# Bias correction in multivariate extremes

Anne-Laure Fougères, Laurens de Haan, Cécile Mercadier

# Bias correction in multivariate extremes

Anne-Laure Fougères[1], Laurens de Haan[2], and Cécile Mercadier[1]
fougeres@math.univ-lyon1.fr, ldehaan@ese.eur.nl, mercadier@math.univ-lyon1.fr

1. Université de Lyon, CNRS, Université Lyon 1, Institut Camille Jordan,
43 blvd du 11 novembre 1918, F-69622 Villeurbanne-Cedex, France

2. Department of Economics, Erasmus university,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

January 10, 2014

**Abstract**

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.

**AMS 2010 subject classification:** Primary 62G32, 62G05, 62G20. Secondary 60F05, 60G70.

**Keywords:** Multivariate extreme value theory, tail dependence, bias correction, threshold choice.

## 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 $L$, firstly introduced by Huang [1992]. For any arbitrary dimension $d$, consider a multivariate vector $(X^{(1)}, \ldots, X^{(d)})$ with continuous marginal cumulative distribution functions (cdf) $F_1, \ldots, F_d$. The stable tail dependence function is defined for each positive reals $x_1, \ldots, x_d$ as

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

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

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

where $F$ denotes the multivariate distribution function of the vector $(X^{(1)},\ldots,X^{(d)})$. 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 $k/n \to 0$. Let us denote $\mathbf{x} = (x_1,\ldots,x_d)$ a vector of the positive quadrant $\mathbb{R}_+^d$ and $X_{k,n}^{(j)}$ the $k$th order statistics among $n$ realisations of the margins $X^{(j)}$. The empirical estimator of $L(\mathbf{x})$ is then obtained from (1), replacing $F$ by its empirical version, $t$ by $n/k$, and $F_j^{-1}(1-t^{-1}x_j)$ for $j = 1,\ldots,d$ by its empirical counterpart $X^{(j)}_{n-[nt^{-1}x_j],n}$, so that

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

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, $\alpha$ 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 $\lambda$, 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 substracted to 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 et al. [2013]. Refer also to Peng [1998], Gomes and Martins [2002] and Gomes et al. [2008] for previous contributions on this problem. Note finally that the case of dependent sequences has been recently studied by de Haan et al. [2013].

The nonparametric estimation of the extremal dependence structure has been widely studied in the bivariate case, see for instance Huang [1992], Einmahl et al. [1997], Capéraà and Fougères [2000], Abdous and Ghoudi [2005], Guillotte et al. [2011] and Bücher et al. [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 et al. [2011] which considers 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 et al. [1992], Einmahl et al. [2008] and Einmahl et al. [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. Two families of asymptotically unbiased estimators of the stable tail dependence function 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. The estimation of the second order parameter is postponed up to Section 4. Several theoretical models are exhibited in Section 5, that satisfy the required assumptions. Section 6 illustrates the performance of the new estimators on both simulated and real data. 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 cumulative distribution function $F$ and continuous marginal distribution functions $F_j$ for $j = 1, \ldots, d$. Suppose $F$ is in the domain of attraction of an extreme value distribution $G$. We recall that it supposes the existence for $j = 1, \ldots, d$ of sequences $a_n^{(j)} > 0$, $b_n^{(j)}$ of real numbers and a distribution function $G$ with nondegenerate marginals such that

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

for all points $\mathbf{x}$ where $G$ is continuous. Denote by $G_j$ the $j$th marginal distribution function 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

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

Let $\mu$ 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 in assuming that the limit given in (1) exists, and is uniform on any $[0, T]^d$, for $T > 0$.
- the *second order condition* consists in assuming the existence of a positive function $\alpha$, 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,

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

uniformly on any $[0, T]^d$, for $T > 0$.

- the *third order condition* consists in assuming the existence of a positive function $\beta$, 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,

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

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*

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

*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 $\rho$ and $\rho'$ such that $\alpha$ (resp. $\beta$) is a regularly varying function of order $\rho$ (resp. $\rho'$), i.e. $\alpha(tz)/\alpha(t) \to z^\rho$ and $\beta(tz)/\beta(t) \to z^{\rho'}$ when $t \to \infty$, for each positive $z$. Besides, the limit function $M$ 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. The limit function $N$ is homogeneous of order $1-\rho-\rho'$, so $N(r\mathbf{x}) = r^{1-\rho-\rho'}N(\mathbf{x})$, for each positive $r$ and $\mathbf{x}$ with positive coordinates.*

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 $\rho$ of the function $\alpha$ defined in (6) is negative;
- the function $M$ defined in (6) is continuous.

These hypotheses allow to get the asymptotic uniform behavior 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 cumulative distribution function $F$ and continuous marginal distribution functions $F_j$ 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 $\{\mathbf{x} \in \mathbb{R}_+^d : x_j > 0\}$.*
*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 $\rho$ and $\rho'$ of the functions $\alpha$ and $\beta$ defined in (6) and (7) are negative;
- the function $M$ defined in (6) is differentiable and $N$ defined in (7) is continuous.

**Proposition 2.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent multivariate random vectors in $\mathbb{R}^d$ with common joint cumulative distribution function $F$ and continuous marginal distribution functions $F_j$ for $j = 1, \ldots, d$. Assume that the set of conditions* **(A3)** *hold. Suppose further that the first order partial derivatives of $L$ (denoted by $\partial_1 L$ for $j = 1, \ldots, d$) exist and that $\partial_1 L$ is continuous on the set of points $\{\mathbf{x} \in \mathbb{R}_+^d : x_j > 0\}$.*
*Consider $\hat{L}_k$ the estimator of $L$ defined by (2) where $k$ is such that $\sqrt{k}\alpha(n/k) \to \infty$ 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*

$$\mathbb{E}[W_L(\mathbf{x}) W_L(\mathbf{y})] = \mu\{R(\mathbf{x}) \cap R(\mathbf{y})\} \, . \tag{10}$$

*given in terms of the measure $\mu$ defined by (5) and of*

$$R(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}_+^d : \text{there exists } j \text{ such that } 0 \le u_j \le x_j\} \, .$$

**Remark 4.** *A difference between the previous result and Theorem 7.2.2 of de Haan and Ferreira [2006] consists in 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 5.** *The conditions on $k$, $\alpha$ and $\beta$ required in Proposition 2 are not too restrictive: because of the regular variation of $\alpha$ and $\beta$, they are implied by the choice $k(n) = n^\kappa$, with $\kappa \in \left( -\dfrac{2\rho}{1-2\rho}, -\dfrac{2(\rho+\rho')}{1-2(\rho+\rho')} \right)$.*

## 3 Bias reduction procedure

As pointed out in Remark 4, a non-null asymptotic bias $\alpha(n/k)M(\mathbf{x})$ appears from Proposition 2. The bias reduction procedure will consist in substracting 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 introduce a positive scale parameter $a$ which allows to contract or to dilate the observed points. Let denote

$$\hat{L}_{k,a}(\mathbf{x}) := a^{-1}\hat{L}_k(a\mathbf{x}) \;, \tag{11}$$

and

$$\Delta_{k,a}(\mathbf{x}) := \hat{L}_{k,a}(\mathbf{x}) - \hat{L}_k(\mathbf{x}) \;. \tag{12}$$

From (8) one gets

$$\sqrt{k}\left\{\hat{L}_{k,a}(\mathbf{x}) - L(\mathbf{x}) - \alpha(\frac{n}{k})a^{-\rho}M(\mathbf{x})\right\} \xrightarrow{d} a^{-1}Z_L(a\mathbf{x}) \;, \tag{13}$$

in $D([0,T]^d)$ for every $T>0$. Equations (12) and Proposition 1 yield as $n$ tends to infinity,

$$\frac{\Delta_{k,a}(\mathbf{x})}{\alpha(\frac{n}{k})} \xrightarrow{\mathbb{P}} (a^{-\rho}-1)M(\mathbf{x}) \;. \tag{14}$$

Fixing $a$ such that $a^{-\rho}-1=1$, a natural estimator of the asymptotic bias of $\hat{L}_k(\mathbf{x})$ is thus $\Delta_{k,2^{-1/\hat{\rho}}}(\mathbf{x})$, where $\hat{\rho}$ is an estimator of $\rho$. Recall that the unknown parameter $\rho$ is the regular variation index of the function $\alpha$ involved in the second order condition. Let $k_\rho$ 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

$$\mathring{L}_{k,1,k_\rho}(\mathbf{x}) := \hat{L}_k(\mathbf{x}) - \Delta_{k,2^{-1/\hat{\rho}}}(\mathbf{x}) \;. \tag{15}$$

The asymptotic behavior of this estimator is provided in Theorem 3 and Remark 7. We refer the reader to Section 4 for more details concerning the estimation of the second order parameter $\rho$.

### 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 (13) 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 $b$,

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

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 $L$

$$\mathring{L}_{k,a,k_\rho}(\mathbf{x}) := \hat{L}_{k,a}(\mathbf{x}) - \Delta_{k,(a^{-\hat{\rho}}+1)^{-1/\hat{\rho}}}(\mathbf{x}) \; . \tag{16}$$

**Theorem 3.** *Assume that the conditions of Proposition 2 are fulfilled and consider the estimator of $L$ defined by (16). Let $k_\rho$ 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,*

$$\sqrt{k}\left\{\mathring{L}_{k,a,k_\rho}(\mathbf{x}) - L(\mathbf{x})\right\} \xrightarrow{d} \mathring{Y}_a(\mathbf{x}) \; , \tag{17}$$

*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*

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

*Here $b = (a^{-\rho}+1)^{-1/\rho}$.*

**Remark 6.** *The assumption that $\sqrt{k_\rho}\alpha(n/k_\rho)(\hat{\rho}-\rho)$ converges in distribution will be reconsidered in Section 4.*

**Remark 7.** *Theorem 3 remains true when $a = 1$ and thus characterizes the asymptotic behavior of the estimator given in (15). 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 $\rho$. 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 (13). Making use of (14), one gets as $n$ tends to infinity

$$\frac{\Delta_{k_\rho,a}(a\mathbf{x})}{\Delta_{k_\rho,a}(a\mathbf{x}) - a\Delta_{k_\rho,a}(\mathbf{x})} \xrightarrow{\mathbb{P}} \frac{a^{-\rho}}{a^{-\rho}-1} \; ,$$

for any intermediate sequence $k_\rho$ that satisfies $\sqrt{k_\rho}\alpha(n/k_\rho) \to \infty$. As a consequence, the expression

$$\Delta_{k,a}(\mathbf{x})\frac{\Delta_{k_\rho,a}(a\mathbf{x})}{\Delta_{k_\rho,a}(a\mathbf{x}) - a\Delta_{k_\rho,a}(\mathbf{x})}$$

can 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

$$\tilde{L}_{k,a,k_\rho}(\mathbf{x}) = \frac{\hat{L}_k(\mathbf{x})\Delta_{k_\rho,a}(a\mathbf{x}) - \hat{L}_k(a\mathbf{x})\Delta_{k_\rho,a}(\mathbf{x})}{\Delta_{k_\rho,a}(a\mathbf{x}) - a\Delta_{k_\rho,a}(\mathbf{x})} \; , \tag{18}$$

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 (18). Let $k_\rho$ 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,*

$$\sqrt{k}\left\{\tilde{L}_{k,a,k_\rho}(\mathbf{x}) - L(\mathbf{x})\right\} \xrightarrow{d} \tilde{Y}_a(\mathbf{x}) \,, \tag{19}$$

*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 8.** *The covariance function specified above is decreasing with respect to the parameter $a$ for any fixed value of $\rho$. 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 (18). See Section 6 for practical considerations for the choice of $a$.*

## 4 Estimation of the second order parameter $\rho$

Let give additional details on the estimation of the second-order parameter $\rho$. A possible method could be to use on each margin some univariate inference techniques developed e.g. by Gomes et al. [2002] or Ciuperca and Mercadier [2010]. Instead of this, the construction described here makes use of the multivariate structure of the data. On this purpose, the following proposition shows that a variable of interest is the ratio of two terms $\Delta_{k,a}$, defined by (12).

**Proposition 5.** *Assume that the conditions of Proposition 1 are fulfilled and fix positive real numbers $r$ and $a \in (0,1)$. Then, as $n$ tends to infinity,*

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

A first consequence of Proposition 5 is the derivation of a family of consistent estimators of the parameter $\rho$

$$\hat{\rho}_{k,a,r}(\mathbf{x}) := 1 - \frac{1}{\log r}\log\left\{\frac{\Delta_{k,a}(r\mathbf{x})}{\Delta_{k,a}(\mathbf{x})}\right\} \,. \tag{20}$$

A second consequence of Proposition 5 is to derive several estimators of the parameter $\rho$ that are integrals. The family of estimators

$$\tilde{\rho}_{k,a}(\mathbf{x}) := 2 - \frac{\Delta_{k,a}(\mathbf{x})}{\int_0^1 \Delta_{k,a}(r\mathbf{x})dr} \tag{21}$$

is consistent for $\rho$. We can also consider the following consistent estimator

$$\check{\rho}_{k,s}(\mathbf{x}) := \frac{\int_0^1 \Delta_{k,a}(\mathbf{x})da - \int_0^1 \Delta_{k,a}(\mathbf{x})\, s^2\, a^{s-1}\, da}{\int_0^1 \Delta_{k,a}(\mathbf{x})\, s\, a^{s-1}\, da - \int_0^1 \Delta_{k,a}(\mathbf{x})da} \,, \tag{22}$$

where $s$ is a positive real number and $\mathbf{x} \in [0,T]^d$. Making use of the asymptotic expansion given in Proposition 2, one can prove the following asymptotic properties.

**Proposition 6.** *Assume that the conditions of Proposition 2 are fulfilled, and consider the estimators of $\rho$ defined by (20), (21) and (22). Then, as $n$ tends to infinity,*

$$\sqrt{k}\alpha(\frac{n}{k})\{\hat{\rho}_{k,a,r}(\mathbf{x})-\rho\} \xrightarrow{d} \hat{Z}_{\rho,a,r}(\mathbf{x}) \ ,$$
$$\sqrt{k}\alpha(\frac{n}{k})\{\tilde{\rho}_{k,a}(\mathbf{x})-\rho\} \xrightarrow{d} \tilde{Z}_{\rho,a}(\mathbf{x}) \ ,$$
$$\sqrt{k}\alpha(\frac{n}{k})\{\check{\rho}_{k,s}(\mathbf{x})-\rho\} \xrightarrow{d} \check{Z}_{\rho,s}(\mathbf{x}) \ ,$$

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

$$\hat{Z}_{\rho,a,r}(\mathbf{x}) := \frac{a^{-1}Z_L\left(a\mathbf{x}\right)-Z_L(\mathbf{x})}{(a^{-\rho}-1)M(\mathbf{x})\log r} - \frac{a^{-1}Z_L\left(ra\mathbf{x}\right)-Z_L(r\mathbf{x})}{(a^{-\rho}-1)M(\mathbf{x})r^{1-\rho}\log r} \ ,$$
$$\tilde{Z}_{\rho,a}(\mathbf{x}) := \frac{\int_0^1\{a^{-1}Z_L(ra\mathbf{x})-Z_L(r\mathbf{x})\}dr}{(a^{-\rho}-1)M(\mathbf{x})/(2-\rho)^2} - \frac{a^{-1}Z_L(a\mathbf{x})-Z_L(\mathbf{x})}{(a^{-\rho}-1)M(\mathbf{x})/(2-\rho)} \ ,$$

*and where $\check{Z}_{\rho,s}(\mathbf{x})$ is given by*

$$\frac{(-\rho+1)^2}{1-s}\left(\frac{\int_0^1\{a^{-1}Z_L(a\mathbf{x})-Z_L(\mathbf{x})\}da}{\frac{\rho}{-\rho+s}M(\mathbf{x})} - \frac{\int_0^1\{a^{-1}Z_L(a\mathbf{x})-Z_L(\mathbf{x})\}sa^{s-1}da}{\frac{\rho(-\rho+1)}{(-\rho+s)^2}M(\mathbf{x})}\right) \ .$$

In Section 6, for simplicity, the estimator $\hat{\rho}_{k,a,r}$ will be used.

## 5 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 is to get the pointwise convergence is to obtain, for well chosen positive reals $p$ and $q$, an expansion 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

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

### 5.1 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 $\alpha$ and $\beta$ 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 discuss some particular choices of the norm:

- For $L^1$-norm and $\alpha = 1$, the model coincides with the bivariate Pareto of type II distribution, denoted by BPII($\beta$) in this paper, and referred to as $\mathrm{MP}^{(2)}(II)(0,1,\beta-2)$ on page 604 of Kotz et al. [2000]. In this case, $p = q = (\beta-2)^{-1}$, and

  $$L(x,y) = x + y - (x^{-p} + y^{-p})^{-1/p} .$$

  The latter stable tail dependence function 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 $\theta$ arise choosing the norm $\|(x,y)^T\| = \nu^{-1/2}(x^2 - 2\theta xy + y^2)^{1/2}$, for a positive real number $\nu$, $\alpha = 2$, $\beta = (\nu+2)/2$ 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 stable tail dependence function 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 distribution function 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 $\rho$ such that $|\rho| < 1$ and some positive real numbers $\sigma$ and $\tau$.

### 5.2 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 $2\times 2$ 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 bivariate probability density function of $(x,y)$ as

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

A sufficient condition to satisfy (7) is to assume that the distribution of $R$ belongs to the Hall and Welsh class, 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 $\rho$ and $\rho_1$.

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

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

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 stable tail dependence function 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 5.1 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.

### 5.3 Archimax distributions

Consider the bivariate distribution function defined for each $0 \le u,v \le 1$ by

$$F(u,v) = \left\{1 + L(u^{-1}-1, v^{-1}-1)\right\}^{-1} \ , \tag{23}$$

given in terms of a stable tail dependence function $L$. This distribution has standard uniform univariate margins and corresponds to a particular case of Archimax bivariate copulas introduced in Capéraà et al. [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

$$t\left\{1 - F\left(1-t^{-1}x,\, 1-t^{-1}y\right)\right\} = L(x,y) - t^{-1}L^2(x,y) + t^{-2}L^3(x,y) + o\left(t^{-2}\right) \ ,$$

which allows to identify $M=-L^2$ and $N=L^3$, so that $\rho=\rho'=-1$.

### 5.4 Multivariate symmetric logistic distributions

Consider the bivariate distribution function defined for each $x, y \in \mathbb{R}$ by

$$F(x,y) = \exp\left\{-\left(e^{-x/r} + e^{-y/r}\right)^r\right\} , \tag{24}$$

which corresponds to the multivariate extreme value distribution with Gumbel univariate margins $F_1(x) = F_2(x) = \exp\{-e^{-x}\}$ and symmetric logistic stable tail dependence function

$$L(x,y) = (x^{1/r} + y^{1/r})^r ,$$

where $0 < r \leq 1$. 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

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

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

This allows to identify $\rho = \rho' = -1$.

## 6 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 5. Thanks to the homogeneity property, one can focus on the estimation of $t \mapsto L(1-t,t)$ for $0 \leq t \leq 1$, which coincides with the Pickands dependence function $A$ (see e.g. [Beirlant et al., 2004, p. 267]). Focusing first on the estimation at $t = 1/2$ leads to consider aggregated versions of our estimators. These new estimators will be both compared in terms of $L^1$-errors for $L$ or associated level curves.

### 6.1 Estimators in practice

Let 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 5.1 and 5.2. 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), (16) and (18). For the last two estimators, the parameters have been tuned as follows: $a = 0.4$, $k_\rho = 990$ and $\rho$ estimated using (20) 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 corrected efficiently by the two estimators $\mathring{L}_k$ and $\tilde{L}_k$. Since the bias almost vanishes along the range of $k$, one can

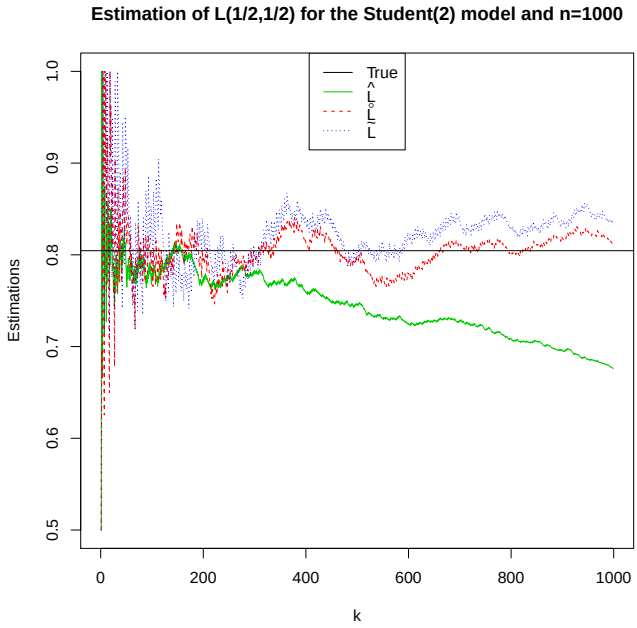



Figure 1: Estimation of $L(1/2, 1/2)$ for the bivariate Student(2) law based on a sample of size 1000.

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

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

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

where $n$ is the sample size and $\kappa_n$ is an appropriate fraction of $n$. Their performances will be compared to those of the family $\{\hat{L}_k, k = 1, \ldots, n-1\}$. Because any stable tail dependence function $L$ satifies $\max(t, 1-t) \leq L(1-t, t) \leq 1$, the competitors have been corrected so that they satisfy the same inequalities.

**Remark 9.** *If $\kappa_n$ 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 the $\mathring{L}_k$ and $\tilde{L}_k$.*

**Remark 10.** *In the following simulation study, $\kappa_n$ 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, $\kappa_n$ should not exceed the size of the subpopulation with heaviest tail.*

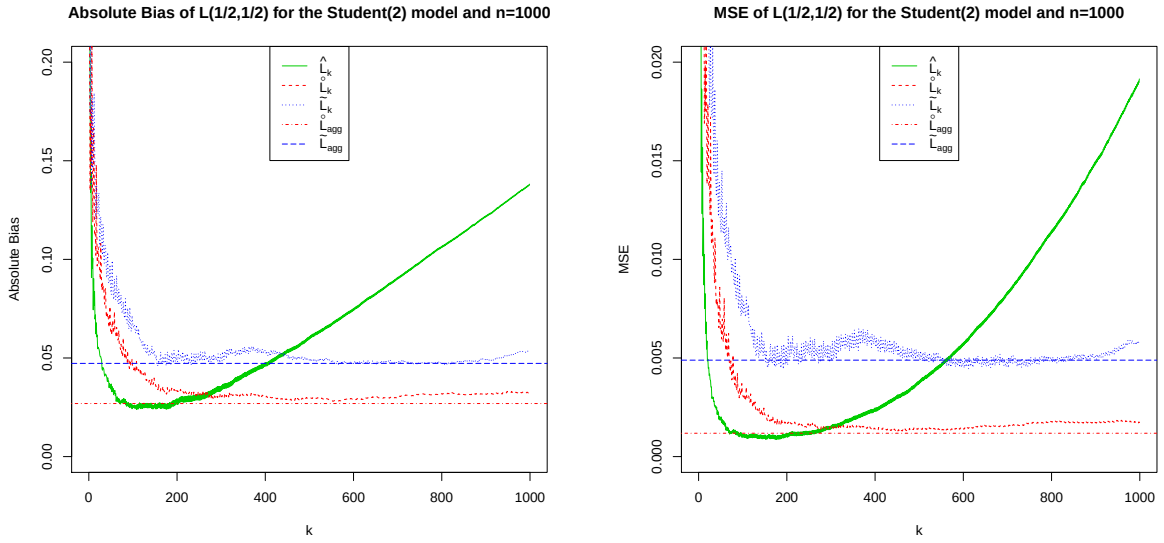
Classical criteria of quality of an estimator $\hat{\theta}$ of $\theta$ are the absolute bias (ABias) and the

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 $i$th sample. 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




(a) (b)

Figure 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 $k$.

the 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 (16) outperforms the estimator $\tilde{L}_k$ defined by (18), 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 $\mathring{L}_{\text{agg}}$ answers surprisingly well to the estimation problem of the stable tail dependence function $L$. 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 10). These comparisons have also be done for five other models obtained from Section 5. The results are very similar to the ones obtained for the bivariate Student(2) distribution and are therefore not presented.

### 6.2 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

$$\frac{1}{T+1}\sum_{t=1}^{T}\left|\hat{L}\left(1-\frac{t}{T},\frac{t}{T}\right)-L\left(1-\frac{t}{T},\frac{t}{T}\right)\right| \tag{25}$$

where $T$ is the size of the subdivision of $[0,1]$. Figure 3 gives the boxplots based on $N=100$ realisations of $\mathring{L}_{\text{agg}}, \tilde{L}_{\text{agg}}$ and $\{\hat{L}_k, k=1,\ldots,n-1\}$ 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 $L(x,y)=(x^2+y^2+xy)/(x+y)$.

As already observed in Figure 2, the estimator $\mathring{L}_{\text{agg}}$ is again very competitive compared to the best element of $\{\hat{L}_k, k=1,\ldots,n-1\}$ 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.

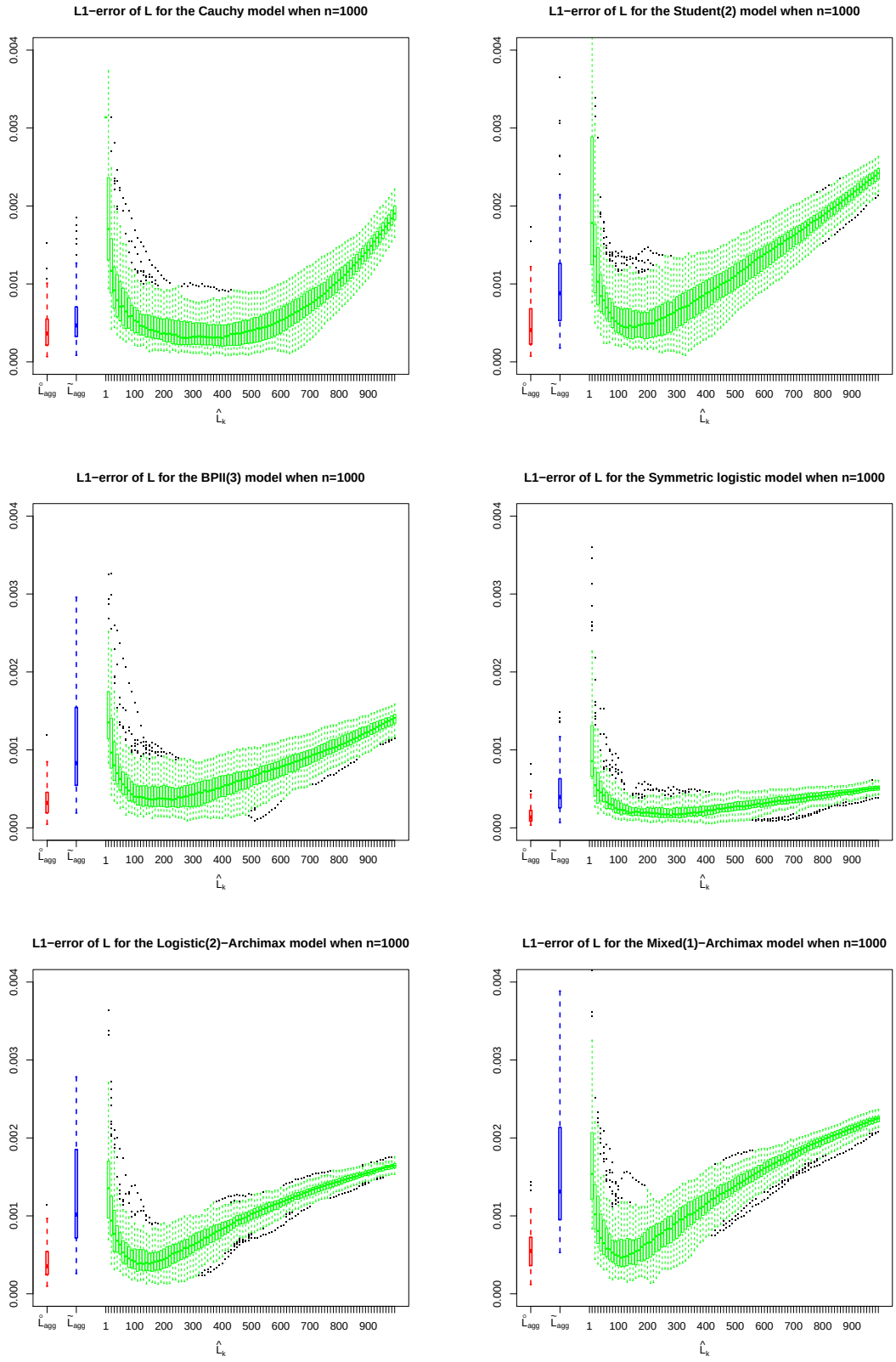



Figure 3: Boxplot of the $L^1$-error of function $L$ for the estimators $\mathring{L}_{\text{agg}}, \tilde{L}_{\text{agg}}$ and $\{\hat{L}_k, k = 1, \ldots, n-1\}$.
First row: bivariate Cauchy model (left) and bivariate Student(2) model (right).
Second row: bivariate BPII(3) model (left) and bivariate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

### 6.3 $Q$-curves

Another nice representation of a function of several variables is through its level sets. In the case of the function $L$, it consists in looking at sets of the form $\{(x,y) \in \mathbb{R}_+^2, L(x,y) \leq c\}$ for any positive real $c$. 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 $\partial Q$ is naturally obtained by replacing $L$ by any estimator, and this is done here for the estimators $\mathring{L}_{\text{agg}}, \tilde{L}_{\text{agg}}$ and $\{\hat{L}_k, k = 1, \ldots, n-1\}$. Figure 4 exhibits the bias phenomenon (as $k$ increases) induced by the empirical estimator $\hat{L}_k$ 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, whereas $\tilde{L}_{\text{agg}}$ suffers of some lack of regularity and a large variability.

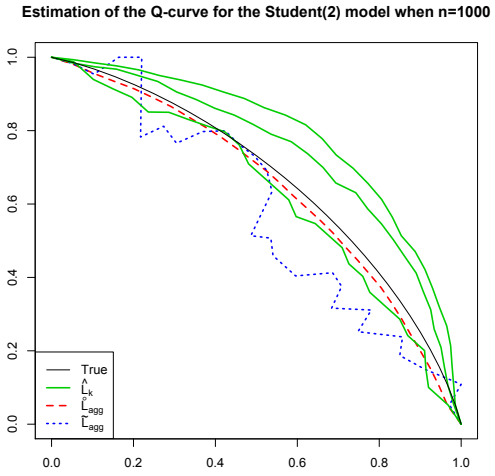



Figure 4: Estimation of the $Q$-curve for the bivariate Student(2) law based on a sample of size 1000.

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

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

Figure 5 displays the boxplot of this measure, based on $N = 100$ realisations and for $T = 30$ under the six bivariate models given in the previous section. 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 $\mathring{L}_{\text{agg}}$

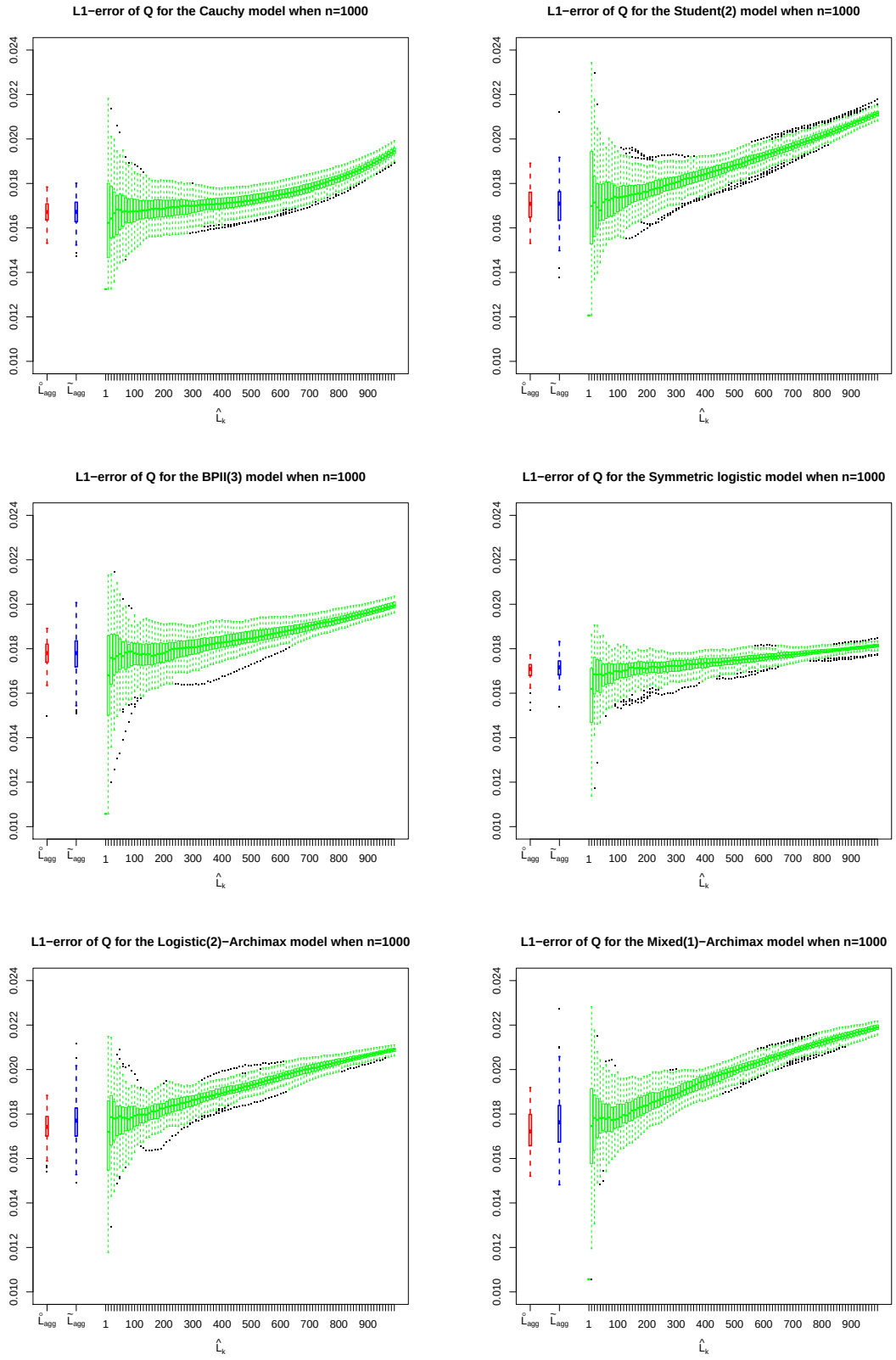



Figure 5: Boxplot of the $L^1$-error of $Q$-curve for the estimators $\mathring{L}_{\mathrm{agg}}, \tilde{L}_{\mathrm{agg}}$ and $\{\hat{L}_k, k = 1, \ldots, n-1\}$.
First row: bivariate Cauchy model (left) and bivariate Student(2) model (right).
Second row: bivariate BPII(3) model (left) and bivariate symmetric logistic model (right).
Third row: bivariate Archimax model with logistic (left) and mixed generator (right).

and $\tilde{L}_{\text{agg}}$ are similar to that of the best $\hat{L}_k$, which is unknown in practice. These features corroborate the conclusions drawn in Section 6.2.

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 6. 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.

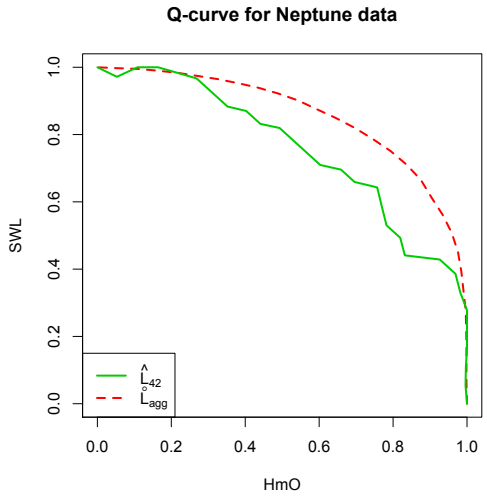



Figure 6: Estimated $Q$-curve for Neptune data set

## Concluding comments

This paper deals with the estimation of the extremal dependence structure in a multivariate context. Focusing on the stable tail dependence function, 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 6.1 appears as a worthy candidate to estimate the stable tail dependence function, even if it doesn't completely solve the problem of the threshold choice as pointed out in Remark 10.

# 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

$$V_k(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\left\{U_i^{(1)} \leq k x_1/n \text{ or } \dots \text{ or } U_i^{(d)} \leq k x_d/n\right\}}$$

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

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

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] we get that for any intermediate sequence

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

in $D([0,T]^d)$ for every $T > 0$ where $W_L$ is a continuous centered Gaussian process with covariance structure specified in (10). Due to the Skorohod construction we can write

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| \sqrt{k} A_{1,k}(\mathbf{x}) - W_L(\mathbf{x}) \right| \to 0 \quad \text{a.s.} , \tag{26}$$

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

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

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

$$\sup_{x \in [0,T]} \left| \sqrt{k}\left(\frac{n}{k} U^{(j)}_{[kx],n} - x\right) + W_L(x\mathbf{e}_j) \right| \to 0 \quad \text{a.s.} , \tag{27}$$

and finally

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| \sqrt{k} A_{3,k}(\mathbf{x}) + \sum_{j=1}^{d} W_L(x_j \mathbf{e}_j) \partial_j L(\mathbf{x}) \right| \to 0 \quad \text{a.s.} . \tag{28}$$

As previously, this yields

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

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

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

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

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

Then the result follows from

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

that is obtained combining (27) 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 (26). Combined with (28), it is sufficient to prove the convergence

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

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 (27) and to the existence of the first-order partial derivatives $\partial_1 M$ and $\partial_2 M$ of the function $M$, we have that

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

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

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

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

$$\mathring{L}_{k,a,k_\rho} - L = \{\hat{L}_{k,a} - L\} + \{\hat{L}_k - L\} - \{\hat{L}_{k,\hat{b}} - L\} \,, \tag{29}$$

which equals, thanks to (13) 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 $\alpha$ 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,*

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

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

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

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

Using the Skorohod construction, it follows from equations (8) and (13) that, as $n$ tends to infinity,

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

tends to 0 almost surely. □

*Proof of Theorem 4.* Note that

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

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

$$\begin{aligned}
\frac{\Delta_{k_\rho,a}(a\mathbf{x})}{\alpha(n/k_\rho)} - a\frac{\Delta_{k_\rho,a}(\mathbf{x})}{\alpha(n/k_\rho)} &= a(a^{-\rho}-1)^2M(\mathbf{x}) \\
&+ \frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\left\{a^{-1}Z_L(a^2\mathbf{x}) - 2Z_L(a\mathbf{x}) + aZ_L(\mathbf{x})\right\} + o\left(\frac{1}{\sqrt{k_\rho}\alpha(n/k_\rho)}\right) \,,
\end{aligned} \tag{30}$$

and

$$\frac{\Delta_{k_\rho,a}(\mathbf{x})\Delta_{k,a}(\mathbf{x})}{\alpha(n/k_\rho)} = \alpha(n/k)a(a^{-\rho}-1)^2M^2(\mathbf{x}) + (a^{-\rho}-1)M(\mathbf{x})\frac{a^{-1}Z_L(a\mathbf{x}) - Z_L(\mathbf{x})}{\sqrt{k}} + O_{\mathbb{P}}\left(\frac{\alpha(n/k)}{\sqrt{k_\rho}\alpha(n/k_\rho)} + \frac{1}{\sqrt{k}\sqrt{k_\rho}\alpha(n/k_\rho)}\right) + o\left(\frac{1}{\sqrt{k}}\right) . \tag{31}$$

on the other hand, both uniformly for $\mathbf{x} \in [0,T]^d$. Combining (30) and (31) with equation (8), one gets

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

Since the last expression and equation (30) 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 $k_\rho$ 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

$$\sup_{0\le x_1,\dots,x_d\le T}\left|\frac{\Delta_{k,a}(\mathbf{x})}{\alpha\left(\frac{n}{k}\right)} - (a^{-\rho}-1)M(\mathbf{x})\right| \xrightarrow{\mathbb{P}} 0 . \tag{32}$$

As a consequence,

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

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

$$\begin{aligned}
\left|\frac{\Delta_{k,a}(r\mathbf{x})}{\alpha(n/k)} - r^{1-\rho}\frac{\Delta_{k,a}(\mathbf{x})}{\alpha(n/k)}\right| &\le \left|\frac{\Delta_{k,a}(r\mathbf{x})}{\alpha(n/k)} - r^{1-\rho}(a^{-\rho}-1)M(\mathbf{x})\right| \\
&\quad + \left|r^{1-\rho}(a^{-\rho}-1)M(\mathbf{x}) - r^{1-\rho}\frac{\Delta_{k,a}(\mathbf{x})}{\alpha(n/k)}\right|
\end{aligned}$$

and using twice equation (32) leads to the conclusion.

□

*Proof of Proposition 6.* Let denote by $Q_{k,a,r}$ the quotient

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

Lemma 7 used twice allows to write

$$\sqrt{k}\alpha(\frac{n}{k})(Q_{k,a,r}(\mathbf{x}) - r^{1-\rho}) \xrightarrow{d} -r^{1-\rho}\log r\hat{Z}_{\rho,a,r}(\mathbf{x}) \ , \tag{33}$$

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 (33) and the Delta method. The last two convergences are obtained by similar arguments. □


## Acknowledgements

This research is 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).


# References


B. Abdous and K. Ghoudi. Non-parametric estimators of multivariate extreme dependence functions. *J. Nonparametr. Stat.*, 17(8):915–935, 2005.

J. Beirlant, Y. Goegebeur, J. L. Teugels, and J. Segers. *Statistics of extremes.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004.

J. Beirlant, G. Dierckx, and A. Guillou. Bias-reduced estimators for bivariate tail modelling. *Insurance Math. Econom.*, 49(1):18–26, 2011. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2011.01.010. URL http://dx.doi.org/10.1016/j.insmatheco.2011.01.010.

A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist.*, 39(4):1963–2006, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS890. URL http://dx.doi.org/10.1214/11-AOS890.

J. J. Cai, L. de Haan, and C. Zhou. Bias reduction in extreme value statistics with index around zero. *Extremes*, 16:173–201, 2013.

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

P. Capéraà, A.-L. Fougères, and C. Genest. Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.*, 72(1):30–49, 2000.

G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13:55–87, 2010.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B*, 53(2):377–392, 1991.

L. de Haan and A. Ferreira. *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. New York, NY: Springer., 2006.

L. de Haan and S. I. Resnick. Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*, 9(2):275–309, 1993.

L. de Haan, C. Mercadier, and C. Zhou. Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Submitted*, 2013.

S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1): 111–129, 2005.

J. H. J. Einmahl, L. de Haan, and A. K. Sinha. Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.*, 70(2):143–171, 1997.

J. H. J. Einmahl, A. Krajina, and J. Segers. A method of moments estimator of tail dependence. *Bernoulli*, 14(4):1003–1026, 2008. ISSN 1350-7265. doi: 10.3150/08-BEJ130. URL http://dx.doi.org/10.3150/08-BEJ130.

J. H. J. Einmahl, A. Krajina, and J. Segers. An $M$-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.*, 40(3):1764–1793, 2012. ISSN 0090-5364. doi: 10.1214/12-AOS1023. URL http://dx.doi.org/10.1214/12-AOS1023.

I. Gomes, L. de Haan, and L. Rodrigues. 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(1):31–52, 2008.

M. I. Gomes and M. J. Martins. "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5(1):5–31, 2002. ISSN 1386-1999. doi: 10.1023/A:1020925908039. URL http://dx.doi.org/10.1023/A:1020925908039.

M. I. Gomes, L. de Haan, and L. Peng. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5(4):387–414 (2003), 2002.

S. Guillotte, F. Perron, and J. Segers. Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(3):377–406, 2011. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2010.00770.x. URL http://dx.doi.org/10.1111/j.1467-9868.2010.00770.x.

P. Hall and A. H. Welsh. Adaptive estimates of parameters of regular variation. *Ann. Statist.*, 13(1):331–341, 1985. ISSN 0090-5364. doi: 10.1214/aos/1176346596. URL http://dx.doi.org/10.1214/aos/1176346596.

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

H. Joe. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.*, 9(1):75–81, 1990. ISSN 0167-7152. doi: 10.1016/0167-7152(90)90098-R. URL http://dx.doi.org/10.1016/0167-7152(90)90098-R.

H. Joe, R. L. Smith, and I. Weissman. Bivariate threshold methods for extremes. *J. Roy. Statist. Soc. Ser. B*, 54(1):171–183, 1992.

S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous multivariate distributions. Vol. 1.* Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, second edition, 2000. ISBN 0-471-18387-3. doi: 10.1002/0471722065. URL http://dx.doi.org/10.1002/0471722065. Models and applications.

A. Krajina. A method of moments estimator of tail dependence in meta-elliptical models. *J. Statist. Plann. Inference*, 142(7):1811–1823, 2012. ISSN 0378-3758. doi: 10.1016/j.jspi.2012.01.020. URL http://dx.doi.org/10.1016/j.jspi.2012.01.020.

L. Peng. Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.*, 38(2):107–115, 1998.

S. I. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18 (1):66–138, 1986.

J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415, 1988.