,+\infty)} \left|V_{\widehat{\xi}_n,\widehat{\sigma}_n}(x) - V_{\xi,\sigma}(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \textit{where } \frac{p_{nx}}{\sqrt{N_x}} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proof: Using Corollary 5.1 we obtain for each point $x \in [0,+\infty)$,

$$p_{nx}\left[V_{\widehat{\xi}_n,\widehat{\sigma}_n}(x) - V_{\xi,\sigma}(x)\right] = p_{nx}\left[\left(1-\frac{\xi\, x}{\sigma}\right)^{\frac{1}{\xi}} - \left(1-\frac{\widehat{\xi}_n\, x}{\widehat{\sigma}_n}\right)^{\frac{1}{\widehat{\xi}_n}}\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0, \tag{56}$$

where $\frac{p_{nx}}{\sqrt{N_x}} \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Finally, applying a stochastic version of Polya's Theorem (see Horowitz, 2001), as $V_{\xi,\sigma}(x)$ is a continuous distribution function, the convergence in (56) holds uniformly on $[0,+\infty)$. □

Acknowledgements This work has been partially supported by the French research national agency (ANR) under the reference ANR-08BLAN-0314-01. The authors thank two anonymous referees for relevant remarks and constructive comments on a previous version of the paper. They also thank Anne-Laure Fougères, Ragnar Norberg and José R. León for fruitful discussions.

References

[1] A. A. Balkema and L. de Haan. Residual life time at great age. *Ann. Probability*, 2:792–804, 1974.

[2] J. Beirlant, G. Dierckx, and A. Guillou. Reduced bias estimators for bivariate tail modelling. *Insurance: Mathematics and Economics*, In Press, Accepted Manuscript, 2011.

[3] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

[4] F. E. Benth and P. Kettler. Dynamic copula models for the spark spread. *Quantitative Finance*, 11(3):407–421, 2011.

[5] P. Capéraà and A.-L. Fougères. Estimation of a bivariate extreme value distribution. *Extremes*, 3(4):311–329 (2001), 2000.

[6] A. Charpentier and A. Juri. Limiting dependence structures for tail events, with applications to credit derivatives. *J. Appl. Probab.*, 43(2):563–586, 2006.

[7] A. Charpentier and J. Segers. Lower tail dependence for Archimedean copulas: characterizations and pitfalls. *Insurance Math. Econom.*, 40(3):525–532, 2007.

[8] S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: an application to structural design. *Appl. Statist.*, 43:1–48, 1994. With discussion.

[9] A. C. Davison. *A statistical model for contamination due to long-range atmospheric transport of radionuclides*. PhD thesis, Department of Mathematics, Imperial College of Science and Technology, London, 1984.

[10] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B*, 52(3):393–442, 1990. With discussion and a reply by the authors.

[11] L. de Haan. *On regular variation and its application to the weak convergence of sample extremes*, volume 32 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1970.

[12] G. Draisma, H. Drees, A. Ferreira, and L. de Haan. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280, 2004.

[13] H. Drees, A. Ferreira, and L. de Haan. On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.*, 14(3):1179–1201, 2004.

[14] H. Drees and X. Huang. Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.*, 64(1):25–47, 1998.

[15] R. M. Dudley. Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.*, 10:109–126, 1966.

[16] J. H. J. Einmahl. The empirical distribution function as a tail estimator. *Statist. Neerlandica*, 44(2):79–82, 1990.

[17] J. H. J. Einmahl, L. de Haan, and D. Li. Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.*, 34(4):1987–2014, 2006.

[18] J. H. J. Einmahl, A. Krajina, and J. Segers. A method of moments estimator of tail dependence. *Bernoulli*, 14(4):1003–1026, 2008.

[19] M.-A. El-Aroui and J. Diebolt. On the use of the peaks over thresholds method for estimating out-of-sample quantiles. *Comput. Statist. Data Anal.*, 39(4):453–475, 2002.

[20] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[21] M. Falk and R.-D. Reiss. On the distribution of Pickands coordinates in bivariate EV and GP models. *J. Multivariate Anal.*, 93(2):267–295, 2005.

[22] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25, 1998.

[23] C. M. Goldie and R. L. Smith. Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford Ser. (2)*, 38(149):45–71, 1987.

[24] J. L. Horowitz. *Handbook of Econometrics*, chapter The Bootstrap, pages 3159–3228. J.J. Heckman and E.E. Leamer, Elsevier Science B.V, 2001.

[25] X. Huang. *Statistics of bivariate extreme values*. PhD thesis, Erasmus University Rotterdam, Tinbergen Institute, 1992.

[26] A. Javid. Limiting tail dependence copulas. *Communications in Statistics-Theory and Methods*, 38(20):3772–3781, 2009.

[27] A. Juri and M. V. Wüthrich. Copula convergence theorems for tail events. *Insurance Math. Econom.*, 30(3):405–420, 2002.

[28] A. Juri and M. V. Wüthrich. Tail dependence from a distributional point of view. *Extremes*, 6(3):213–246 (2004), 2003.

[29] A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

[30] A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, 59(2):475–499, 1997.

[31] A. W. Ledford and J. A. Tawn. Concomitant tail behaviour for extremes. *Adv. in Appl. Probab.*, 30(1):197–215, 1998.

[32] L. Lescourret and C. Y. Robert. Extreme dependence of multivariate catastrophic losses. *Scandinavian Actuarial Journal*, (4):203–225, 2006.

[33] A. J. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27:1117–1137, 1997.

[34] A. J. McNeil. Extreme value theory for risk managers. *preprint, ETH, Zurich*, 1999.

[35] R. Michel. Some notes on multivariate generalized pareto distributions. *J. Multivariate Anal.*, 99(6):1288–1301, 2008.

[36] R. B. Nelsen. *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1999.

[37] L. Peng. Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.*, 43(4):399–409, 1999.

[38] J. Pickands, III. Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131, 1975.

[39] J. Pickands, III. Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49, pages 859–878, 894–902, 1981. With a discussion.

[40] A. Ramos and A. Ledford. A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(1):219–241, 2009.

[41] R.-D. Reiss and M. Thomas. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*. Birkhäuser Verlag, Basel, third edition, 2007. With 1 CD-ROM (Windows).

[42] S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987.

[43] H. Rootzén and N. Tajvidi. Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930, 2006.

[44] M. Schlather. Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statist. Probab. Lett.*, 53(3):325–329, 2001.

[45] R. L. Smith. Estimating tails of probability distributions. *Ann. Statist.*, 15(3):1174–1207, 1987.

[46] M. V. Wüthrich. Bivariate extension of the Pickands-Balkema-de Haan theorem. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(1):33–41, 2004.