Nonparametric Statistics
Vol. 17, No. 8, December 2005, 915–935

Non-parametric estimators of multivariate extreme dependence functions

BELKACEM ABDOUS† and KILANI GHOUDI*‡

†Département de médecine sociale et préventive, Université Laval, Ste-Foy, QC, Canada G1K 7P4
‡Department of Statistics, College of Business and Economics, United Arab Emirates University, P.O. Box 17555, AL Ain, United Arab Emirates

(Received January 2004; revised in final form August 2005)

This article reviews various characterizations of a multivariate extreme dependence function A(·). The most important estimators derived from these characterizations are also sketched. Then, a unifying approach, which puts all these estimators under the same framework, is presented. This unifying approach enables us to construct new estimators and, most importantly, to propose an automatic selection method for an unknown parameter on which all the existing non-parametric estimators of A(·) depend. Constrained smoothing splines and convex hull techniques are used to force the obtained estimators to be extreme dependence functions. A simulation study comparing these estimators on a wide range of extreme dependence functions is provided.

Keywords: Extreme dependence functions; Kernel estimator; Constrained spline; Optimal bandwidth selection

AMS Subject Classification: Primary: 62G32, 62G05

1. Introduction

Modeling of extremes has received considerable attention in the literature. Univariate extremes have a natural definition and much has been written on their finite and asymptotic behavior. For background and a thorough treatment, see ref. [1]. Unlike the one-dimensional case, there are several concepts of multivariate ordering (see, *e.g.* ref. [2]). The most popular is the marginal ordering (or M-ordering) whereby a multivariate extreme is obtained by taking componentwise maxima. Numerous results on modeling the class of these multivariate extremes have been obtained; for a good account on the subject, we refer to refs. [3–5]. In what follows, we will restrict our presentation to the two-dimensional case only. Higher dimensions can be handled in almost a similar manner.

Assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is an independent and identically distributed random sample drawn from an unknown distribution function H with marginal distributions H_1 and

*Corresponding author. Email: kghoudi@uaeu.ac.ae

Nonparametric Statistics
ISSN 1048-5252 print/ISSN 1029-0311 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10485250500336379

H_2. Define the vector of componentwise maxima by

$$(X_n^*, Y_n^*) = (\max(X_1, \ldots, X_n), \max(Y_1, \ldots, Y_n)).$$

Then, the distribution function H is said to belong to the max-domain of attraction of a bivariate extreme value distribution H^* with non-degenerate marginal distributions, if there exist sequences $a_{1n} > 0$, $a_{2n} > 0$, b_{1n} and b_{2n} such that

$$\lim_{n \longrightarrow \infty} P\left[\frac{X_n^* - b_{1n}}{a_{1n}} \leq x, \frac{Y_n^* - b_{2n}}{a_{2n}} \leq y\right]$$
$$= \lim_{n \longrightarrow \infty} H^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) = H^*(x, y).$$

The margins H_1^* and H_2^* of H^* must be univariate generalized extreme value distributions. In this case, the margins H_1 and H_2 are said to belong to the max-domain of attraction of H_1^* and H_2^*, respectively. The class of generalized extreme value distributions constitutes a finite parametric family. For details, see refs. [1, 3, 4]. Unfortunately, there is no finite parametric model for H^*. A common approach to investigate and model the extreme, value distribution H^* is to analyze the margins H_1^* and H_2^* together with the copula (or dependence) function C^*. The copula C^* captures the dependence structure of the model and relates H^* to its margins by means of the relationship

$$H^*(x, y) = C^*(H_1^*(x), H_2^*(y)).$$

From ref. [3], H belongs to the max-domain of attraction of H^* if and only if H_1 and H_2 belong to the max-domain of attraction of H_1^* and H_2^*, respectively, and C, the copula associated with H, satisfies

$$C^*(u, v) = \lim_{n \longrightarrow \infty} C(u^{1/m}, v^{1/m})^m, \quad \text{for all } 0 \leq u, \quad v \leq 1. \tag{1}$$

In what follows, we will mainly focus on the estimation of the dependence structure. The reader can refer to refs. [1, 5] for a good background and details on the estimation of univariate extreme distributions. Thus, hereafter, marginal distributions H_1 and H_2 are of secondary interest and they can be arbitrarily fixed to be in the generalized extreme value distribution family or transformed to any other arbitrary univariate distribution. It will be seen later in the estimation section that this can be achieved without assuming that the marginals are known. To review and define the existing estimators of the copula C^*, we need to recall various characterizations and relationships which link H, C, H^* and C^*. Indeed, the most popular representation of C^* is due to Pickands [6].

$$C^*(u, v) = \exp\left[\log(uv) A\left(\frac{\log(u)}{\log(uv)}\right)\right], \tag{2}$$

where $0 \leq u, v \leq 1$ and $A(\cdot)$ is a convex function on $[0, 1]$ satisfying $\max(t, 1 - t) \leq A(t) \leq 1$. The function A is commonly called the dependence function associated with C^*. There exist several characterizations of $A(\cdot)$. For instance, upon transforming the random vector, under investigation (X, Y) to a r.v. $\mathbf{Z} = (Z_1, Z_2)$ with unit Fréchet margins, *i.e.*, $Pr(Z_i \leq z) = \exp(-1/z)$ for all $z > 0$, one can show that

$$A(t) = \int_{\mathcal{B}} \max\{tw_1, (1 - t)w_2\} S(\mathrm{d}w), \tag{3}$$

where S is a finite measure on the set $\mathcal{B} = \{w \in \mathbb{R}_+^2 : \| w \| = 1\}$ with $\| \cdot \|$ being an arbitrary norm on $\mathbb{R}^2$. For a thorough treatment of this characterization, we refer to ref. [4, Chapter 5].

If we denote by $\triangle$ an arbitrary Borel set on $\mathcal{B}$ and we use the following polar transformation

$$R = \|\mathbf{Z}\| \quad \text{and} \quad \mathbf{W} = (W_1, W_2) = \frac{\mathbf{Z}}{\|\mathbf{Z}\|}, \tag{4}$$

then, it follows that the measure S satisfies

$$\lim_{s \longrightarrow \infty} sP\{\mathbf{W} \in \triangle, R > s\} = S(\triangle),$$

and

$$\int_{\mathcal{B}} w_1 S\,(\mathrm{d}w) = \int_{\mathcal{B}} w_2 S\,(\mathrm{d}w) = 1. \tag{5}$$

The previous expressions take simpler forms upon adopting the L_p-norm $\|(x, y)\|_p = (|x|^p + |y|^p)^{1/p}$ with $p \in [1, \infty]$. Indeed, for $p \neq \infty$, standard algebra yields that (3) is equivalent to

$$A(t) = \int_0^1 \max\{tw, (1-t)(1-w^p)^{1/p}\} S\,(\mathrm{d}w), \tag{6}$$

where

$$S(w) = S([0, w]) = \lim_{s \longrightarrow \infty} sP\left\{\frac{Z_1}{\|\mathbf{Z}\|_p} \leq w, \|\mathbf{Z}\|_p > s\right\}.$$

Also, equation (5) reduces to

$$\int_0^1 wS\,(\mathrm{d}w) = \int_0^1 (1-w^p)^{1/p} S\,(\mathrm{d}w) = 1.$$

Whereas, for $p = \infty$, one has

$$A(t) = \int_0^1 \max\{tw, 1-t\} S_1\,(\mathrm{d}w) + \int_0^1 \max\{t, (1-t)w\} S_2\,(\mathrm{d}w), \tag{7}$$

where

$$S_1(w) = S_1([0, w]) = \lim_{s \longrightarrow \infty} sP\left\{\frac{Z_1}{Z_2} \leq w; Z_1 \leq Z_2, Z_2 > s\right\},$$

and

$$S_2(w) = S_2([0, w]) = \lim_{s \longrightarrow \infty} sP\left\{\frac{Z_2}{Z_1} \leq w; Z_1 \leq Z_2; Z_1 > s\right\}.$$

In addition, (5) is equivalent to

$$\int_0^1 wS_1\,(\mathrm{d}w) + \int_0^1 S_2\,(\mathrm{d}w) = \int_0^1 S_1\,(\mathrm{d}w) + \int_0^1 wS_2\,(\mathrm{d}w) = 1.$$

Alternative characterizations of the function $A(\cdot)$ have been discussed in the literature and are outlined next. Joe *et al.* [7] showed that

$$A(t) = G(t), \quad \text{with } G(t) = \lim_{s \longrightarrow \infty} sP\{T_t > s\} \tag{8}$$

where $T_t = \max\{tZ_1, (1-t)Z_2\}$ for all $t \in [0, 1]$.

Using an approach based on copula functions, Abdous *et al.* [8] showed that for any $t \in [0, 1]$

$$A(t) = \lim_{u \longrightarrow 0} \frac{1 - C[(1-u)^t, (1-u)^{1-t}]}{u}. \tag{9}$$

Upon letting (U, V) denote a random vector with distribution function C, one sees that the previous relation is equivalent to

$$A(t) = \lim_{u \longrightarrow 0} \frac{G_t(u)}{u}, \quad \forall t \in [0, 1], \tag{10}$$

where $G_t(\cdot)$ is the cumulative distribution function of the following random variable

$$W_t = \begin{cases} 1 - \max\{U^{1/t}, V^{1/(1-t)}\} & \text{for } t \in (0, 1), \\ 1 - V & \text{for } t = 0, \\ 1 - U & \text{for } t = 1. \end{cases}$$

At this stage, let us point out that all the previous characterizations link the dependence function $A(\cdot)$ to the distribution H solely. The next two characterizations link $A(\cdot)$ to the limit distribution H^* itself. Let (Z_1^*, Z_2^*) stand for an extreme random vector with dependence function A and unit Fréchet margins. Then, according to Pickands [6],

$$P\{Z^*(t) \geq z\} = \exp\{-zA(t)\}, \tag{11}$$

where $Z^*(t) = \min[\{tZ_1^*\}^{-1}, \{(1-t)Z_2^*\}^{-1}]$. Tiago de Oliveira [9] observed that

$$\begin{aligned} A(t) &= \exp\left\{\int_0^t \frac{L(x) - x}{x(1-x)}\,\mathrm{d}x\right\} \\ &= \exp\left\{-\int_t^1 \frac{L(x) - x}{x(1-x)}\,\mathrm{d}x\right\}, \end{aligned} \tag{12}$$

where $L(\cdot)$ is the distribution function of the random variable $Z_2^*/(Z_1^* + Z_2^*)$.

Heretofore, we summarized the essential results concerning the various characterizations of $A(\cdot)$. In section 2, we review the most important estimators of $A(\cdot)$ derived from the previous characterizations. Section 3 provides a unified presentation of a large class of non-parametric estimators, and at the same time, introduces some new non-parametric estimators. Fine tuning techniques such as optimal bandwidth selection and procedures to ensure that the estimate is an extreme dependence function are discussed in sections 4 and 5, respectively. A comparative study by means of Monte-Carlo simulations is provided in section 6. Finally, in section 7, we apply the previous estimators to two sets of real data.

2. Review of the existing estimators of $A(\cdot)$

As usual, to estimate the function A$(\cdot)$, one has the option of using parametric or non-parametric methods. The parametric approach will not be tackled in this work [see refs. [10–12] for details and references]. The non-parametric approaches can be classified into two categories. The first category makes the assumption that the sample at hand has an extreme value distribution (*i.e.* one disposes of a sample of componentwise maxima). The second category assumes that the sample is drawn from a distribution function belonging to the domain of attraction of the extreme value distribution under investigation. Several non-parametric estimators of $A(\cdot)$ have been proposed in the literature; hereafter, we give a succinct review of the most popular of them.

2.1 *Estimators based on non-extreme bivariate sample*

In the sequel, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote a sample drawn from a distribution function H which belongs to the domain of attraction of the extreme value distribution H^*. As mentioned earlier, the focus being on the estimation of the dependence structure, the margins are treated as nuisance parameters. For instance, one produces Fréchet marginal by letting $\mathbf{Z}_i = (Z_{1i}, Z_{2i})$ where $Z_{1i} = -1/\log\{H_1(X_i)\}$ and $Z_{2i} = -1/\log\{H_2(Y_i)\}$ for $i = 1, \ldots, n$. If H_1 and H_2 are unknown, one uses the following transformation $Z_{1i} = -1/\log\{n/(n+1)H_{1n}(X_i)\}$ and $Z_{2i} = -1/\log\{n/(n+1)H_{2n}(Y_i)\}$ where H_{1n} is empirical distribution of the Xs and H_{2n} empirical distribution of the Ys. The coefficient $n/(n+1)$ is added to avoid division by zero. This transformation is shown to be valid for the estimation of extreme dependence function by Huang [13] and Abdous *et al.* [8].

Clearly, characterizations (6) and (7) show that the estimation of $A(\cdot)$ is achieved via that of the spectral measures S, S_1 and S_2. Empirical estimators of these measures can be defined as follows

$$S_n(w) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{Z_{1i}}{\|\mathbf{Z}_i\|_p} \le w;\ \|\mathbf{Z}_i\|_p > \frac{n}{k_n}\right\},$$

$$S_{1n}(w) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{Z_{1i}}{Z_{2i}} \le w;\ Z_{1i} \le Z_{2i};\ Z_{2i} > \frac{n}{k_n}\right\},$$

$$S_{2n}(w) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{Z_{2i}}{Z_{1i}} \le w;\ Z_{1i} \le Z_{2i};\ Z_{1i} > \frac{n}{k_n}\right\},$$

where k_n is a sequence of positive real numbers satisfying

$$\lim_{n\longrightarrow\infty} k_n = \infty \quad \text{and} \quad \lim_{n\longrightarrow\infty} \frac{k_n}{n} = 0.$$

Substituting S_n, S_{1n} or S_{2n} in the definition of A yields the following estimators

$$\hat{A}_p(t) = \frac{1}{k_n}\sum_{i=1}^{n} \max\left\{t\frac{Z_{1i}}{\|\mathbf{Z}_i\|_p}, (1-t)\frac{Z_{2i}}{\|\mathbf{Z}_i\|_p}\right\} \mathbb{1}\left\{\|\mathbf{Z}_i\|_p > \frac{n}{k_n}\right\} \tag{13}$$

for $1 \le p \le \infty$. Up to some minor adjustments, these estimators correspond to those proposed by Einmahl *et al.* [14] for $p = 2$, by Einmahl *et al.* [15, 16] for $p = \infty$ and by Capéraà and Fougères [17] for $p = 1$. Quite in the same spirit, Joe *et al.* [7] used equation (8) to obtain the following estimator

$$\hat{A}_J(t) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{T_{t,i} > \frac{n}{k_n}\right\} \tag{14}$$

where $T_{t,i} = \max\{Z_{1i}, (1-t)Z_{2i}/t\}$ for $i = 1, \ldots, n$. To close the review of this kind of estimator, we mention that Abdous *et al.* [8] exploited (10) to construct a kernel-based estimator of $A(\cdot)$. For the moment, we postpone the precise expression of that estimator and content ourselves with the following naive version

$$\hat{A}_A(t) = \frac{1}{k_n} \sum_{i=1}^{n} \mathbb{1}\left\{W_{t,i} \le \frac{k_n}{n}\right\} \tag{15}$$

where

$$W_{t,i} = \begin{cases} 1 - \max\{\mathrm{e}^{-(1/tZ_{1i})}, \mathrm{e}^{-[1/(1-t)Z_{2i}]}\}, & \text{for } t \in (0,1), \\ 1 - \mathrm{e}^{-(1/Z_{2i})} & \text{for } t = 0, \\ 1 - \mathrm{e}^{-(1/Z_{1i})} & \text{for } t = 1. \end{cases}$$

Before giving a unified treatment and formulation of all the estimators presented so far, we review estimators based on a sample drawn from the extreme value distribution itself.

2.2 *Estimators based on extreme bivariate sample*

Let $(X^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$ denote a sample drawn from the extreme distribution H^*. To get exactly or approximately Fréchet margins, we need to set for $j = 1, 2$ and $i = 1, \ldots, n$

$$Z_{ji} = \begin{cases} -\dfrac{1}{\ln(H_j^*(X_i^*))} & \text{for known marginals } H_1^* \text{ and } H_2^* \\ -\dfrac{1}{\ln(n/(n+1)H_{jn}^*(X_i^*))}, & \text{for unknown marginals} \end{cases}$$

where H_{1n}^* and H_{2n}^* are the empirical counterparts of H_1^* and H_2^*, respectively.

A straightforward consequence of characterization (11) is that $A(t)$ can be viewed as a location parameter of an exponential distribution. Pickands [6] exploited this fact and used maximum likelihood to estimate $A(t)$. Then, several authors modified and adapted the proposal of Pickands [6]), [see, *e.g.* refs. 9, 11, 18, 19]. In the following simulation study, we will retain the estimator proposed by Hall and Tajvidi [19], namely

$$\hat{A}_{HT}(t) = n\left\{\sum_{i=1}^{n} \min\left[\frac{\tilde{Z}_{1i}}{t}, \frac{\tilde{Z}_{2i}}{1-t}\right]\right\}^{-1},$$

where

$$\tilde{Z}_{ji} = \frac{Z_{ji}^{-1}}{n^{-1}\sum_{k=1}^{n} Z_{jk}^{-1}} \qquad \text{for } j = 1, 2 \quad \text{and } i = 1, \ldots, n.$$

Finally, Capéraa *et al.* [20] used (12) and proposed the following estimator

$$\tilde{A}_c(t) = \exp[p(t)\log\{A_n^0(t)\} + \{1 - p(t)\}\log\{A_n^1(t)\}],$$

where $p(t)$ is an appropriate weight function, L_n the empirical distribution function of $\{Z_{2i}/(Z_{1i} + Z_{2i})\}_{i=1}^n$ and

$$A_n^0(t) = \exp\left\{\int_0^t \frac{L_n(u) - u}{u(1-u)}\,\mathrm{d}u\right\} \quad \text{and} \quad A_n^1(t) = \exp\left\{-\int_t^1 \frac{L_n(u) - u}{u(1-u)}\,\mathrm{d}u\right\},$$

A discussion about the choice of the weight function $p(\cdot)$ can be found in ref. [20]. A modification of the above proposal, which adjusts for boundary problems, can be found in ref. [21].

3. A unifying approach

In this section, we provide a unifying approach that allows us to introduce new estimators in addition to giving a common presentation of all the estimators presented in section 2.1. Moreover, recall that these estimators depend on an unknown parameter k_n which is directly related to the threshold level. In practice, such a parameter plays an important role in the performance of the proposed estimators and it has to be selected in an appropriate manner. Some *ad hoc* selection methods have been proposed in the references mentioned previously. But, hereafter, we will provide an automatic selection technique for the parameter k_n.

First, notice that each of the characterizations (6), (7), (8) and (10) involves one or two unknown functionals. All these functionals can be cast in the following form:

$$\Lambda(w) = \lim_{u \longrightarrow 0^+} \frac{F(w, u)}{u}, \quad \forall w \in [0, 1],$$

where $F(\cdot, \cdot)$ is a distribution function satisfying $F(w, 0) = 0$. Indeed, for the characterization (6), one has

$$\Lambda(w) = S(w) \quad \text{and} \quad F(w, u) = P\left\{\frac{Z_1}{\|\mathbf{Z}\|_p} \leq w, \frac{1}{\|\mathbf{Z}\|_p} < u\right\},$$

for equation (7), one has two functionals

$$\Lambda(w) = S_1(w) \quad \text{and} \quad F(w, u) = P\left\{\frac{Z_1}{Z_2} \leq w;\ Z_1 \leq Z_2;\ \frac{1}{Z_2} < u\right\},$$

and

$$\Lambda(w) = S_2(w) \quad \text{and} \quad F(w, u) = P\left\{\frac{Z_2}{Z_1} \leq w;\ Z_1 > Z_2;\ \frac{1}{Z_1} < u\right\},$$

whereas for (8) and (10), one has $\Lambda(w) = A(w)$ for both cases and

$$F(w, u) = \begin{cases} P\left\{\dfrac{1}{T_w} < u\right\} & \text{for (8)} \\ P\{W_w < u\} & \text{for (10)} \end{cases}$$

Consequently, all the methods presented in section 2.1 are essentially estimating the derivative of a distribution function at zero and can be unified under this framework.

This functional estimation problem can be tackled in several ways. We will adopt the weighted polynomial fitting approach subsequently [see, *e.g.*, ref. 22, for details]. In this way, we will be able to estimate the unknown functional in addition to constructing an automatic selection technique for the smoothing parameter. Also, we can either fix w in $F(w, u)$ and try to estimate the derivative $f_w(u) = (\partial F(w, u))/(\partial u)$ by fitting a local univariate polynomial to the empirical counterpart of $f_w(u)$ or we can view the problem as a two dimensional one and estimate $F(w, u)$ and its partial derivatives by fitting a bivariate polynomial to the empirical counterpart of $F(w, u)$. For simplicity, we take the first option: the second approach has been investigated by Abdous and Bensaid [23] in the context of multivariate probability distribution functions.

First, for any fixed $w \in [0, 1]$ and any $u \geq 0$, let $F_n(w, u)$ stand for the empirical version of $F(w, u)$, *i.e.*,

$$F_n(w,u) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{1}\left\{\dfrac{Z_{1i}}{\| \mathbf{Z}_i \|_p} \leq w\right\} \mathbb{1}\{\| \mathbf{Z}_i \|_p^{-1} < u\} & \text{for } S \text{ in (6)} \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{1}\left\{\dfrac{Z_{1i}}{Z_{2i}} \leq w;\ Z_{1i} \leq Z_{2i}\right\} \mathbb{1}\{Z_{2i}^{-1} < u\} & \text{for } S_1 \text{ in (7)} \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{1}\left\{\dfrac{Z_{2i}}{Z_{1i}} \leq w;\ Z_{1i} > Z_{2i}\right\} \mathbb{1}\{Z_{1i}^{-1} < u\} & \text{for } S_2 \text{ in (7)} \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{1}\{T_{w,i}^{-1} < u\} & \text{for (8)} \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{1}\{W_{w,i} < u\} & \text{for (10).} \end{cases}$$

Then, the local polynomial fitting technique consists in using a weighted least squares criterion to fit a polynomial to the empirical estimator $F_n(w, u)$ for u in the vicinity of zero. To this end, fix w in [0,1], take $m \geq 0$ to be an arbitrary integer and seek for a univariate polynomial $P_m(x) = \sum_{j=1}^{m} \alpha_j x^j$ whose coefficients $\alpha_1, \ldots, \alpha_m$ minimize the following criterion

$$\int_0^b \frac{1}{h} K\left(\frac{v}{h}\right) \left\{ F_n(w, v) - \sum_{j=1}^{m} \alpha_j v^j \right\}^2 \mathrm{d}v, \tag{16}$$

where $b = \infty$ for characterizations (6) to (8) and $b = 1$ for (10). The smoothing parameter h controls the size of the neighborhood of 0, the kernel K acts as an arbitrary weight function. Note that the constant term in the fitted polynomial is omitted as $F(w, 0) = 0$. In addition, the integral is restricted to the support of $F(w, \cdot)$ to avoid the well-known boundary effects problem. More details on this fact will be given shortly.

Now, assume that the kernel K is a probability density having as many finite moments as needed. Let $v_j(h) = \int_0^b v^j K_h(v)\,\mathrm{d}v$, where $K_h(v) = K(v/h)/h$ and define S_m to be the $m \times m$ non-singular matrix with entries $S_m(i, j) = v_{i+j}(h)$ for $1 \leq i, j \leq m$. Then (16) is minimized at

$$\hat{\alpha}_j(w) = \mathrm{e}_j^{\mathrm{T}} S_m^{-1} \mu_n(w), \quad j = 1, \ldots, m, \tag{17}$$

where μ_n is the vector with components given by $\mu_{ni} = \int_0^b v^i F_n(w, v) K_h(v)\,\mathrm{d}v$, for $i = 1, \ldots, m$ and where $\mathrm{e}_j = [0, 0, \ldots, 0, 1, 0, \ldots, 0]^{\mathrm{T}}$, with 1 being at the jth position. The previous expression can be also viewed as a kernel estimator by writing

$$\hat{\alpha}_j(w) = \int_0^b K_j^{[m]}(v, h) F_n(w, v)\,\mathrm{d}v, \quad j = 1, \ldots, m, \tag{18}$$

where $K_j^{[m]}(v, h) = \mathrm{e}_j^{\mathrm{T}} S_m^{-1} [v, \ldots, v^m]^{\mathrm{T}} K_h(v)$. The kernel $K_j^{[m]}$ is called an equivalent kernel. It has the feature of reducing the bias of $\hat{\alpha}_j(\cdot)$ as m increases. Indeed, under mild regularity conditions, one can show that the bias of $\hat{\alpha}_j(w)$ is of order h^{m+1}. This stems from the fact that

$$\int_0^b v^i K_j^{[m]}(v,h)\,\mathrm{d}v = \begin{cases} 0 & 1 \le i \le m \text{ and } i \ne j, \\ 1 & i = j, \\ C \ne 0 & i = m+1. \end{cases}$$

In the sequel, any kernel satisfying these moments constraints will be a higher order kernel with order $(j, m+1)$. Note that this important property is not valid if one does not restrict the integral in (16) to the support of $F(w,\cdot)$ in which case the obtained equivalent kernels will not adapt to boundary regions and will yield biased estimators.

The estimator given by (18) inherits all the well-established asymptotic properties of kernel smoothers. Such aspects will not be investigated here. However, given the practical importance of the selection of the bandwidth parameter h, the next section will be devoted to the presentation of an automatic bandwidth selection procedure. In fact, as we are mainly interested in the estimation of $\Lambda(w) = \partial F(w,u)/\partial u|_{u=0}$, we will only consider

$$\Lambda_n(w) = \hat{\alpha}_1(w) = \int_0^b K^{[m]}(v,h) F_n(w,v)\,\mathrm{d}v, \tag{19}$$

where $K^{[m]}$ stands for $K_1^{[m]}$.

Now, upon replacing $\Lambda(\cdot)$ by its estimator $\Lambda_n(\cdot)$ in characterizations (6) to (10), one gets the following estimators for $A(t)$

$$\hat{A}_p^{[m]}(t) = \frac{1}{n}\sum_{i=1}^n \frac{\max\{tZ_{1i},(1-t)Z_{2i}\}}{\|\mathbf{Z}_i\|_p} \int_{\|\mathbf{Z}_i\|_p^{-1}}^{\infty} K^{[m]}(v,h)\,\mathrm{d}v,$$

with $p \in [1,\infty)$ for equation (6) and $p = \infty$ for equation (7)

$$\hat{A}_J^{[m]}(t) = \frac{1}{n}\sum_{i=1}^n \int_{T_{t,i}^{-1}}^{1} K^{[m]}(v,h)\,\mathrm{d}v, \quad \text{for equation (8)}$$

$$\hat{A}_A^{[m]}(t) = \frac{1}{n}\sum_{i=1}^n \int_{W_{t,i}}^{1} K^{[m]}(v,h)\,\mathrm{d}v, \quad \text{for equation (10)}$$

Remark 3.1 Note that upon choosing $m = 1$, $h = k_n/n$ and K the Dirac measure at 1, all the estimators of section 3.1 can be seen as special cases of these local polynomial fitting estimators.

4. Optimal bandwidth selection

Many classical bandwidth selection methods such as 'rules of thumb', 'solve-the-equation', 'plug-in', 'double kernel' and 'smoothed bootstrap' could be adapted to the present situation [for a review, we refer to refs. 24, 25]. All these techniques make use of an error criterion to select the optimal bandwidth. Owing to its tractability and its simple decomposition into variance and bias terms, the mean squared error (MSE) is by far the most popular criterion. In the classical kernel estimation settings, minimizing the asymptotic MSE produces a closed expression for the optimal bandwidth h. Unfortunately, due to the complex structure of the equivalent kernel $K_j^{[m]}$ and its dependence on h, we failed to find a satisfactory and closed form of the optimal bandwidth associated to the estimator (19). However, it is always possible to use the reference density together with the solve-the-equation techniques to approximate the optimal bandwidth. We will not carry on with this idea; we will rather propose a fully automatic technique based ‘on the double kernel approach. The L_1-double kernel method is a fully automatic technique which has proved to be powerful in density estimation context. It has been proposed first by Devroye [26]. Then, some improvements and modifications have been proposed by Berlinet and Devroye [27], Devroye and Lugosi [28] and Devroye and Lugosi [29]. The L_2 version of this technique has been investigated by Jones [30] and Abdous [31].

For kernel density estimation, the L_1-double kernel method seeks for the bandwidth which minimizes the criterion $\int |\hat{f}_n(x) - \hat{g}_n(x)|\,dx$, where $\hat{f}_n(\cdot)$ and $\hat{g}_n(\cdot)$ are two kernel estimates of the density f. Among other standard regularity assumptions, the two kernels must have different orders. This way, the previous criterion can be seen as a rough estimate of the L_1 error associated with either $\hat{f}_n(\cdot)$ or $\hat{g}_n(\cdot)$. To adapt this idea to the present problem, let $\hat{A}^{[m]}(\cdot)$ stand for either $\hat{A}_p^{[m]}(\cdot)$ or $\hat{A}_J^{[m]}(\cdot)$ or $\hat{A}_p^{[m]}(\cdot)$. Then, as mentioned earlier, the bias of $\hat{A}^{[m]}(\cdot)$ is of order h^{m+1}. Consequently, if one takes l to be an arbitrary integer satisfying $l > m$ and denote by $K^{[l]}$ the associated kernel which is computed in the same way as $K^{[m]}$, then as the kernel $K^{[l]}$ is of order $l + 1 > m + 1$, the bias of the estimator $\hat{A}^{[l]}$ should be negligible in comparison to that of $\hat{A}^{[m]}$. Hence, a rough estimate of $\int_0^1 |\hat{A}^{[m]}(t) - A(t)|\mathrm{d}t$ could be

$$\mathrm{DK}(h) = \int_0^1 \left|\hat{A}^{[m]}(t) - \hat{A}^{[l]}(t)\right| \,\mathrm{d}t. \tag{20}$$

Minimizing (20) with respect to h will yield the double kernel bandwidth h_{Dk} which is an approximation of the optimal bandwidth h_m^*:

$$h_m^* = \mathrm{argmin}_h \int_0^1 \left|\hat{A}^{[m]}(t) - A(t)\right| \,\mathrm{d}t.$$

In density estimation context, the double kernel technique provides L_1 universally consistent estimators, its associated bandwidth is asymptotically optimal in L_1 sense and it performs very well in practice. As highlighted in the subsequent comparative study, the previous, adaptation of the L_1 double kernel to extreme dependence functions also performs extremely well.

Remark 4.1 With regard to the estimators of section 3.1, the double kernel technique could be adapted to automatically select the parameter k_n. Indeed, if we set $m = 1, h = k_n/n$ and $K(x) = \delta_1(x)$ in the expression of $\hat{A}^{[m]}(t)$, then we find again estimators (13), (14) and (15) and obtain a kernel of order (1, 1). While, if we take $l = 1, h = k_n/n$ and $K(x) = 2\delta_1(x) - \delta_2(x)$ in the expression of $\hat{A}^{[l]}(t)$ then we get a kernel of order (1, 2) together with the following analogues of (13), (14) and (15):

$$\hat{A}_p(t) = \frac{1}{k_n}\sum_{i=1}^{n} \frac{\max\{Z_{1i}, (1-t)Z_{2i}\}}{\|\mathbf{Z}_i\|_p} \mathbb{1}\left\{\frac{n}{2k_n} < \|\mathbf{Z}_i\|_p \le \frac{n}{k_n}\right\}$$

$$\hat{A}_J(t) = \frac{t}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{n}{2k_n} < T_{t,i} \le \frac{n}{k_n}\right\}$$

$$\hat{A}_A(t) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{k_n}{n} < W_{t,i} \le \frac{2k_n}{n}\right\}.$$

Then, it suffices to plug these estimates in (20) and minimize the obtained expression with respect to k_n.

Note that the estimates $\tilde{A}_p(\cdot)$ and $\tilde{A}_J(\cdot)$ exclude the largest values of $\| \mathbf{Z}_i \|_p$s and $T_{t,i}$s, respectively, whereas the estimate $\hat{A}_A(\cdot)$ excludes the smallest $W_{t,i}$s. Incidently, these estimators resemble some proposals of Joe *et al.* [7] and Capéraà and Fougères [17].

We close this section by stressing the fact that none of the estimators proposed so far satisfies the properties of a dependence function. The next section deals with some solutions to this drawback.

5. Constraining the estimator to be an extreme dependence function

Though the estimators obtained in the previous sections are consistent, they are not necessarily dependence functions. Each of them is either not convex or it does not fall between the lower and upper bounds max $(t, 1-t)$ and 1. There exist several ways to circumvent this problem. For instance, the following *ad hoc* transformation can be applied to the estimator $\hat{A}_p^{[1]}(\cdot)$

$$\tilde{A}_p^{[1]}(t) = \max\left\{\Psi(t), \hat{A}_p^{[1]}(t) + (1-t)(1-\hat{A}_p^{[1]}(0)) + t(1-\hat{A}_p^{[1]}(1))\right\}$$

where $\Psi(t) = \max(t, 1-t)$. By mimicking the arguments of Capèraá and Fougères [17], one can see that, as long as $h + (nh)^{-1} \to 0$ as $n \to \infty$, the estimator $\hat{A}_p^{[1]}(\cdot)$ is consistent for $A(\cdot)$ in addition to enjoying dependence functions properties.

Unfortunately, this *ad hoc* approach can not be adapted to the other estimators. For this reason, henceforth, we will adopt two alternative methods. The first one is based on the convex hull technique, whereas the second one makes use of the constrained smoothing splines approach. In both cases, the proposed estimation scheme is a two-step procedure. First, we pick one of the estimators $\hat{A}_p^{[m]}(\cdot)$, $\hat{A}_J^{[m]}(\cdot)$, $\hat{A}_A^{[m]}(\cdot)$, $\hat{A}_{\mathrm{HT}}(\cdot)$, and $\hat{A}_C(\cdot)$. When needed, the associated optimal bandwidth is also selected. Then, the obtained estimate is modified to fulfill dependence function properties. We shall use the generic notation $\hat{A}(\cdot)$ to designate any of the previous estimators subsequently.

5.1 *Convex hull technique*

This subsection uses a convex hull technique similar to the one described in ref. [6]. To be specific, define the cross-over points as follows

$$\begin{cases} t_0 &= 0 \\ t_j &= \dfrac{\log\{H_1(X_j)\}}{\log\{H_1(X_j)\} + \log\{H_2(Y_j)\}}, \quad \text{for } j = 1, \dots, n \\ t_{n+1} &= 1, \end{cases}$$

where $H_l; l = 1, 2$, is replaced by H_{ln} if the marginals are unknown. By reordering the sample, we can assume without loss of generality that $t_0 \le t_1 \le \cdots \le t_{n+1}$. Note that these cross-over points represent the points where the estimate possibly changes expression. In particular, one can see that the estimate $\hat{A}_p^{[m]}(\cdot)$ is piecewise linear with possible slope change at the points t_j, $j = 1, \dots, n$. Let $\bar{A}$ represent the following modification of the estimator $\hat{A}$

$$\begin{cases} \bar{A}(t_0) = \bar{A}(t_{n+1}) = 1 \\ \bar{A}(t_j) = \min\left[1, \max\{t_j, 1-t_j, \hat{A}(t_j)\}\right]; \quad \text{for} j = 1, \dots, n, \end{cases}$$

$\bar{A}$ is linear in the intervals (t_j, t_{j+1}) for $j = 0, \dots, n$. The convex hull technique consists in defining the new estimator $\tilde{A}$ of A as the sup of all convex functions f satisfying $f(t) \le \bar{A}(t)$

for all $t \in [0, 1]$. As noted by Pickands [6], $\tilde{A}$ is the pointwise maximum of all straight lines which are less or equal to $\bar{A}(t_j)$ for all j. To implement such an estimator, it suffices to consider the supremum of all tangent lines. As pointed out by Pickands [6], this amounts to considering at most $(n + 1)$ lines.

In the earlier procedure, the cross-over points t_j could be replaced by any partition of the interval [0, 1]. However, to guarantee consistency of the convex estimate one needs the partition mesh to go to zero as the sample size goes to infinity.

5.2 *Constrained smoothing splines technique*

The convex hull approach can possibly produce non-smooth estimates, so as an alternative we propose the constrained smoothing splines approach. To be specific, let $0 = t_1 < t_2 < \cdot < t_q = 1$ be a partition of the interval [0, 1] and put $y_i = \hat{A}(t_i)$. Define the natural-cubic C^2-spline by

$$A_s(t) = \sum_{i=1}^{q-1} \mathbb{1}[t_i, t_{i+1})(t) S_i(t)$$

with

$$S_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3, \quad 1 \le i \le q - 1$$

where the coefficients fulfill the following continuity conditions:

$$S_{i-1}^{(r)}(t_i) = S_i^{(r)}(t_i), \quad \text{for } r = 0, 1, 2 \text{ and } i = 2, \ldots, q - 1,$$

with $S_i^{(r)}(t)$ denoting the rth derivative of $S_i(t)$ with respect to t. Then, we seek the spline which fulfills dependence function properties together with being smooth enough and as close as possible to the basic estimate $\hat{A}(\cdot)$. Thus, we have to minimize with respect to $A_s(\cdot)$ the following penalized least squares problem

$$\sum_{i=1}^{q} \{y_i - A_s(t_i)\}^2 + \lambda \int_0^1 (A_s^{(2)}(t))^2 \mathrm{d}t \tag{21}$$

$$\text{subject to} \begin{cases} A_s^{(2)}(t) \ge 0 \\ A_s(t) \le 1 \\ A_s(t) \ge \max(t, 1 - t) \end{cases} \quad \text{for all } t \in [0, 1], \tag{22}$$

where $A_s^{(2)}(t)$ denotes the second derivative of $A_s(t)$ with respect to t. As usual, the smoothing parameter λ acts as a trade-off between the solution's roughness and its fidelity to the basic estimate $\hat{A}(\cdot)$. However, for the actual problem, the issue of choosing the most appropriate λ for the pseudo-data $y_1, \ldots, y_q$ is not very important. In fact, the observations $\hat{A}(t_1), \ldots, \hat{A}(t_q)$ being themselves consistent estimates of $A(t_1), \ldots, A(t_q)$ and almost not variable, we just need to take λ relatively small or equal to 0. From our experience, the smoothness of the spline solution does not deteriorate when the smoothing parameter λ is set to 0. Now, despite this fact, one can also automatically select λ by adapting any standard technique, such as cross-validation or generalized cross-validation or Akaike's information criterion, to the actual problem [see, *e.g.*, ref. 32]. Constrained smoothing splines have been considered by several authors, and we refer to refs. [33–35] for details and references.

Finally, as pointed out by Turlach [33], a direct implementation of the previous minimization problem would lead to an unnecessarily large quadratic programming problem. To get around this problem and speed up computations, we used the value-second derivative representation [see, ref. 32]. together with the algorithm discussed by Turlach [33].

6. Monte-Carlo simulations

To compare the estimators presented in the previous sections, we carried out the Monte-Carlo experiment. A wide range of dependence functions have been considered. Indeed, we considered three sets of distributions. The first set consists of distributions in the max-domain of attraction of independence, the second set contains distributions belonging to the max-domain of attraction of extreme dependence and the third set consists of extreme value distributions. The tail dependence measures $\chi = \lim_{u\to 1} 2 - \{1 - C(u,u)\}/(1-u) = \lim_{u\to 1} 2 - \log C(u,u)/\log(u)$ and $\bar{\chi} = \lim_{u\to 1} 2\log(1-u)/\log \bar{C}(u,u) - 1$ introduced by Coles *et al.* [36] are used to select the set of parameters of each distribution. As pointed out by Coles *et al.* [36], the couple $(\chi, \bar{\chi})$ provides a good summary for extremal dependence. In fact, Coles *et al.* [36] showed that for distributions in max-domain of attraction of independence $\chi = 0$ and the degree of dependence in the extreme is appropriately measured by $\bar{\chi}$, whereas for distributions in max-domain of attraction of extreme dependence $\bar{\chi} = 1$ and χ measures the strength of dependence in the extremes [see ref. 36 for details]. Note that, using (9), one verifies that $\chi = 2\{1 - A(1/2)\}$ holds for any copula C in the max-domain of attraction of an extreme value copula with dependence function A.

In the first set, we chose the Gaussian copula

$$C_\rho(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) \mathrm{d}s\mathrm{d}t$$

where $-1 < \rho < 1$ and Φ is the standard normal distribution function. Recall that $A(t) = 1$, for all $t \in [0,1]$ and $\chi = 0$. We also have $\chi = \rho$ see ref. [36]. We also simulated the Clayton (Cook-Johnson) family of Archimedean copulas given by

$$C(u,v) = \{u^{-\alpha} + v^{-\alpha} - 1\}^{-1/\alpha}, \quad \alpha > 0, \quad u, v \in [0,1],$$

For this family $A(t) = 1$, for all $t \in [0,1]$, the dependence measures $\chi = 0$ and $\bar{\chi} = 0$. To vary the degree of dependence and to allow for comparison the parameter is selected using values of Kendall tau comparable to those used in the Gaussian copula.

For the second set, we simulated from the t-copula,

$$C_{\nu,\rho}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left\{1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} \mathrm{d}s\,\mathrm{d}t$$

where $-1 < \rho < 1$ and $t_\nu^{-1}(\cdot)$ is the inverse of the cumulative t-distribution with v degrees of freedom. Using characterization (9) and some algebraic manipulations, one gets

$$A(t) = 1 - \frac{\Gamma(\nu + 2/2)(1-\rho^2)^{(\nu+1)/2}}{\nu\sqrt{\pi}\Gamma(\nu+1)/2} \int_0^1 \min\{t(1-x), (1-t)x\}[x(1-x)]^{-(2\nu+1)/\nu}$$
$$\times \{x^{-2/\nu} - 2\rho[x(1-x)]^{-1/\nu} + (1-x)^{-2/\nu}\}^{-(\nu+2)/2}\,\mathrm{d}x.$$

As a special case, one gets the extreme dependence function of the Cauchy copula by letting $\nu = 1$. In which case, the extreme attractor of this distribution has $A(t) = (1 + \sqrt{1 - 2(1+\rho)t(1-t)}/2$. These distributions belong to the max-domain of attraction of dependence and have $\chi = 2\{1 - A(1/2)\}$ and $\bar{\chi} = 1$.

We also included, in this second set, a mixture of Clayton archimedean copula and Gumbel extreme value copula. We used the mixture procedure discussed in ref. [37]. This procedure is summarized as follows: for the copulas C_1 and C_2 and for $r_1, r_2 \in [0,1]$, we generate (U_1, V_1) from C_1 and (U_2, V_2) from C_2 and we set $(U, V) = (\max\{U_1^{1/r_1}, U_2^{1/(1-r_1)}\}, \max\{V_1^{1/r_2}, V_2^{1/(1-r_2)}\})$. Then, it is easy to verify that the extreme attractor of (U, V) has dependence function A given by (23), where A_1 (respectively A_2) is the dependence function of the attractor of C_1 (respectively C_2). In our case, C_1 is the Clayton copula with parameter α and C_2 is the Gumbel copula, with parameter $r \geq 1$, given by $C_2(u, v) = \exp[\{(-\log u)^r + (-\log v)^r\}^{1/r}]$. This yields

$$\chi = (2 - r_1 - r_2)\left[1 - \left\{\left(\frac{1-r_1}{2-r_1-r_2}\right)^r + \left(\frac{1-r_2}{2-r_1-r_2}\right)^r\right\}^{1/r}\right]$$

and $\bar{\chi} = 1$.

For the third set, we simulated the following extreme value copulas:

- Gumbel: $A(t) = \{(1-t)^r + t^r\}^{1/r}, \quad r \geq 1$
- Galambos: $A(t) = 1 - \{(1-t)^{-r} + t^{-r}\}^{-1/r}, \quad r \geq 0$
- Marshall–Olkin: $A(t) = \max(1 - r_1 t, 1 - r_2(1-t)), \quad r_1, r_2 \in [0, 1]$

Furthermore, we enriched this selection of models by using the asymmetrization technique [see ref. 37]. Given two dependence functions A_1 and A_2, one obtains a new dependence function by setting

$$A(t) = [r_1 t + r_2(1-t)]A_1\left(\frac{r_1 t}{r_1 t + r_2(1-t)}\right) + [(1-r_1)t + (1-r_2)(1-t)]A_2\left(\frac{(1-r_1)t}{(1-r_1)t + (1-r_2)(1-t)}\right) \quad (23)$$

where $r_1,\ r_2 \in [0, 1]$. This dependence function $A(\cdot)$ is symmetric for $r_1 = r_2$ and asymmetric otherwise. For instance, if we take $A_1(t) = \{(1-t)^r + t^r\}^{1/r}$ and $A_2(t) = 1$, we get the well-known asymmetric logistic family [see ref. 10]:

$$A(t) = \{\phi^r t^r + \vartheta^r(1-t)^r\}^{1/r} + (\vartheta - \phi)t + 1 - \vartheta,$$

with $0 \leq \vartheta, \phi \leq 1, r \geq 1$. For all distributions in this class $\bar{\chi} = 1$ and $\chi = 2\{1 - A(1/2)\}$.

The subsequent comparisons involve the estimators $\hat{A}_1^{[1]}$, $\hat{A}_2^{[1]}$, $\hat{A}_\infty^{[1]}$, $\hat{A}_J^{[1]}$, $\hat{A}_A^{[1]}$, $\hat{A}_{\mathrm{HT}}$ and $\hat{A}_C$. As a weight function $K(\cdot)$, we took the Epanechnikov kernel $K(x) = 3/4(1-x)^2 \mathbb{1}\{|x| \leq 1\}$. For the double kernel technique, we used kernels $K^{[1]}$ and $K^{[2]}$. For each selected model, we run 200 simulations with sample size of 100. In each case, we computed the L_1-error and the integrated error between estimates and the theoretical dependence function under investigation. Both convex hull and constrained smoothing techniques were applied and yielded quite similar results, therefore only the results for constrained smoothing are presented.

For the set of distributions belonging to the max-domain of attraction of independence ($\chi = 0$), table 1 gives the mean and the standard deviation of $100\times$ L_1-errors of the seven estimators outlined previously, whereas table 2 provides the mean and the standard deviation of $100\times$ the integrated error. First note that all estimators are underestimating the dependence function. This is expected here as the dependence function corresponds to $A(t) = 1$ for all $t \in [0, 1]$, which is the maximum value that a dependence function can take. We also observe that the estimation error increases with the degree of dependence measured via the Kendall's tau. For the same Kendall's tau, the error increases with the tail dependence measured by $\bar{\chi}$.

For distributions that are not extreme value distributions and that present tail dependence property ($\chi > 0$), table 3 outlines the results of the simulation about the L_1-error and table 4 provides the results for the integrated error. Both tables show that the estimation is of better

Table 1. Mean and standard deviation of 100× L_1-error using constrained smoothing for distributions in the domain of attraction of extreme independence.

Copula	τ	$\bar{\chi}$	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{\text{HT}}$	$\hat{A}_C$
Gaussian	0.0	0	6.02	9.63	10.85	11.82	1.57	2.20	2.37
			(3.95)	(2.95)	(2.32)	(1.75)	(2.23)	(2.33)	(1.97)
	0.2	0.31	9.66	12.47	13.51	14.32	6.50	10.42	9.09
			(5.18)	(3.91)	(3.14)	(2.55)	(3.04)	(2.47)	(2.19)
	0.4	0.59	14.09	16.17	16.92	17.41	12.12	17.07	15.23
			(5.79)	(4.28)	(3.53)	(3.01)	(3.18)	(1.57)	(1.63)
	0.6	0.81	19.53	20.55	20.90	21.05	17.33	21.76	20.38
			(4.74)	(3.55)	(3.05)	(2.45)	(2.68)	(0.70)	(0.91)
	0.8	0.95	23.79	24.11	24.15	24.21	21.72	24.25	23.73
			(1.95)	(1.33)	(1.08)	(0.96)	(1.61)	(0.16)	(0.30)
Clayton	0.2	0	6.03	9.86	11.33	12.45	4.11	12.05	8.31
			(3.96)	(3.11)	(2.42)	(1.89)	(2.21)	(2.32)	(2.01)
	0.4	0	7.29	11.07	12.74	13.62	7.45	18.45	13.57
			(4.93)	(3.80)	(3.14)	(2.31)	(2.47)	(1.35)	(1.74)
	0.6	0	9.47	12.98	14.63	15.53	11.80	22.20	18.13
			(5.63)	(4.41)	(3.55)	(2.75)	(2.19)	(0.68)	(1.26)
	0.8	0	12.79	16.15	18.10	18.17	16.94	24.23	21.90
			(6.00)	(4.76)	(3.81)	(2.81)	(2.10)	(0.23)	(0.84)

Note: The standard deviation is given between parentheses.

quality than the case ($\chi = 0$) and that the quality of estimation improves as the tail dependence (χ) increases.

Next, for the set of extreme value distributions, tables 5 and 6 provide the mean and standard deviation of the L_1 and the integrated errors, respectively. The results confirm the intuitive fact that dependence functions of extreme value distributions should be easier to estimate than the attractor's dependence function of non-extreme value distributions. They also show that

Table 2. Mean and standard deviation of 100× integrated error using constrained smoothing for distributions in the domain of attraction of extreme independence.

Copula	τ	$\bar{\chi}$	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{\text{HT}}$	$\hat{A}_C$
Gaussian	0.0	0	−6.02	−9.63	−10.85	−11.82	−1.57	−2.20	−2.37
			(3.95)	(2.95)	(2.32)	(1.75)	(2.23)	(2.33)	(1.97)
	0.2	0.31	−9.66	−12.47	−13.51	−14.32	−6.50	−10.42	−9.09
			(5.18)	(3.91)	(3.14)	(2.55)	(3.04)	(2.47)	(2.19)
	0.4	0.59	−14.09	−16.17	−16.92	−17.41	−12.12	−17.07	−15.23
			(5.79)	(4.28)	(3.53)	(3.01)	(3.18)	(1.57)	(1.63)
	0.6	0.81	−19.53	−20.55	−20.90	−21.05	−17.33	−21.76	−20.38
			(4.74)	(3.55)	(3.05)	(2.45)	(2.68)	(0.70)	(0.91)
	0.8	0.95	−23.79	−24.11	−24.15	−24.21	−21.72	−24.25	−23.73
			(1.95)	(1.33)	(1.08)	(0.96)	(1.61)	(0.16)	(0.30)
Clayton	0.2	0	−6.03	−9.86	−11.33	−12.45	−4.11	−12.05	−8.31
			(3.96)	(3.11)	(2.42)	(1.89)	(2.21)	(2.32)	(2.01)
	0.4	0	−7.29	−11.07	−12.74	−13.62	−7.45	−18.45	−13.57
			(4.93)	(3.80)	(3.14)	(2.31)	(2.47)	(1.35)	(1.74)
	0.6	0	−9.47	−12.98	−14.63	−15.53	−11.80	−22.20	−18.13
			(5.63)	(4.41)	(3.55)	(2.75)	(2.19)	(0.68)	(1.26)
	0.8	0	−12.79	−16.15	−18.10	−18.17	−16.94	−24.23	−21.90
			(6.00)	(4.76)	(3.81)	(2.81)	(2.10)	(0.23)	(0.84)

Note: The standard deviation is given between parentheses.

Table 3. Mean and standard deviation of 100× L_1-error using constrained smoothing for distributions in the domain of attraction of extreme dependence.

Copula	Parameters	χ	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{\mathrm{HT}}$	$\hat{A}_C$
Cauchy	0	0.3	5.65	6.72	6.44	7.03	5.80	5.28	4.68
			(2.99)	(3.21)	(2.92)	(2.85)	(3.21)	(2.64)	(2.42)
	0.68	0.6	4.69	4.43	4.28	4.46	3.85	2.11	1.94
			(2.04)	(1.69)	(1.60)	(1.55)	(2.21)	(0.87)	(1.18)
	0.95	0.84	1.58	1.50	1.38	1.40	1.87	0.89	0.78
			(1.49)	(0.97)	(0.59)	(0.45)	(1.39)	(0.28)	(0.42)
Mixture	(0.7, 0.1)†	0.3	6.69	7.19	7.05	7.46	3.36	3.28	2.98
Clayton ($\alpha = 2$)			(3.20)	(2.94)	(2.51)	(2.54)	(2.33)	(1.72)	(1.60)
Gumbel ($r = 10$)	(0.4, 0.1)	0.6	5.08	4.99	4.86	4.85	2.95	2.32	1.96
			(1.92)	(1.66)	(1.42)	(1.60)	(1.64)	(1.09)	(1.04)
	(0.1, 0.1)	0.84	2.71	2.44	2.39	2.36	2.27	1.15	1.17
			(1.76)	(0.68)	(0.43)	(0.42)	(1.30)	(0.49)	(0.49)

†Asymmetry parameters (r_1, r_2).
Note: The standard deviation is given between parentheses.

even within extreme value distributions the quality of estimation improves with increasing tail dependence (χ).

Note also that as seen from tables 3–6, the asymmetry effect is not noticeable. In fact, it is mainly the effect of the dependence structure which dominates the behavior of the proposed estimators. This can be seen mainly when looking at the Marshall–Olkin part in table 5.

In general, looking at the integrated errors, it is worth noting that most of the time the estimators $\hat{A}_1^{[1]}$, $\hat{A}_2^{[1]}$, $\hat{A}_\infty^{[1]}$ and $\hat{A}_J^{[1]}$ are underestimating the dependence function. The estimator $\hat{A}_A^{[1]}$ has a tendency of being larger than $\hat{A}_1^{[1]}$, $\hat{A}_2^{[1]}$, $\hat{A}_\infty^{[1]}$ or $\hat{A}_J^{[1]}$, which is why it has in general a smaller integrated error. One notices that in terms of the L_1-errors, and for non-extreme value distributions, the estimator $\hat{A}_A^{[1]}$ has the best performance followed by $\hat{A}_1^{[1]}$. As expected for extreme value distributions, the estimators $\hat{A}_{\mathrm{HT}}$ and $\hat{A}_C$ outperform the rest of the estimators. Nonetheless, even for extreme-value distributions, the estimator $\hat{A}_A^{[1]}$ has quite adequate behavior even for moderate sample sizes ($n = 100$). It is also worth mentioning that the estimators $\hat{A}_{\mathrm{HT}}$ and $\hat{A}_C$ have lower variances, as they use all the data for estimation. Unfortunately, they are not consistent for non-extreme distributions unless one modifies them to work with block

Table 4. Mean and standard deviation of 100× integrated error using constrained smoothing for distributions in the domain of attraction of extreme dependence.

Copula	Parameters	χ	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{\mathrm{HT}}$	$\hat{A}_C$
Cauchy	0	0.3	−4.01	−6.25	−6.29	−6.94	2.77	4.96	4.51
			(4.97)	(4.03)	(3.22)	(3.01)	(6.00)	(3.15)	(2.67)
	0.68	0.6	−2.89	−3.75	−3.81	−4.04	0.55	−0.41	0.41
			(4.19)	(2.83)	(2.48)	(2.27)	(4.33)	(1.95)	(2.07)
	0.95	0.84	−0.90	−1.04	−1.18	−1.18	1.31	−0.42	−0.03
			(1.95)	(1.42)	(0.89)	(0.81)	(1.89)	(0.63)	(0.75)
Mixture	(0.7, 0.1)†	0.3	−5.27	−6.69	−6.65	−7.37	−0.79	−3.00	−2.61
Clayton ($\alpha = 2$)			(4.99)	(3.75)	(3.09)	(2.71)	(3.90)	(2.12)	(2.09)
Gumbel ($r = 10$)	(0.4, 0.1)	0.6	−3.18	−4.26	−4.36	−4.61	0.16	−2.02	−1.47
			(4.32)	(2.92)	(2.33)	(2.07)	(3.27)	(1.49)	(1.56)
	(0.1, 0.1)	0.84	−1.32	−1.86	−2.22	−2.09	0.86	−0.77	−0.44
			(2.88)	(1.58)	(0.87)	(0.99)	(2.33)	(0.84)	(1.02)

†Asymmetry parameters (r_1, r_2).
Note: The standard deviation is given between parentheses.

Table 5. Mean and standard deviation of 100× L_1-error using constrained smoothing for a set of extreme value distributions.

Copula	Parameters	χ	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{HT}$	$\hat{A}_C$
Gumbel	1.475	0.4	6.22	6.28	5.99	6.40	3.53	2.11	1.90
			(2.68)	(2.59)	(2.52)	(2.51)	(2.38)	(1.01)	(1.08)
	3.802	0.8	1.43	1.39	1.29	1.30	1.74	0.43	0.39
			(0.94)	(0.55)	(0.47)	(0.40)	(1.42)	(0.23)	(0.23)
Galambos	0.756	0.4	8.04	8.37	8.27	8.52	4.77	5.36	5.37
			(2.72)	(2.44)	(2.21)	(2.04)	(2.88)	(1.38)	(1.36)
	3.106	0.8	3.04	2.50	2.56	2.29	6.31	5.43	5.43
			(3.77)	(2.94)	(2.67)	(2.09)	(3.57)	(1.48)	(1.39)
Marshall–Olkin	(0.4, 0.4)†	0.4	6.63	7.45	7.86	8.02	3.63	2.56	2.39
			(3.55)	(3.48)	(3.07)	(2.75)	(2.07)	(1.43)	(1.38)
	(0.4, 0.8)	0.4	6.23	6.40	6.22	6.45	3.64	2.05	2.08
			(2.73)	(2.50)	(2.13)	(2.31)	(2.35)	(1.23)	(1.19)
	(0.8, 0.8)	0.8	4.04	3.90	4.03	3.94	2.91	1.49	1.56
			(1.66)	(1.04)	(0.86)	(0.88)	(1.72)	(0.69)	(0.79)
	(0.8, 0.9)	0.8	3.32	3.18	3.18	3.12	2.86	1.26	1.31
			(1.72)	(1.05)	(0.69)	(0.71)	(1.59)	(0.64)	(0.66)
Asymmetric logistic ($r = 6.6$)	(0.4, 0.9)†	0.4	6.38	6.40	6.20	6.37	3.59	1.99	1.94
			(2.63)	(2.51)	(2.12)	(2.36)	(2.23)	(1.18)	(1.20)
	(0.9, 0.9)	0.8	2.63	2.55	2.47	2.50	2.45	1.18	1.19
			(1.08)	(0.57)	(0.49)	(0.45)	(1.49)	(0.58)	(0.56)

†Parameters (r_1, r_2).
‡Asymmetry parameters (ϕ, φ).
Note: The standard deviation is given between parentheses.

Table 6. Mean and standard deviation of 100× integrated error using constrained smoothing for a set of extreme value distributions.

Copula	Parameters	χ	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$	$\hat{A}_{HT}$	$\hat{A}_C$
Gumbel	1.475	0.4	−4.90	−5.77	−5.74	−6.26	−0.18	−0.62	−0.85
			(4.65)	(3.55)	(3.00)	(2.79)	(4.19)	(2.07)	(1.91)
	3.802	0.8	−1.00	−1.11	−1.04	−1.14	1.21	−0.08	0.02
			(1.38)	(0.99)	(0.90)	(0.73)	(1.88)	(0.44)	(0.43)
Galambos	0.756	0.4	−7.67	−8.29	−8.25	−8.52	−4.37	−5.35	−5.37
			(3.62)	(2.68)	(2.25)	(2.05)	(3.43)	(1.41)	(1.36)
	3.106	0.8	2.48	1.99	2.12	1.78	6.22	5.43	5.43
			(4.15)	(3.28)	(3.00)	(2.49)	(3.73)	(1.48)	(1.39)
Marshall-Olkin	(0.4, 0.4)†	0.4	−4.96	−7.10	−7.73	−7.91	0.14	−0.83	−1.04
			(5.63)	(4.11)	(3.32)	(2.96)	(4.06)	(2.66)	(2.41)
	(0.4, 0.8)	0.4	−4.15	−5.52	−5.53	−6.16	0.22	−0.55	−0.63
			(5.26)	(3.92)	(3.24)	(2.84)	(4.21)	(2.07)	(2.11)
	(0.8, 0.8)	0.8	−2.35	−3.32	−3.81	−3.66	0.96	−0.46	−0.37
			(3.57)	(2.10)	(1.36)	(1.42)	(3.05)	(1.34)	(1.50)
	(0.8, 0.9)	0.8	−1.81	−2.47	−2.81	−2.78	1.11	−0.16	−0.06
			(3.15)	(2.08)	(1.44)	(1.37)	(2.94)	(1.17)	(1.25)
Asymmetric logistic ($r = 6.6$)	(0.4, 0.9)‡	0.4	−4.56	−5.59	−5.59	−6.15	−0.05	−0.85	−0.91
			(5.05)	(3.78)	(3.05)	(2.76)	(4.13)	(1.95)	(1.95)
	(0.9, 0.9)	0.8	−1.72	−2.16	−2.19	−2.25	1.04	−0.16	−0.09
			(2.17)	(1.30)	(1.11)	(0.97)	(2.56)	(1.12)	(1.10)

†Parameters (r_1, r_2).
‡Asymmetry parameters (ϕ, ϑ).
Note: The standard deviation is given between parentheses.

Table 7. Mean and standard deviation of the optimal bandwidth using Epanechnikov kernel.

Copula	Parameters	τ	χ	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$
Clayton	0.5	0.2	0	0.16	0.17	0.16	0.30	0.79
				(0.05)	(0.05)	(0.04)	(0.05)	(0.19)
	3	0.6	0	0.12	0.15	0.14	0.25	0.73
				(0.04)	(0.05)	(0.05)	(0.05)	(0.23)
	8	0.8	0	0.10	0.13	0.14	0.22	0.62
				(0.04)	(0.05)	(0.07)	(0.05)	(0.20)
Cauchy	0	0	0.3	0.25	0.26	0.23	0.36	0.62
				(0.11)	(0.10)	(0.09)	(0.11)	(0.31)
	0.68	0.48	0.6	0.21	0.22	0.22	0.30	0.43
				(0.09)	(0.10)	(0.09)	(0.08)	(0.19)
	0.95	0.8	0.84	0.14	0.18	0.22	0.28	0.42
				(0.06)	(0.07)	(0.08)	(0.05)	(0.12)
Mixture	(0.7, 0.1)†		0.3	0.18	0.21	0.19	0.32	0.59
Clayton ($\alpha = 2$)				(0.08)	(0.09)	(0.07)	(0.08)	(0.21)
Gumbel ($r = 10$)	(0.4, 0.1)		0.6	0.16	0.20	0.21	0.30	0.49
				(0.08)	(0.09)	(0.08)	(0.07)	(0.16)
	(0.1, 0.1)		0.84	0.14	0.17	0.20	0.28	0.43
				(0.06)	(0.07)	(0.07)	(0.05)	(0.12)

†Asymmetry parameters (r_1, r_2).
Note: The standard deviation is given between parentheses.

maxima. The variances of $\hat{A}_\infty^{[1]}$, $\hat{A}_J^{[1]}$ and $\hat{A}_A^{[1]}$ are, in general, smaller than those of $\hat{A}_1^{[1]}$ and $\hat{A}_2^{[1]}$. As a general conclusion, we recommend the estimate $\hat{A}_A^{[1]}$, unless it is known that the distribution is of the extreme value type, in which case it is advisable to use the estimate $\hat{A}_{\mathrm{HT}}$ or $\hat{A}_C$.

The optimal bandwidth is determined for each simulation run. Tables 7 and 8 report the variation of the optimal bandwidth with the distribution function and with the sample size. For table 7, the sample size is fixed and equals 100. The table reports the mean and standard deviation of the optimal bandwidth for a few non-extreme value distributions, obtained via 200 simulation replicates. It shows, in particular, that the optimal bandwidth decreases with increasing degree of dependence or tail-dependence. Table 8 studies the effect of sample size on the optimal bandwidth. It reports the mean and standard deviation of 200 simulation replicates from a bivariate Cauchy distribution with $\rho = 0.5$ and sample sizes equal 50, 100 and 200. As expected the optimal bandwidth decreases with the sample size. It is worth mentioning that the optimal bandwidth for the estimator $\hat{A}_A^{[1]}$ is in general bigger than that of the rest of the estimators.

Table 8. Mean and standard deviation of the optimal bandwidth using Epanechnikov kernel for a bivariate Cauchy distribution with ($\rho = 0.5$).

Sample size	$\hat{A}_1^{[1]}$	$\hat{A}_2^{[1]}$	$\hat{A}_\infty^{[1]}$	$\hat{A}_J^{[1]}$	$\hat{A}_A^{[1]}$
50	0.26	0.30	0.30	0.46	0.65
	(0.11)	(0.13)	(0.11)	(0.13)	(0.24)
100	0.22	0.24	0.22	0.32	0.53
	(0.10)	(0.11)	(0.09)	(0.09)	(0.23)
200	0.18	0.18	0.18	0.24	0.38
	(0.08)	(0.08)	(0.07)	(0.08)	(0.20)

Note: The standard deviation is given between parentheses.

7. Examples

As a first application, we consider the insurance company indemnity claims data set discussed in ref. [38] and also revisited by Genest *et al.* [37]. The data contains 1500 claims, each claim comprising the indemnity payment (the loss, X) and the allocated loss adjustment expense (ALAE, Y). Extreme value analysis is becoming a standard technique in risk analysis for insurance companies as large values of claims and other expenses are the main contributors to the company's risk factor. For a discussion on the use of extreme value in risk analysis, see refs. [39, 40]. Furthermore, for the previous data set, Frees and Valdez [38] and later Genest *et al.* [37] argued that the Gumbel copula provides the best fit describing the dependence structure between the loss and ALAE, therefore, the data fits an extreme value copula. We decided to verify this claim through the estimation of the extreme dependence function of its attractor and see if it coincides with the suggested Gumbel model. The results of the estimation are shown in figure 1. It shows the dependence function of the Gumbel family with $r = 1.44$, which Genest *et al.* [37] showed to be the max-pseudo-likelihood estimate for the copula associated with these data. For the figure to be readable, we only represented the two most prominent estimates, that is $\hat{A}_1^{[1]}$ and $\hat{A}_A^{[1]}$ computed using Epanechnikov kernel. The optimal bandwidths are 0.122 and 0.347, respectively. We also included the 2.5% and 97.5% percentiles of $\hat{A}_A^{[1]}$, obtained using non-parametric bootstrap. The figure shows that the suggested dependence function is quite close to $\hat{A}_1^{[1]}$ and falls within the limits of the 95% bootstrap confidence interval most of the times. The other estimates ($\hat{A}_2^{[1]}$, $\hat{A}_\infty^{[1]}$ and $\hat{A}_J^{[1]}$) fall between $\hat{A}_1^{[1]}$ and $\hat{A}_A^{[1]}$. Though not included in figure 1, $\hat{A}_{\mathrm{HT}}$ and $\hat{A}_C$ are very close to the suggested Gumbel model with $r = 1.44$ and they fall between the Gumbel curve and that of $\hat{A}_A^{[1]}$. As these two estimates are only consistent for extreme value distributions, this reaffirms that the copula associated with this data might be an extreme value copula.

The second application consists of the acid rain data discussed by Joe *et al.* [7]. The data consist of 504 observations of sulfate and nitrate concentrations measured over a period of about 6 years for station 65 (Penn State). As argued by Joe *et al.* [7], there is no time trend, therefore we may treat the data as a series of independent and identically distributed random variables. The aim is to study dependence in the extreme of the two pollutants (sulfate and

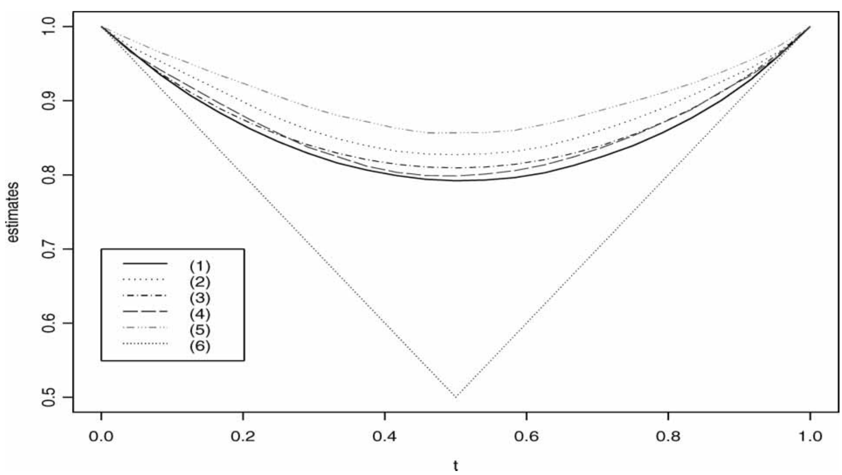

Figure 1. Estimates of the function A for the Loss and ALAE data. (1) $\hat{A}_1^{[1]}$; (2) $\hat{A}_A^{[1]}$; (3) $A(t)$ of the Gumbel family with $r = 1.44$; (4) lower limit of 95% bootstrap confidence interval obtained from $\hat{A}_A^{[1]}$; (5) upper limit of 95% bootstrap confidence interval obtained from $\hat{A}_A^{[1]}$ and (6) $\max(t, 1 - t)$.

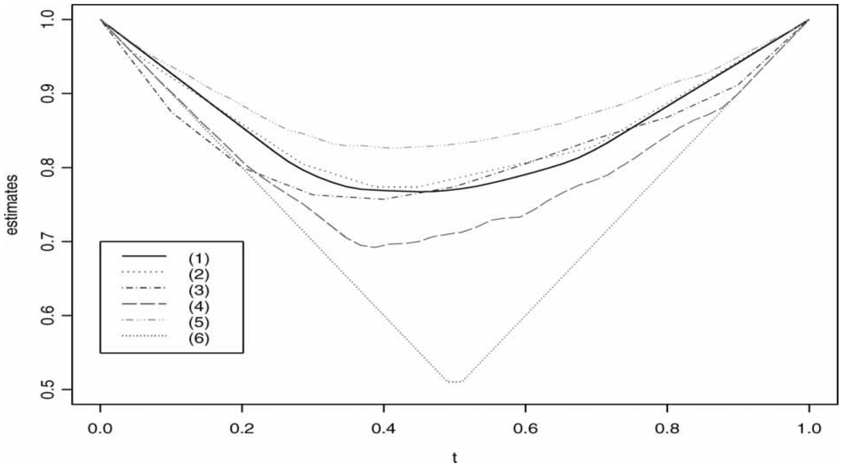

Figure 2. Estimates of the function A for the acid rain data. (1) $\hat{A}_1^{[1]}$; (2) $\hat{A}_A^{[1]}$; (3) the non-parametric estimate given in Joe *et al.* [7]; (4) lower limit of 95% bootstrap confidence interval obtained from $\hat{A}_A^{[1]}$; (5) upper limit of 95% bootstrap confidence interval obtained from $\hat{A}_A^{[1]}$ and (6) $\max(t, 1-t)$.

nitrate). By estimating the extreme dependence function, one gets a model that will be used in the computation of the probabilities of exceeding dangerous levels of pollutants.

We are also using this example to illustrate the use of the double kernel technique in the selection of the parameter k_n, discussed in Remark 4.1. The results presented in figure 2 were obtained using the double kernel technique outlined in Remark 4.1 for the selection of k_n. We also used the convex hull technique to force the estimate to be a dependence function. The optimal bandwidths ($h_n = k_n/n$) were 0.069 for $\tilde{A}_1$ and 0.259 for $\tilde{A}_A$. For clarity, we only included, in the figure, the estimates $\tilde{A}_A$ and $\tilde{A}_1$ and the limits of the 95% non-parametric bootstrap confidence interval for $\tilde{A}_A$. We also included the results of the non-parametric estimate obtained by Joe *et al.* [7] for comparison. The figure shows that the non-parametric estimate given in ref. [7] is close to both $\tilde{A}_1$ and $\tilde{A}_A$. This non-parametric estimate is not consistent with the properties of dependence functions for all $t \in (0, 1)$. Namely, it goes below $\max(t, 1-t)$ for small t. As shown in figure 2, when greater than $\max(t, 1-t)$, the Joe *et al.* [7] non-parametric estimate falls within the 95% bootstrap confidence interval obtained from $\tilde{A}_A$. It also falls within the 95% bootstrap confidence interval obtained from $\tilde{A}_1$, which is not included in the figure.

Though not included in the graphic, the estimates $\tilde{A}_{\mathrm{HT}}$ and $\tilde{A}_C$ differ substantially from the rest, which could be perceived as an indication that the extreme value family is not a good model for these data. It is also worth noting that all the estimates indicate that the extreme attractor is not symmetric, in opposition to the previous example where all curves indicate symmetry.

References

[1] Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997, *Modelling Extremal Events for Insurance and Finance* (Berlin: Springer).

[2] Barnett, V., 1976, The ordering of multivariate data (with discussion). *Journal of Royal Statistics Society Series A*, **139**, 318–354.

[3] Galambos, J., 1987, *The Asymptotic Theory of Extreme Order Statistics* (2nd edn) (Robert E. Melbourne, FL: Krieger Publishing Co. Inc.).

[4] Resnick, S.I., 1987, *Extreme Values, Regular Variation, and Point Processes* (New York: Springer-Verlag).

[5] Joe, H., 1997, *Multivariate Models and Dependence Concepts. Monographs on Statistics and Applied Probability*, Vol. 73 (London: Chapman & Hall).

[6] Pickands, J., 1981, Multivariate extreme value distributions. *Proceedings of the 43rd Session International Statistical Institute*, 859–878.
[7] Joe, H. Smith, R.L. and Weissman, I., 1992, Bivariate threshold methods for extremes. *Journal of Royal Statistical Society Series B*, **54**, 171–183.
[8] Abdous, B., Ghoudi, K. and Khoudraji, A., 1999, Non-parametric estimation of the limit dependence function of multivariate extremes. *Extremes*, **2:3**, 245–268.
[9] Tiago de Oliveira, J., 1989, Intrinsic estimation of the dependence structure for bivariate extremes. *Statistical Probability Letters*, **8**, 213–218.
[10] Tawn, J.A., 1988, Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
[11] Smith, R.L., Tawn, J.A. and Yuen, H.K., 1990, Statistics of multivariate extremes. *International Statistical Review*, **58**, 47–58.
[12] Stephenson, A. and Tawn, J.A., 2004, Exploiting occurence times in likelihood inference for componentwise maxima. *Biometrika*, **92**, 213–217.
[13] Huang, X., 1992, Statistics of bivariate extreme values. Ph.D. thesis, Erasmus University, Rotterdam, Tinbergen Institute Research Series 22.
[14] Einmahl, J., de Haan, L. and Huang, X., 1993, Estimating a multidimensional extreme value distribution. *Journal of Multivariate Analysis*, **47**, 35–47.
[15] Einmahl, J., de Haan, L. and Sinha, A.K., 1997, Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, **70**, 143–171.
[16] Einmahl, J., de Haan, L. and Piterbarg, V., 2001, Non-parametric estimation of the spectral a measure of an extreme value distribution. *Annals of Statistics*, **29**(5), 1401–1423.
[17] Capéraà, P. and Fougéres, A.-L., 2000, Estimation of a bivariate extreme value distribution. *Extremes*, **3**(4), 311–329.
[18] Deheuvels, P., 1991, On the limiting behavior of the Pickands estimator for bivariate extremevalue distributions. *Statistical Probability Letters*, **12**, 429–439.
[19] Hall, P. and Tajvidi, N., 2000, Distribution and dependence function estimation for bivariate extreme-value distributions. *Bernoulli*, **6**(5), 835–844.
[20] Capéraà, P. Fougères, A.-L. and Genest, C., 1997, A non-parametric estimation procedure for bivariate extreme value copulas. *Biometrika*, **84**(3), 567–577.
[21] Rojo, J.J., Vi1latDiharce, E., and Flores, M., 2001, Non-parametric estimation of the dependence function in bivariate extreme value distributions. *Journal of Multivariate Analysis*, **76**, 159–191.
[22] Fan, J. and Gijbels, L., 1996, *Local Polynomial Modelling and its Applications. Monographs on Statistics and Applied Probability*, Vol. 66 (London: Chapman & Hall).
[23] Abdous, B. and Bensaid, E., 2001, Multivariate local polynomial fitting for a probability distribution function and its partial derivatives. *Journal of Nonparametric Statistics*, **13**(1), 77–94.
[24] Wand, M.P. and Jones, M.C., 1995, *Kernel Smoothing. Monographs on Statistics and Applied Probability*, 60 (London: Chapman & Hall).
[25] Devroye, L., 1997, Universal smoothing factor selection in density estimation: theory and practice. *Test*, **6**, 223–320.
[26] Devroye, L., 1989, The double kernel method in density estimation. *Annales de l'Institut Henri Poincaré*, **25**, 533–580.
[27] Berlinet, A. and Devroye, L., 1994, A comparison of kernel density estimates. *Publications de l'Institut de Statistique de l'Université de Paris*, **38**(3), 3–59.
[28] Devroye, L. and Lugosi, G., 1996, A universally acceptable smoothing factor for kernel density estimates. *Annals of Statistics*, **24**, 2499–2512.
[29] Devroye, L. and Lugosi, G., 1997, Nonasymptotic universal smoothing factors, kernel complexity, and Yatracos classes. *Annals of Statistics*, **25**, 2626–2637.
[30] Jones, M.C., 1998, On some kernel density estimation bandwidth selectors related to the double kernel method. *Sankhyā Series A*, **60**(2), 249–264.
[31] Abdous, B., 1999, L_2 version of the double kernel method. *Statistics*, **32**(3), 249–266.
[32] Green, P.J. and Silverman, B.W., 1994, *Non-parametric Regression and Generalized Linear Models. A Roughness Penalty Approach. Monographs on Statistics and Applied Probability*, Vol. 58 (London: Chapman & Hall).
[33] Turlach, B., 1997, Constrained smoothing splines revisited. *Statistics Research Report SRR 008-97*. Centre for Mathematics and its Applications, School of Mathematical Sciences. Australian National University.
[34] Mammen, E. and Thomas-Agnan, C., 1999, Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, **26**(2), 239–252.
[35] Rodriguez Poo, J.M., 1999, Constrained smoothing splines. *Econometric Theory*, **15**, 114–138.
[36] Coles, S., Heffernan, J. and Tawn, J., 1999, Dependence measures for extreme value analyses. *Extremes*, **2**(4), 339–365.
[37] Genest, C., Ghoudi, K. and Rivest, L.-P., 1998, Discussion of "Understanding relationships using copulas, by Edward Frees and Emiliano Valdez". *North American Actuarial Journal*, **3**, 543–552.
[38] Frees, W.E. and Valdez, E.A., 1998, Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–25.
[39] Longin, F.M., 2000, From value at risk to stress testing: The extreme value approach. *Journal of Banking and Finance*, **24**, 1097–1130.
[40] McNeil, A.J. and Frey, R., 2000, Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, **7**, 271–300.