Submitted to the Annals of Statistics

BIAS CORRECTION IN MULTIVARIATE EXTREMES

By Anne-Laure Fougères* Laurens de Haan† and Cécile Mercadier*

Université Lyon 1 and Erasmus University†*

The estimation of the extremal dependence structure is spoiled by the impact of the bias, which increases with the number of observations used for the estimation. Already known in the univariate setting, the bias correction procedure is studied in this paper under the multivariate framework. New families of estimators of the stable tail dependence function are obtained. They are asymptotically unbiased versions of the empirical estimator introduced by Huang (1992). Since the new estimators have a regular behaviour with respect to the number of observations, it is possible to deduce aggregated versions so that the choice of the threshold is substantially simplified. An extensive simulation study is provided as well as an application on real data.

1. Introduction. Estimating extreme risks in a multivariate framework is highly connected with the estimation of the extremal dependence structure. This structure can be described *via* the stable tail dependence function (stdf) L, firstly introduced by Huang (1992). For any arbitrary dimension d, consider a multivariate vector $(X^{(1)}, \dots, X^{(d)})$ with continuous marginal cumulative distribution functions (cdf) $F_1, \dots, F_d$. The stdf is defined for each positive reals $x_1, \dots, x_d$ as

$$\lim_{t\to\infty} t\,\mathbb{P}\{1 - F_1(X^{(1)}) \le t^{-1}x_1 \text{ or } \dots \text{ or } 1 - F_d(X^{(d)}) \le t^{-1}x_d\} = L(x_1, \dots, x_d)\,.$$

Assuming that such a limit exists and is non degenerate is equivalent to the classical assumption of existence of a multivariate domain of attraction for the componentwise maxima (see e.g. de Haan and Ferreira (2006, Chapter 7)). The previous limit can be rewritten as

$$\lim_{t\to\infty} t\,[1 - F\{F_1^{-1}(1 - t^{-1}x_1), \dots,\, F_d^{-1}(1 - t^{-1}x_d)\}] = L(x_1, \dots, x_d)\,, \tag{1}$$

where F denotes the multivariate cdf of the vector $(X^{(1)}, \dots, X^{(d)})$, and for $j = 1, \dots, d$, $F_j^{-1}(t) = \inf\{z \in \mathbb{R} : F_j(z) \ge t\}$. Consider a sample of size n drawn from F and an intermediate sequence, that is to say a sequence $k = k(n)$ tending to infinity as $n \to \infty$, with $k/n \to 0$. Denote by $\mathbf{x} = (x_1, \dots, x_d)$ a vector of the positive quadrant $\mathbb{R}_+^d = \{(x_1, \dots, x_d) : x_j \ge 0, j = 1, \dots, d\}$ and by $X_{k,n}^{(j)}$ the kth order statistics among n realisations of the margins $X^{(j)}$. The empirical estimator of $L(\mathbf{x})$ is obtained from (1), replacing F by its empirical version, t by n/k, and $F_j^{-1}(1 - t^{-1}x_j)$ for $j = 1, \dots, d$ by its empirical counterpart $X_{n-[nt^{-1}x_j],n}^{(j)}$, so that

$$\hat{L}_k(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\left\{X_i^{(1)} \ge X_{n-[kx_1]+1,n}^{(1)} \text{ or } \dots \text{ or } X_i^{(d)} \ge X_{n-[kx_d]+1,n}^{(d)}\right\}}\,. \tag{2}$$

MSC 2010 subject classifications: Primary 62G32, 62G05, 62G20; secondary 60F05, 60G70

Keywords and phrases: Multivariate extreme value theory, tail dependence, bias correction, threshold choice

See Huang (1992) for pioneer works on this estimator. Under suitable conditions, it can be shown (see Section 2) that the estimator $\hat{L}_k(\mathbf{x})$ has the following asymptotic expansion

$$\hat{L}_k(\mathbf{x}) - L(\mathbf{x}) \approx \frac{Z_L(\mathbf{x})}{\sqrt{k}} + \alpha(n/k)M(\mathbf{x}) \; , \tag{3}$$

where Z_L is a continuous centered Gaussian process, α is a function that tends to 0 at infinity, and M is a continuous function. In particular $\sqrt{k}\{\hat{L}_k(\mathbf{x}) - L(\mathbf{x})\}$ can be approximated in distribution by $Z_L(\mathbf{x})$, provided that $\sqrt{k}\alpha(n/k)$ tends to 0 as n tends to infinity. This condition imposes a slow rate of convergence of the estimator $\hat{L}_k(\mathbf{x})$, so one would be interested in relaxing this hypothesis. As a counterpart, as soon as $\sqrt{k}\alpha(n/k)$ tends to a non null constant λ, an asymptotic bias appears and is explicitely given by $\lambda M(\mathbf{x})$. The aim of this paper is to provide a procedure that reduces the asymptotic bias. The latter will be estimated and then subtracted from the empirical estimator. This kind of approach has been considered in the univariate setting for the bias correction of the extreme value index with unknown sign by Cai, de Haan and Zhou (2013). Refer also to Peng (1998), Fraga Alves, de Haan and Lin (2003), Gomes, de Haan and Rodrigues (2008), Caeiro, Gomes and Rodrigues (2009) and Peng (2010) for previous contributions on this problem. Note finally that the case of dependent sequences has been recently studied by de Haan, Mercadier and Zhou (2014).

The nonparametric estimation of the extremal dependence structure has been widely studied in the bivariate case, see for instance Huang (1992), Einmahl, de Haan and Sinha (1997), Capéraà and Fougères (2000), Abdous and Ghoudi (2005), Guillotte, Perron and Segers (2011) and Bücher, Dette and Volgushev (2011). Bias correction problems in the bivariate context received less attention than in the univariate setting. To the best of our knowledge, it seems to be reduced to Beirlant, Dierckx and Guillou (2011) and Goegebeur and Guillou (2013), which consider the estimation of bivariate joint tails, so differs slightly from our task.

As for the multivariate framework, de Haan and Resnick (1993) introduces the empirical estimator. General approaches under parametric assumptions on the function L have been developed e.g. by Coles and Tawn (1991), Joe, Smith and Weissman (1992), Einmahl, Krajina and Segers (2008) and Einmahl, Krajina and Segers (2012). Apparently, no procedure correcting the bias can be found in the literature for dimension greater than two. The objective of this article is to fill this gap. Note that our method does not only consists of applying the univariate bias procedure at several points. Indeed, the bias is not anymore a parametric function, so that the new feature is mainly the fact that we are able to estimate and then subtract a function with an unknown form. Two families of asymptotically unbiased estimators of the stdf are proposed and their theoretical behaviours are studied. A practical advantage of these new estimators is that they can be aggregated, reducing that way the variability.

The paper is organized as follows: Section 2 contains hypotheses and first results. The bias reduction procedure is described in Section 3, and the main theoretical results are presented therein. Several theoretical models are exhibited in Section 4, that satisfy the required assumptions. Section 5 illustrates the performance of the new estimators on both simulated and real data. The estimation of side components are postponed up to Section 6. The proofs are relegated to Section 7.

2. Notation, assumptions and first results. Let $\mathbf{X}_1 = (X_1^{(1)}, \ldots, X_1^{(d)}), \ldots, \mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(d)})$ be independent and identically distributed multivariate random vectors with cdf F and continuous marginal cdfs F_j for $j = 1, \ldots, d$. Suppose F is in the domain of attraction of an extreme value distribution with cdf G. We recall that it supposes the existence for $j = 1, \dots, d$ of sequences $a_n^{(j)} > 0$, $b_n^{(j)}$ of real numbers and a cdf G with nondegenerate marginals such that

$$\lim_{n\to\infty} \mathbb{P}(\max\{X_1^{(1)}, \dots, X_n^{(1)}\} \le a_n^{(1)} x_1 + b_n^{(1)}, \dots, \max\{X_1^{(d)}, \dots, X_n^{(d)}\} \le a_n^{(d)} x_d + b_n^{(d)}) = G(\mathbf{x})$$

for all points $\mathbf{x}$ where G is continuous. Denote by G_j the jth marginal cdf of G. It is possible to show that the domain of attraction condition can be expressed as the condition (1) along with the convergence of the marginal distributions to the G_j's, and that

$$L(\mathbf{x}) = -\log G\left(\{-\log G_1\}^{-1}(x_1), \dots, \{-\log G_d\}^{-1}(x_d)\right) . \tag{4}$$

Let μ be the measure defined by

$$\mu\{A(\mathbf{x})\} := L(\mathbf{x}) , \tag{5}$$

where $A(\mathbf{x}) := \{\mathbf{u} \in \mathbb{R}_+^d : \text{ there exists } j \text{ such that } u_j > x_j\}$ for any vector $\mathbf{x} \in \mathbb{R}_+^d$.

Several conditions are now described. The first two have been introduced by de Haan and Resnick (1993).

- the *first order condition* consists of assuming that the limit given in (1) exists, and that the convergence is uniform on any $[0, T]^d$, for $T > 0$.
- the *second order condition* consists of assuming the existence of a positive function α, such that $\alpha(t) \to 0$ as $t \to \infty$, and a non null function M such that for all $\mathbf{x}$ with positive coordinates,

$$\lim_{t\to\infty} \frac{1}{\alpha(t)} \left\{t\,[1 - F\{F_1^{-1}(1 - t^{-1}x_1), \dots,\, F_d^{-1}(1 - t^{-1}x_d)\}] - L(\mathbf{x})\right\} = M(\mathbf{x}) , \tag{6}$$

  uniformly on any $[0, T]^d$, for $T > 0$.
- the *third order condition* consists of assuming the existence of a positive function β, such that $\beta(t) \to 0$ as $t \to \infty$, and a non null function N such that for all $\mathbf{x}$ with positive coordinates,

$$\lim_{t\to\infty} \frac{1}{\beta(t)} \left\{ \frac{t\,[1 - F\{F_1^{-1}(1 - t^{-1}x_1), \dots,\, F_d^{-1}(1 - t^{-1}x_d)\}] - L(\mathbf{x})}{\alpha(t)} - M(\mathbf{x}) \right\} = N(\mathbf{x}) , \tag{7}$$

  uniformly on any $[0, T]^d$, for $T > 0$. It implicitly requires that N is not a multiple of the function M, see Remark 2.

REMARK 1. *The function L defined by (1) and that appears in (6) and (7) is homogeneous of order 1. We refer for instance to de Haan and Ferreira (2006, pages 213 and 236). Most of the estimators constructed in this paper use the homogeneity property. Note that pointwise convergence in (1) entails uniform convergence on the square $[0, T]^d$. See for instance de Haan and Ferreira (2006, page 237).*

REMARK 2. *If $N = c \cdot M$ for some constant c, the relation can be reformulated as*

$$\lim_{t\to\infty} \frac{1}{\beta(t)} \left\{ \frac{t\,[1 - F\{F_1^{-1}(1-t^{-1}x_1), \ldots, F_d^{-1}(1-t^{-1}x_d)\}] - L(\mathbf{x})}{\alpha(t)(1+c\beta(t))} - M(\mathbf{x}) \right\} = 0\,,$$

which we want to exclude. We refer to de Haan and Ferreira (2006, page 385) to see that the same complication turns up in the one-dimensional case.

REMARK 3. *The functions M and N involved in the second and third order conditions satisfy some usual properties, see e.g. de Haan and Resnick (1993). More specifically, one can show that there exists non positive reals ρ and ρ' such that α (resp. β) is a regularly varying function of order ρ (resp. ρ'), i.e. $\alpha(tz)/\alpha(t) \to z^\rho$ when $t \to \infty$, for each positive z. Besides, M is homogeneous of order $1-\rho$, that is to say $M(r\mathbf{x}) = r^{1-\rho} M(\mathbf{x})$, for each positive r and $\mathbf{x}$ with positive coordinates. Finally, the function N is homogeneous of order $1-\rho-\rho'$.*

REMARK 4. *An interesting situation is when the cdf F is in the domain of attraction of an extreme value distribution G with independent components, i.e. $G = \prod_{j=1}^d G_j$. Such a cdf is said to have the property of* asymptotic independence. *In this case, the function M is the limit of the joint tail of the distribution, and in dimension 2, the coefficient of tail dependence η introduced by Ledford and Tawn (1996, 1997) equals $1/(1-\rho)$, where ρ is defined in Remark 3.*

In this paper, we will handle two sets of assumptions. First consider

(A2) - the second order condition is satisfied, so that (6) holds;
- the coefficient of regular variation ρ of the function α defined in (6) is negative;
- the function M defined in (6) is continuous.

These hypotheses allow to get the asymptotic uniform behaviour of $\hat{L}_k$, the empirical estimator of L defined by (2), as detailed in the following proposition.

PROPOSITION 1. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent multivariate random vectors in $\mathbb{R}^d$ with common joint cdf F and continuous marginal cdfs F_j for $j = 1, \ldots, d$. Assume that the set of conditions* **(A2)** *hold. Suppose further that the first order partial derivatives of L (denoted by $\partial_j L$ for $j = 1, \ldots, d$) exist and that $\partial_j L$ is continuous on the set of points $\{\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}_+^d : x_j > 0\}$. Consider $\hat{L}_k$ the estimator of L defined by (2) where k is such that $\sqrt{k}\alpha(n/k) \to \infty$. Then as n tends to infinity, we get*

$$\sup_{0 \le x_1, \ldots, x_d \le T} \left| \frac{1}{\alpha(n/k)} \left\{ \hat{L}_k(\mathbf{x}) - L(\mathbf{x}) \right\} - M(\mathbf{x}) \right| \xrightarrow{\mathbb{P}} 0\,.$$

Under stronger assumptions, and for some choice of the intermediate sequence, the asymptotic distribution of the previous stochastic process can be obtained after multiplication by the rate $\sqrt{k}\alpha(n/k)$. For a positive T, let $D([0,T]^d)$ be the space of real valued functions that are right-continuous with left-limits. Now, introduce the conditions

(A3) - the third order condition is satisfied, so that (6) and (7) hold;
- the coefficients of regular variation ρ and ρ' of the functions α and β defined in (6) and (7) are negative;
- the function M defined in (6) is differentiable and N defined in (7) is continuous.

PROPOSITION 2. *Assume that the conditions of Proposition 1 are fulfilled and that the set of conditions* **(A3)** *hold. Consider $\hat{L}_k$ the estimator of L defined by (2) where k is such that $\sqrt{k}\alpha(n/k) \to \infty$ and $\sqrt{k}\alpha(n/k)\beta(n/k) \to 0$. Then as n tends to infinity,*

$$\sqrt{k}\left\{\hat{L}_k(\mathbf{x}) - L(\mathbf{x}) - \alpha(\frac{n}{k})M(\mathbf{x})\right\} \xrightarrow{d} Z_L(\mathbf{x}) \,, \tag{8}$$

in $D([0,T]^d)$ for every $T > 0$ where

$$Z_L(\mathbf{x}) := W_L(\mathbf{x}) - \sum_{j=1}^{d} W_L(x_j\mathbf{e}_j)\partial_j L(\mathbf{x}) \,. \tag{9}$$

The process W_L above is a continuous centered Gaussian process with covariance structure $\mathbb{E}[W_L(\mathbf{x})W_L(\mathbf{y})] = \mu\{R(\mathbf{x}) \cap R(\mathbf{y})\}$ given in terms of the measure μ defined by (5) and of $R(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}_+^d :$ there exists j such that $0 \le u_j \le x_j\}$.

REMARK 5. *A difference between the previous result and Theorem 7.2.2 of de Haan and Ferreira (2006) consists of the choice of the intermediate sequence that is larger here. Indeed, we suppose $|\sqrt{k}\alpha(n/k)| \to \infty$ whereas they choose $k(n) = o\left(n^{-2\rho/(1-2\rho)}\right)$ which implies $\sqrt{k}\alpha(n/k) \to 0$. Our choice requires the more informative second order condition (6). A non-null asymptotic bias appears in our framework.*

REMARK 6. *The conditions on k, α and β required in Proposition 2 are not too restrictive: because of the regular variation of α and β, they are implied by the choice $k(n) = n^\kappa$, with $\kappa \in \left(-\frac{2\rho}{1-2\rho},\, -\frac{2(\rho+\rho')}{1-2(\rho+\rho')}\right)$.*

3. Bias reduction procedure. As pointed out in Remark 5, a non-null asymptotic bias $\alpha(n/k)M(\mathbf{x})$ appears from Proposition 2. The bias reduction procedure will consist in subtracting the estimated asymptotic bias obtained in Section 3.1. The key ingredient is the homogeneity of the functions L and M mentioned in Remarks 1 and 3. This homogeneity will also provide other constructions to get rid of the asymptotic bias.

3.1. *Estimation of the asymptotic bias of $\hat{L}_k$.* Equation (8) suggests a natural correction of $\hat{L}_k$ as soon as an estimator of $\alpha(n/k)M(\mathbf{x})$ is available. In order to take advantage of the homogeneity of L, let us introduce a positive scale parameter a which allows to contract or to dilate the observed points. We denote

$$\hat{L}_{k,a}(\mathbf{x}) := a^{-1}\hat{L}_k(a\mathbf{x}) \,, \tag{10}$$

and

$$\hat{\Delta}_{k,a}(\mathbf{x}) := \hat{L}_{k,a}(\mathbf{x}) - \hat{L}_k(\mathbf{x}) \,. \tag{11}$$

From (8) one gets

$$\sqrt{k}\left\{\hat{L}_{k,a}(\mathbf{x}) - L(\mathbf{x}) - \alpha(\frac{n}{k})a^{-\rho}M(\mathbf{x})\right\} \xrightarrow{d} a^{-1}Z_L(a\mathbf{x}) \,, \tag{12}$$

in $D([0,T]^d)$ for every $T>0$. Equations (11) and Proposition 1 yield as n tends to infinity,

$$\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(\frac{n}{k})} \xrightarrow{\mathbb{P}} (a^{-\rho}-1)M(\mathbf{x}) \; . \tag{13}$$

Fixing a such that $a^{-\rho}-1=1$, a natural estimator of the asymptotic bias of $\hat{L}_k(\mathbf{x})$ is thus $\hat{\Delta}_{k,2^{-1/\hat{\rho}}}(\mathbf{x})$, where $\hat{\rho}$ is an estimator of ρ. Recall that the unknown parameter ρ is the regular variation index of the function α involved in the second order condition. Let k_ρ be an intermediate sequence that represents the number of order statistics used in the estimator $\hat{\rho}$. Assume that $k_\rho \gg k$ where $k=k(n)$ is the sequence used in Proposition 2. A first asymptotically unbiased estimator of $L(\mathbf{x})$ can be defined as

$$\mathring{L}_{k,1,k_\rho}(\mathbf{x}) := \hat{L}_k(\mathbf{x}) - \hat{\Delta}_{k,2^{-1/\hat{\rho}}}(\mathbf{x}) \; . \tag{14}$$

The asymptotic behaviour of this estimator is provided in Theorem 3 and Remark 8. We refer the reader to Section 6 for more details concerning the estimation of ρ.

3.2. *Estimation of the asymptotic bias of $\hat{L}_{k,a}$.* The previous construction can be easily generalized by correcting the estimator $\hat{L}_{k,a}$ instead of $\hat{L}_k$. Indeed, from (12) one can see that the asymptotic bias of $\hat{L}_{k,a}(\mathbf{x})$ is $\alpha(\frac{n}{k})a^{-\rho}M(\mathbf{x})$. Recall that when n tends to infinity, one has for any positive real b,

$$\frac{\hat{\Delta}_{k,b}(\mathbf{x})}{\alpha(\frac{n}{k})} \xrightarrow{\mathbb{P}} (b^{-\rho}-1)M(\mathbf{x}) \; .$$

Thus, fixing b such that $b^{-\rho}-1=a^{-\rho}$ will help for canceling the asymptotic bias. It yields the following asymptotically unbiased estimator of L

$$\mathring{L}_{k,a,k_\rho}(\mathbf{x}) := \hat{L}_{k,a}(\mathbf{x}) - \hat{\Delta}_{k,(a^{-\hat{\rho}}+1)^{-1/\hat{\rho}}}(\mathbf{x}) \; . \tag{15}$$

THEOREM 3. *Assume that the conditions of Proposition 2 are fulfilled and consider the estimator of L defined by (15). Let k_ρ be an intermediate sequence such that $\sqrt{k_\rho}\alpha(n/k_\rho)(\hat{\rho}-\rho)$ converges in distribution. Suppose also that k is such that $k=o(k_\rho)$, $\sqrt{k}\alpha(n/k)\to\infty$ and $\sqrt{k}\alpha(n/k)\beta(n/k) \to 0$. Under these assumptions, as n tends to infinity,*

$$\sqrt{k}\left\{\mathring{L}_{k,a,k_\rho}(\mathbf{x}) - L(\mathbf{x})\right\} \xrightarrow{d} \mathring{Y}_a(\mathbf{x}) \; , \tag{16}$$

in $D([0,T]^d)$ for every $T>0$, where $\mathring{Y}_a$ is a continuous centered Gaussian process defined by

$$\mathring{Y}_a(\mathbf{x}) := Z_L(\mathbf{x}) - b^{-1}Z_L(b\mathbf{x}) + a^{-1}Z_L(a\mathbf{x})$$

with covariance $\mathbb{E}[\mathring{Y}_a(\mathbf{x})\mathring{Y}_a(\mathbf{y})] = \mathbb{E}[Z_L(\mathbf{x})Z_L(\mathbf{y})]\left(1-b^{-1/2}+a^{-1/2}\right)^2$ and $b=(a^{-\rho}+1)^{-1/\rho}$.

REMARK 7. *The assumption that $\sqrt{k_\rho}\alpha(n/k_\rho)(\hat{\rho}-\rho)$ converges in distribution will be reconsidered in Section 6.*

REMARK 8. *Theorem 3 remains true when $a=1$ and thus characterizes the asymptotic behaviour of the estimator given in (14). For this particular choice of a, the covariance reduces to $\mathbb{E}[Z_L(\mathbf{x})Z_L(\mathbf{y})](2-2^{1/2\rho})^2$.*

3.3. *An alternative estimation of the asymptotic bias of $\hat{L}_{k,a}$.* The procedure of bias reduction introduced in the previous section requires the estimation of the second order parameter ρ. It is actually possible to avoid it, making use of combinations of estimators of L. The asymptotic bias of $\hat{L}_{k,a}(\mathbf{x})$ is $\alpha(\frac{n}{k})a^{-\rho}M(\mathbf{x})$, as already noted from (12). Making use of (13) and homogeneity of M, one gets as n tends to infinity

$$\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\hat{\Delta}_{k_\rho,a}(a\mathbf{x}) - a\hat{\Delta}_{k_\rho,a}(\mathbf{x})} \xrightarrow{\mathbb{P}} \frac{a^{-\rho}}{a^{-\rho}-1} ,$$

for any intermediate sequence k_ρ that satisfies $\sqrt{k_\rho}\alpha(n/k_\rho) \to \infty$. The expression

$$\hat{\Delta}_{k,a}(\mathbf{x})\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\hat{\Delta}_{k_\rho,a}(a\mathbf{x}) - a\hat{\Delta}_{k_\rho,a}(\mathbf{x})}$$

can thus be used as an estimator of the asymptotic bias of $\hat{L}_{k,a}(\mathbf{x})$. After simplifications, this leads to a new family of asymptotically unbiased estimators of $L(\mathbf{x})$ by substracting the estimated bias from $\hat{L}_{k,a}(\mathbf{x})$, namely

$$\tilde{L}_{k,a,k_\rho}(\mathbf{x}) = \frac{\hat{L}_k(\mathbf{x})\hat{\Delta}_{k_\rho,a}(a\mathbf{x}) - \hat{L}_k(a\mathbf{x})\hat{\Delta}_{k_\rho,a}(\mathbf{x})}{\hat{\Delta}_{k_\rho,a}(a\mathbf{x}) - a\hat{\Delta}_{k_\rho,a}(\mathbf{x})} , \tag{17}$$

which is well defined for any real number a such that $0 < a < 1$.

THEOREM 4. *Assume that the conditions of Proposition 2 are fulfilled and consider the estimator of L defined by (17). Let k_ρ be an intermediate sequence such that $\sqrt{k_\rho}\alpha(n/k_\rho)(\hat{\rho}-\rho)$ converges in distribution. Suppose also that k is such that $k = o(k_\rho)$, $\sqrt{k}\alpha(n/k) \to \infty$, $\sqrt{k} = O(\sqrt{k_\rho}\alpha(n/k_\rho))$ and $\sqrt{k}\alpha(n/k)\beta(n/k) \to 0$. Then, as n tends to infinity,*

$$\sqrt{k}\left\{\tilde{L}_{k,a,k_\rho}(\mathbf{x}) - L(\mathbf{x})\right\} \xrightarrow{d} \tilde{Y}_a(\mathbf{x}) , \tag{18}$$

in $D([0,T]^d)$ for every $T > 0$, where $\tilde{Y}_a$ is a continuous centered Gaussian process with covariance $\mathbb{E}[\tilde{Y}_a(\mathbf{x})\tilde{Y}_a(\mathbf{y})]$ given by $\mathbb{E}[Z_L(\mathbf{x})Z_L(\mathbf{y})](a^{-\rho}-1)^{-2}(a^{-\rho}-a^{-1/2})^2$.

REMARK 9. *The covariance function specified above is decreasing with respect to the parameter a for any fixed value of ρ. This suggests at first glance to choose a close to 1 in order to reduce the asymptotic variance of $\tilde{Y}_a$, but this would give a degenerate form of (17). See Section 5 for practical considerations for the choice of a.*

4. Theoretical examples. The aim of this section is to furnish several multivariate distributions that satisfy the third order condition (7). For the sake of simplicity, expressions are displayed in the bivariate setting. We start by focusing on heavy tailed margins. In this case, a first possible step to get the pointwise convergence is to obtain, for well chosen positive reals p and q, an expansion (for t tending to infinity) of the form

$$t\mathbb{P}(X > t^p x \,\text{or}\, Y > t^q y) = T_1(x,y) + \alpha(t)T_2(x,y) + \alpha(t)\beta(t)T_3(x,y) + o(\alpha(t)\beta(t)) ,$$

with $T_1(1,1) > 0$. One can then identify each term involved in (7) as follows

$$L(x,y) = T_1(a(x),b(y)), \quad M(x,y) = T_2(a(x),b(y)), \quad \text{and } N(x,y) = T_3(a(x),b(y)),$$

where

$$a(x) = x^{-p}\{T_1(1,+\infty)\}^p, \quad b(x) = x^{-q}\{T_1(+\infty,1)\}^q \, .$$

Applying Resnick (1986, Corollary 5.18), one can check that in such a framework a form of the bivariate extreme value distribution G is given by

$$G(x,y) = \exp\left(-\frac{T_1(x,y)}{T_1(1,1)}\right) \, .$$

4.1. *Powered norm densities.* Following the idea of Resnick (1986, page 276 and 286), consider first a norm $\|\cdot\|$, and a cone $\mathcal{D}$ of $\mathbb{R}^2$, that is to say a set such that if $(x,y) \in \mathcal{D}$, then $(tx,ty) \in \mathcal{D}$ for every positive t. Without loss of generality, suppose that $(1,1) \in \mathcal{D}$. Let (X,Y) be a bivariate random vector with probability density function given by

$$f(x,y) := \frac{c\mathbf{1}_{\mathcal{D}}(x,y)}{(1+\|(x,y)^T\|^{\alpha})^{\beta}} \, ,$$

where c is a normalizing positive constant and where α and β are some positive real numbers such that $\alpha\beta > 2$. Set $A_{\mathcal{D}}(x,y) := \{(u,v) \in \mathcal{D} : u > x \text{ or } v > y\}$ and define $p := (\alpha\beta-2)^{-1}$. One can check that, for $j = 1,2,3$,

$$T_j(x,y) = \iint_{A_{\mathcal{D}}(x,y)} \frac{c\, c_j\, dudv}{\|(u,v)^T\|^{\alpha(\beta+j-1)}} \, ,$$

where $c_1 = 1$, $c_2 = -\beta$ and $c_3 = \beta(\beta+1)/2$. The functions M and N are homogeneous with order given through $\rho = \rho' = -\alpha p$.

Let us discuss some particular choices of the norm:

- For the L^1-norm and $\alpha = 1$, the model coincides with the bivariate Pareto of type II distribution, denoted by BPII(β) in this paper, and referred to as $\mathrm{MP}^{(2)}(II)(0,1,\beta-2)$ in (Kotz, Balakrishnan and Johnson, 2000, p. 604). In this case, $p = q = (\beta-2)^{-1}$, and $L(x,y) = x + y - (x^{-p}+y^{-p})^{-1/p}$. The latter stdf is known as the negative logistic model, introduced by Joe (1990), see also (Beirlant et al., 2004, p. 307).
- When the Euclidean norm is chosen, one recovers the bivariate Cauchy distribution for $\alpha = 2$, $\beta = 3/2$ and $p = 1$. On the positive quadrant, that means for $\mathcal{D} = \mathbb{R}_+^2$, we have $c = 2/\pi$, $T_1(u,v) = c(u^{-2}+v^{-2})^{1/2}$ and $a(x) = b(x) = c/x$. On the whole plane, which means that $\mathcal{D} = \mathbb{R}^2$, we get $c = 1/(2\pi)$, $T_1(u,v) = c\left\{u^{-1}+v^{-1}+(u^{-2}+v^{-2})^{1/2}\right\}$ and $a(x) = b(x) = 2c/x$. This can also be seen as a particular case of the following item.
- The Student distributions with Pearson correlation coefficient θ arise choosing the norm $\|(x,y)^T\| = \nu^{-1/2}(x^2-2\theta xy+y^2)^{1/2}$, for a positive real number ν, $\alpha = 2$, $\beta = (\nu+2)/2$ and $p = \nu^{-1}$. In this case, the integral form of the function T_1 can not be totally simplified, and one classically writes the stdf as

$$L(x,y) = (x+y)\left[\frac{y}{x+y}F_{\nu+1}\left\{\frac{(y/x)^{1/\nu}-\theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right\} + \frac{x}{x+y}F_{\nu+1}\left\{\frac{(x/y)^{1/\nu}-\theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right\}\right] \, ,$$

  where $F_{\nu+1}$ is the cdf of the univariate Student distribution with $\nu+1$ degrees of freedom. This dependence structure is also obtained for some elliptical models, see e.g. (Krajina, 2012, p. 1813) and next subsection.
- Other choices for the norm would lead to other distributions. Note that one can also relax the symmetry condition, considering for instance the Mahalanobis pseudo-norm defined by $\|(x,y)^T\|^2 = (x/\sigma)^2 - 2\rho(x/\sigma)(y/\tau) + (y/\tau)^2$ for a real number ρ such that $|\rho| < 1$ and some positive real numbers σ and τ.

4.2. *Elliptical distributions.* Consider the usual representation of the centered elliptical distribution $(X,Y)^T = R\mathbf{A}\mathbf{U}$, in terms of a positive random variable R, a 2×2 matrix $\mathbf{A}$ such that $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$ is of full rank, and a bivariate random vector $\mathbf{U}$ independent of R, uniformly distributed on the unit circle of the plane. Assume that R has a probability density function denoted by g_R. One can then express the probability density function of (X,Y) as

$$f(x,y) := \frac{1}{|\det \mathbf{A}|} g_R \left\{ (x,y)\mathbf{\Sigma}^{-1}(x,y)^T \right\} .$$

A sufficient condition to satisfy (7) is to assume that the distribution of R belongs to the Hall and Welsch class (Hall and Welsh (1985)), viz.

$$\mathbb{P}(R > r) = cr^{-1/\gamma} \left\{ 1 + D_1 r^{\rho/\gamma} + D_2 r^{(\rho+\rho_1)/\gamma} + o(r^{(\rho+\rho_1)/\gamma}) \right\} ,$$

with positive real c, non null reals D_1 and D_2, and negative reals ρ and ρ_1.

One can check that, for $j = 1,2,3$,

$$T_j(x,y) = \frac{c}{2\pi\gamma|\det \mathbf{A}|} \iint_{\{(u,v):u>x \text{ or } v>y\}} \frac{dudv}{\{(u,v)\mathbf{\Sigma}^{-1}(u,v)^T\}^{1+1/(2\gamma)+p_j}} ,$$

where $p_1 = 0, p_2 = -\rho/(2\gamma)$ and $p_3 = -(\rho+\rho_1)/(2\gamma)$.

Assuming for simplicity that $\mathbf{\Sigma} = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$, the stdf can be written as

$$L(x,y) = (x+y)\left[\frac{y}{x+y} F_{1/\gamma+1} \left\{ \frac{(y/x)^\gamma - \theta}{\sqrt{1-\theta^2}} \sqrt{1/\gamma+1} \right\} + \frac{x}{x+y} F_{1/\gamma+1} \left\{ \frac{(x/y)^\gamma - \theta}{\sqrt{1-\theta^2}} \sqrt{1/\gamma+1} \right\} \right] ,$$

which is the form already obtained for the Student distribution in Subsection 4.1 for $\nu = 1/\gamma$. See Demarta and McNeil (2005) for more details. Note finally that for a general matrix $\mathbf{\Sigma}$ and the special case $g_R(r) = c(1+r^\alpha)^{-\beta}$, one recovers the Mahalanobis pseudo-norm already mentioned in the previous subsection.

When dealing with margins that are *not* heavy tailed, the calculus are done directly from (6). The last two examples of bivariate distributions have short and light tailed margins respectively.

4.3. *Archimax distributions.* Consider the bivariate cdf defined for each $0 \le u, v \le 1$ by

$$F(u,v) = \left\{ 1 + L(u^{-1}-1, v^{-1}-1) \right\}^{-1} , \tag{19}$$

given in terms of a stdf L. This distribution has standard uniform univariate margins and corresponds to a particular case of Archimax bivariate copulas introduced in Capéraà, Fougères and Genest (2000), in which the function $\phi(t) = t^{-1} - 1$ is the Clayton Archimedean generator with index 1. Expanding the left-hand side term of (6) leads to, as t tends to infinity,

$$t\left\{ 1 - F\left(1 - t^{-1}x,\, 1 - t^{-1}y\right) \right\} = L(x,y) + t^{-1}M(x,y) + t^{-2}N(x,y) + o\left(t^{-2}\right) ,$$

where

$$\begin{aligned} M(x,y) &:= x^2\partial_1 L(x,y) + y^2 \partial_2 L(x,y) - L^2(x,y) \\ N(x,y) &:= x^4/2\partial_{11}^2 L(x,y) + x^2y^2\partial_{12}^2 L(x,y) + y^4/2\partial_{22}^2 L(x,y) \\ &\quad - L(x,y)\left(2x^2\partial_1 L(x,y) + 2y^2\partial_2 L(x,y) - L^2(x,y)\right) . \end{aligned}$$

This allows to identify $\rho = \rho' = -1$. Above, the notation $\partial_{ij}L$ stands for $\partial^2 L/(\partial x_i \partial x_j)$.

4.4. *Multivariate Symmetric logistic distributions.* Consider the cdf defined by

$$F(x,y) = \exp\left\{-\left(e^{-x/r} + e^{-y/r}\right)^r\right\}\ , \tag{20}$$

for each $x, y \in \mathbb{R}$, which corresponds to the bivariate extreme value distribution with Gumbel univariate margins $F_1(x) = F_2(x) = \exp\{-e^{-x}\}$ and symmetric logistic stdf $L(x,y) = (x^{1/r} + y^{1/r})^r$, where $0 < r \le 1$. This distribution has been introduced in Tawn (1988), see e.g. (Beirlant et al., 2004, p. 304). Expanding $t\left[1 - F\left\{F_1^{-1}(1-t^{-1}x),\, F_2^{-1}(1-t^{-1}y)\right\}\right]$ leads to

$$L(x,y) + t^{-1}M(x,y) + t^{-2}N(x,y) + o\left(t^{-2}\right)\ ,$$

where

$$\begin{aligned}
M(x,y) &:= \frac{1}{2}\ (xx^{1/r} + yy^{1/r})\{L(x,y)\}^{1-1/r} - \frac{1}{2}\{L(x,y)\}^2 \\
N(x,y) &:= \frac{1}{3}(x^2x^{1/r} + y^2y^{1/r})\{L(x,y)\}^{1-1/r} + \frac{1-r}{8r}(xy)^{1/r}(x-y)^2\{L(x,y)\}^{1-2/r} \\
&\quad + \frac{1}{3!}\{L(x,y)\}^3 - \frac{1}{2}\ (xx^{1/r} + yy^{1/r})\{L(x,y)\}^{2-1/r}\ .
\end{aligned}$$

This allows to identify $\rho = \rho' = -1$. The identification of second and third order terms has previously be derived by Ledford and Tawn (1997).

5. Finite sample performances. The purpose of this section is to evaluate the performance of the estimators of L introduced in Section 3. For simplicity, we will focus on dimension 2, and simulate samples from the distributions presented in Section 4. Thanks to the homogeneity property, one can focus on the estimation of $t \mapsto L(1-t,t)$ for $0 \le t \le 1$, which coincides with the Pickands dependence function A (see e.g. Beirlant et al. (2004), p. 267). Considering first the estimation at $t = 1/2$ leads to define aggregated versions of our estimators. These new estimators will be both compared in terms of L^1-errors for L or associated level curves.

5.1. *Estimators in practice.* Let us start with the estimation of $L(1/2,1/2)$ for the bivariate Student distribution with 2 degrees of freedom. This model is a particular case of Sections 4.1 and 4.2. For one sample of size 1000, Figure 1 gives, as functions of k, the estimation of L at point $(1/2,1/2)$ by $\hat{L}_k$, $\mathring{L}_k$ and $\tilde{L}_k$ respectively defined by (2), (15) and (17). For the last two estimators, the parameters have been tuned as follows: $a = 0.4$, $k_\rho = 990$ and ρ estimated using (22) with $a = r = 0.4$. These values have been empirically selected based on intensive simulation, and will be kept throughout the paper. One can check from Figure 1 that the empirical estimator $\hat{L}_k$ behaves fairly well in terms of bias for small values of k. Besides, the bias is efficiently corrected by the two estimators $\mathring{L}_k$ and $\tilde{L}_k$. Since the bias almost vanishes along the range of k, one can think about reducing the variance through an aggregation in k (via mean or median) of $\mathring{L}_k$ or $\tilde{L}_k$. This leads to consider the two following estimators

$$\begin{aligned}
\mathring{L}_{\text{agg}} &:= \text{Median}(\mathring{L}_k, k = 1, \cdots, \kappa_n)\ , \\
\tilde{L}_{\text{agg}} &:= \text{Median}(\tilde{L}_k, k = 1, \cdots, \kappa_n)\ ,
\end{aligned}$$

where n is the sample size and κ_n is an appropriate fraction of n. Their performances will be compared to those of the family $\{\hat{L}_k, k = 1, \ldots, n-1\}$. Simplified notation $\{\hat{L}_k, k\}$ will be used instead of $\{\hat{L}_k, k = 1, \ldots, n-1\}$. Because any stdf L satisfies $\max(t, 1-t) \le L(1-t,t) \le 1$, the competitors have been corrected so that they satisfy the same inequalities.

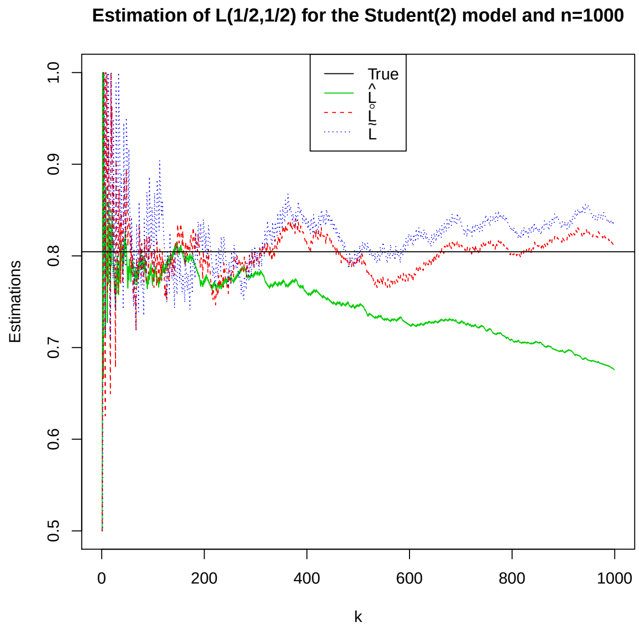


Fig 1: Estimation of $L(1/2, 1/2)$ for the bivariate Student(2) law based on a sample of size 1000.

Remark 10. *If κ_n satisfies the condition imposed on k_n in Theorem 3 and 4, then the aggregated estimators $\mathring{L}_{\text{agg}}$ and $\tilde{L}_{\text{agg}}$ would inherit the asymptotic properties of $\mathring{L}_k$ and $\tilde{L}_k$. Indeed, all the estimators jointly converge, since they are based on a single process.*

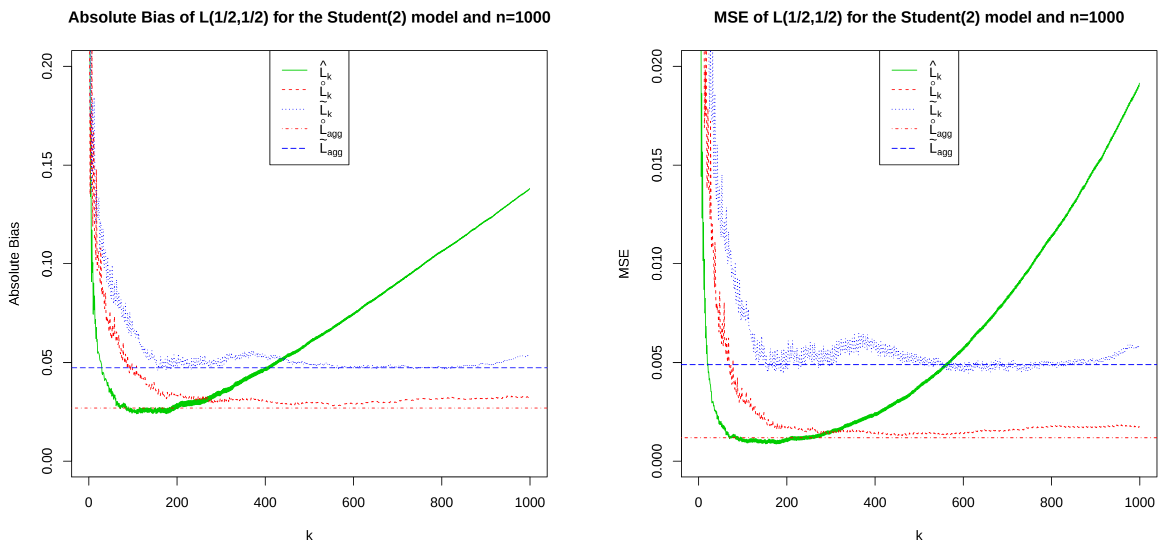
Remark 11. *In the following simulation study, κ_n is arbitrarily fixed to $n-1$. Such a choice is open to criticism since it does not satisfy the theoretical assumptions mentioned in the previous remark. But it is motivated here by the fact that the bias happened to be efficiently corrected even for very large values of k, as already illustrated on Figure 1. Note however that such a choice would not be systematically the right one. In presence of more complex models such as mixtures, κ_n should not exceed the size of the subpopulation with heaviest tail. To illustrate this point, take e.g. the bivariate cdf $F = pG + (1-p)H$, where G is the cdf of the bivariate BPII(3) model and H is the uniform cdf on $[0,1]^2$. Then the stdf is $L(x,y) = x + y - (1/x + 1/y)^{-1}$, and only $p\%$ of the data belong to the targeted domain of attraction, so κ_n should not exceed pn.*

Classical criteria of quality of an estimator $\hat{\theta}$ of θ are the absolute bias (ABias) and the mean square error (MSE) defined by

$$\text{ABias} = \frac{1}{N}\sum_{i=1}^{N} |\hat{\theta}^{(i)} - \theta| \; ,$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (\hat{\theta}^{(i)} - \theta)^2 \; ,$$

where N is the number of replicates of the experiment and $\hat{\theta}^{(i)}$ is the estimate from the ith sample. Note that what we call here Abias is also referred as MAE (for Mean Absolute Error) in the literature. Figure 2 plots these criteria in the estimation of $L(1/2, 1/2)$ for the bivariate Student(2) model when $n = 1000$ and $N = 200$. Figure 2 exhibits the strong dependence of the



(a) (b)

Fig 2: (a) ABias (b) MSE for the estimation of $L(1/2, 1/2)$ in the bivariate Student(2) model when $n = 1000$ as a function of k.

behaviour of $\hat{L}_k$ in terms of k, as well as the efficiency of the bias correction procedures. The estimator $\mathring{L}_k$ given by (15) outperforms the estimator $\tilde{L}_k$ defined by (17), no matter the value of k. Moreover, the ABias and MSE curves associated to $\mathring{L}_k$ almost reach the minimum of those of $\hat{L}_k$. Finally, the aggregated version $\mathring{L}_{\rm agg}$ answers surprisingly well to the estimation problem of the stdf L. First, its performances are similar to the best reachable from the original estimator $\hat{L}_k$. Second, it gets rid of the delicate choice of a threshold k (or would at least simplify this choice, see Remark 11). These comparisons have also been done for five other models obtained from Section 4. The results are very similar to the ones obtained for the bivariate Student(2) distribution and are therefore not presented.

5.2. *Comparisons in terms of L^1-error for L.* The comparisons are now handled not only at a single point but for the whole function using an L^1-error defined as follows

$$\frac{1}{T+1}\sum_{t=1}^{T}\left|\hat{L}\left(1-\frac{t}{T},\frac{t}{T}\right)-L\left(1-\frac{t}{T},\frac{t}{T}\right)\right| \tag{21}$$

where T is the size of the subdivision of $[0, 1]$. Figure 3 gives the boxplots based on $N = 100$ realisations of $\mathring{L}_{\rm agg}$, $\tilde{L}_{\rm agg}$ and $\{\hat{L}_k, k\}$ for $T = 30$ in the case of six bivariate models:

- First row: Cauchy and Student(2) models;
- Second row: BPII(3) model and Symmetric logistic model with $r = 1/3$;
- Third row: Archimax model with logistic generator $L(x, y) = (x^2 + y^2)^{1/2}$ and mixed generator $L(x, y) = (x^2 + y^2 + xy)/(x + y)$.

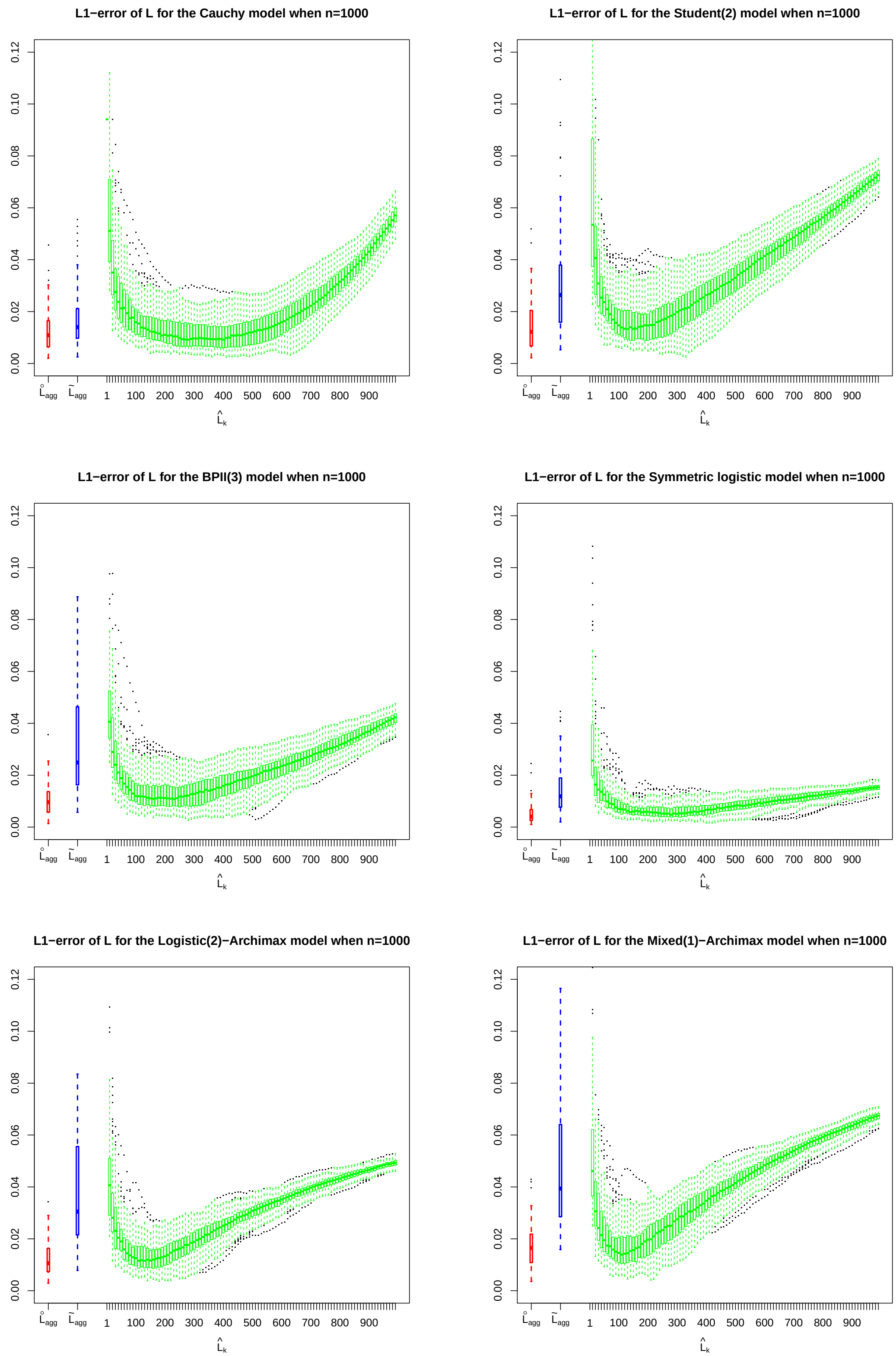


Fig 3: Boxplot of the L^1-error of function L for the estimators $\mathring{L}_{\text{agg}}, \tilde{L}_{\text{agg}}$ and $\{\hat{L}_k, k\}$.
First row: bivariate Cauchy model (left) and bivariate Student(2) model (right).
Second row: bivariate BPII(3) model (left) and bivariate Symmetric logistic model (right).
Third row: bivariate Archimax model with logistic (left) and mixed generator (right).

As already observed in Figure 2, the estimator $\mathring{L}_{\text{agg}}$ is again very competitive compared to the best element of $\{\hat{L}_k, k\}$ no matter the choice of model. Recall that the value of k leading to the best $\hat{L}_k$ depends crucially on the model, and is consequently unknown in practice, which invites any users to apply this new procedure.

The estimator $\tilde{L}_{\text{agg}}$ is definitely less competitive compared to $\mathring{L}_{\text{agg}}$. Given these results we will not pursue with the $\tilde{L}_{\text{agg}}$ estimator in the rest of this paper, and will focus our attention on the behaviour of $\mathring{L}_{\text{agg}}$.

5.3. *Comparisons between $\mathring{L}_{\text{agg}}$, a convex version of $\mathring{L}_{\text{agg}}$, and Peng's estimator.* A natural step is now to compare the performance of our best estimator $\mathring{L}_{\text{agg}}$ with an existing competitor, recently introduced by Peng (2010). In his work, Peng provides a data-driven method which chooses the threshold via estimating a stdf. Another interesting task is to compare $\mathring{L}_{\text{agg}}$ with a convexified version of itself, since any stdf is a convex function (see e.g. (Beirlant et al., 2004, Section 8.2.2) or de Haan and Ferreira (2006, Section 6.1.5)). In the sequel these two estimates will be denoted by $\hat{L}_P$, representing Peng's candidate and by $\mathring{L}_{\text{aggc}}$ for the convexified version of our best estimate.

In order to take maximal advantage from this simulation study, the three different models implemented have been considered in two versions for each: the first model is the Gaussian one, simulated with Pearson's correlation coefficient ± 0.5. The Gaussian model is a particular case of elliptical distributions (see Section 4.2) which illustrates the asymptotic independent situation (cf. Remark 4). The second model is the bivariate Symmetric logistic one, introduced in Section 4.4, with two different strengths of dependence (close to independence on the left column, stronger dependence on the right column). The third model is the bivariate Student family, introduced in Sections 4.1 and 4.2 as a particular case. Two strengths of dependence have also been chosen, close to asymptotic independence on the left column, stronger dependence on the right column.

Our results, summarized in Figure 4, will thus exhibit in particular how the performance in the estimation of the stdf depends on the distance to the asymptotic independence case. The y-axis scale has been fixed for all the six cases so that one can measure that the estimation of the stdf is a more ambitious problem under asymptotic independence. However, our estimator $\mathring{L}_{\text{agg}}$ has still nice properties when comparing it to the empirical estimator $\hat{L}_k$.
The convex version $\mathring{L}_{\text{aggc}}$ performs quite equivalently as $\mathring{L}_{\text{agg}}$. A reason for this is that by construction our estimator is actually not far from a convex function. So balancing the cost of convexifying with the benefit in the performance motivates the simple use of $\mathring{L}_{\text{agg}}$.
Finally, regarding Peng's estimator $\hat{L}_P$, one observes that this estimator is an interesting alternative to the original family $\{\hat{L}_k, k\}$, which however never outperforms our proposal.

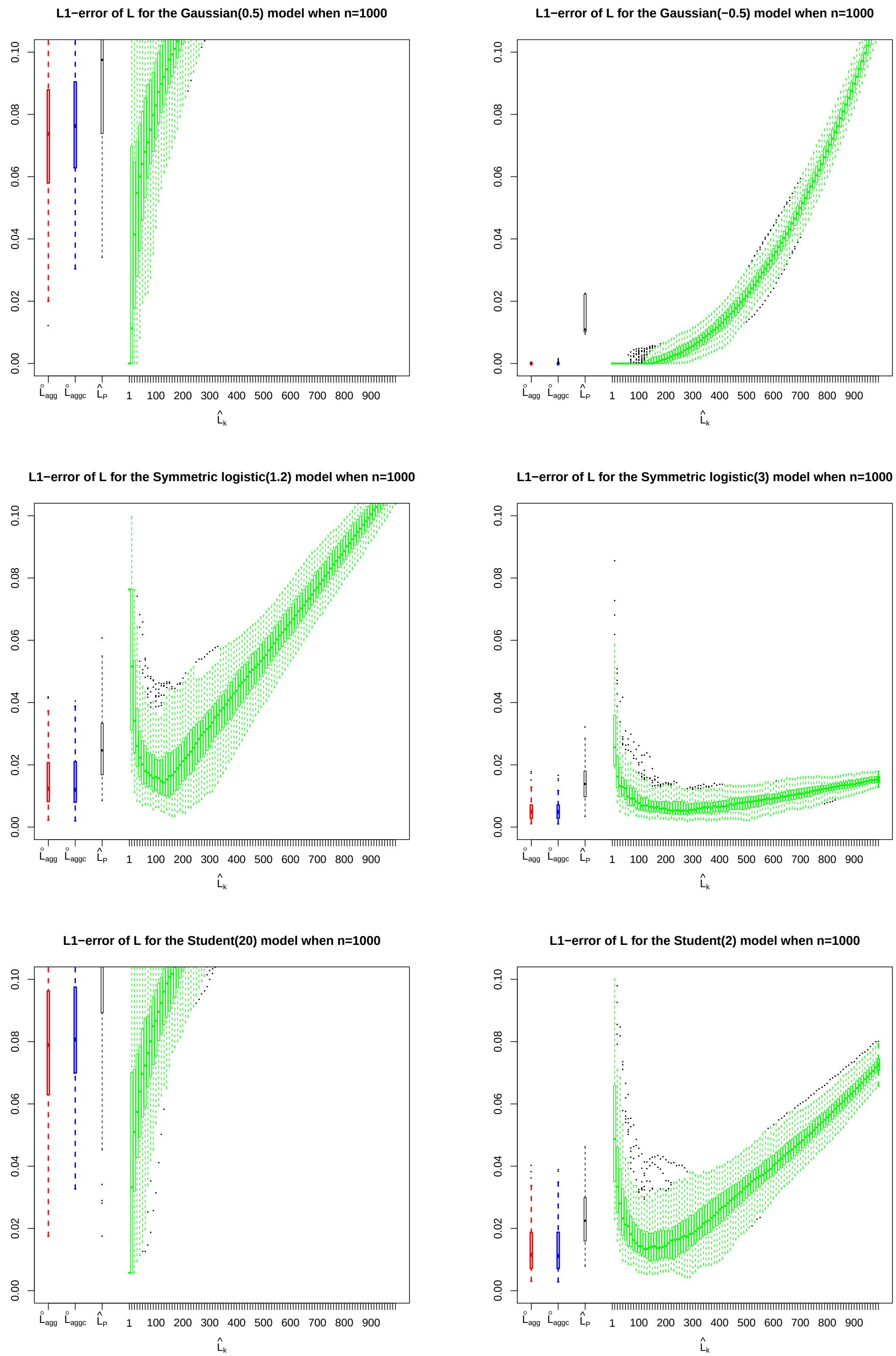


Fig 4: Boxplot of the L^1-error of function L for the estimators $\mathring{L}_{\text{agg}}, \mathring{L}_{\text{aggc}}, \hat{L}_P$ and $\{\hat{L}_k, k\}$.
First row: bivariate Normal model with correlation τ: $\tau = 0.5$ (left) and $\tau = -0.5$ (right).
Second row: bivariate Symmetric logistic(r) model: $r = 1.2$ (left) and $r = 3$ (right).
Third row: bivariate Student(ν) model: $\nu = 20$ (left) and $\nu = 2$ (right).

5.4. *Estimating a failure probability.* Let us illustrate in this subsection the question of estimating an arbitrarily chosen failure probability $P(X > 10^4 \,\text{or}\, Y > 2 \cdot 10^4)$, where (X, Y) comes from the BPII(3) model. This probability is estimated via an intensive simulation study from 200 repetitions of samples of size $n = 10^7$, and using the empirical estimator

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i > 10^4 \,\text{or}\, Y_i > 2 \cdot 10^4\}}.$$

Now, let us consider samples of size $n = 1000$. Empirical estimation will be useless for evaluating the probability of exceeding such extreme values for X or Y, and an extrapolation based on Extreme Value Theory is thus needed.
First assume that it is known that the margins are standard Pareto. This probability can be approximated by

$$P(X > 10^4 \,\text{or}\, Y > 2 \cdot 10^4) \simeq \left(10^{-4} + 5 \cdot 10^{-5}\right) L(2/3, 1/3)$$

that naturally comes from (1), the projection on the simplex and the homogeneity of L. Estimating the unknown parameter $L(2/3, 1/3)$ with our candidate $\mathring{L}_{\text{agg}}$ and the original family $\{\hat{L}_k, k\}$ gives several boxplots (based on 500 replicates) that are presented in Figure 5. The comparison of these estimates is again favourable to $\mathring{L}_{\text{agg}}$.

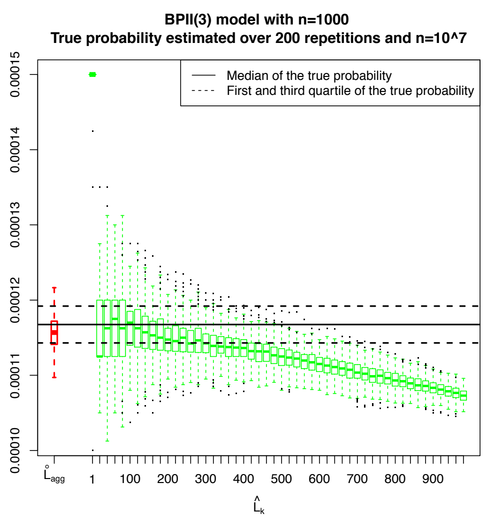


Fig 5: Boxplot (based on 500 replicates) for the estimation of $P(X > 10^4 \,\text{or}\, Y > 2 \cdot 10^4)$ when (X, Y) is drawn from the BPII(3) model with sample size $n = 1000$ and assuming margins to be known.

Second, when the margins are not assumed to be known, the estimation of $p_1 = 1 - F_1(10^4)$ and $p_2 = 1 - F_2(2 \cdot 10^4)$ can be reached by the POT method (see e.g. (Beirlant et al., 2004, Section 7.4)) for several values of a threshold. After the study of Mean Residual Life Plots and Quantile Plots, the thresholds have been fixed to be $X_{n-k,n}$ and $Y_{n-k,n}$ for $k = 200$. The POT estimates deduced with these thresholds are respectively denoted by $\hat{p}_1$ and $\hat{p}_2$. The probability $P(X > 10^4 \,\text{or}\, Y > 2 \cdot 10^4)$ is then approximated by

$$P(X > 10^4 \,\text{or}\, Y > 2 \cdot 10^4) \simeq (\hat{p}_1 + \hat{p}_2)\, L\left(\frac{\hat{p}_1}{\hat{p}_1 + \hat{p}_2}, \frac{\hat{p}_2}{\hat{p}_1 + \hat{p}_2}\right) .$$

Estimating on each repetition the unknown parameter $L\left(\frac{\hat{p}_1}{\hat{p}_1+\hat{p}_2}, \frac{\hat{p}_2}{\hat{p}_1+\hat{p}_2}\right)$ with our candidate $\mathring{L}_{\text{agg}}$ and the original family $\{\hat{L}_k, k\}$ gives several boxplots (based on 500 replicates) presented in Figure 6. It seems clear that the uncertainty on the margins F_1 and F_2 is much more influent

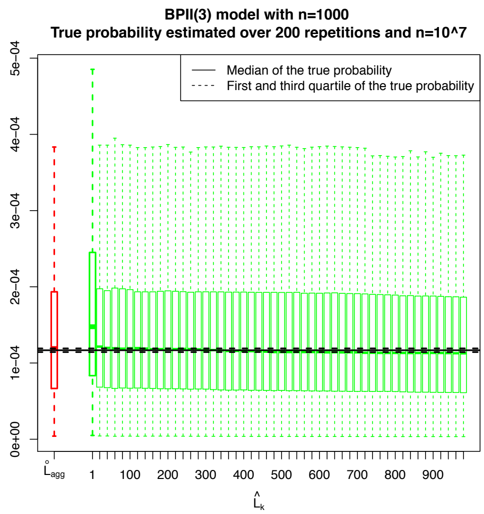


Fig 6: Boxplot (based on 500 replicates) of the estimation of $P(X > 10^4 \text{ or } Y > 2 \cdot 10^4)$ when (X, Y) is drawn from the BPII(3) model with sample size $n = 1000$ and estimating margins by POT method.

than that of the stdf L. Such findings corroborate previous studies, see e.g. Bruun and Tawn (1998) and de Haan and Sinha (1999).

5.5. *Q-curves.* Another nice representation of a function of several variables is through its level sets. In the case of the function L, it consists of looking (for any positive real c) at sets of the form $\{(x, y) \in \mathbb{R}_+^2, L(x, y) \leq c\}$. From homogeneity property, it is characterized by

$$Q := \{(x, y) \in \mathbb{R}_+^2, L(x, y) \leq 1\} .$$

Following de Haan and Ferreira (2006, page 245), the boundary of this set can be written as

$$\partial Q = \left\{ \Big(b(\theta) \cos\theta, b(\theta) \sin\theta \Big) : \quad b(\theta) = \frac{1}{L(\cos\theta, \sin\theta)}, \theta \in [0, \pi/2] \right\} .$$

The estimation of ∂Q is naturally obtained by replacing L by any estimator, and this is done here for the estimators $\mathring{L}_{\text{agg}}$ and $\{\hat{L}_k, k\}$. Figure 8 (left) exhibits the bias phenomenon (as k increases) induced by $\hat{L}_k$ in the estimation of the Q-curve. The bias factor on $\hat{L}_k$ is illustrated with $k = 50, k = 100$ and $k = 800$. The correction of the bias with $\mathring{L}_{\text{agg}}$ is effective. As done in the previous section, the comparison of the different estimators is provided in terms of a global criterium based on the L^1-norm, given by

$$\frac{\pi}{2(T+1)} \sum_{t=0}^{T} \left| \hat{b}\left(\frac{\pi t}{2T}\right) - b\left(\frac{\pi t}{2T}\right) \right| \left\{ \cos\left(\frac{\pi t}{2T}\right) + \sin\left(\frac{\pi t}{2T}\right) \right\} .$$

Figure 7 displays the boxplots of this measure, based on $N = 100$ realisations and for $T = 30$ under the six bivariate models given in the previous section.

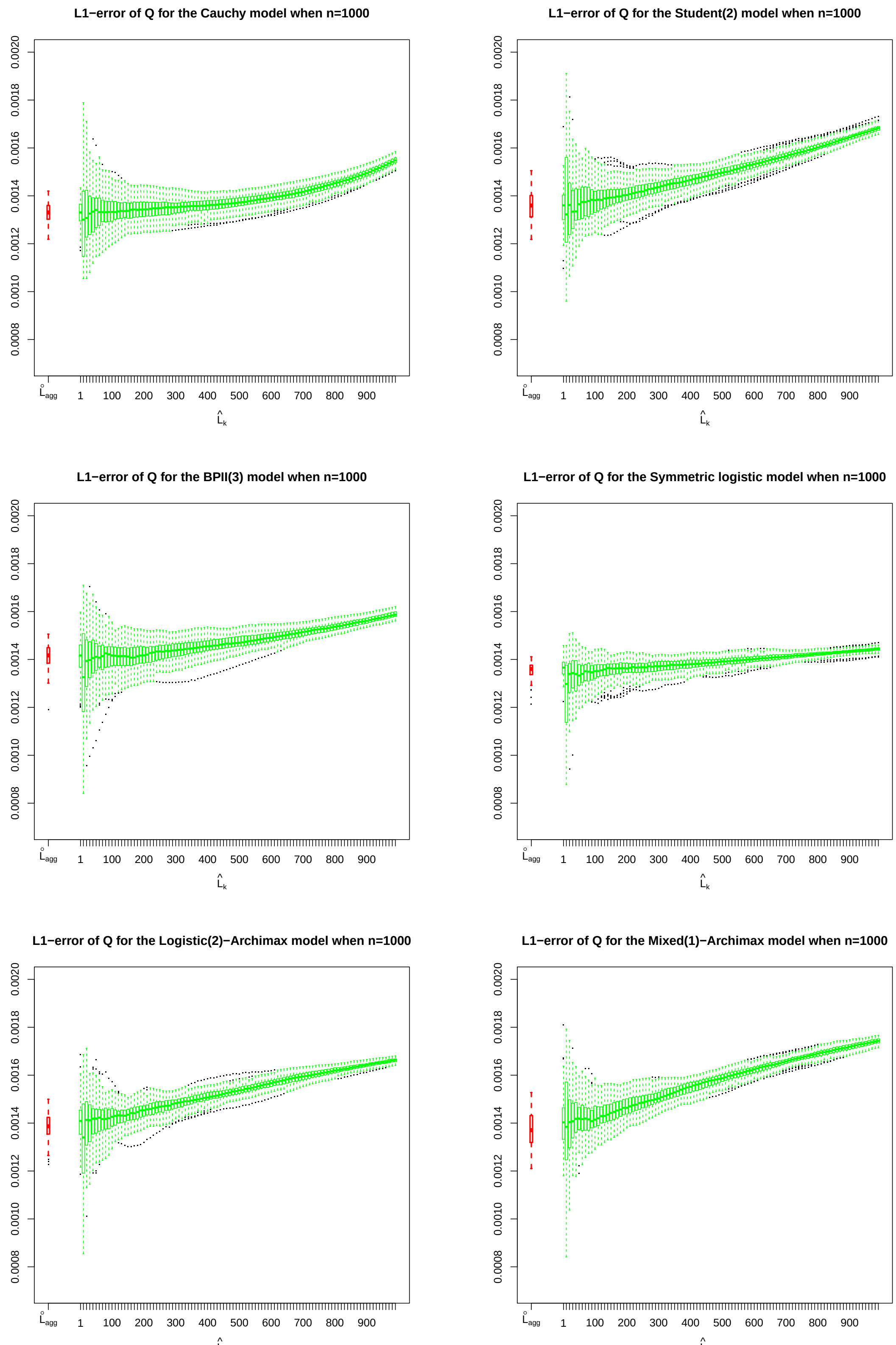


Fig 7: Boxplot of the L^1-error of Q-curve for the estimators $\mathring{L}_{\text{agg}}$ and $\{\hat{L}_k, k\}$.
First row: bivariate Cauchy model (left) and bivariate Student(2) model (right).
Second row: bivariate BPII(3) model (left) and bivariate Symmetric logistic model (right).
Third row: bivariate Archimax model with logistic (left) and mixed generator (right).

The estimation of the Q-curve based on the original estimator $\hat{L}_k$ is strongly sensitive to the choice of k: the bias (resp. the variability) is an increasing (resp. decreasing) function of k. The performances of $\mathring{L}_{\text{agg}}$ is similar to that of the best $\hat{L}_k$, which is unknown in practice. These features corroborate the conclusions drawn in Section 5.2.

To close this section, let us illustrate the Q-curve estimation on the data set of de Haan and Ferreira (2006, page 207). As explained therein, wave height (HmO) and still water level (SWL) have been recorded during 828 storm events on the Dutch coast. The analogous of Figure 7.2 from de Haan and Ferreira (2006) is reported in Figure 8 (right). Even if the two curves are not so close, the conclusion remains the same: the estimated boundary is concave, which indicates that the high values of the two variables are dependent.

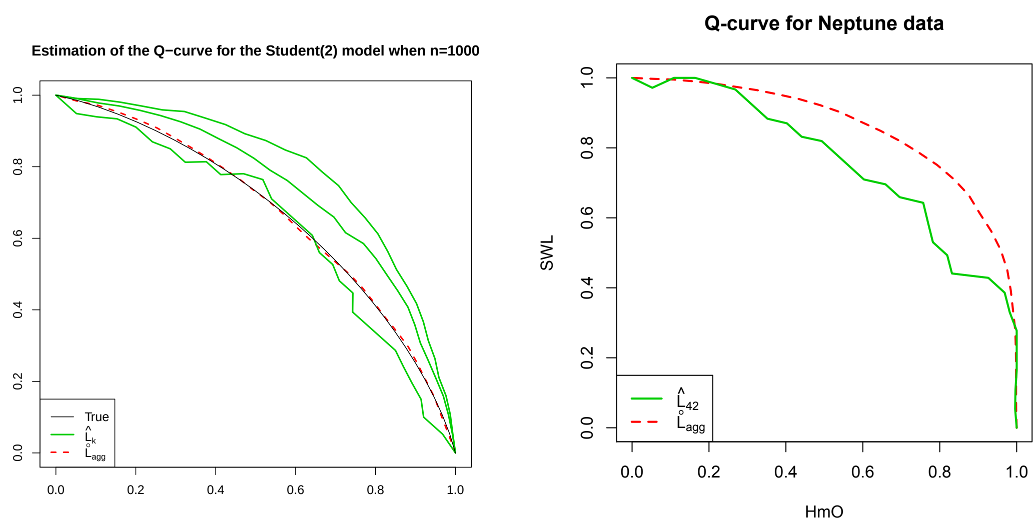


Fig 8: Left: Estimation of the Q-curve for the bivariate Student(2) law based on a sample of size 1000. Right: Estimated Q-curve for Neptune data set.

6. Estimation of second order components ρ and M. In this section, we focus on the estimation of the function M coming from the second order condition (6) and on the estimation of its homogeneity parameter $1 - \rho$.

6.1. *Second order parameter ρ.* A possible way to estimate ρ is to use on each margin one of the techniques developed in the univariate setting, see e.g. Gomes, de Haan and Peng (2002) or Ciuperca and Mercadier (2010). Other methods make use of the multivariate structure of the data, see e.g. Peng (2010) and also Goegebeur and Guillou (2013) in a slightly different framework. The construction described here takes likewise advantage of the multivariate information of the sample. On this purpose, the following proposition shows that a variable of interest is the ratio of two terms $\hat{\Delta}_{k,a}$, defined by (11).

PROPOSITION 5. *Assume that the conditions of Proposition 1 are fulfilled and fix positive real numbers r and $a \in (0, 1)$. Assume moreover that the function M never vanishes except* *on the axes. Then, as n tends to infinity,*

$$\sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\hat\Delta_{k,a}(r\mathbf{x})}{\hat\Delta_{k,a}(\mathbf{x})}-r^{1-\rho}\right|\xrightarrow{\mathbb{P}}0\,.$$

REMARK 12. *In case the requirement that the function M is either positive or negative in the positive quadrant does not hold, one could consider the integral of $(\hat\Delta_{k,a}(\mathbf{x}))^2$ over the set $\{\mathbf{x}=(x_1,\dots,x_d)\quad\text{s.t.}\quad x_1^2+\ldots+x_d^2=1\}$ and prove a result like Lemma 7 for this statistic. Then the integral of M^2 appears in the denominator in Proposition 5 instead of M itself and the sign of M does not matter. This will be part of a future work.*

A family of consistent estimators of the parameter ρ can be derived from Proposition 5.

$$\hat\rho_{k,a,r}(\mathbf{x}):=\left(1-\frac{1}{\log r}\log\left\{\frac{\hat\Delta_{k,a}(r\mathbf{x})}{\hat\Delta_{k,a}(\mathbf{x})}\right\}\right)\wedge 0\,. \tag{22}$$

The following property can be obtained from the asymptotic expansion given in Proposition 2.

PROPOSITION 6. *Assume that the conditions of Proposition 2 are fulfilled and fix positive real numbers r and $a\in(0,1)$. Consider the estimator of ρ defined by (22). Assume moreover that the function M never vanishes except on the axes. Then, as n tends to infinity,*

$$\sqrt{k}\alpha(\frac{n}{k})\{\hat\rho_{k,a,r}(\mathbf{x})-\rho\}\xrightarrow{d}\hat Z_{\rho,a,r}(\mathbf{x})\,,$$

in $D([0,T]^d)$ for every $T>0$, with

$$\hat Z_{\rho,a,r}(\mathbf{x}):=\frac{a^{-1}Z_L\,(a\mathbf{x})-Z_L(\mathbf{x})}{(a^{-\rho}-1)M(\mathbf{x})\log r}-\frac{a^{-1}Z_L\,(ra\mathbf{x})-Z_L(r\mathbf{x})}{(a^{-\rho}-1)M(\mathbf{x})r^{1-\rho}\log r}\,.$$

Figure 9 illustrates the finite sample behaviour of this estimator of ρ for a collection of bivariate models introduced in Section 4, for which the true value of ρ is equal to -1. These

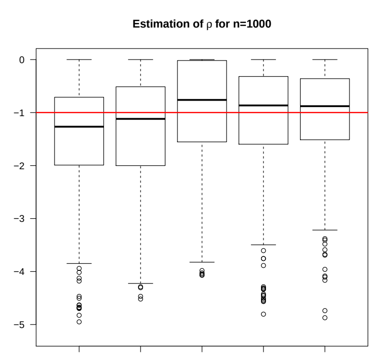


Fig 9: Boxplot of 500 estimations of ρ given by (22) using samples of size 1000 drawn from six models: (a) Student(2); (b) BPII(3); (c) Symmetric Logistic with $r=1/3$; (d) Archimax model with logistic generator with $r=1/2$; (e) Archimax model with mixed generator. Red line indicates the true value of $\rho=-1$.

boxplots show that the estimator performs reasonably well, no matter the choice of model.

6.2. *Second order function M.* Recall that from (12) the asymptotic bias of $\hat{L}_{k,a}(\mathbf{x})$ is given by $\alpha(\frac{n}{k})a^{-\rho}M(\mathbf{x})$. In order to circumvent an estimation of the term $\alpha(n/k)$, a renormalisation is needed, focusing for instance on the estimation of $M(\mathbf{x})/M(\mathbf{1/2})$ where $\mathbf{1/2} = (1/2,\dots,1/2)$. Thanks to (13), this ratio can be consistently estimated by

$$\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\hat{\Delta}_{k,a}(\mathbf{1/2})}$$

as soon as k is a well chosen intermediate sequence. The asymptotic normality can also be derived from analogous arguments to those used in the proof of Proposition 6. Details are not presented here for the sake of simplicity.

Figure 10 summarizes the behaviour of the estimator of the curve $t \mapsto M(t,1-t)/M(1/2,1/2)$ through boxplots of the L^1-error, defined as in (21). We observe from this figure that the best estimation is reached for large values of k. This feature does not depend on the degree of asymptotic dependence in the Symmetric logistic model, nor on the strength of the bias of the original estimator $\hat{L}_k$ detected on Figure 3. These graphs confirm that the asymptotic bias is remarkably well estimated for large values of k. This helps to understand why the bias subtraction is accurate for large or very large choices of k, as also commented in Section 5.1.

Concluding comments. This paper deals with the estimation of the extremal dependence structure in a multivariate context. Focusing on the stdf, the empirical counterpart is the non parametric reference. A common feature when modelling extreme events is the delicate choice of the number of observations used in the estimation, and it spoils the good performance of this estimator. The aim of this paper has been to correct the asymptotic bias of the empirical estimator, so that the choice of the threshold becomes less sensitive. Two asymptotically unbiased estimators have been proposed and studied, both theoretically and numerically. The estimator defined in Section 3.2 proves to outperform the original estimator, whatever the model considered. Its aggregated version defined in Section 5.1 appears as a worthy candidate to estimate the stdf.

7. Proofs.

PROOF OF PROPOSITION 1. Denote by $U_i^{(j)}$ the uniform random variables $U_i^{(j)} = 1 - F_j(X_i^{(j)})$ for $j = 1,\dots,d$. Introducing

$$V_k(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\left\{U_i^{(1)} \le kx_1/n \text{ or } \dots \text{ or } U_i^{(d)} \le kx_d/n\right\}}$$

allows to rewrite $\hat{L}_k$ as the following

$$\hat{L}_k(\mathbf{x}) = V_k\left(\frac{n}{k}U^{(1)}_{[kx_1],n},\dots,\frac{n}{k}U^{(d)}_{[kx_d],n}\right) .$$

Write

$$\begin{aligned}\hat{L}_k(\mathbf{x}) - L(\mathbf{x}) &= V_k\left(\frac{n}{k}U^{(1)}_{[kx_1],n},\dots,\frac{n}{k}U^{(d)}_{[kx_d],n}\right) - \frac{n}{k}\left[1 - F\{F_1^{-1}(1-U^{(1)}_{[kx_1],n}),\dots,F_d^{-1}(1-U^{(d)}_{[kx_d],n})\}\right]\\ &+ \frac{n}{k}\left[1 - F\{F_1^{-1}(1-U^{(1)}_{[kx_1],n}),\dots,F_d^{-1}(1-U^{(d)}_{[kx_d],n})\}\right] - L\left(\frac{n}{k}U^{(1)}_{[kx_1],n},\dots,\frac{n}{k}U^{(d)}_{[kx_d],n}\right)\\ &+ L\left(\frac{n}{k}U^{(1)}_{[kx_1],n},\dots,\frac{n}{k}U^{(d)}_{[kx_d],n}\right) - L(\mathbf{x}) ,\end{aligned}$$

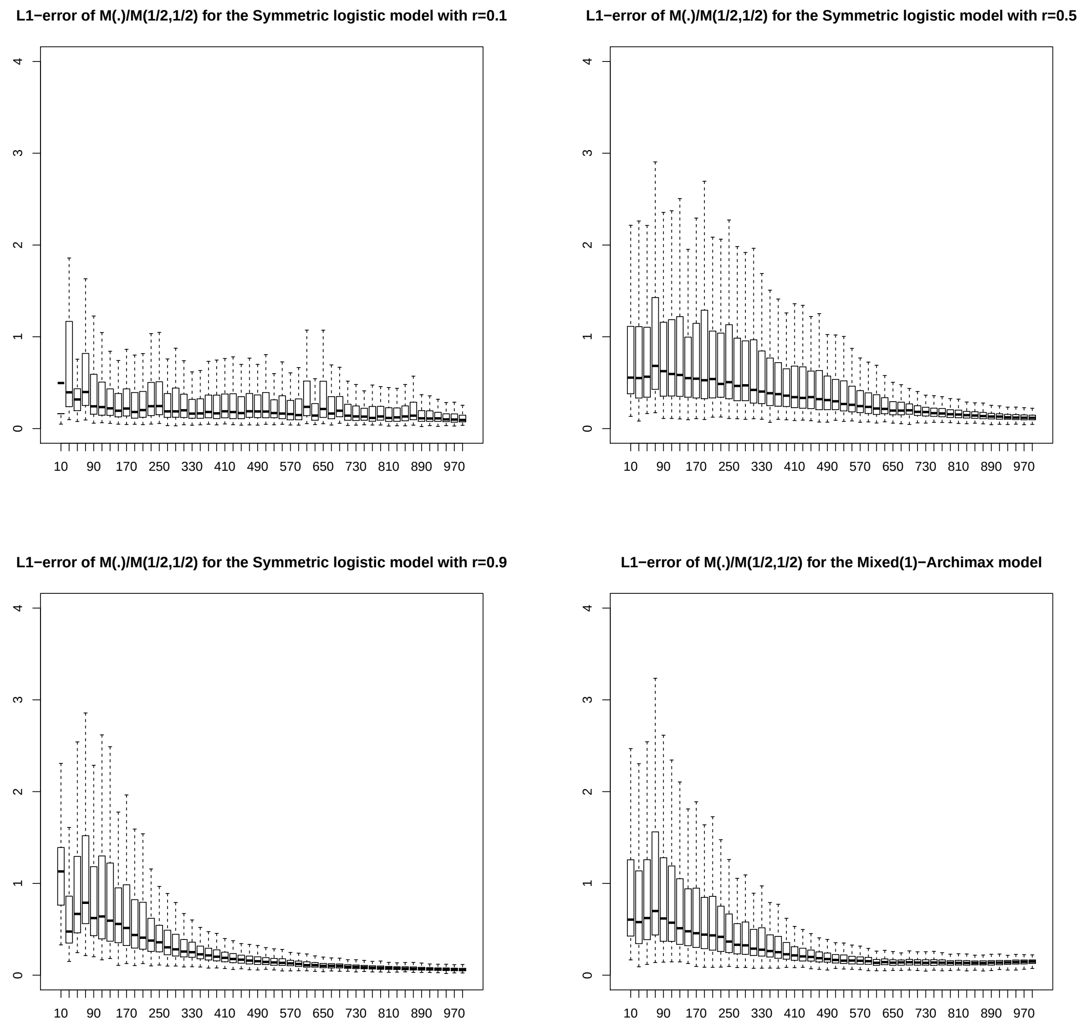


Fig 10: Boxplot of the L^1-error of $M(\cdot)/M(1/2,1/2)$-curve.
First row: Bivariate logistic model with $r = 0.1$ (left) and with $r = 0.5$ (right).
Second row: Bivariate logistic model with $r = 0.9$ (left) and bivariate Archimax with mixed generator (right).

and denote $A_{1,k}(\mathbf{x})$ (resp. $A_{2,k}(\mathbf{x})$ and $A_{3,k}(\mathbf{x})$) the first line (resp. second and third lines) of the right-hand side.
Applying de Haan and Ferreira (2006, Proposition 7.2.3) leads to

$$\sqrt{k}A_{1,k}(\mathbf{x}) \xrightarrow{d} W_L(\mathbf{x}) \ ,$$

in $D([0,T]^d)$ for every $T > 0$ and for any intermediate sequence, where W_L is a continuous centered Gaussian process with covariance structure specified in Proposition 2. Due to the Skorohod construction we can write

$$\sup_{0\le x_1,\ldots,x_d\le T} \left|\sqrt{k}A_{1,k}(\mathbf{x}) - W_L(\mathbf{x})\right| \to 0 \quad \text{a.s. ,} \tag{23}$$

which implies, since $\sqrt{k}\alpha(n/k) \to \infty$,

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\frac{A_{1,k}(\mathbf{x})}{\alpha\left(\frac{n}{k}\right)}\right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}\alpha\left(\frac{n}{k}\right)}\right) .$$

Again for any intermediate sequence, the proof of de Haan and Ferreira (2006, Theorem 7.2.2) ensures the convergence for $j = 1, \ldots, d$

$$\sup_{x\in[0,T]} |\sqrt{k}\left(\frac{n}{k}U^{(j)}_{[kx],n} - x\right) + W_L(x\mathbf{e}_j)| \to 0 \quad \text{a.s.} , \tag{24}$$

and finally

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\sqrt{k}A_{3,k}(\mathbf{x}) + \sum_{j=1}^{d} W_L(x_j\mathbf{e}_j)\partial_j L(\mathbf{x})\right| \to 0 \quad \text{a.s.} . \tag{25}$$

As previously, this yields

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\frac{A_{3,k}(\mathbf{x})}{\alpha\left(\frac{n}{k}\right)}\right| = O\left(\frac{1}{\sqrt{k}\alpha\left(\frac{n}{k}\right)}\right) .$$

Since the intermediate sequence satisfies $\sqrt{k}\alpha\left(\frac{n}{k}\right) \to \infty$, it thus remains to prove that

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\frac{A_{2,k}(\mathbf{x})}{\alpha(\frac{n}{k})} - M(\mathbf{x})\right| \to 0 \quad \text{a.s.} .$$

The second order condition that holds uniformly on $[0,T]^d$ in (6) yields

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\frac{A_{2,k}(\mathbf{x})}{\alpha(\frac{n}{k})} - M\left(\frac{n}{k}U^{(1)}_{[kx_1],n}, \ldots, \frac{n}{k}U^{(d)}_{[kx_d],n}\right)\right| \to 0 \quad \text{a.s.} .$$

Then the result follows from

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|M(\mathbf{x}) - M\left(\frac{n}{k}U^{(1)}_{[kx_1],n}, \ldots, \frac{n}{k}U^{(d)}_{[kx_d],n}\right)\right| \to 0 \quad \text{a.s.} ,$$

that is obtained combining (24) and the continuity of the function M. □

PROOF OF PROPOSITION 2. We use the notation introduced in the proof of Proposition 1. Thanks to the Skorohod construction, we can start from (23). Combined with (25), it is sufficient to prove the convergence

$$\sup_{0\le x_1,\ldots,x_d\le T}\left|\sqrt{k}\left\{A_{2,k}(\mathbf{x}) - \alpha(\frac{n}{k})M(\mathbf{x})\right\}\right| \to 0 \quad \text{a.s.} .$$

Note that the third order condition, the uniformity on $[0,T]^d$ of the convergence in (7) and the continuity of N yield

$$A_{2,k}(\mathbf{x}) = \alpha(\frac{n}{k})M\left(\frac{n}{k}U^{(1)}_{[kx_1],n}, \ldots, \frac{n}{k}U^{(d)}_{[kx_d],n}\right) + O_{\mathbb{P}}\left(\alpha(\frac{n}{k})\beta(\frac{n}{k})\right) .$$

Thanks to (24) and to the existence of the first-order partial derivatives $\partial_j M$ $(j = 1, \ldots, d)$ of the function M, we have that

$$\sup_{0 \le x_1, \ldots, x_d \le T} \left| \sqrt{k} \left\{ M\left(\frac{n}{k} U^{(1)}_{[kx_1],n}, \ldots, \frac{n}{k} U^{(d)}_{[kx_d],n} \right) - M(\mathbf{x}) \right\} + \sum_{j=1}^{d} W_L(x_j \mathbf{e}_j) \partial_j M(\mathbf{x}) \right|$$

converges to 0 in probability, as n tends to infinity. This implies that

$$\sup_{0 \le x_1, \ldots, x_d \le T} \left| \sqrt{k} \left\{ A_{2,k}(\mathbf{x}) - \alpha(\frac{n}{k}) M(\mathbf{x}) \right\} \right| = O_{\mathbb{P}} \left(|\sqrt{k} \alpha(\frac{n}{k}) \beta(\frac{n}{k}) + \alpha(\frac{n}{k})| \right) ,$$

that completes the proof thanks to the choice of the intermediate sequence. □

Proof of Theorem 3. Recall that $b = (a^{-\rho}+1)^{-1/\rho}$ and denote $\hat{b} = (a^{-\hat{\rho}}+1)^{-1/\hat{\rho}}$. Write

$$\mathring{L}_{k,a,k_\rho} - L = \{\hat{L}_{k,a} - L\} + \{\hat{L}_k - L\} - \{\hat{L}_{k,\hat{b}} - L\} , \tag{26}$$

which equals, thanks to (12) and under Skorohod's construction,

$$\begin{aligned}
&\alpha\left(\frac{n}{k}\right)(a^{-\rho}+1)M(\mathbf{x}) + \frac{1}{\sqrt{k}}\left(a^{-1}Z_L(a\mathbf{x}) + Z_L(\mathbf{x})\right) - \alpha\left(\frac{n}{k}\right)\hat{b}^{-\rho}M(\mathbf{x}) - \frac{b^{-1}}{\sqrt{k}}Z_L(b\mathbf{x}) + o\left(\frac{1}{\sqrt{k}}\right)\\
&= \alpha\left(\frac{n}{k}\right)\left((a^{-\rho}+1) - b^{-\rho}\right)M(\mathbf{x}) + \frac{1}{\sqrt{k}}\mathring{Y}_a(\mathbf{x}) + \alpha\left(\frac{n}{k}\right)\left(b^{-\rho} - \hat{b}^{-\rho}\right)M(\mathbf{x}) + o\left(\frac{1}{\sqrt{k}}\right)\\
&= \alpha\left(\frac{n}{k}\right)\left((a^{-\rho}+1) - b^{-\rho}\right)M(\mathbf{x}) + \frac{1}{\sqrt{k}}\mathring{Y}_a(\mathbf{x}) + \alpha\left(\frac{n}{k}\right)O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\right) + o\left(\frac{1}{\sqrt{k}}\right) .
\end{aligned}$$

The first term is zero. Since both $k = o(k_\rho)$ and α is regularly varying with negative index, the one but last term can be put into the term $o\left(\frac{1}{\sqrt{k}}\right)$. Finally, the covariance function follows from the equality in law as processes between $Z_L(a\mathbf{x})$ and $\sqrt{a}Z_L(\mathbf{x})$. □

The proofs of Theorem 4 and Proposition 6 are based on the following auxiliary result.

Lemma 7. *Assume that the conditions of Proposition 2 are fulfilled. Then for any positive real r, one has as n tends to infinity,*

$$\sqrt{k}\alpha(\frac{n}{k}) \left\{ \frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\alpha(\frac{n}{k})} - (a^{-\rho} - 1) r^{1-\rho} M(\mathbf{x}) \right\} \xrightarrow{d} a^{-1} Z_L(ra\mathbf{x}) - Z_L(r\mathbf{x}) ,$$

in $D([0,T]^d)$ for every $T > 0$.

Proof of Lemma 7. Making use of the homogeneity of the function L, write

$$\hat{\Delta}_{k,a}(r\mathbf{x}) = \{\hat{L}_{k,a}(r\mathbf{x}) - L(r\mathbf{x})\} - \{\hat{L}_k(r\mathbf{x}) - L(r\mathbf{x})\} .$$

Using the Skorohod construction, it follows from equations (8) and (12) that

$$\sup_{0 \le x_1, \ldots, x_d \le T/r} \left| \sqrt{k}\alpha(\frac{n}{k}) \left\{ \frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\alpha(\frac{n}{k})} - (a^{-\rho} - 1) r^{1-\rho} M(\mathbf{x}) \right\} - a^{-1} Z_L(ra\mathbf{x}) + Z_L(r\mathbf{x}) \right|$$

tends to 0 almost surely, as n tends to infinity. □

PROOF OF THEOREM 4. Note that

$$\hat{L}_k(\mathbf{x})\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\alpha(n/k_\rho)} - \hat{L}_k(a\mathbf{x})\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})}{\alpha(n/k_\rho)} = \hat{L}_k(\mathbf{x})\left(\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\alpha(n/k_\rho)} - a\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})}{\alpha(n/k_\rho)}\right) - a\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(n/k_\rho)} \, .$$

Under a Skorohod construction, Lemma 7 allows to write the expansions of the terms $\hat{\Delta}_{k,a}(\mathbf{x})$, $\hat{\Delta}_{k_\rho,a}(\mathbf{x})$ and $\hat{\Delta}_{k_\rho,a}(a\mathbf{x})$, which implies on the one hand

$$\begin{aligned}\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\alpha(n/k_\rho)} - a\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})}{\alpha(n/k_\rho)} &= a(a^{-\rho}-1)^2 M(\mathbf{x}) \\ &+ \frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\left\{a^{-1}Z_L(a^2\mathbf{x}) - 2Z_L(a\mathbf{x}) + aZ_L(\mathbf{x})\right\} + o\left(\frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\right) \, ,\end{aligned} \tag{27}$$

and

$$\begin{aligned}\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(n/k_\rho)} &= \alpha(n/k)(a^{-\rho}-1)^2M^2(\mathbf{x}) + (a^{-\rho}-1)M(\mathbf{x})\frac{a^{-1}Z_L(a\mathbf{x}) - Z_L(\mathbf{x})}{\sqrt{k}} \\ &+ O_{\mathbb{P}}\left(\frac{\alpha(n/k)}{\sqrt{k_\rho}\alpha(n/k_\rho)} + \frac{1}{\sqrt{k}\sqrt{k_\rho}\alpha(n/k_\rho)}\right) + o\left(\frac{1}{\sqrt{k}}\right) \, .\end{aligned} \tag{28}$$

on the other hand, both uniformly for $\mathbf{x} \in [0,T]^d$. Combining (27) and (28) with equation (8), one gets

$$\begin{aligned}&\hat{L}_k(\mathbf{x})\frac{\hat{\Delta}_{k_\rho,a}(a\mathbf{x})}{\alpha(n/k_\rho)} - \hat{L}_k(a\mathbf{x})\frac{\hat{\Delta}_{k_\rho,a}(\mathbf{x})}{\alpha(n/k_\rho)} \\ &= a(a^{-\rho}-1)^2M(\mathbf{x})L(\mathbf{x}) + \frac{1}{\sqrt{k}}M(\mathbf{x})(a^{-\rho}-1)\left(a^{1-\rho}Z_L(\mathbf{x}) - Z_L(a\mathbf{x})\right) \\ &+ \frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}L(\mathbf{x})\left\{a^{-1}Z_L(a^2\mathbf{x}) - 2Z_L(a\mathbf{x}) + aZ_L(\mathbf{x})\right\} \\ &+ o\left(\frac{1}{\sqrt{k}}\right) + o\left(\frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\right) \, .\end{aligned}$$

Since the last expression and equation (27) are respectively the numerator and denominator of $\tilde{L}_{k,k_\rho,a}(\mathbf{x})$, one obtains after simplifications

$$\sqrt{k}(\tilde{L}_{k,k_\rho,a}(\mathbf{x}) - L(\mathbf{x})) = \frac{a^{-\rho}Z_L(\mathbf{x}) - a^{-1}Z_L(a\mathbf{x})}{a^{-\rho}-1} + o\left(\frac{\sqrt{k}}{\sqrt{k_\rho}\alpha(n/k_\rho)}\right) + o(1) \, .$$

The choice of the sequences k and k_ρ allows to conclude since $\sqrt{k} = O\left(\sqrt{k_\rho}\alpha(n/k_\rho)\right)$. □

PROOF OF PROPOSITION 5. Applying Lemma 7, we have

$$\sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha\left(\frac{n}{k}\right)} - (a^{-\rho}-1)M(\mathbf{x})\right| \xrightarrow{\mathbb{P}} 0 \, . \tag{29}$$

As a consequence,

$$\begin{aligned}\sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\hat{\Delta}_{k,a}(\mathbf{x})} - r^{1-\rho}\right| &= \sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\hat{\Delta}_{k,a}(r\mathbf{x})/\alpha(n/k)}{\hat{\Delta}_{k,a}(\mathbf{x})/\alpha(n/k)} - r^{1-\rho}\right| \\ &= O_{\mathbb{P}}\left(\sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\alpha(n/k)} - r^{1-\rho}\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(n/k)}\right|\right) \, ,\end{aligned}$$

since $(a^{-\rho}-1)M(\mathbf{x}) \neq 0$ by assumption. Writing

$$\left|\frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\alpha(n/k)} - r^{1-\rho}\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(n/k)}\right| \leq \left|\frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\alpha(n/k)} - r^{1-\rho}(a^{-\rho}-1)M(\mathbf{x})\right| + \left|r^{1-\rho}(a^{-\rho}-1)M(\mathbf{x}) - r^{1-\rho}\frac{\hat{\Delta}_{k,a}(\mathbf{x})}{\alpha(n/k)}\right|$$

and using twice equation (29) leads to the conclusion.

□

PROOF OF PROPOSITION 6. Let denote by $Q_{k,a,r}$ the quotient

$$Q_{k,a,r}(\mathbf{x}) := \frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\hat{\Delta}_{k,a}(\mathbf{x})} .$$

Lemma 7 used twice allows to write

$$\sqrt{k}\alpha(\frac{n}{k})(Q_{k,a,r}(\mathbf{x}) - r^{1-\rho}) \xrightarrow{d} -r^{1-\rho}\log r \hat{Z}_{\rho,a,r}(\mathbf{x}) , \tag{30}$$

where $\hat{Z}_{\rho,a,r}(\mathbf{x})$ is defined in the Proposition 6. Since $\hat{\rho}_{k,a,r}(\mathbf{x}) = 1 - \log(Q_{k,a,r}(\mathbf{x}))/\log r$, the result follows straightforwardly from (30) and the Delta method. □

Acknowledgements. This research has been partially supported by the Agence Nationale de la Recherche through the AST&RISK project (ANR-08-BLAN-0314-01), and by FCT/PTDC/MAT/112770/2009 (Portugal).

We wish to thank Armelle Guillou for pointing out a deficiency in the original version of the paper. We thank the referees for very helpful comments.

References.

ABDOUS, B. and GHOUDI, K. (2005). Non-parametric estimators of multivariate extreme dependence functions. *J. Nonparametr. Stat.* **17** 915–935.

BEIRLANT, J., DIERCKX, G. and GUILLOU, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance Math. Econom.* **49** 18–26. MR2811890 (2012f:62092)

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. L. and SEGERS, J. (2004). *Statistics of extremes. Wiley Series in Probability and Statistics.* John Wiley & Sons Ltd., Chichester.

BRUUN, J. T. and TAWN, J. A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *Applied Statistics.*

BÜCHER, A., DETTE, H. and VOLGUSHEV, S. (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist.* **39** 1963–2006. MR2893858

CAEIRO, F., GOMES, M. I. and RODRIGUES, L. H. (2009). Reduced-bias tail index estimators under a third-order framework. *Comm. Statist. Theory Methods* **38** 1019–1040.

CAI, J. J., DE HAAN, L. and ZHOU, C. (2013). Bias reduction in extreme value statistics with index around zero. *Extremes* **16** 173–201.

CAPÉRAÀ, P. and FOUGÈRES, A. L. (2000). Estimation of a bivariate extreme value distribution. *Extremes* **3** 311–329 (2001).

CAPÉRAÀ, P., FOUGÈRES, A. L. and GENEST, C. (2000). Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.* **72** 30–49.

CIUPERCA, G. and MERCADIER, C. (2010). Semi-parametric estimation for heavy tailed distributions. *Extremes* **13** 55–87.

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392.

DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. New York, NY: Springer.

de Haan, L., Mercadier, C. and Zhou, C. (2014). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Submitted*.

de Haan, L. and Resnick, S. I. (1993). Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models* **9** 275–309.

de Haan, L. and Sinha, A. K. (1999). Estimating the probability of a rare event. *The Annals of Statistics* **27** 732–759.

Demarta, S. and McNeil, A. J. (2005). The t Copula and Related Copulas. *International Statistical Review* **73** 111–129.

Einmahl, J. H. J., de Haan, L. and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70** 143–171.

Einmahl, J. H. J., Krajina, A. and Segers, J. (2008). A method of moments estimator of tail dependence. *Bernoulli* **14** 1003–1026. MR2543584 (2011b:62143)

Einmahl, J. H. J., Krajina, A. and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793.

Fraga Alves, M. I., de Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods Statist.* **12** 155–176.

Goegebeur, Y. and Guillou, A. (2013). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand. J. Stat.* **40** 174–189.

Gomes, M. I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes* **5** 387–414 (2003).

Gomes, I. M., de Haan, L. and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 31–52.

Guillotte, S., Perron, F. and Segers, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 377–406. MR2815781 (2012k:62049)

Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13** 331–341. MR773171 (86h:62058)

Huang, X. (1992). Statistics of bivariate extremes. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* **9** 75–81. MR1035994 (91c:62051)

Joe, H., Smith, R. L. and Weissman, I. (1992). Bivariate threshold methods for extremes. *J. Roy. Statist. Soc. Ser. B* **54** 171–183.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous multivariate distributions. Vol. 1*, second ed. *Wiley Series in Probability and Statistics: Applied Probability and Statistics*. Wiley-Interscience, New York Models and applications. MR1788152 (2001h:62001)

Krajina, A. (2012). A method of moments estimator of tail dependence in meta-elliptical models. *J. Statist. Plann. Inference* **142** 1811–1823. MR2903392

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

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B.* **59** 475–499.

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.* **38** 107–115.

Peng, L. (2010). A practical way for estimating tail dependence functions. *Statist. Sinica* **20** 365–378.

Resnick, S. I. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probability* **18** 66–138.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75** 397–415.

Université de Lyon, CNRS, Université Lyon 1
Institut Camille Jordan
43 blvd du 11 novembre 1918
F-69622 Villeurbanne-Cedex
France
E-mail: fougeres@math.univ-lyon1.fr
mercadier@math.univ-lyon1.fr

Department of Economics
Erasmus university
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
E-mail: ldehaan@ese.eur.nl