Multivariate Nonparametric Estimation of the Pickands Dependence Function using Bernstein Polynomials

G. Marcon, S. A. Padoan, P. Naveau and P. Muliere *

August 27, 2014

Abstract

Many applications in risk analysis, especially in environmental sciences, require the estimation of the dependence among multivariate maxima. A way to do this is by inferring the so-called Pickands dependence function used in multivariate Extreme Value Theory. In this context, a clear advantage of a nonparametric approach over a parametric one is its flexibility and theoretical generality. Beyond the bivariate case, nonparametric estimation of the dependence function remains a challenging task and an active research field. In this article, we propose a new nonparametric approach for estimating the Pickands dependence function and we insure that it obeys all Pickands’ constraints by taking advantage of a specific type of Bernstein polynomials representation. We discuss the properties of the proposed estimation method and illustrate its performance with a simulation study. For moderate dimension sizes, we illustrate our approach by analyzing clusters made of seven weather stations that have recorded weekly maxima of hourly rainfall in France from 1993 to 2011.

*Marcon is PhD student, Muliere and Padoan work at the Department of Decision Sciences, Bocconi University of Milan, via Roentgen 1, 20136 Milano, Italy, E-mail: giulia.marcon@phd.unibocconi.it, pietro.muliere@unibocconi.it, simone.padoan@unibocconi.it. Naveau is a CNRS researcher at the Laboratoire des Sciences du Climat et l’Environnement, Gif-sur-Yvette, France, E-mail: naveau@lsce.ipsl.fr
Keywords: Bernstein polynomials, Extremal dependence, Heavy rainfall, Non-parametric estimation, Multivariate max-stable distribution, Pickands dependence function.

1 Introduction and background

In recent years, inference methods for assessing the extremal dependence have increasingly been in demand. This is especially due to growing requests for multivariate analyses of extreme values in the fields of environmental and economic sciences. In these contexts, the dimension of the multivariate vector under study is often greater than two. For example, Figure 1 displays a map of clusters made of seven French weather stations (see Bernard, Naveau, Vrac, and Mestre (2013) for details on how to construct the clusters). Here, the dimension of our multivariate vector of weekly maxima of hourly rainfall is seven and it would be of interest for hydrologists to infer

![Figure 1: Analysis of French weekly precipitation maxima from 1993-2011. Clusters of 49 weather stations and their estimated extremal coefficients in dimension $d = 7$, see Section 5 for details.](image-url)
and compare the dependence among these multivariate vector of maxima with seven components. Such an endeavor represents the main motivation of this work.

We assume that a \( d \)-dimensional random vector of maxima, say \( \mathbf{X} = \{X_i\}_{i=1,\ldots,d} \), with unit Fréchet marginal distributions (\( \mathbb{P}\{X \leq x\} = e^{-1/x} \) for \( x > 0 \)) follows a multivariate max-stable distribution (e.g., de Haan and Ferreira 2006, Ch. 6)

\[
G(\mathbf{x}) = \exp\{-V(\mathbf{x})\}, \quad V(\mathbf{x}) = \left( \frac{1}{x_1} + \cdots + \frac{1}{x_d} \right) A(\mathbf{w}), \quad (1)
\]

where \( \mathbf{x} = (x_1, \ldots, x_d) \) and \( \mathbf{w} = (w_1, \ldots, w_{d-1}) \) with \( w_i = x_i / (x_1 + \cdots + x_d) \) for \( i = 1, \ldots, d \). Specifically, the function \( V : [0, \infty)^d \to [0, \infty) \), named the exponent measure, has to be continuous, convex, homogeneous of order \(-1\), i.e. \( V(ax) = a^{-1}V(x) \) for any \( a > 0 \), bounded by \( \max(x_1, \ldots, x_d) \leq V(\mathbf{x}) \leq x_1 + \cdots + x_d \), and the end points must satisfy \( V(x, 0, \ldots, 0) = \cdots = V(0, \ldots, 0, x) = x \) for all positive \( x \) (for more details see Falk, Hüsler, and Reiss, 2010, Ch. 4; de Haan and Ferreira, 2006, Ch. 6). The homogeneity property of \( V \) means that it can be rewritten in terms of \( A \), the Pickands dependence function, which is the restriction of \( V \) on the \( d \)-dimensional unit simplex (Pickands 1981). Simplifying, the Pickands dependence function can be seen as a function defined on the space

\[
\mathcal{S}_{d-1} := \left\{ (w_1, \ldots, w_{d-1}) \in [0, 1]^{d-1} : \sum_{i=1}^{d-1} w_i \leq 1 \right\}, \quad (2)
\]

as we can always define a component through the others, e.g. \( w_d = 1 - w_1 - \cdots - w_{d-1} \).

The Pickands inherits the properties of the exponent measure function. Let \( \mathcal{A} \) be the family of functions, \( A : \mathcal{S}_{d-1} \to [1/d, 1] \), satisfying the conditions:

C1) \( A(\mathbf{w}) \) is continuous and convex, that is \( A(a\mathbf{w}_1 + (1-a)\mathbf{w}_2) \leq aA(\mathbf{w}_1) + (1-a)A(\mathbf{w}_2) \), for \( a \in [0, 1] \) and \( \mathbf{w}_1, \mathbf{w}_2 \in \mathcal{S}_{d-1} \);

C2) \( A(\mathbf{w}) \) has lower and upper bounds

\[
1/d \leq \min(\mathbf{w}) \leq A(\mathbf{w}) \leq 1,
\]

for any \( \mathbf{w} = (w_1, \ldots, w_{d-1}) \in \mathcal{S}_{d-1} \) with \( w_d = 1 - w_1 - \cdots - w_{d-1} \);

C3) \( A(\mathbf{0}) = 1 \) and \( A(\mathbf{e}_i) = 1 \), for the boundary points of \( \mathcal{S}_{d-1} \), \( \mathbf{0} = (0, \ldots, 0) \) and \( \mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \) for \( i = 1, \ldots, d-1 \).
A function $\mathcal{A}$ in order to be a valid Pickands dependence function has to be a member of $\mathcal{A}$ (Falk, Hüsler, and Reiss, 2010, Ch. 4). But, conditions C1-C2-C3 are not sufficient. For instance, an ad hoc example of a function that satisfies C1-C2-C3 and is not a valid Pickands dependence function is discussed in Beirlant et al. (2004, p. 257). In condition C2, the lower and upper bounds represent the cases of complete dependence or independence, respectively.

Many parametric models have been introduced for modelling the extremal dependence for a variety of applications, summaries can be found in Kotz and Nadarajah (2000) and Padoan (2013). In theory, the dependence of multivariate max-stable distributions cannot be parameterised (e.g. de Haan and Ferreira, 2006, Ch. 6) and thus any arbitrary choice of a parametric model for $\mathcal{A}$ represents a loss of generality. For this reason, several nonparametric estimators of the Pickands have been proposed, examples are: Pickands (1981), Capéraà et al. (1997), Hall and Tajvidi (2000), Zhang et al. (2008), Genest and Segers (2009), Gudendorf and Segers (2011, 2012) to name a few. However, it has not been possible to show that the nonparametric estimators proposed so far, especially in their multivariate version, fulfill, without further adjustments, all three conditions C1, C2 and C3 together. So, for practical applications it is already a considerable result to provide an estimator that meets such conditions.

In this paper we propose a method to non-parametrically estimate the multivariate Pickands dependence function. In the bivariate case, a fast-to-compute and easy-to-interpret estimator based on a type of madogram was introduced by Naveau et al. (2009) but it has two drawbacks. It was only defined for the bivariate case and it does not necessarily satisfy the three conditions C1, C2 and C3. Our first task is to propose a new type of madogram for a multivariate setting with $d \geq 2$, see also the work of Fonseca et al. (2013). Still, conditions C1, C2 and C3 are not necessarily checked for this extension and another aim of this paper is to “regularise” it by projecting it onto $\mathcal{A}$ through a Bernstein polynomials based method. Hence, by construction, the new regularized estimator based on a multivariate madogram estimate will belong to $\mathcal{A}$. In the bivariate case, the regularization strategy has been initially investigated by Pickands (1981) proposing, in order to satisfy the convexity condition, to consider the greatest convex minorant (see also Capéraà et al. 1997, Hall and Tajvidi 2000). Smith et al. (1990) proposed to modify a first guess estimator $\hat{\mathcal{A}}_n$ using kernel methods and Hall and Tajvidi (2000) to approximate $\hat{\mathcal{A}}_n$ using constrained smoothing splines. However, as discussed by Fils-Villetard et al. (2008) the
impact of these adjustments on the asymptotic properties of the estimator changes
from one case to another, while a general result is unknown. The projection estima-
tor approach introduced by Fils-Villetard et al. (2008) provides a general theoretical
framework that is suitable for estimating the Pickands dependence function and it
has recently been extended in the multivariate case by Gudendorf and Segers (2012).
In any case, only approximate projection estimators belonging to finite-dimensional
subsets \( A_k \subseteq \mathcal{A} \), which are increasingly accurate with increasing \( k \), can be obtained.
Typically, approximate projection estimators of the Pickands dependence function
have been carried out by piecewise linear functions (Fils-Villetard et al. 2008; Gu-
dendorf and Segers 2012). This leads to estimators that may lack some smoothness
while the true Pickands dependence function has to be continuous and convex, see
condition C1. More importantly from a practical point of view, the computation of
multivariate integrals involved in the estimation procedure (Gudendorf and Segers 2012)
can be time consuming for large sample sizes and challenging to implement in
high dimensions (e.g. \( d > 3 \)). This may explain why applications of these approaches
have been, to our knowledge, limited to the bivariate case.

To bypass these computational hurdles, our strategy is to project any Pickands
dependence function estimators via a sequence of restricted Bernstein polynomials
(Lorentz, 1986; Sauer, 1991). In virtue of their optimal shape restriction proper-
ties (Carnicer and Peña, 1993), Bernstein polynomials are suitable for nonparametric
curve estimation (e.g. Petrone 1999; Chang et al. 2005) and shape-preserving re-
gressions (Wang and Ghosh 2012). Here we show their utilities when estimating
the Pickands dependence function. Our inference approach displays nice asymptotic
properties, it is computationally efficient and it is feasible to implement in moderately
high dimensions with an acceptable computational cost, and the uncertainties of the
estimates can be easily assessed through bootstrap confidence bands. We stress that
with our approach, in practice we are able to nonparametrically estimate the Pickands
dependence function, for instance, up to the dimension seven. This, it seems, has not
already been done in a real application before.

Throughout the paper we use the following notation. Given \( \mathcal{X} \subset \mathbb{R}^n \) and \( n \in \mathbb{N} \),
\( \mathcal{C}(\mathcal{X}) \) and \( \ell^\infty(\mathcal{X}) \) denote the space of continuous and bounded real-valued functions
on \( \mathcal{X} \), respectively. For \( f : \mathcal{X} \to \mathbb{R} \), \( \| f \|_\infty = \sup_{x \in \mathcal{X}} |f(x)| \). "\( \xrightarrow{a.s.} \)"", "\( \xrightarrow{d} \)"", "\( \xrightarrow{w} \)"
 denote the almost sure, in distribution and weak convergence, respectively. \( L^2(\mathcal{X}) \)
denotes the Hilbert space endowed with the \( n \)-dimensional Lebesgue measure and
with $L^2$-norm $\|f\|_2 = (\int_X f^2(x)dx)^{1/2}$.

The paper is organised as follows. In Section 2, we introduce our multivariate nonparametric madogram estimator and we discuss its properties. In Section 3, we describe the projection method based on the Bernstein polynomials and we illustrate its properties. In Section 4, we investigate by means of simulation the performance of our estimation method. Finally, we apply our approach to French weekly maxima of hourly rainfall in Section 5.

### 2 Nonparametric estimators of the multivariate Pickands dependence function

Let $X$ be a max-stable distributed random vector and let $\{X_m\}_{m=1}^n$ be independent and identically distributed (i.i.d.) replicates of it. For the rest of the paper, we allow the continuous marginal cumulative distribution functions (cdf) to be different for each coordinate and we denote them by $F_i(x) = \mathbb{P}\{X_i \leq x\}$. For comparison purposes, we briefly discuss the inference proposed by Gudendorf and Segers (2012) and we denote with HT, the estimator of Hall and Tajvidi (2000). Let $Y_{mi} = -\log F_i(X_{mi})$ be the standard exponential random variables for $m = 1, \ldots, n$ and $i = 1, \ldots, d$. Define for each $m = 1, \ldots, n$,

$$Y^*_m(w) = \bigwedge_{i=1,...,d} \left( \frac{Y_{mi}}{w_i} \right), \quad w \in S_{d-1},$$

where $\overline{Y}_{mi} = n Y_{mi}/(Y_{i1} + \cdots + Y_{ni})$. Then, the multivariate HT estimator is

$$\hat{A}_{HT}^n(w) = n/\sum_{m=1}^n Y^*_m(w)$$

(3)

Some properties of $\hat{A}_{HT}^n$ had been derived by noticing that $\hat{A}_{HT}^n(w) = \hat{A}_n^P(w)/\hat{A}_n^P(e_i)$, where $\hat{A}_n^P$ is the Pickands dependence function estimator (Gudendorf and Segers 2012). For the case of unknown margins, a weak convergence result has been shown by Gudendorf and Segers (2012). In the bivariate case $\hat{A}_{HT}^n$ satisfies conditions C2 and C3, but not C1. In the multivariate case also C2 has not been shown.

Now, we can introduce our multivariate nonparametric estimator. Basically, it is an extension of the bivariate madogram (MD) estimator introduced by Naveau et al. (2009), see also Fonseca et al. (2013).
**Definition 1.** The multivariate madogram called \( \nu(w) \) is defined as the expected distance between the componentwise maximum and the componentwise mean of the rescaled \( \{ F_{i}^{1/w_{i}}(X_{i}) \}_{i=1,...,d} \)

\[
\nu(w) = \mathbb{E} \left( \bigvee_{i=1,...,d} \{ F_{i}^{1/w_{i}}(X_{i}) \} - \frac{1}{d} \sum_{i=1,...,d} F_{i}^{1/w_{i}}(X_{i}) \right),
\]

with \( 0 < w_{i} < 1 \) and \( w_{d} = 1 - (w_{1} + \cdots + w_{d-1}) \).

**Proposition 1.** If \( X \) represents a max-stable distributed random vector with exponent measure \( V \) and Pickands dependence function \( A \), then

\[
\nu(w) = \frac{V(1/w_{1},\ldots,1/w_{d})}{1 + V(1/w_{1},\ldots,1/w_{d})} - c(w),
\]

where \( c(w) = d^{-1} \sum_{i=1}^{d} w_{i}/(1 + w_{i}) \). It follows that

\[
V(1/w_{1},\ldots,1/w_{d}) = \frac{\nu(w) + c(w)}{1 - \nu(w) - c(w)},
\]

and

\[
A(w) = \frac{\nu(w) + c(w)}{1 - \nu(w) - c(w)}.
\]

All proofs are reported in the Appendix.

**Remark 1.** One advantage of \( \nu(w) \) is that it can be interpreted as an L1-distance, \( \rho(u,v) = \mathbb{E}|u - v| \), between the maximum \( u \) and the mean \( v \), see (4). We have that \( \rho(u,v) \geq 0 \), if all components of \( X \) are equal (in probability), then the distance is null and the converse is also true. In other words, \( \nu(w) \) tells us how far away \( X \) is from the complete dependence case.

For the bivariate case, our definition of \( \nu(w) \) is slightly different from the one proposed by Naveau et al. (2009). Here, we use the bivariate vector \( \{ F_{1}^{1/w_{1}}(X_{1}) , F_{2}^{1/w_{2}}(X_{2}) \} \) instead of the bivariate couple \( \{ F_{1}^{w}(X_{1}) , F_{2}^{1-w}(X_{2}) \} \). This new version has the advantage that conditions C2 and C3 are automatically satisfied for any Pickands dependence function, this was not the case in Naveau et al. (2009).

From Equation (4), a natural estimator of the multivariate madogram is given by

\[
\hat{\nu}_{n}(w) = \frac{1}{n} \sum_{m=1}^{n} \left( \bigvee \{ F_{1}^{1/w_{1}}(X_{m,1}) , \ldots , F_{d}^{1/w_{d}}(X_{m,d}) \} - \frac{1}{d} \sum_{i=1,...,d} F_{i}^{1/w_{i}}(X_{m,i}) \right).
\]

(5)
The Pickands dependence function can be then estimated with

$$\hat{A}_n^{MD}(w) = \frac{\hat{\nu}_n(w) + c(w)}{1 - \hat{\nu}_n(w) - c(w)} , \quad w \in S_{d-1}. \quad (6)$$

The coming proposition summarizes the basic asymptotic properties of $\hat{A}_n^{MD}$.

**Proposition 2.** Let $\hat{A}_n^{MD}$ be the $w$-madogram based estimator of the Pickands dependence function. Then,

a) For any fixed $w \in S_{d-1}$, $\hat{A}_n^{MD}(w) \xrightarrow{a.s.} A(w)$ and $\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{d} N(0,\sigma^2(w))$, with an appropriate variance.

b) With known margins, $\|\hat{A}_n^{MD} - A\| \xrightarrow{a.s.} 0$ and in $C(S_{d-1})$, $\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{w} Z(w)$, $w \in S_{d-1}$, as $n \to \infty$, where $Z$ is a zero-mean Gaussian process with an appropriate covariance function.

c) In (5) if the empirical distribution functions $\hat{F}_{n,i}$, $i = 1, \ldots, d$, are used, then $\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{w} Z(w)$, $w \in S_{d-1}$, as $n \to \infty$, where $Z$ is a zero-mean Gaussian process with an appropriate covariance function.

Finally, we stress that the three conditions C1-C2-C3 are not necessarily satisfied for $\hat{A}_n^{MD}(w)$. A projection method is needed to regularize this estimator.

### 3 Estimation based on Bernstein polynomials

#### 3.1 Bernstein polynomials on the simplex $S_{d-1}$

Multivariate Bernstein polynomials, defined on the $d$-dimensional hypercube and simplex have been widely discussed in mathematics and statistics, see for example Ditzian (1986) and the references therein, and for a comprehensive description see Petrone (2004). Here our focus is on any bounded function $f(w)$ defined on the simplex $S_{d-1}$. In the univariate case, the shape features of the original function are preserved by its Bernstein polynomials approximation. For higher dimensions, shape properties like convexity may no longer be retained, e.g. Sauer (1991). The Bernstein-Bézier polynomials (Sauer 1991) solves this issue and preserves the desired shape properties (Li 2011, Lai 1993). They are defined as follows.
Definition 2. Let $k \in \mathbb{N}$ be the order of the Bernstein-Bézier polynomial and $J = \{0, \ldots, k\}$ be the index set. Given a vector $x$ of dimension $d$, we define $\Sigma_x := \sum_{i=1}^{d} x_i$. Let $\alpha_l = (j_i)_{i=1, \ldots, d-1}$, where each $j_i \in J$, is an index of $d-1$ digits. Thus, for a given index $\alpha_l$, $\Sigma_{\alpha_l}$ defines the sum of its digits. For any $k$ and $d-1$, each $\alpha_l$ is an ordered selection with repetition of elements in $J$. Denote by $\alpha = \{\alpha_l\}_{l \in L}$ the set of all such vectors, with index set $L = \{1, \ldots, D_{k,d-1}\}$ of cardinality $D_{k,d-1} = (k + 1)^{d-1}$.

We define the $l$th Bernstein basis polynomial of degree $k$ as the following continuous function with values in $[0,1]$,

$$ b_l(w; k) := \binom{k}{\alpha_l} w^{\alpha_l} (1 - \Sigma w)^{k-\Sigma_{\alpha_l}}, \quad w \in S_{d-1} $$

where

$$ \binom{k}{\alpha_l} = \frac{k!}{\alpha_l!(k-\Sigma_{\alpha_l})!}, \quad \alpha_l! = j_1! \cdots j_{d-1}!, \quad w^{\alpha_l} = \prod_{i=1}^{d-1} w_i^{j_i}. $$

Therefore, the Bernstein-Bézier polynomial representation of the function $f$ is given by

$$ B_f(w; k) = \sum_{l \in L_k} \beta_l b_l(w; k), \quad w \in S_{d-1}, \quad (7) $$

where $\beta_l \in \mathbb{R}$, for $l \in L_k$ and $L_k = \{l \in L : \Sigma_{\alpha_l} \leq k\} \subset L$ are $f$-dependent coefficients. The number of coefficients involved in (7) is denoted by $p$ and it is determined as follows.

Proposition 3. Given $d-1 \in \mathbb{N}$ and a polynomial degree $k \in \mathbb{N}$, the number of coefficients in (7) are equal to $p$, where

$$ p = \begin{cases} k + 1 & \text{if } d = 2, \\ \frac{1}{k+1} \sum_{m=1}^{k+1} (k + 2 - m) & \text{if } d = 3, \\ \frac{1}{k+1} \sum_{m=1}^{k+1} \left( d - 1 + m - 4 \right) \frac{(k + 2 - m)^2 + (k + 2 - m)}{2} & \text{if } d \geq 4. \end{cases} \quad (8) $$

Note that $p$ depends on $k$ and $d-1$, however, for simplicity we will not write such dependence explicitly. Polynomial (7) can be expressed in matrix form, $B_f(w; k) = b_k(w) \beta_k$, for any $k = 1, 2, \ldots$, where $b_k(w)$ and $\beta_k$ are the $p$-dimensional row and column vectors of polynomial bases and coefficients, respectively.
Remark 2. If $d = 2$, then $\alpha_i = (j), \; j \in J$ and $\Sigma_{\alpha_i} = j$. Therefore the $l^{th}$ Bernstein basis polynomial of degree $k$ simplifies to

\[ b_l(w; k) = \binom{k}{j} w^j (1 - w)^{k-j}, \quad w \in [0, 1] \]

and the polynomial representation of $f$ becomes simply

\[ B_f(w; k) = \sum_{j=0}^{k} \beta_j b_j(w; k), \quad w \in [0, 1] \]

Additionally, if $f(w) \in C(S_{d-1})$, then $B_f(w, k)$ converges uniformly to $f$ as $k$ goes to infinity and $\|B_f(w, k) - f(w)\|_\infty = O(k^{-1})$ (Li 2011).

3.2 Shape-preserving estimator

In this section, we describe how to use the Bernstein-Bézier polynomial functional representation of the Pickands dependence function in order to obtain a projection estimator (Fils-Villetard et al. 2008) that satisfies C1-C2-C3. This method can be applied to nonparametric estimators such as those discussed in Section 2, as well as others (e.g., Capéraà et al. 1997, Zhang et al. 2008). The projection estimator based on a first guess, say $\hat{A}_n$, is obtained as the solution of the optimization problem

\[ \hat{A}_n = \arg \min_{A \in \mathcal{A}} \|\hat{A}_n - A\|_2, \]

where the minimum is taken among all the functions in $\mathcal{A}$, the latter is a closed and convex subset of $L^2(S_{d-1})$. There is no closed form to the above equation, and so an approximation based on the sieves method (Geman and Hwang 1982) is explored. Consider a nested sequence $\mathcal{A}_k \subseteq \mathcal{A}$ of constrained multivariate Bernstein-Bézier polynomial families

\[ \mathcal{A}_k = \{ B_A(w; k) = b_k(w) \beta_k : R_k \beta_k \geq r_k \}, \quad (9) \]

where $\beta_k \in B_k \subseteq B \subseteq \mathbb{R}^p$ with $B$ and $B_k$ closed and convex. $R_k = [R_k^{(1)}, R_k^{(2)}, R_k^{(3)}]^\top$ and $r_k = [r_k^{(1)}, r_k^{(2)}, r_k^{(3)}]^\top$ are a $(q \times p)$ full row rank matrix and a $(q \times 1)$ vector respectively such that the restriction on $B$ given by $R_k \beta_k \geq r_k$ ensures that each member of $\mathcal{A}$ satisfies all conditions C1-C2-C3. For each condition, the details for deriving the block matrices and vectors of restrictions are provided below.
A sufficient condition to guarantee that the function $B_A(w; k)$, with $w$ varying in $S_{d-1}$, is convex is that the Hessian matrix be positive semi-definite. In order to show this, we resort to the following relations. First, for any two direction vectors $e_t, e_q, t \neq q \in \{1, \ldots, d-1\}$, the directional gradient of $B_A$ with respect to $e_t e_q$ is

$$\nabla_{tq} B_A(w; k) = k \sum_{l \in L_{k-1}} \Delta_{t,q} \beta_l b_l(w; k), \quad w \in S_{d-1}$$

where

$$\Delta_{t,q} \beta_l = (\beta_{lt} - \beta_{lq})$$

and $l_t, l_q \in L_k$. Second, for any direction $q = \sum_{i \in \{1, \ldots, d-1\}} \eta_i (e_i - 0)$ for some $\eta = (\eta_1, \ldots, \eta_{d-1})^T$,

$$\nabla_{q}^2 B_A(w; k) = k(k-1) \sum_{l \in L_{k-2}} \eta^T Q_l \eta b_l(w; k), \quad w \in S_{d-1},$$

where, for all $l \in L_{k-2}$, $Q_l$ is a symmetric $(d-1 \times d-1)$ matrix with form

$$Q_l = \begin{pmatrix}
\Delta^2_{1,0} \beta_t & \Delta_{1,0} \Delta_{2,0} \beta_t & \cdots & \cdots & \Delta_{1,0} \Delta_{d-1,0} \beta_l \\
\Delta^2_{2,0} \beta_t & \Delta_{2,0} \Delta_{3,0} \beta_t & \cdots & \cdots & \Delta_{2,0} \Delta_{d-1,0} \beta_l \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\Delta^2_{d-1,0} \beta_t & \cdots & \cdots & \cdots & \Delta^2_{d-1,0} \beta_l
\end{pmatrix}.$$  

Then, by the weak diagonal dominance criterion (Lai 1993) in order to guarantee that $Q_l$ is positive semi-definite, it is sufficient to check, for all $l \in L_{k-2}$ and $t \in \{1, \ldots, d-1\}$, the conditions

$$\Delta^2_{t,0} \beta_l - \sum_{t \neq q \in \{1, \ldots, d-1\}} |\Delta_{q,0} \Delta_{t,0} \beta_l| \geq 0.$$  

This can be synthesized in matrix form $R_k(1) \beta_k \succeq r_k(1)$ when $R_k(1)$ is a $((d - 1)^2 N_{L_{k-2}}, d)$-dimensional matrix and $r_k(1)$ is the corresponding null vector. For
example, with $d = 3$ and $k = 3$,

\[
R(1)_3 = \begin{pmatrix}
0 & 1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 \\
2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 & 0 \\
0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 \\
2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 \\
0 & 2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 \\
0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 \\
0 & 2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\
0 & 0 & 0 & 0 & 2 & -1 & 0 & -3 & 1 \\
0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 2 & -3 & 1 & -1 & 1 
\end{pmatrix}, \quad r^{(1)}_3 = \begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}.
\]

R2) The lower limit for $B_A$ is guaranteed if for all $w \in S_{d-1}$ the condition $b_k(w)\beta_k \geq \max(w_1, \ldots, w_{d-1}, 1 - (w_1 + \cdots + w_{d-1}))$ holds. For example, for a finite set \{w_s, s = 1, \ldots, t\}, $w_s \in S_{d-1}$, the $(t \times p)$ matrix and $p$-dimensional vector of restrictions are equal to

\[
R_k^{(2)} = \begin{pmatrix} b_k(w_1) \\ \vdots \\ b_k(w_t) \end{pmatrix}, \quad r_k^{(2)} = \begin{pmatrix} \lor_{i=1}^{d}(w_{1i}) \\ \vdots \\ \lor_{i=1}^{d}(w_{ti}) \end{pmatrix}.
\]

R3) The restrictions on the vertices are guaranteed if $\beta_l = 1$ for the set of coefficients $\{\beta_l, l \in L_k : \alpha_l = 0 \text{ or } \alpha_l = k e_i, \forall i = 1, \ldots, d - 1\}$. Therefore, the $(2d \times p)$ matrix and $2d$-dimensional vector of restrictions are equal to

\[
R_k^{(3)} = \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
-1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & -1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
0 & 0 & \cdots & 0 & -1 & \cdots & 0
\end{pmatrix}, \quad r_k^{(3)} = \begin{pmatrix}
1 \\
-1 \\
1 \\
-1 \\
\vdots \\
1 \\
-1
\end{pmatrix}.
\]

Finally, the approximate projection estimator is given by the solution to the equation

\[
\tilde{A}_{n,k} = \arg \min_{B_A \in A_k} \| \tilde{A}_n - B_A \|_2.
\]

In practice, the estimator $\tilde{A}_{n,k}$ can be only evaluated on a finite set of points \{w_s, s = 1, \ldots, t\}, with $t \in \mathbb{N}$ and $w_s \in S_{d-1}$. Therefore, an approximation of the above solution is given by

\[
\tilde{A}_{n,k}(w_s) = b_k(w_s)\tilde{\beta}_k, \quad w_s \in S_{d-1}, \quad s = 1, \ldots, t, \quad (10)
\]
where $\hat{\beta}_k$ is the minimizer of the constrained least-squares problem

$$\hat{\beta}_k = \arg\min_{\beta_k \in B_k} \frac{1}{t} \sum_{s=1}^{t} (b_k(w_s)\beta_k - \hat{A}_n(w_s))^2 \quad \text{s.t.} \quad R_k\beta_k \geq r_k.$$ 

This is a quadratic programming problem, which solution is

$$\hat{\beta}_k = \beta'_k - (b_k^\top b_k)^{-1}r_k^\top \gamma$$

where $\gamma$ is a vector of Lagrange multipliers and $\beta'_k = (b_k^\top b_k)^{-1}b_k^\top \hat{A}_n$ is the unconstrained least square estimator. $\hat{\beta}_k$ and $\gamma$ can be efficiently computed with an iterative quadratic programming algorithm (e.g. Goldfarb and Idnani 1983). Clearly, the accuracy of the approximate solution (10) improves with increasing values of $m$.

In practice, estimates are obtained using the R package quadprog (Turlach and Weingessel 2010).

The asymptotic distribution of the Bernstein-projection estimator based on our multivariate madogram estimator $\hat{A}_n^{MD}$ is established in the following.

**Proposition 4.** Assume that the polynomial degree $k$ increases with the sample size $n$, so we write $k_n$. If $k_n$ goes to infinity satisfying the condition $k_n/n^{1/2} \to \infty$ as $n \to \infty$, then for $w \in S_{d-1}$

$$\sqrt{n}(\hat{A}_{n,k_n}^{MD}(w) - A(w)) \xrightarrow{w} \arg\min_{Z' \in T_A(A)} \|Z' - Z\|_2, \quad n \to \infty,$$

where $T_A(A)$ is the tangent cone of $A$ at $A$, given by the set of limits of all the sequences $a_n(A_n - A)$, $a_n \geq 0$ and $A_n \in A$.

This results from the work of Fils-Villetard et al. (2008). In particular, from Proposition 3 it follows that $\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{w} Z(w)$ in $L^2(S_{d-1})$ as $n \to \infty$ where $Z$ is a Gaussian process. Then, by applying Theorem 1 in Fils-Villetard et al. (2008) it follows that

$$\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{w} \arg\min_{Z' \in T_A(A)} \|Z' - Z\|_2, \quad n \to \infty$$

From Li (2011) we have that $\|B_A(w; k) - A(w)\|_\infty = O(k^{-1})$ as $k \to \infty$. Hence, by the hypothesis on $k_n$ and applying Lemma 1 in Fils-Villetard et al. (2008) we have that $\|\hat{A}_{n,k_n}^{MD} - \hat{A}_n^{MD}\| = o_p(n^{1/2})$. Therefore, $\sqrt{n}(\hat{A}_{n,k_n}^{MD}(w) - A(w)) = \sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) + o_p(1)$ as $n \to \infty$ and the asymptotic distribution of the approximate Bernstein-projection estimator follows from that of our madogram estimator.
3.3 Confidence bands

We construct confidence bands using classical bootstrap methods. For \( w \in S_{d-1} \) and \( 0 < \alpha < 1 \), the bootstrap \( (1 - \alpha) \) pointwise confidence band, based on the estimates \( \tilde{A}^{(r)}_{n,k}(w) \), \( r = 1, 2, \ldots \), obtained from the bootstrapped sample \( X_n = (X_1, \ldots, X_n) \), has the drawback that the lower and upper limits of the band are rarely convex and continuous. To bypass this hurdle, we followed the strategy to work with the estimated Bernstein polynomials’ coefficients themselves. Specifically, let \( \hat{\beta}^{(r)}_{k} \) be the Bernstein polynomials’ coefficient estimator based on the bootstrap sample \( X^{(r)}_n \), \( r = 1, 2, \ldots \), we define a bootstrap simultaneous \( (1 - \alpha) \) confidence band specifying the lower \( \tilde{A}^{(r)}_{L,n,k}(w) \) and upper \( \tilde{A}^{(r)}_{U,n,k}(w) \) limits as

\[
\left[ \sum_{l \in L_k} \hat{\beta}^{(r)[\alpha/2]}_{l} b_l(w; k); \sum_{l \in L_k} \hat{\beta}^{(r)[1-\alpha/2]}_{l} b_l(w; k) \right], \quad w \in S_{d-1},
\]

where \( \hat{\beta}^{(r)[\alpha/2]}_{l} \) and \( \hat{\beta}^{(r)[1-\alpha/2]}_{l} \), for all \( l \in L_k \), correspond to the \( [r(\alpha/2)] \) and \( [r(1-\alpha/2)] \) ordered statistics respectively and \( b_l(w; k) \) is \( l^{th} \) Bernstein basis polynomial of degree \( k \), see (7). This approach, although does not guarantee convex confidence bands, it works very well in our simulations, where we find that the convexity is not satisfied only for very weak dependences. This point will be illustrated in the coming section.

4 Simulations

To visually illustrate the gain in implementing our Bernstein-Bézier projection approach, Figure 2 compares the madogram (MD) estimator \( \tilde{A}^{MD}_{n} \) defined by (6) with its Bernstein-projection (BP) version defined by (10) for the special case of the symmetric logistic model with \( d = 3 \) and \( \alpha = 0.3 \) (Tawn 1990). For all sample sizes \( (n = 20, 50, 100) \), an improvement can be observed by comparing the estimated contour lines (dotted) in the top and bottom panels. This is particularly true for a small sample size like \( n = 20 \), the corrected version provides much smoother and realistic contour lines.

To guarantee a good approximation of \( A \) with \( \tilde{A}_{n,k} \), Proposition 4 suggested to set a large polynomial degree \( k \) for large sample sizes, see also Fils-Villetard et al. (2008), Gudendorf and Segers (2011), Gudendorf and Segers (2012). But computational time
limits restrict the choice of $k$. Figure 3 explores this issue for the logistic model with $\alpha = 3$ and $n = 100$. As expected from the theory, the choice of $k$ is not anecdotal. A shift in the contour lines appears for the small value $k = 5$, see the left panel of Figure 3. This undesirable feature disappears for a moderate value of $k$, see the right panel with $k = 14$.

To go beyond these visual checks, we also compute the mean integrated squared error

$$\text{MISE}(\hat{A}_n, A) = \mathbb{E} \left\{ \int_{S_{d-1}} \left( \hat{A}(w) - A(w) \right)^2 \, dw \right\},$$

for a variety of setups. The MISE is obtained by repeating 1000 times a given inference method for three different sample sizes $n = 50, 100, 200$. To explore the influence of the dependence strength on the choice of $k$, the value of $\alpha$ in the logistic model takes three values, 0.3, 0.6 and 0.9, i.e. strong, mild and weak dependence, respectively. Table 1 compares four estimators: the MD defined (4), its Bernstein-projection (BP) version defined by (10), and HT defined by (3) and its BP version. Note that each

\[ \begin{array}{ccc}
 n = 20 & n = 50 & n = 100 \\
 \text{MD estimates} & \text{BP-MD estimates} & \text{BP-MD estimates}
\end{array} \]

Figure 2: Inference (dashed lines) of the Pickands dependence function for the symmetric logistic model with $\alpha = 0.3$ (solid line) with two different estimators: MD defined (4) and its MD Bernstein-projection (BP) version defined by (10) with $k = 14$, see each row. Each column represents a different sample size, $n = 20, 50$ and 100 from left to right.
Figure 3: Same as Figure 2 but for a fixed $n = 100$ and three different values of $k = 5, 11, 14$ for the MD-BP estimator defined by (10).

MISE here corresponds to an “optimal” value of $k$, that is the $k$ value chosen in a such way that the MISE does not decrease significantly for larger values of $k$. Systematically, Table 1 indicates that the lowest MISE values are obtained for the MD-BP method. The second best is the HT-BP and this confirms that our Berstein-Beziers projection approach improves the inference. Concerning the value of $k$, a strong dependence like $\alpha = 0.3$ implies a larger $k$, compared to $\alpha = 0.9$. This makes sense if we view a dependence structure as an added complexity, especially with respect to the independence case, the simplest possible model. In such a framework, the polynomial degree has to be higher to capture this extra information.

To explore the validity of our procedure to derive a bootstrap simultaneous $(1 - \alpha)$ confidence band given by (12), Table 2 displays 95% coverage probabilities from 1000 independent samples and $r = 500$ bootstrap resampling. The parametric setup is identical to the one used in Table 1 but the sample size is equal to $n = 50$ instead of $n = 100$. Again, the BP-MD estimator appears to slightly outperform the BP-HT. The coverage probabilities are closer to the true value (0.95), for the strong dependence case ($\alpha = 0.3$) that needs a large polynomial degree ($k = 17$).

To close this small simulation study, we increase the dimension to $d = 4$ and we extend the class of parametric families to the asymmetric logistic (AL, Tawn 1990) with $\theta = 0.6$, $\phi = 0.3$, $\psi = 0$, the Hüsler–Reiss model (HR, Hüsler and Reiss 1989)\(^2\)

\(^1\)The case $d = 2$ has also been considered. The results have been omitted for brevity, since they arrive at the same conclusion.

\(^2\)Tables like Table 1 and 2 are available upon request for the HR and EST families and brings the same overall message.
Table 1: MISE defined by (4) for the logistic model with three values of $\alpha$ (strong, mild and weak dependence, see first column). The second column lists four estimators: MD defined (4), its Bernstein-projection (BP) version defined by (10), and HT defined by (3) and its BP version. The third column reports the polynomial degree used with our BP method.

with three cases ($\gamma_1 = 0.8, \gamma_2 = 0.3, \gamma_3 = 0.7$), ($\gamma_1 = 0.49, \gamma_2 = 0.51, \gamma_3 = 0.03$), ($\gamma_1 = 0.24, \gamma_2 = 0.23, \gamma_3 = 0.11$) and the extremal skew-t (EST, Padoan 2011) with three setups ($\alpha = 7, -10, 1$, $\nu = 3, \omega = 0.9$), ($\alpha = -2, 9, -15$, $\nu = 2, \omega = 0.9$),

Table 2: 95% coverage probabilities from (12) with 1000 independent samples and $r = 500$ bootstrap resampling. The parametric setup is identical to the one used in Table 1 but the sample size is equal to $n = 50$ instead of $n = 100$. 

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimator</th>
<th>Poly’s degree</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td></td>
<td>$k$</td>
<td>50</td>
</tr>
<tr>
<td>0.3</td>
<td>HT</td>
<td>5.71 x 10^{-4}</td>
<td>5.27 x 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>MD</td>
<td>4.94 x 10^{-4}</td>
<td>2.70 x 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>BP-HT</td>
<td>6.77 x 10^{-5}</td>
<td>6.39 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>BP-MD</td>
<td>5.66 x 10^{-5}</td>
<td>4.68 x 10^{-5}</td>
</tr>
<tr>
<td>0.6</td>
<td>HT</td>
<td>3.80 x 10^{-3}</td>
<td>2.85 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>MD</td>
<td>2.73 x 10^{-3}</td>
<td>1.36 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>BP-HT</td>
<td>7.15 x 10^{-4}</td>
<td>5.08 x 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>BP-MD</td>
<td>6.74 x 10^{-4}</td>
<td>3.53 x 10^{-4}</td>
</tr>
<tr>
<td>0.9</td>
<td>HT</td>
<td>6.98 x 10^{-3}</td>
<td>4.32 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>MD</td>
<td>4.81 x 10^{-3}</td>
<td>3.17 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>BP-HT</td>
<td>1.57 x 10^{-3}</td>
<td>1.04 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>BP-MD</td>
<td>1.12 x 10^{-3}</td>
<td>5.39 x 10^{-4}</td>
</tr>
</tbody>
</table>
Figure 4: Estimates of Pickands dependence function for $d = 4$ (light grey shade) and bootstrap variability bands (dark grey shade) for the SL, AL, HR, EST (left-right) models with strong, mild and weak dependence (top-bottom) ($\alpha = -0.5, -0.5, -0.5$, $\nu = 3$, $\omega = 0.9$). Figure 4 shows that, for all these cases, the lower and upper limits of the variability bands are always convex functions and they always contain the true Pickands dependence function. The variability bands of weaker dependence structures are typically wider than those of stronger dependence structures. The same is true for asymmetric versus symmetric dependence structures.

5 Weekly maxima of hourly rainfall in France

Coming back to Figure 1 introduced in Section 1, our goal here is to measure the dependence within each cluster of size $d = 7$. The clusters were obtained by running the algorithm proposed by Bernard et al. (2013) on weekly maxima of hourly rainfall recorded in the Fall season from 1993 to 2011, i.e. $n = 228$ for each station. Climatologically, extreme precipitation that affect the Mediterranean coast in the Fall are caused by the interaction of southern and mountains winds coming from the Pyrénées, Cévennes and Alps regions. In the north of France, heavy rainfall is often produced
by mid-latitude perturbations in Brittany or in the north of France and Paris.

For each cluster, we compute our Bernstein-projection estimator based on the madogram. To summarize this seven-dimensional dependence structure, we take advantage of the extremal coefficient (Smith 1990) defined by

$$\theta = d A(1/d, \ldots, 1/d).$$

It satisfies the condition $1 \leq \theta \leq d$, where the lower and upper bound represent the case of complete dependence and independence among the extremes, respectively. Thus, in each cluster the extremal coefficient is estimated using the equation $\hat{\theta} = 7 \tilde{A}_{n,k}^{MD}(1/7, \ldots, 1/7)$ and so $\hat{\theta}$ always belongs to the interval $[1, 7]$.

As climatologically expected, we can see in Figure 1 a latitudinal gradient in the estimated extremal coefficients. They are smaller in the northern regions and higher in the south. This can be explained by westerly fronts above the $46^\circ$ latitude that affect large regions, while extreme precipitation in the south is more likely driven by localised convective storms with weak spatial dependence structures.

Acknowledgements

We thank Sonia Petrone for her useful suggestions. The work of Philippe Naveau has been supported by the French National Research Agency (ANR-09-RISK-007, MOPERA, ANR McSim), the LEFE-Multirisk and the Extremoscope projects. The authors acknowledge Meteo France for precipitation time series. The authors would also like very much to credit the contributors of the R project and LaTeX.

Appendix A.1: Proofs

Proof of Proposition 1

By the definition of the $w$-madogram (4) it follows that

$$\nu(w) = \mathbb{E}(M_d) - \mathbb{E} (\bar{U}_d),$$

where $M_d = \max\{F^{1/w_1}(X_1), \ldots, F^{1/w_d}(X_d)\}$, $\bar{U}_d = d^{-1} \sum_{i=1}^{d} F^{1/w_i}(X_i)$, $0 < w_i < 1$ with $i = 1, \ldots, d$. The expression of $\nu(w)$ follows by noting that

$$\mathbb{E}\{M_d\} = \frac{V(1/w_1, \ldots, 1/w_d)}{1 + V(1/w_1, \ldots, 1/w_d)}, \quad \mathbb{E}\{F^{1/w_i}(X_i)\} = \frac{w_i}{1 + w_i}, \quad i = 1, \ldots, d,$$
where the left-hand side is easily obtained knowing the distribution of $M_d$ that is

$$\mathbb{P}\{M_d \leq u\} = u^{V(1/w_1,\ldots,1/w_d)},$$

whereas the right-hand side is obtained from the facts that $U = F(X_i)$ is uniformly distributed in $[0, 1]$ and $\mathbb{E}\{U^{1/a}\} = a/(1 + a)$. Putting this all together we obtain

$$\nu(w) = \frac{V(1/w_1,\ldots,1/w_d)}{1 + V(1/w_1,\ldots,1/w_d)} - \frac{1}{d} \sum_{i=1}^{d} \frac{w_i}{1 + w_i}.$$

Finally, the expression of $V(1/w_1,\ldots,1/w_d)$ is simply obtained inverting the above expression.

**Proof of Proposition 2**

For $w \in S_{d-1}$, define the function $\nu_w : [0, 1]^d \to [0, 1]$ by

$$\nu_w(u) = \bigvee_{i=1,\ldots,d} (u_i^{1/w_i}) - \frac{1}{d} \sum_{i=1}^{d} u_i^{1/w_i}, \quad u \in [0, 1]^d.$$

Then, the Pickands dependence function estimator based on the $w$-madogram is defined by

$$\hat{A}_n^{MD}(w) = g_w(\hat{\nu}_n(w)), \quad \hat{\nu}_n(w) = \frac{1}{n} \sum_{m=1}^{n} \nu_w(U_m),$$

where for $w \in S_{d-1}$, $g_w : [0, 1] \to [0, 1]$ is defined by $g_w(\nu_w(u)) = (\nu_w(u) + c(w))/(1 - \nu_w(u) - c(w))$ with $c(w) = \sum_{i=1}^{d} w_i/(1 + w_i)$, $U_m$ is the $m$th i.i.d. copy of the random vector $U = (F_1(X_1),\ldots,F_d(X_d))$. The three statements are results of the following facts.

a) For any fixed $w \in S_{d-1}$, for the strong law of large numbers and the central limit theorem, then $\hat{\nu}_n(w) \xrightarrow{a.s.} \nu(w)$ and $\sqrt{n}(\hat{\nu}_n(w) - \nu(w)) \xrightarrow{d} N(0, \hat{\sigma}^2(w))$, as $n \to \infty$, where $\hat{\sigma}^2(w) = \mathbb{E}[\nu_w(U)^2] - \nu(w)^2$. Since $g_w$ is continuous and differentiable, applying the continuous mapping theorem and the delta-method (van der Vaart 2000, Ch. 3), it follows that $\hat{A}_n^{MD}(w) \xrightarrow{a.s.} A(w)$ and $\sqrt{n}(\hat{A}_n^{MD}(w) - A(w)) \xrightarrow{d} N(0, \sigma^2(w))$, where $\sigma^2(w) = \hat{\sigma}^2(w)/(1 - \nu(w) - c(w))^4$.

b) The map $w \mapsto \nu_w(u)$ is continuous in $w \in S_{d-1}$ for each $u \in [0, 1]^d$, so the results in a) are also valid in the functional sense. Applying Theorem 1 of Jain
and Marcus (1975), then \(\sqrt{n}(\hat{\nu}_n(w) - \nu(w)) \to^w Z\) as \(n \to \infty\), where \(Z\) is a Gaussian process in \(S_{d-1}\) with covariance function

\[
\text{Cov}(w, v) = \mathbb{E}[\nu_w(U)\nu_v(U)] - \nu(w)\nu(v), \quad w, v \in S_{d-1}.
\]

However, we need to show as in Lemma 2.1 of Deheuvels (1991) that

\[
|\nu_w(U) - \nu_v(U)| \leq Y \rho(w, v), \quad w, v \in S_{d-1}, \tag{14}
\]

for a random variable \(Y\) such that \(\mathbb{E}[Y^2] < \infty\) and a metric \(\rho\) on \(S_{d-1}\). Applying the definition of \(\nu_w(u)\) it can be verified that

\[
|\nu_w(U) - \nu_v(U)| \leq \Bigg| \bigvee_{i=1,\ldots,d} (U_i^{1/w_i} - U_i^{1/v_i}) \Bigg| + \frac{1}{d} \sum_{i=1}^d (U_i^{1/w_i} - U_i^{1/v_i}) \leq \bigvee_{i=1,\ldots,d} (|U_i^{1/w_i} - U_i^{1/v_i}|) + \frac{1}{d} \sum_{i=1}^d |U_i^{1/w_i} - U_i^{1/v_i}|.
\]

For any \(i \in I\), \(|w_i^* - v_i^*| \geq |u_i^{w_i^-} - u_i^{v_i^-}|\), where \(w_i^* = 1/w_i, v_i^* = 1/v_i \geq 1\) and \(u_i \in [0,1]\). Indeed, applying the relation \(|a - b| = 2 \max(a, b) - (a + b)\) we obtain the inequality \(\max(w_i^* - u_i^{w_i^-}, v_i^* - u_i^{v_i^-}) \geq \{(w_i^* - u_i^{w_i^-}) + (v_i^* - u_i^{v_i^-})\}/2\). Therefore,

\[
|\nu_w(U) - \nu_v(U)| \leq \bigvee_{i=1,\ldots,d} (|1/w_i - 1/v_i|) + \sum_{i=1}^d |1/w_i - 1/v_i|,
\]

and so (14) follows with \(Y \equiv 1\) and \(\rho(w, v) = \Delta_\infty(w, v) + d_1(w, v)\). Since the map \(g_w : C(S_{d-1}) \to C(S_{d-1}) : \tilde{\nu} \mapsto g_w(\tilde{\nu})\) is continuous and differentiable, the weak convergence of \(\sqrt{n} (\hat{A}_{\nu}^{MD}(w) - A(w))\) is obtained applying the functional delta-method (van der Vaart 2000, Ch. 20). In conclusion, applying Theorem 2.2 in Deheuvels (1991), which is also true in our case since \(\|\hat{\nu}_n\|_{\infty} \leq 1\), it follows that \(\|\hat{\nu}_n - \nu\|_{\infty} \to^{a.s.} 0\) as \(n \to \infty\) and for the continuity of the map \(g_w\) also \(\|\hat{A}_{\nu}^{MD} - A\|_{\infty} \to^{a.s.} 0\) holds.

c) In (13) if the continuous marginal c.d.f.s are estimated by the empirical distribution functions \(\hat{F}_{n,i}, i = 1, \ldots, d\), then the \(w\)-madogram estimator becomes

\[
\hat{\nu}_n(w) = \frac{1}{n} \sum_{m=1}^n \nu_w(U_m), \quad \hat{U}_m = (\hat{U}_{m,1}, \ldots, \hat{U}_{m,d}),
\]
where
\[
\hat{U}_{t,i} = \hat{F}_{n,i}(X_{t,i}) = \frac{1}{n+1} \sum_{m=1}^{n} \mathbb{I}(X_{m,i} \leq X_{t,i}).
\]

We denote with $C$ the cdf of the random vector $\mathbf{U}$ and with $\hat{C}_n$ the empirical copula
\[
\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}(\hat{U}_{m,1} \leq u_1, \ldots, \hat{U}_{m,d} \leq u_d), \quad \mathbf{u} \in [0,1]^d
\]

Under appropriate conditions (Fermanian et al. 2004; Segers 2012; Gudendorf and Segers 2012) $\sqrt{n}(\hat{C}_n - C) \xrightarrow{w} Z$ as $n \to \infty$ in $\ell^\infty([0,1]^d)$, where $Z$ is a zero-mean Gaussian process defined as
\[
Z(\mathbf{u}) = Z'(\mathbf{u}) - \sum_{i=1}^{d} Z'(1, \ldots, u_i, \ldots, 1) \frac{\partial C}{\partial u_i}(\mathbf{u}), \quad \mathbf{u} \in [0,1]^d
\]

and where $Z'$ is a Brownian bridge on $[0,1]^d$ with covariance function
\[
\text{Cov}(\mathbf{u}, \mathbf{v}) = C(\mathbf{u} \land \mathbf{v}) - C(\mathbf{u})C(\mathbf{v}) \quad \mathbf{u}, \mathbf{v} \in [0,1]^d,
\]

where $(\mathbf{u} \land \mathbf{v}) = \{\min(u_i, v_i), i = 1, \ldots, d\}$. Now, $\nu_w$ is of bounded variation on $[0,1]^d$. Indeed, for all $w \in S_{d-1}$ the functions $u_i^{1/w_i}$, $i = 1, \ldots, d$, are of bounded variation since they are monotone. Thus, of bounded variation are also the maximum and the sum of those functions and therefore it is so for the difference between them. Since $\nu_w$ is a continuous function then applying Theorem 6 of Fermanian et al. (2004) (see also Naveau et al. 2009) the $\mathbf{w}$-madogram estimator defined as
\[
\sqrt{n} [\hat{\nu}_n(\mathbf{w}) - \mathbb{E} \{\nu_w(\mathbf{U})\}], \quad \mathbf{w} \in S_{d-1}
\]

converges weakly to $\int_{[0,1]^d} Z(\mathbf{u}) \, d\nu_w(\mathbf{u})$ in $\ell^\infty(S_{d-1})$. Finally, since $g_w$ is continuous and differentiable, the weak convergence of $\hat{A}$ is obtained applying the functional delta-method.

**Proof of Proposition 3**

$p = \text{card}(L_k)$. In order to determine $p$ we organize the $D_{k,d-1}$ elements of $\mathbf{a}$ in $(k+1)^{d-1-\min(d-1,2)}$ matrices of size $(k+1)^{\min(d-1,2)}$. The indexes of $L_k$ correspond to those elements $\mathbf{a}_l$ of the $(r \times c)$ matrices such that
\[
r + c \leq \begin{cases} k + 2 & \text{if } d - 1 = 1, \\ k + 2 - m & \text{if } d - 1 \geq 2, \end{cases}
\] (15)
where $m = 0, \ldots, k$. The number of matrices with at least one element that satisfies the inequality (15) is, for $d - 1 = 1, 2$ equals $k + 1$ and for $d - 1 \geq 3$ described by the following Pascal’s triangle

\[
\begin{array}{cccc}
  & & & k + 1 \\
  & d - 1 = 3 & & \\
  & d - 1 = 4 & (0) & k \\
  & d - 1 = 5 & (1) & (1) \\
  & & & k - 1 \\
\end{array}
\]

\[
\ddots
\]

\[
\begin{array}{cccc}
  (2) & (2) & (2) & \ddots \\
  (3) & (2) & (3) & (3) \\
  (3) & (3) & (3) & (3) \\
\end{array}
\]

Therefore, we distinguish three cases.

1. If $d - 1 = 1$, then $\boldsymbol{\alpha}_t = \{j_1\}, j_1 \in J$ and hence we have $k + 1$ matrices of one element all satisfying inequality (15).

2. If $d - 1 = 2$, then $\boldsymbol{\alpha}_t = \{j_1, j_2\}, j_i \in J, i = 1, 2$ and hence we have $k + 1$ matrices with size $r = 1$ and $c = 1, \ldots, k + 1$. The elements of these vectors, satisfying inequality (15), changes according to $m$, that is

\[
p = \sum_{m=1}^{k+1} \sum_{c=1}^{k+1} \mathbb{I}_{\{1+c\leq k+3-m\}} = \sum_{m=1}^{k+1} (k+2-m).
\]

3. If $d - 1 \geq 3$, then $\boldsymbol{\alpha}_t = \{j_i\}_{i=1,\ldots,d}, j_i \in J$ and hence we have $(k+1)^{d-1-\min(d-1,2)}$ matrices of size $r, c = 1, \ldots, (k + 1)$. Given $k$ and $d - 1$, the number of matrices with at least one element satisfying inequality (15) is the sum of the binomial coefficients in the Pascal’s triangle. Each of those coefficients is the number of matrices that have at most the elements of the secondary diagonal and its upper triangular part that satisfies the inequality (15), these are $[(k + 2 - m)^2 + (k + 2 - m)]/2$ elements, $m = 1, \ldots, k + 1$. Specifically,

\[
p = \sum_{m=1}^{k+1} \binom{d - 1 + m - 4}{m - 1} \sum_{r=1}^{k+1} \sum_{c=1}^{k+1} \mathbb{I}_{\{r+c\leq k+3-m\}}
\]

\[
= \sum_{m=1}^{k+1} \binom{d - 1 + m - 4}{m - 1} \frac{(k+2-m)^2(k+1-m)}{2} + (k+2-m).
\]
References


