Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials

G. Marcon, S.A. Padoan, P. Naveau, P. Muliere, J. Segers

Multivariate Nonparametric Estimation of the Pickands Dependence Function using Bernstein Polynomials

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

April 18, 2016

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 Pickands dependence function of the underlying extreme-value copula. A nonparametric estimator is constructed as the sample equivalent of a multivariate extension of the madogram. Shape constraints on the family of Pickands dependence functions are taken into account by means of a representation in terms of a specific type of Bernstein polynomials. The large-sample theory of the estimator is developed and its finite-sample performance is evaluated with a simulation study. The approach is illustrated by analyzing clusters consisting of seven weather stations that have recorded weekly maxima of hourly rainfall in France from 1993 to 2011.

Keywords: Bernstein polynomials, Extremal dependence, Extreme-value copula, Heavy rainfall, Nonparametric estimation, Multivariate max-stable distribution, Pickands dependence function.

1 Introduction and background

In recent years, inference methods for assessing the extremal dependence have been in increasing demand. This is especially due to growing requests for multivariate analyses of extreme values in the fields of environmental and economic sciences. The dimension of the random vector under study is often greater than two. For example, Figure 1 displays a map of clusters containing seven weather stations in France each; see Bernard, Naveau, Vrac, and Mestre (2013) for details on the construction of the clusters. The data consist of weekly

*Marcon is Post-doc at university of Pavia, Italy, E-mail: giulia.marcon@phd.unibocconi.it. Muliere and Padoan work at the Department of Decision Sciences, Bocconi University of Milan, via Roentgen 1, 20136 Milano, Italy. E-mail: 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. Segers is a Professor at the Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: Johan.Segers@uclouvain.be.

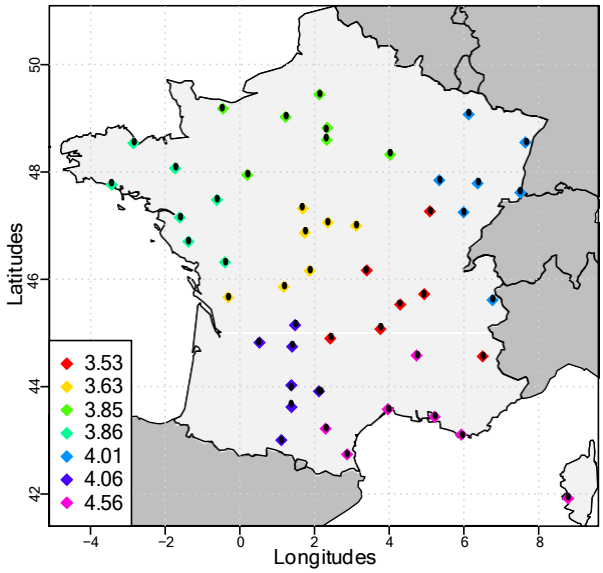

Figure 1: Analysis of French weekly precipitation maxima in the period 1993–2011. Clusters of 49 weather stations and their estimated extremal coefficients in dimension $d = 7$ obtained with the projected version of the madogram estimator, see Section 5 for details.

maxima of hourly rainfall recorded at each station[1]. It would be of interest to hydrologists to infer the dependence within each of the seven-dimensional vectors of component-wise maxima and to compare the dependence structures among clusters. Such an endeavor represents the main motivation of this work.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-dimensional random vector of maxima that follows a multivariate max-stable distribution G; for more background on univariate and multivariate extreme-value theory, see for instance Beirlant et al. (2004), de Haan and Ferreira (2006), or Falk, Hüsler, and Reiss (2010). The margins of G, denoted by $F_i(x) = \mathrm{P}\{X_i \leq x\}$, for all $x \in \mathbb{R}$ and $i = 1, \ldots, d$, are univariate max-stable distributions. The joint distribution takes the form

$$G(\boldsymbol{x}) = C\big(F_1(x_1), \ldots, F_d(x_d)\big), \qquad \boldsymbol{x} \in \mathrm{R}^d, \tag{1.1}$$

where C is an extreme-value copula:

$$C(u_1, \ldots, u_d) = \exp\big\{-l(-\log u_1, \ldots, -\log u_d)\big\}, \qquad \boldsymbol{u} \in (0, 1]^d, \tag{1.2}$$

with $l : [0, \infty)^d \to [0, \infty)$ the so-called stable tail dependence function. The latter function is homogeneous of order one and is therefore determined by its restriction on the unit simplex, the restriction itself being called the Pickands dependence function, denoted here by A. Formally, we have

$$l(\boldsymbol{z}) = (z_1 + \cdots + z_d)\, A(\boldsymbol{w}), \qquad \boldsymbol{z} \in [0, \infty)^d, \tag{1.3}$$

[1] Data provided by Météo–France and published within the R package ClusterMax, freely available from the homepage of Philippe Naveau, http://www.lsce.ipsl.fr/Pisp/philippe.naveau/.

where $w_i = z_i/(z_1 + \cdots + z_d)$ for $i = 1, \ldots, d-1$ and $w_d = 1 - w_1 - \cdots - w_{d-1}$. We view A as a function defined on the $(d-1)$-dimensional unit simplex

$$S_{d-1} := \left\{ (w_1, \ldots, w_{d-1}) \in [0,1]^{d-1} : \sum_{i=1}^{d-1} w_i \leq 1 \right\}. \tag{1.4}$$

Let $\mathbb{A}$ be the family of functions $A : S_{d-1} \to [1/d, 1]$ that satisfy the following conditions:

(C1) $A(\boldsymbol{w})$ is convex, i.e., $A(\alpha \boldsymbol{w}_1 + (1-\alpha)\boldsymbol{w}_2) \leq \alpha A(\boldsymbol{w}_1) + (1-\alpha)A(\boldsymbol{w}_2)$, for $\alpha \in [0,1]$ and $\boldsymbol{w}_1, \boldsymbol{w}_2 \in S_{d-1}$;

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

$$1/d \leq \max(w_1, \ldots, w_{d-1}, w_d) \leq A(\boldsymbol{w}) \leq 1,$$

for any $\boldsymbol{w} = (w_1, \ldots, w_{d-1}) \in S_{d-1}$ with $w_d = 1 - w_1 - \ldots - w_{d-1}$;

Any Pickands dependence function belongs to the class $\mathbb{A}$ (Falk, Hüsler, and Reiss, 2010, Ch. 4). The converse is not true, however; see Beirlant et al. (2004, p. 257) for a counterexample. A characterization of the class of stable tail dependence functions has been given in Ressel (2013). In condition (C2), the lower and upper bounds represent the cases of complete dependence and independence, respectively.

Many parametric models have been introduced for modelling the extremal dependence for a variety of applications, with summaries to be found in Kotz and Nadarajah (2000) and Padoan (2013). However, such finite-dimensional parametric models can never cover the full class of Pickands dependence functions. For this reason, several nonparametric estimators of the Pickands dependence function have been proposed: see for instance Pickands (1981), Capéraà et al. (1997), Hall and Tajvidi (2000), Zhang et al. (2008), Genest and Segers (2009), Bücher et al. (2011), Gudendorf and Segers (2011, 2012), and Berghaus et al. (2013). All of these estimators require further adjustments to ensure they are genuine Pickands dependence functions.

Given an independent random sample from a multivariate distribution with continuous margins and whose copula is an extreme-value copula, we propose a nonparametric estimator for its 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). It has two drawbacks, however: it was only defined for the bivariate case and it is not necessarily a Pickands dependence function itself. Our first contribution is to propose a new type of madogram in the multivariate setting, see also Fonseca et al. (2015). A second contribution is to regularise the estimator by projecting it onto the space $\mathbb{A}$, imposing the necessary constraints (C1)–(C2). To do so, we make use of Bernstein polynomials. We admit that the resulting estimator still need not be a Pickands dependence function. Still, simulation results show that imposing (C1)–(C2) already greatly improves the estimation accuracy.

Many regularization strategies have already been considered in the literature. In the bivariate case, Pickands (1981) suggested the use of the greatest convex minorant. Smith

et al. (1990) proposed to modify a pilote estimator using kernel methods, while Hall and Tajvidi (2000) advocated constrained smoothing splines. However, as discussed in 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 estimator approach developed in Fils-Villetard et al. (2008) and Gudendorf and Segers (2012) provides a general framework based on projections of a pilote estimate onto an increasing sequence of finite-dimensional subsets $\mathbb{A}$. The approximation space they proposed consists of piecewise linear functions, yielding computational challenges in higher dimensions.

To bypass these computational hurdles, our strategy is to replace piecewise linear functions by Bernstein polynomials (Lorentz, 1986; Sauer, 1991). In virtue of their optimal shape restriction properties (Carnicer and Peña, 1993), Bernstein polynomials are suitable for nonparametric curve estimation (e.g. Petrone 1999; Chang et al. 2005) and shape-preserving regression (Wang and Ghosh 2012). We provide the asymptotic theory for our estimator and we demonstrate its practical use in dimension seven, which seems to be higher than what has been possible hitherto with nonparametric methods. The estimation uncertainty can be assessed through a resampling procedure.

Throughout the paper we use the following notation. Given $\mathrm{X} \subset \mathrm{R}^n$ and $n \in \mathrm{N}$, let $l^\infty(\mathrm{X})$ denote the spaces of bounded real-valued functions on X. For $f : \mathrm{X} \to \mathrm{R}$, let $\|f\|_\infty = \sup_{\boldsymbol{x} \in \mathrm{X}} |f(\boldsymbol{x})|$. The arrows "$\xrightarrow{\text{a.s.}}$", "$\rightarrow$", "$\Rightarrow$" and "$\rightsquigarrow$" denote almost sure convergence, convergence in distribution of random vectors (see van der Vaart 2000, Ch. 2) and weak convergence of functions in $l^\infty(\mathrm{X})$ (see van der Vaart 2000, Ch. 18–19), respectively. Let $L^2(\mathrm{X})$ denote the Hilbert space of square-integrable functions $f : \mathrm{X} \to \mathrm{R}$, with X equipped with n-dimensional Lebesgue measure; the L^2-norm is denoted by $\|f\|_2 = (\int_{\mathrm{X}} f^2(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x})^{1/2}$. For analytical reasons, we view the unit simplex S_{d-1} as a subset of R^{d-1}, see (1.4), although geometrically, it is perhaps more natural to consider it as a subset of R^d. A similar convention applies to our use of the multi-index $\boldsymbol{\alpha}$ in Section 3.

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. In Section 4, we investigate the finite-sample performance of our estimation method by means of Monte Carlo simulations. Finally, we apply our approach to French weekly maxima of hourly rainfall in Section 5. All proofs are deferred to the appendices.

2 Madogram estimator

Let $\boldsymbol{X}$ be a random vector with continuous marginal distribution functions $F_1, \ldots, F_d$ and whose copula C is an extreme-value copula with stable tail dependence function l and Pickands dependence function A; see above. Our estimator is based on the sample version of the multivariate madogram, extending Naveau et al. (2009), see also Fonseca et al. (2015).

Definition 2.1. For $\boldsymbol{w} \in S_{d-1}$, the multivariate $\boldsymbol{w}$-madogram, denoted by $\nu(\boldsymbol{w})$, is defined as the expected distance between the componentwise maximum and the componentwise mean of the variables $F_1^{1/w_1}(X_1), \ldots, F_d^{1/w_d}(X_d)$, that is,

$$\nu(\boldsymbol{w}) = \mathrm{E}\left[\max_{i=1}^{d}\left(F_i^{1/w_i}(X_i)\right) - \frac{1}{d}\sum_{i=1}^{d} F_i^{1/w_i}(X_i)\right]. \tag{2.1}$$

For $w_i = 0$ and $0 < u < 1$, we put $u^{1/w_i} = 0$ by convention.

Proposition 2.2. *If the random vector $\boldsymbol{X}$ has continuous margins and extreme-value copula with Pickands dependence function A, then, for all $\boldsymbol{w} \in S_{d-1}$,*

$$\begin{aligned} \nu(\boldsymbol{w}) &= \frac{A(\boldsymbol{w})}{1 + A(\boldsymbol{w})} - c(\boldsymbol{w}), \\ A(\boldsymbol{w}) &= \frac{\nu(\boldsymbol{w}) + c(\boldsymbol{w})}{1 - \nu(\boldsymbol{w}) - c(\boldsymbol{w})}, \end{aligned} \tag{2.2}$$

where $c(\boldsymbol{w}) = d^{-1}\sum_{i=1}^{d} w_i/(1 + w_i)$.

The madogram can be interpreted as the L_1 distance between the maximum and the average of the random variables $F_1^{1/w_1}(X_1), \ldots, F_d^{1/w_d}(X_d)$. If $w_1 = \ldots = w_d = 1/d$, then the L_1 distance is zero if and only if all components $F_i(X_i)$ are equal with probability one, that is, in case of complete dependence.

In the bivariate case, Definition 2.1 is slightly different from the one proposed by Naveau et al. (2009). Here, we use the vector $F_1^{1/w_1}(X_1), F_2^{1/w_2}(X_2)$ instead of $F_1^{w_1}(X_1), F_2^{w_2}(X_2)$. This new version has the advantage that the sample equivalent of (2.2) will automatically satisfy condition (C2).

Assume first that the marginal distributions $F_1, \ldots, F_d$ are known; below, we will estimate them by the empirical distribution functions. Equation (2.1) suggests the statistic

$$\nu_n(\boldsymbol{w}) = \frac{1}{n}\sum_{m=1}^{n}\left[\max_{i=1}^{d}\left(F_i^{1/w_i}(X_{m,i})\right) - \frac{1}{d}\sum_{i=1}^{d} F_i^{1/w_i}(X_{m,i})\right]. \tag{2.3}$$

The Pickands dependence function can then be estimated through

$$A_n^{\mathrm{MD}}(\boldsymbol{w}) = \frac{\nu_n(\boldsymbol{w}) + c(\boldsymbol{w})}{1 - \nu_n(\mathrm{w}) - c(\boldsymbol{w})}, \qquad \boldsymbol{w} \in S_{d-1}. \tag{2.4}$$

Next, we estimate the unknown marginal distributions $F_1, \ldots, F_d$ by the empirical distribution functions

$$F_{n,i}(x) = \frac{1}{n}\sum_{m=1}^{n} 1(X_{m,i} \le x), \qquad i = 1, \ldots, d, \tag{2.5}$$

where $\mathbb{1}(E)$ is the indicator function of the event E. Replacing F_i by $F_{n,i}$ in Equation (2.3) yields our nonparametric estimators $\hat{\ell}_n$ and $\hat{A}_n^{\mathrm{MD}}$ of the multivariate madogram and of the Pickands dependence function, respectively:

$$\hat{\nu}_n(\boldsymbol{w}) = \frac{1}{n}\sum_{m=1}^{n}\max_{i=1}^{d} w_i F_{n,i}^{1/w_i}(X_{m,i}) - \frac{1}{d}\sum_{i=1}^{d} F_{n,i}^{1/w_i}(X_{m,i}),$$

$$\hat{A}_n^{\mathrm{MD}}(\boldsymbol{w}) = \frac{\hat{\nu}_n(\boldsymbol{w}) + c(\boldsymbol{w})}{1 - \hat{\ell}_n(\boldsymbol{w}) - c(\boldsymbol{w})}.$$

Other estimators of the margins could be inserted as well. However, the use of the empirical distribution functions requires minimal assumptions and yields an estimator for A which is invariant under monotone transformations.

The next theorem summarizes the asymptotic properties related to A_n^{MD} and $\hat{A}_n^{\mathrm{MD}}$. The asymptotic normality requires a smoothness condition on the extreme-value copula C, see Example 5.3 in Segers (2012).

Condition 2.3. For every $i \in \{1, \ldots, d\}$, the partial derivative of C with respect to u_i exists and is continuous on the set $\{\boldsymbol{u} \in [0,1]^d : 0 < u_i < 1\}$.

Let $\mathbb{D}$ be a C-Brownian bridge, that is, a zero-mean Gaussian process on $[0,1]^d$ with continuous sample paths and with covariance function given by

$$\mathrm{Cov}(\mathbb{D}(\boldsymbol{u}), \mathbb{D}(\boldsymbol{v})) = C(\boldsymbol{u} \wedge \boldsymbol{v}) - C(\boldsymbol{u})\,C(\boldsymbol{v}), \qquad \boldsymbol{u}, \boldsymbol{v} \in [0,1]^d, \tag{2.6}$$

where the minimum is considered componentwise. Further, provided Condition 2.3 is satisfied, define the Gaussian process $\hat{\mathbb{D}}$ on $[0,1]^d$ by

$$\hat{\mathbb{D}}(\boldsymbol{u}) = \mathbb{D}(\boldsymbol{u}) - \sum_{i=1}^{d} \frac{\partial C}{\partial u_i}(\boldsymbol{u})\,\mathbb{D}(1, \ldots, 1, u_i, 1, \ldots, 1), \qquad \boldsymbol{u} \in [0,1]^d. \tag{2.7}$$

Theorem 2.4. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent and identically distributed random vectors whose common distribution has continuous margins and extreme-value copula C with Pickands dependence function A. Then:*

a) $\|A_n^{\mathrm{MD}} - A\|_\infty \xrightarrow{\text{a.s.}} 0$ *as* $n \to \infty$ *and in* $l^\infty(S_{d-1})$, *as* $n \to \infty$,

$$\sqrt{n}(A_n^{\mathrm{MD}} - A) \rightsquigarrow (1 + A(\boldsymbol{w}))^2 \sum_{i=1}^{d} \frac{1}{d} \int_0^1 \mathbb{D}(1, \ldots, 1, x^{w_i}, 1, \ldots, 1) - \mathbb{D}(x^{w_1}, \ldots, x^{w_d})\,\mathrm{d}x \quad ;\ \boldsymbol{w} \in S_{d-1}$$

b) $\|\hat{A}_n^{\mathrm{MD}} - A\|_\infty \xrightarrow{\text{a.s.}} 0$ *as* $n \to \infty$. *Moreover, if Condition 2.3 is satisfied, then, in* $l^\infty(S_{d-1})$, *as* $n \to \infty$,

$$\sqrt{n}(\hat{A}_n^{\mathrm{MD}} - A) \rightsquigarrow -(1 + A(\boldsymbol{w}))^2 \int_0^1 \hat{\mathbb{D}}(x^{w_1}, \ldots, x^{w_d})\,\mathrm{d}x \quad .\ \boldsymbol{w} \in S_{d-1}$$

The two conditions (C1)–(C2) are not necessarily satisfied by $\hat{A}_n^{\mathrm{MD}}$. To ensure both conditions, we propose a projection method based on Bernstein polynomials.

3 Estimation based on Bernstein polynomials

3.1 Bernstein polynomials on the simplex

Multivariate Bernstein polynomials, defined on a cube or on a simplex, have been widely discussed in mathematics and statistics, see for example Ditzian (1986) and Petrone (2004). Here our focus is on approximating a bounded function f on the simplex S_{d-1}. In the univariate case, the shape features of the original function are preserved by its Bernstein approximation. For higher dimensions, shape properties like convexity may no longer be retained. The Bernstein–Bézier polynomials (Sauer 1991) solve this issue and preserve various shape properties (Li 2011, Lai 1993).

Fix the dimension $d \geq 2$. For positive integer k, let Γ_k be the set of multi-indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{d-1}) \in \{0, 1, \ldots, k\}^{d-1}$ such that $\alpha_1 + \cdots + \alpha_{d-1} \leq k$. The cardinality of Γ_k is equal to the number of multi-indices $\boldsymbol{\alpha} \in \{0, 1, \ldots, k\}^d$ such that $\alpha_1 + \cdots + \alpha_d = k$; just set $\alpha_d = k - \alpha_1 - \cdots - \alpha_{d-1}$. Replacing each α_j by $\alpha_j + 1$, we find that the number of such multi-indices is also equal to the number of compositions of the integer $k + d$ into d positive integer parts. The number of such compositions is equal to

$$p_k = \binom{k+d-1}{d-1}, \tag{3.1}$$

and so is the cardinality of Γ_k. Define the Bernstein basis polynomial $b_{\boldsymbol{\alpha}}(\cdot; k)$ on S_{d-1} of degree k by

$$b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k) = \binom{k}{\boldsymbol{\alpha}} \boldsymbol{w}^{\boldsymbol{\alpha}}, \qquad \boldsymbol{w} \in S_{d-1} \tag{3.2}$$

where

$$\binom{k}{\boldsymbol{\alpha}} = \frac{k!}{\alpha_1! \ldots \alpha_d!}, \qquad \boldsymbol{w}^{\boldsymbol{\alpha}} = w_1^{\alpha_1} \cdots w_d^{\alpha_d}.$$

The k-th degree Bernstein polynomial associated to A is defined as

$$B_A(\boldsymbol{w}; k) = \sum_{\boldsymbol{\alpha} \in \Gamma_k} A(\boldsymbol{\alpha}/k) b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k), \qquad \boldsymbol{w} \in S_{d-1}. \tag{3.3}$$

Proposition 3.1. *For every $A \in \mathcal{A}$ and every $k = 1, 2, \ldots$,*

$$\sup_{\boldsymbol{w} \in S_{d-1}} |B_A(\boldsymbol{w}; k) - A(\boldsymbol{w})| \leq \frac{d}{2\sqrt{k}}$$

The family of Bernstein–Bézier polynomials of degree k is defined as the set

$$\mathcal{B}_k = \left(\sum_{\boldsymbol{\alpha} \in \Gamma_k} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\cdot; k) : \boldsymbol{\beta} \in [0, 1]^{p_k} \right).$$

For $\boldsymbol{w} \in S_{d-1}$, let $\boldsymbol{b}_k(\boldsymbol{w})$ be the row vector $(b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k), \boldsymbol{\alpha} \in \{0, 1, \ldots, k\}^d : \alpha_1 + \cdots + \alpha_d = k)$. In matrix notation, we have $\sum_{\boldsymbol{\alpha} \in \Gamma_k} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k) = \boldsymbol{b}_k(\boldsymbol{w}) \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is viewed as a column vector.

3.2 Shape-preserving estimator

In this section, we describe how to use Bernstein–Bézier polynomials to obtain a projection estimator (Fils-Villetard et al. 2008) that satisfies (C1)–(C2). Given a pilot estimator, say $\hat{A}_n$, the idea is to seek approximate solutions to the constrained optimization problem

$$\tilde{A}_n = \arg\min_{A \in \mathcal{A}} \|\hat{A}_n - A\|_2.$$

There is no closed-form solution to the above equation, and so an approximation based on the sieves method is explored. Consider a sequence $\mathcal{A}_k \subseteq \mathcal{A}$ of constrained multivariate Bernstein–Bézier polynomial families on S_{d-1} given by

$$\mathcal{A}_k = \{\boldsymbol{w} \mapsto B(\boldsymbol{w}; k) = \boldsymbol{b}_k(\boldsymbol{w})\boldsymbol{\beta}_k : \boldsymbol{\beta}_k \in [0, 1]^{p_k} \text{ such that } \boldsymbol{R}_k\boldsymbol{\beta}_k \geq \boldsymbol{r}_k\}. \tag{3.4}$$

Here, $\boldsymbol{R}_k = [\boldsymbol{R}_k^{(1)}, \boldsymbol{R}_k^{(2)}, \boldsymbol{R}_k^{(3)}]^{\mathrm{T}}$ and $\boldsymbol{r}_k = [\boldsymbol{r}_k^{(1)}, \boldsymbol{r}_k^{(2)}, \boldsymbol{r}_k^{(3)}]^{\mathrm{T}}$ are a $(q \times p_k)$ full row rank matrix and a $(q \times 1)$ vector respectively such that the constraint $\boldsymbol{R}_k\boldsymbol{\beta}_k \geq \boldsymbol{r}_k$ on the coefficient vector $\boldsymbol{\beta}_k$ ensures that each member of $\mathcal{A}$ satisfies (C1)–(C2). Details for deriving the block matrices and vectors of constraints are provided below.

R1) A sufficient condition to guarantee that the function $\boldsymbol{w} \mapsto B(\boldsymbol{w}; k)$ on S_{d-1} is convex is that its Hessian matrix be positive semi-definite. In order to enforce the latter, we resort by applying Theorem 1 in Lai (1993). First, for $s \neq r \in \{0, \ldots, d-1\}$ and two vectors $\boldsymbol{v}_r$ and $\boldsymbol{v}_s$, where $\boldsymbol{v}_r = \boldsymbol{0}$ if $r = 0$ and $\boldsymbol{v}_r = \boldsymbol{e}_r$ if $r > 0$ with $\boldsymbol{e}_r$ the canonical unit vector (same for $\boldsymbol{v}_s$), the directional derivative of B with respect to the direction $\overrightarrow{\boldsymbol{v}_r\boldsymbol{v}_s}$ is

$$D_{\boldsymbol{v}_s - \boldsymbol{v}_r} B(\boldsymbol{w}; k) = k \sum_{\boldsymbol{\alpha} \in \Gamma_{k-1}} \Delta_{s,r}\beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k-1), \qquad \boldsymbol{w} \in S_{d-1}$$

where $\Delta_{s,r}\beta_{\boldsymbol{\alpha}} = (\beta_{\boldsymbol{\alpha}+\boldsymbol{v}_s} - \beta_{\boldsymbol{\alpha}+\boldsymbol{v}_r})$. Second, the second directional derivative of B with respect to the directions $\overrightarrow{\boldsymbol{v}_r\boldsymbol{v}_s}$ and $\overrightarrow{\boldsymbol{v}_r\boldsymbol{v}_t}$ is

$$D^2_{\boldsymbol{v}_s - \boldsymbol{v}_r, \boldsymbol{v}_t - \boldsymbol{v}_r} B(\boldsymbol{w}; k) = k(k-1) \sum_{\boldsymbol{\alpha} \in \Gamma_{k-2}} \Delta_{t,r}\Delta_{s,r}\beta_{\boldsymbol{\alpha}}\, b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k-2), \qquad \boldsymbol{w} \in S_{d-1}.$$

Then, the Hessian matrix of $B(\boldsymbol{w}; k)$, $\boldsymbol{w} \in S_{d-1}$, is $H_B = [D^2_{\boldsymbol{v}_s, \boldsymbol{v}_t} B(\boldsymbol{w}; k)]_{s,t \in \{1,\ldots,d-1\}, r=0}$, and it can be written as

$$H_B = k(k-1) \sum_{\boldsymbol{\alpha} \in \Gamma_{k-2}} \Sigma_{\boldsymbol{\alpha}}\, b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k-2), \qquad \boldsymbol{w} \in S_{d-1},$$

where, for all $\boldsymbol{\alpha} \in \Gamma_{k-2}$, $\Sigma_{\boldsymbol{\alpha}}$ is a symmetric $(d-1) \times (d-1)$ matrix given by

$$\Sigma_{\boldsymbol{\alpha}} = \begin{bmatrix} \Delta^2_{1,0}\beta_{\boldsymbol{\alpha}} & \Delta_{1,0}\Delta_{2,0}\beta_{\boldsymbol{\alpha}} & \cdots & \Delta_{1,0}\Delta_{d-1,0}\beta_{\boldsymbol{\alpha}} \\ \Delta_{2,0}\Delta_{1,0}\beta_{\boldsymbol{\alpha}} & \Delta^2_{2,0}\beta_{\boldsymbol{\alpha}} & \cdots & \Delta_{2,0}\Delta_{d-1,0}\beta_{\boldsymbol{\alpha}} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{d-1,0}\Delta_{1,0}\beta_{\boldsymbol{\alpha}} & \Delta_{d-1,0}\Delta_{2,0}\beta_{\boldsymbol{\alpha}} & \cdots & \Delta^2_{d-1,0}\beta_{\boldsymbol{\alpha}} \end{bmatrix}.$$

By the weak diagonal dominance criterion (Lai 1993) in order to guarantee that $\Sigma_{\boldsymbol{\alpha}}$ is positive semi-definite, it is sufficient to check, for all $\boldsymbol{\alpha} \in \Gamma_{k-2}$ and $i \in \{1, \ldots, d-1\}$, the conditions

$$\Delta_{i,0}^2 \beta_{\boldsymbol{\alpha}} - \sum_{j \neq i} |\Delta_{i,0}\Delta_{j,0}\beta_{\boldsymbol{\alpha}}| \geq 0.$$

Such conditions produce constraints that are more severe than necessary. The above conditions can be synthesized in matrix form as $\boldsymbol{R}_k^{(1)}\boldsymbol{\beta}_k \geq \boldsymbol{r}_k^{(1)}$ where $\boldsymbol{R}_k^{(1)}$ is a $(p_{k-2}(d-1)2^{d-2} \times p_k)$ matrix and $\boldsymbol{r}_k^{(1)}$ is the corresponding null vector. For example, with $d = 3$ and $k = 3$,

$$\boldsymbol{R}_3^{(1)} = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 2 & -1 & 0 & 0 & -3 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 2 & -3 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 2 & -1 & 0 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 & -3 & 1 & -1 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{r}_3^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

A consequence of this approach is that

R2) B satisfies the upper bound condition in (C2) if $\beta_{\boldsymbol{\alpha}} = 1$ for the set of coefficients $\{\beta_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} = \mathbf{0}$ or $\boldsymbol{\alpha} = k\boldsymbol{e}_i, \ \forall i = 1, \ldots, d-1\}$. Thus, the $(2d \times p_k)$ matrix and $2d$-dimensional vector of restrictions are equal to

$$\boldsymbol{R}_k^{(2)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & -1 & \cdots & 0 \end{bmatrix}, \quad \boldsymbol{r}_k^{(2)} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ \vdots \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

R3) B satisfies the lower bound condition in (C2) if the restrictions R1)-R2) hold and the following constraints are fulfilled. Specifically, for all $(i, j) \in \{0, \ldots, d-1\}^2$, $i = j$, the first directional derivatives with respect to $\overrightarrow{\boldsymbol{v}_i\boldsymbol{v}_j}$, evaluated at the vertices of the simplex, are compared with the first directional derivatives of the planes $z_0 = 1$, $z_1 = w_1$, $z_2 = w_2$, ..., $z_d = 1 - w_1 - w_2 - \cdots - w_{d-1}$, with respect to the same directions. So, it is sufficient to check the conditions

$$D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B(\boldsymbol{v}_j; k) \geq -1, \qquad \forall\, (i, j) \in \{0, \ldots, d-1\}^2,\ i \neq j.$$

As a consequence, it is sufficient to check the conditions $\beta_{\boldsymbol{\alpha}} > 1 - 1/k$ for the set of coefficients $\{\beta_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} = \boldsymbol{e}_i$ or $\boldsymbol{\alpha} = (k-1)\boldsymbol{e}_i$ or $\boldsymbol{\alpha} = (k-1)\boldsymbol{e}_i + \boldsymbol{e}_j, \ \forall j \quad i = 1, \ldots, d-1\}$. This can be synthesized in matrix form as $\boldsymbol{R}_k^{(3)}\boldsymbol{\beta}_k \geq \boldsymbol{r}_k^{(3)}$ where $\boldsymbol{R}_k^{(3)}$ is a $(d(d-1) \times p_k)$

matrix and $\boldsymbol{r}_k^{(3)}$ is the corresponding vector of $1 - 1/k$ vaules. For example, when $d = 3$ and $k = 3$, the constraint matrix is the following:

$$\boldsymbol{R}_3^{(3)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \boldsymbol{r}_3^{(3)} = \begin{bmatrix} 1 - 1/k \\ 1 - 1/k \\ 1 - 1/k \\ 1 - 1/k \\ 1 - 1/k \\ 1 - 1/k \end{bmatrix}.$$

The use of the third restriction is justified by the following result.

Proposition 3.2. *Let B_A be the polynomial (3.3). Assume that B_A is convex on the simplex and $B_A(\boldsymbol{v}_j; k) = 1$ for all $j \in \{0, \ldots, d-1\}$. Then, for all $\boldsymbol{w} \in \mathbb{S}_{d-1}$*

$$B_A(\boldsymbol{w}; k) \geq \max(w_1, \ldots, w_d) \qquad \Longleftrightarrow \qquad D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B_A(\boldsymbol{v}_j; k) \geq -1,$$

for all $(i, j) \in \{0, \ldots, d-1\}^2$, $i \neq j$.

Recall that the approximate projection estimator of A based on a pilot estimator $\hat{A}_n$ is given by the solution to the optimization problem

$$\tilde{A}_{n,k} = \operatorname*{arg\,min}_{B \in \mathbb{A}_k} \|\hat{A}_n - B\|_2. \tag{3.5}$$

In case the pilot estimator is the madogram estimator $\hat{A}_n^{\mathrm{MD}}$, the corresponding projection estimator is denoted by $\tilde{A}_{n,k}^{\mathrm{MD}}$.

In practice, the estimator $\tilde{A}_{n,k}$ is evaluated on a finite set of points $\{\boldsymbol{w}_q : q = 1, \ldots, Q\}$, with $Q \in \mathbb{N}$ and $\boldsymbol{w}_q \in \mathbb{S}_{d-1}$. The discretized version of the above solution is given by

$$\tilde{A}_{n,k}(\boldsymbol{w}_q) = \boldsymbol{b}_k(\boldsymbol{w}_q)\hat{\boldsymbol{\beta}}_k, \qquad \boldsymbol{w}_q \in \mathbb{S}_{d-1}, \quad q = 1, \ldots, Q, \tag{3.6}$$

where $\hat{\boldsymbol{\beta}}_k$ is the minimizer of the constrained least-squares problem

$$\hat{\boldsymbol{\beta}}_k = \operatorname*{arg\,min}_{\boldsymbol{\beta}_k \in [0,1]^{p_k} : \boldsymbol{R}_k \boldsymbol{\beta}_k \geq \boldsymbol{r}_k} \frac{1}{Q} \sum_{q=1}^{Q} \left| \boldsymbol{b}_k(\boldsymbol{w}_q)\boldsymbol{\beta}_k - \hat{A}_n(\boldsymbol{w}_q) \right|^2.$$

This is a quadratic programming problem, whose solution is

$$\hat{\boldsymbol{\beta}}_k = \boldsymbol{\beta}_k^{\perp} - (\boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{b}_k)^{-1} \boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{\gamma}, \tag{3.7}$$

where $\boldsymbol{\gamma}$ is a vector of Lagrange multipliers and $\boldsymbol{\beta}_k^{\perp} = (\boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{b})^{-1} \boldsymbol{b}^{\mathrm{T}} \hat{A}_n$ is the unconstrained least squares estimator. The vectors $\hat{\boldsymbol{\beta}}_k$ and $\boldsymbol{\gamma}$ can be efficiently computed with an iterative quadratic programming algorithm (e.g. Goldfarb and Idnani 1983). A high resolution of (3.6) is obtained with increasing values of Q. Numerical experiments showed that a close approximation of the true Pickands dependence function is already reached with moderate values of Q. However, Q should not be seen as an additional parameter of the projection estimator. The solution (3.6) provides better approximations of the true Pickands dependence function for increasing sample sizes n and polynomial degrees k.

In order to state the asymptotic distribution of the projection estimator, the following result is required.

Proposition 3.3. $\mathbb{A}_k$, $k = 1, 2, \ldots$ *is a nested sequence in* $\mathbb{A}$. *Furthermore, if* $A \in \mathbb{A}$ *satisfies the condition*

$$\Delta^2_{i,0}A(\boldsymbol{\alpha}/k) - \sum_{j/=i} |\Delta_{i,0}\Delta_{j,0}A(\boldsymbol{\alpha}/k)| \geq 0, \quad \forall\, k, \boldsymbol{\alpha} \in \Gamma_{k-2},\ i \in \{1, \ldots, d-1\}, \tag{3.8}$$

then there exist polynomials $A_k \in \mathbb{A}_k$ *such that* $\lim_{k\to\infty} \sup_{\boldsymbol{w}\in S_{d-1}} |A_k(\boldsymbol{w}) - A(\boldsymbol{w})| = 0$.

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

Proposition 3.4. *Assume that the polynomial's degree,* k_n, *increases with the sample size* n *in such a way that* $k_n/n \to \infty$ *as* $n \to \infty$. *If the Pickands dependence function* A *satisfies the condition* (3.8), *then, for some Gaussian process* Z,

$$\sqrt{n}(\tilde{A}^{\mathrm{MD}}_{n,k_n} - A) \rightsquigarrow \arg\min_{Z' \in T_{\mathbb{A}}(A)} \| Z' - Z \|_2, \quad n \to \infty,$$

in $L^2(S_{d-1})$, *where* $T_{\mathbb{A}}(A)$ *is the tangent cone of* $\mathbb{A}$ *at* A, *given by the set of limits of all the sequences* $a_n(A_n - A)$, $a_n \geq 0$ *and* $A_n \in \mathbb{A}$.

It remains an open problem to establish the asymptotic behaviour of the projection estimator without condition (3.8). Moreover, if the Pickands dependence function is sufficiently smooth, we conjecture that the approximation rate in Proposition 3.1 can be improved, leading a slower growth rate needed for the degree of the Bernstein polynomial in Proposition 3.4. The simulation results in Section 4 confirm that polynomial degrees k much lower than n are already sufficient to achieve good results.

Finally, note that Proposition 3.4 and, in fact, everything else in this section applies to any estimator of the Pickands dependence function which satisfies a suitable functional central limit theorem. We are grateful to an anonymous Referee for having pointed this out.

3.3 Confidence bands

We construct confidence bands using a resampling method. For $\boldsymbol{w} \in S_{d-1}$ and $0 < \tilde{\alpha} < 1$, the bootstrap $(1 - \tilde{\alpha})$ pointwise confidence band, based on the estimates $\tilde{A}^{*(r)}_{n,k}(\boldsymbol{w})$, $r = 1, 2, \ldots$, obtained from the bootstrapped sample $\boldsymbol{X}^{(r)}_n = (\boldsymbol{X}^{(r)}_1, \ldots, \boldsymbol{X}^{(r)}_d)$, 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 $\hat{\boldsymbol{\beta}}^{*(r)}_k$ be the Bernstein polynomials' coefficient estimator based on the bootstrap sample $\boldsymbol{X}^{(r)}_n$, $r = 1, 2, \ldots$, we define a bootstrap simultaneous $(1 - \tilde{\alpha})$ confidence band specifying the lower $\tilde{A}^{L}_{n,k}(\boldsymbol{w})$ and upper $\tilde{A}^{U}_{n,k}(\boldsymbol{w})$ limits as

$$\left(\sum_{\boldsymbol{\alpha}\in\Gamma_k} \hat{\beta}^{*[r(\tilde{\alpha}/2)]}_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k); \sum_{\boldsymbol{\alpha}\in\Gamma_k} \hat{\beta}^{*[r(1-\tilde{\alpha}/2)]}_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k) \right), \quad \boldsymbol{w} \in S_{d-1}, \tag{3.9}$$

where $\hat{\theta}^{*}_{\boldsymbol{\alpha}}{}^{\lfloor r(\tilde{\alpha}/2)\rfloor}$ and $\hat{\theta}^{*}_{\boldsymbol{\alpha}}{}^{\lfloor r(1-\tilde{\alpha}/2)\rfloor}$, for all $\boldsymbol{\alpha} \in \Gamma_k$, correspond to the $\lfloor r(\tilde{\alpha}/2)\rfloor$ and $\lfloor r(1-\tilde{\alpha}/2)\rfloor$ ordered statistics respectively and $b_{\boldsymbol{\alpha}}(\boldsymbol{w}; k)$ is the Bernstein basis polynomial of degree k, see (3.2). Although this approach does not guarantee convex confidence bands, it works very well in our simulations, where we find that the convexity is violated only when dependence is weak. Another possibility, that can be considered, is to bootstrap bands for unconstrained estimators and then apply projection to the lower and upper bound, as pointed out by an anonymous Referee. Our specific simulations results indicate that our method performs slightly better than this valuable alternative.

4 Simulations

To visually illustrate the gain in implementing our Bernstein-Bézier projection approach, Figure 2 compares the madogram (MD) estimator $\hat{A}_n^{\mathrm{MD}}$ defined by (2.4) with its Bernstein-Bézier-projection (BP) version defined by (3.6) for the special case of the symmetric logistic model (SL, Tawn 1990) with $d = 3$ and $\alpha^{\mathrm{J}} = 0.3$. 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 providing smoother and more realistic contour lines.

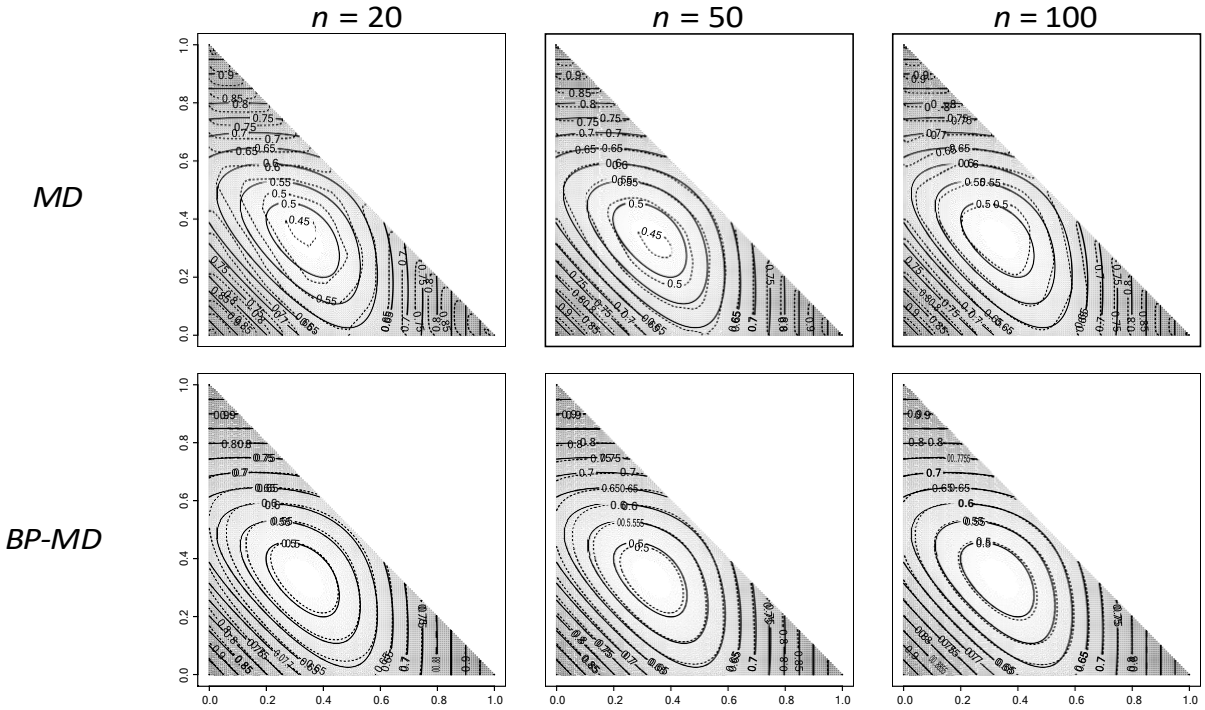

Figure 2: Estimates (dashed lines) of the Pickands dependence function obtained with the MD estimator (top row) and its BP version (bottom row) with polynomial degree $k = 14$. The solid line is the true Pickands dependence function. Each column represents a different sample size.

To guarantee a good approximation of A with $\tilde{A}_{n,k}$, Proposition 3.4 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 $\alpha^J = 0.3$ and $n = 100$. As expected from the theory, the choice of k is not anecdotical. 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 $k = 14$.

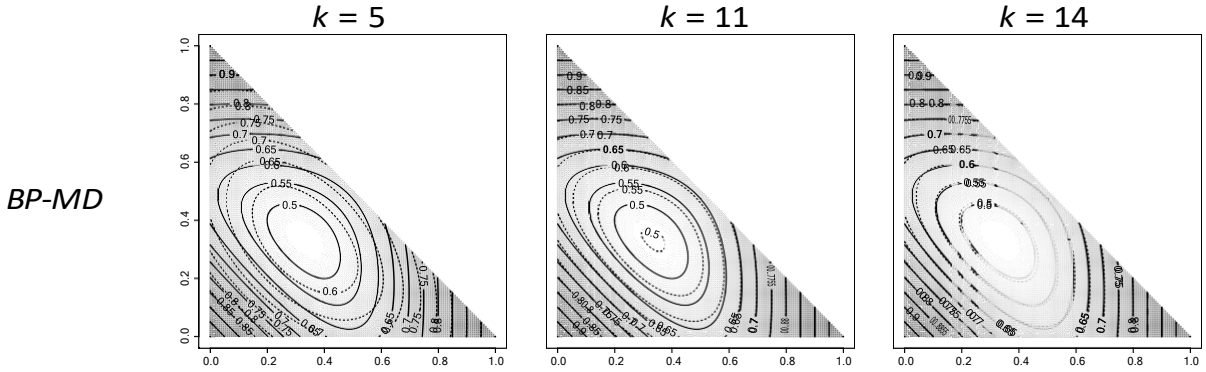

Figure 3: Same as Figure 2 but with $n = 100$ and three different values of the polynomial degree.

To go beyond these visual checks, we also compute a Monte-Carlo approximation of the mean integrated squared error

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

for a variety of setups. The approximate MISE is obtained by repeating 1000 times a given inference method for three different sample sizes $n = 50, 100, 200$. Different dependence strength of the logistic model has been explored setting the parameter α^J between 0.3 (strong dependence) and 1 (independence). Table 1 compares four non-parametric estimators introduced in Section 2: the madogram estimator (MD), the Pickands (1981) estimator (P), the multivariate version of the Hall and Tajvidi (2000) estimator (HT), and finally the multivariate extension of the Capéraà, Fougères, and Genest (1997) estimator. For comparison purposes we have also considered the weighted and endpoint-corrected versions of the P and CFG estimators as discussed in Gudendorf and Segers (2012), denoted by Pw and CFGw respectively. We can see that the MD estimator provided the best results if compared with the other classical non-parametric estimators. Taking into account also the weighted versions, it turns out that the CFGw estimator performs the best, especially for small sample sizes ($n = 50$). With a medium dependence ($\alpha^J = 0.5, 0.7$) the estimators provide similar results. With a weak dependence or in the independence case ($\alpha^J = 0.9, 1$), the MD estimator still provides the best results, especially for small and moderate sample sizes ($n = 50, 100$).

Sample size n	Estimator	Parameter α'				
		0.3	0.5	0.7	0.9	1
50	P	4.25×10^{-4}	8.06×10^{-4}	1.47×10^{-3}	2.45×10^{-3}	2.50×10^{-3}
	Pw	1.45×10^{-4}	5.13×10^{-4}	1.26×10^{-3}	2.53×10^{-3}	2.81×10^{-3}
	CFG	2.36×10^{-4}	6.92×10^{-4}	1.87×10^{-3}	4.07×10^{-3}	5.02×10^{-3}
	CFGw	9.17×10^{-5}	4.45×10^{-4}	1.24×10^{-3}	2.66×10^{-3}	3.07×10^{-3}
	HT	2.64×10^{-4}	8.54×10^{-4}	2.59×10^{-3}	5.13×10^{-3}	5.65×10^{-3}
	MD	1.80×10^{-4}	8.66×10^{-4}	1.91×10^{-3}	3.02×10^{-3}	2.87×10^{-3}
100	P	1.53×10^{-4}	3.16×10^{-4}	6.98×10^{-4}	1.20×10^{-3}	1.39×10^{-3}
	Pw	6.36×10^{-5}	2.38×10^{-4}	6.51×10^{-4}	1.25×10^{-3}	1.51×10^{-3}
	CFG	9.54×10^{-5}	3.27×10^{-4}	8.66×10^{-4}	1.78×10^{-3}	2.15×10^{-3}
	CFGw	4.32×10^{-5}	2.21×10^{-4}	6.35×10^{-4}	1.24×10^{-3}	1.39×10^{-3}
	HT	2.61×10^{-4}	7.66×10^{-4}	2.16×10^{-3}	4.24×10^{-3}	5.27×10^{-3}
	MD	7.02×10^{-5}	3.18×10^{-4}	7.91×10^{-4}	1.19×10^{-3}	1.09×10^{-3}
200	P	5.87×10^{-5}	1.54×10^{-4}	3.40×10^{-4}	6.25×10^{-4}	7.24×10^{-4}
	Pw	3.01×10^{-5}	1.31×10^{-4}	3.28×10^{-4}	6.60×10^{-4}	7.59×10^{-4}
	CFG	3.87×10^{-5}	1.58×10^{-4}	4.00×10^{-4}	8.31×10^{-4}	8.52×10^{-4}
	CFGw	2.12×10^{-5}	1.23×10^{-4}	3.24×10^{-4}	6.36×10^{-4}	5.90×10^{-4}
	HT	2.55×10^{-4}	7.31×10^{-4}	2.05×10^{-3}	3.82×10^{-3}	5.85×10^{-3}
	MD	3.17×10^{-5}	1.58×10^{-4}	3.70×10^{-4}	5.81×10^{-4}	4.91×10^{-4}

Table 1: MISE of four estimators of the Pickands dependence function, and some weighted version, based on a trivariate symmetric logistic dependence model for different parameter values and sample sizes.

Table 2 shows how an initial estimate of the Pickands dependence function improves using the projection method. The improvement is computed by

$$\frac{MISE_N - MISE_P}{MISE_N} \times 100,$$

and is reported in columns 3–6, where $MISE_N$ and $MISE_P$ are the MISE obtained with a non-parametric estimator and its projection respectively. As before, MISE provides a Monte-Carlo approximation of MISE($\hat{A}_n$, A) obtained with 1000 random samples. The true dependence structure is still the symmetric logistic model. α^j denotes the model parameter, and n and k are the sample size and the polynomial degree respectively. Estimates obtained with the initial non-parametric are regularized using the BP method. The order of the polynomial exploited is 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. It turns out that with a weak dependence a small value of k is enough, conversely with a strong dependence a large value of k is needed. 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. The improvements obtained with the classical estimators, sorted from largest to smallest, are: MD, CFG, P and HT. As expected, with Pw and CFGw the improvements are the

			Projection method							
			Bernstein-Bézier						Discrete spectral measure	
			% Improvement							
			Estimator						Estimator	
n	α'	k	P	Pw	CFG	CFGw	HT	MD	Pw	CFGw
50	0.3	23	18.11	13.34	76.84	18.97	51.53	8.50	2.14	0.82
	0.5	20	8.19	5.44	13.98	1.46	12.52	2.22	5.51	1.17
	0.7	16	15.60	11.01	4.43	2.10	9.05	6.48	11.03	3.17
	0.9	6	44.70	25.92	3.98	6.51	16.93	48.72	22.39	4.37
	1	3	69.95	34.53	4.92	9.04	34.68	93.60	29.07	4.89
100	0.3	23	16.59	13.36	59.75	13.43	45.45	7.41	1.27	0.40
	0.5	20	5.85	3.83	7.59	0.63	9.78	1.23	3.52	0.84
	0.7	16	9.89	8.15	2.21	0.95	6.42	2.74	7.51	1.37
	0.9	6	34.95	23.98	3.48	6.50	8.33	26.72	23.13	4.07
	1	3	68.00	39.35	5.93	11.50	19.22	87.46	36.10	8.11
200	0.3	23	15.16	10.63	37.73	5.66	44.72	5.05	0.60	0
	0.5	20	3.06	2.51	3.80	0.41	9.06	0.13	2.10	0
	0.7	16	5.70	5.22	0.90	0	5.60	0.76	4.85	0.88
	0.9	6	25.22	20.48	3.43	6.07	5.53	13.39	18.52	3.28
	1	3	69.17	46.32	8.63	16.06	10.88	81.99	40.99	11.89

Table 2: Percentage improvement of the MISE gained with the projection method.

smallest. For each estimator, the improvements sorted from largest to smallest, are obtained with: independence ($\alpha^J = 1$), strong dependence ($\alpha^J = 0.3$), weak dependence ($\alpha^J = 0.9$) and medium dependence ($\alpha^J = 0.5, 0.7$). These results are compared with those provided in Gudendorf and Segers (2012) that are obtained with the discrete spectral measure projection method proposed by the same authors (see columns 7,8). We can conclude that overall the BP method provides a better percentage improvement.

To explore the validity of our procedure to derive a bootstrap pointwise and simultaneous $(1 - \tilde{\alpha})$ confidence band described in Section 3.3, Table 3 displays 95% coverage probabilities from 1000 independent samples and $r = 500$ bootstrap resampling. The parametric setup is identical to the one used in Table 2 but with fixed sample size equal to $n = 100$. Overall, excluded the independence case, the simultaneous method (3.9) outperforms the pointwise method, since the coverage probabilities are always larger.

To close this small simulation study[2], we extend the class of parametric families to the asymmetric logistic (AL, Tawn 1990) with $\vartheta = 0.6$, $\varphi = 0.3$, $\psi = 0$, the Hüsler–Reiss model (HR, Hüsler and Reiss 1989) 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 ($\alpha^* = 7, -10, 1$, $\nu = 3$, $\omega = 0.9$), ($\alpha^* = -2, 9, -15$, $\nu = 2$, $\omega = 0.9$),

[2]The case $d = 2$ has also been considered. The results have been omitted for brevity, since they arrive at the same conclusion. Tables like Table 1, 2 and 3 are available upon request for the HR and EST families and brings the same overall message.

Estimator	Confident bands' type	Parameter α'				
		0.3	0.5	0.7	0.9	1
BP-P	Pointwise	41.53	35.92	50.66	72.23	83.05
	Simultaneous	73.34	69.13	68.79	75.11	84.82
BP-CFG	Pointwise	26.89	42.65	42.60	57.68	57.30
	Simultaneous	62.24	61.92	60.67	66.54	57.32
BP-HT	Pointwise	29.33	28.92	49.38	65.20	10.42
	Simultaneous	51.26	54.22	60.91	81.33	10.68
BP-MD	Pointwise	54.63	70.15	66.13	73.43	94.63
	Simultaneous	76.40	80.48	80.26	81.36	94.65

Table 3: 95% coverage probabilities of the BP method with four non-parametric estimators for the symmetric logistic model.

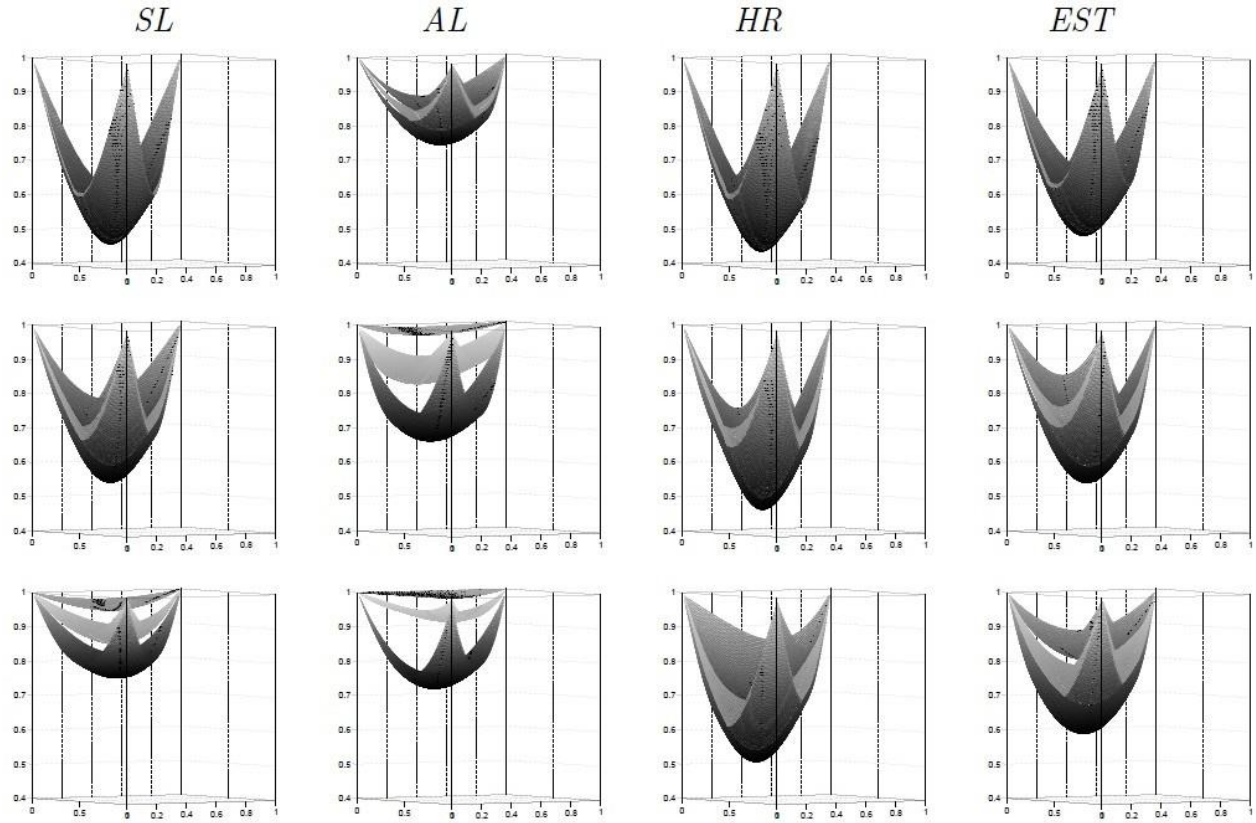

Figure 4: Estimates of Pickands dependence function for d = 3 (light grey shade) and bootstrap variability bands (dark grey shade) for the SL, AL, HR, EST (left-right) models with strong, mild and weak dependence (top-bottom)

($\alpha^* = -0.5, -0.5, -0.5$, $v = 3$, $\omega = 0.9$). 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. In the first place, the aim of clustering was to describe the dependence of locations, with homogeneous climatology characteristics within a cluster and heterogeneous characteristics between clusters. Climatologically, extreme precipitation that affects the Mediterranean coast in the fall is 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. It can be checked that extremes within clusters are indeed strongly dependent.

For each cluster, we compute our Bernstein projection estimator based on the madogram and fixed the polynomial's order k equal to 7. To summarize this seven-dimensional dependence structure, we take advantage of the *extremal coefficient* (Smith 1990) defined by

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

It satisfies the condition $1 \leq \vartheta \leq d$ where the lower and upper bounds represent the cases of complete dependence and independence among the extremes, respectively. In each cluster, the extremal coefficient is estimated using the equation $\hat{\vartheta} = 7\,\tilde{A}^{\mathrm{MD}}_{n,k}(1/7, \ldots, 1/7)$, so that $\hat{\vartheta}$ always belongs to the interval $[1, 7]$. The range of the estimated coefficients is between 3.5, indicating strong dependence, and 4.6, indicating medium dependence.

As climatologically expected, we can detect 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 46° latitude that affect large regions, whereas extreme precipitation in the south is more likely to be driven by localised convective storms with weak spatial dependence structures. Finally, in the center of the country, away from the coasts, there is the highest degree of dependence among extremes, as they are the result of the meeting between different densities of air masses.

For all possible pairs of locations we have estimated the bivariate Pickands dependence function using the madogram estimator and its Bernstein projection. The left-hand panel of Figure 5 shows the pairwise extremal coefficients versus the Euclidean distance between sites, computed through the estimated Pickands dependence functions. We have $\hat{\vartheta} \leq 1.5$ for the locations that are less than 200 km far apart, meaning that the extremes are strongly or at least mildly dependent, while for sites more than 200 km far apart, we have $\hat{\vartheta} > 1.5$,

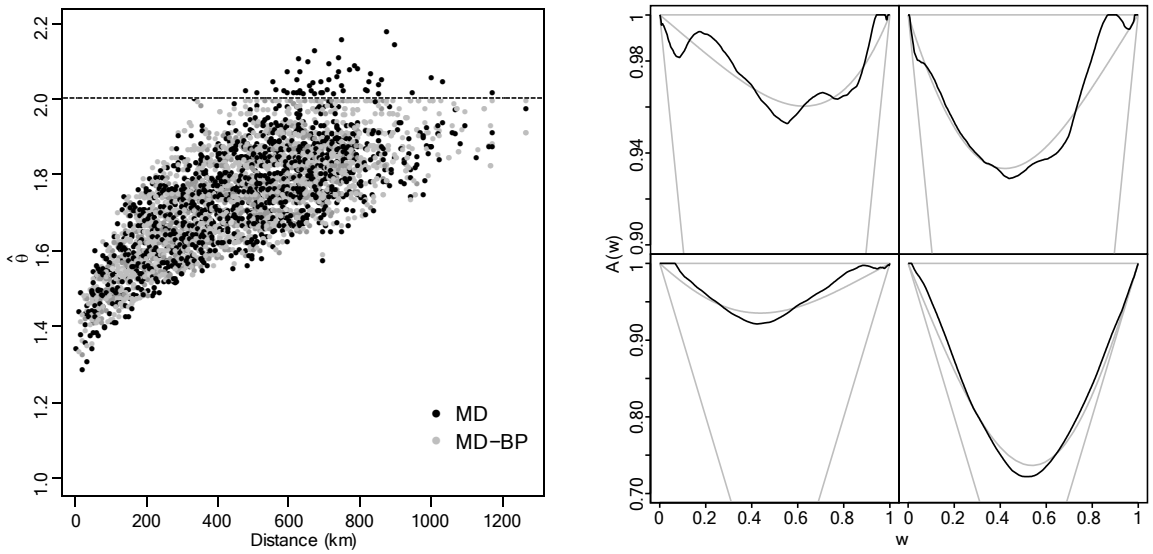

Figure 5: French precipitation data. Left: pairwise extremal coefficients as a function of distance between weather stations. Right: estimates of Pickands dependence functions for four pairs of stations at decreasing distances (black: raw madogram estimator; gray: Bernstein projection madogram estimator).

meaning that the extremes at most mildly dependent up to independent. The graph also shows the benefits of the projection method: after projection, the extremal coefficients fall within the admissible range [1, 2], whereas they can be larger than 2 without the projection method.

The right-hand plot of Figure 5 shows four examples of estimated Pickands dependence functions obtained with pairs of sites whose distances are 979.8, 505.9, 390.1 and 158.1 km, respectively (top-left to bottom-right panels). The madogram estimator provides estimates (black lines) that are not convex functions and hence are not Pickands dependence functions themselves. Contrarily, the estimates (gray lines) obtained with the projection estimator are valid Pickands dependence functions.

6 Computational Details

Simulations and data analysis were performed using the R package `ExtremalDep` (https://r-forge.r-project.org/R/?group_id=1998)

Acknowledgements

We thank three anonymous Referees for their valuable suggestions that have improved the presentation of this paper. We also thank Catia Scricciolo, Sabrina Vettori and Sonia Petrone for their valuable support. The work of Philippe Naveau has been supported by ANR-DADA,

LEFE-INSU-Multirisk, AMERISKA, A2C2, and Extremoscope projects and was completed during his visit at the IMAGE-NCAR group in Boulder, CO, USA. The work of Johan Segers was funded by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). The authors acknowledge Meteo France for the precipitation time series. The authors would also very much like to credit the contributors to the R project and LaTeX.

A Proofs

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

$$v_{\boldsymbol{w}}(\boldsymbol{u}) = \min_{i=1}^{d} (u_i^{1/w_i}) - \frac{1}{d} \sum_{i=1}^{d} u_i^{1/w_i}, \qquad \boldsymbol{u} \in [0,1]^d, \tag{A.1}$$

where, by convention, $u^{1/w} = 0$ whenever $w = 0$ and $u \in [0,1]$.

Lemma A.1. *For any cumulative distribution function H on $[0,1]^d$ and for any $\boldsymbol{w} \in S_{d-1}$, we have*

$$\int_{[0,1]^d} v_{\boldsymbol{w}}(\boldsymbol{u}) \, \mathrm{d}H(\boldsymbol{u}) = \frac{1}{d} \sum_{i=1}^{d} \int_0^1 H(1, \ldots, 1, x^{w_i}, 1, \ldots, 1) \, \mathrm{d}x - \int_0^1 H(x^{w_1}, \ldots, x^{w_d}) \, \mathrm{d}x.$$

Proof. Fix $\boldsymbol{w} \in S_{d-1}$. For every $\boldsymbol{u} \in [0,1]^d$ we have

$$\begin{aligned} \min_{i=1}^{d} u_i^{1/w_i} &= 1 - \int_0^1 1(\forall i = 1, \ldots, d : u_i^{1/w_i} \le x) dx \\ &= 1 - \int_0^1 1(\forall i = 1, \ldots, d : u_i \le x^{w_i}) dx \end{aligned}$$

and

$$\frac{1}{d} \sum_{i=1}^{d} u_i^{1/w_i} = 1 - \frac{1}{d} \sum_{i=1}^{d} \int_0^1 1(u_i \le x^{w_i}) dx.$$

Subtracting both expressions and integrating over H yields

$$\begin{aligned} \int_{[0,1]^d} v_{\boldsymbol{w}}(\boldsymbol{u}) \, \mathrm{d}H(\boldsymbol{u}) = \frac{1}{d} \sum_{i=1}^{d} & \int_{[0,1]^d} \int_0^1 1(u_i \le x^{w_i}) \, \mathrm{d}x \, \mathrm{d}H(\boldsymbol{u}) \\ & - \int_{[0,1]^d} \int_0^1 1(\forall i = 1, \ldots, d : u_i \le x^{w_i}) \, \mathrm{d}x \, \mathrm{d}H(\boldsymbol{u}). \end{aligned}$$

Applying Fubini's theorem to both double integrals yields the stated formula. □

Proof of Proposition 2.2. The marginal distribution functions being continuous, the copula C is the joint distribution function of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. For $\boldsymbol{w} \in S_{d-1}$, the multivariate $\boldsymbol{w}$-madogram can thus be written as

$$\nu(\boldsymbol{w}) = \int_{[0,1]^d} \nu_{\boldsymbol{w}}(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}).$$

Next, apply Lemma A.1. Since C is an extreme-value copula with Pickands dependence function A, we find, after some elementary calculations using (1.2) and (1.3),

$$C(x^{w_1}, \ldots, x^{w_d}) = x^{A(\boldsymbol{w})}$$

for all $x \in (0, 1)$. We obtain

$$\begin{aligned} \nu(\boldsymbol{w}) &= \frac{1}{d} \sum_{i=1}^{d} \int_0^1 C(1, \ldots, 1, x^{w_i}, 1, \ldots, 1) \, \mathrm{d}x - \int_0^1 C(x^{w_1}, \ldots, x^{w_d}) \, \mathrm{d}x \qquad \text{(A.2)} \\ &= \frac{1}{d} \sum_{i=1}^{d} \int_0^1 x^{w_i} \, \mathrm{d}x - \int_0^1 x^{A(\boldsymbol{w})} \, \mathrm{d}x, \end{aligned}$$

yielding the first formula stated in the proposition. Solve for $A(\boldsymbol{w})$ to obtain (2.2). Since $\nu(\boldsymbol{w}) + c(\boldsymbol{w}) = A(\boldsymbol{w})/(1 + A(\boldsymbol{w}))$, necessarily $\nu(\boldsymbol{w}) + c(\boldsymbol{w}) < 1$, so that the right-hand side of (2.2) is well-defined. □

Proof of Theorem 2.4. The proof proceeds by expressing the statistics and empirical $\boldsymbol{w}$-madogram $\nu_n(\boldsymbol{w})$ and $\hat{\nu}_n(\boldsymbol{w})$ in terms of the empirical distribution and empirical copula and exploiting known results thereon. For $i = 1, \ldots, d$ and $j = 1, \ldots, n$, let

$$\begin{aligned} \boldsymbol{U}_j &= (U_{j,1}, \ldots, U_{j,d}), \quad U_{j,i} = F_i(X_{j,i}), \\ \hat{\boldsymbol{U}}_j &= (\hat{U}_{j,1}, \ldots, \hat{U}_{j,d}), \quad \hat{U}_{j,i} = F_{n,i}(X_{j,i}) = \frac{1}{n} \sum_{m=1}^{n} 1(X_{m,i} \le X_{j,i}). \end{aligned}$$

Recall $\nu_{\boldsymbol{w}}$ in (A.1). The statistics and empirical $\boldsymbol{w}$-madogram are equal to

$$\nu_n(\boldsymbol{w}) = \frac{1}{n} \sum_{m=1}^{n} \nu_{\boldsymbol{w}}(\boldsymbol{U}_m) = \int_{[0,1]^d} \nu_{\boldsymbol{w}}(\boldsymbol{u}) \, \mathrm{d}C_n(\boldsymbol{u}), \quad \hat{\nu}_n(\boldsymbol{w}) = \int_{[0,1]^d} \nu_{\boldsymbol{w}}(\boldsymbol{u}) \, \mathrm{d}\hat{C}_n(\boldsymbol{u}),$$

respectively, where C_n and $\hat{C}_n$ are the empirical distribution and copula:

$$C_n(\boldsymbol{u}) = \frac{1}{n} \sum_{m=1}^{n} 1(\boldsymbol{U}_m \le \boldsymbol{u}), \quad \hat{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{m=1}^{n} 1(\hat{\boldsymbol{U}}_m \le \boldsymbol{u}), \qquad \boldsymbol{u} \in [0, 1]^d,$$

(component-wise inequalities). By Lemma A.1 we obtain

$$\nu_n(\boldsymbol{w}) = \frac{1}{d} \sum_{i=1}^{d} \int_0^1 C_n(1, \ldots, 1, x^{w_i}, 1, \ldots, 1) \, \mathrm{d}x - \int_0^1 C_n(x^{w_1}, \ldots, x^{w_d}) \, \mathrm{d}x, \qquad \text{(A.3)}$$

and a similar expression is attained for $\hat{v}_n(\boldsymbol{w})$ but with C_n replaced by $\hat{C}_n$. Comparing the latter equation with (A.2) yields

$$\|v_n - v\|_\infty \leq 2\|C_n - C\|_\infty.$$

Standard empirical process arguments yield uniform strong consistency of the empirical copula (Deheuvels 1991). We come to a similar inequality for $\hat{v}_n$. Uniform strong consistency of A_n and $\hat{A}_n$ follows.

Next, consider the empirical processes

$$\mathbb{D}_n = \sqrt{n}(C_n - C), \qquad \hat{\mathbb{D}}_n = \sqrt{n}(\hat{C}_n - C).$$

Combining Equations (A.2) and (A.3) we obtain

$$\sqrt{n}\,\{v_n(\boldsymbol{w}) - v(\boldsymbol{w})\} = \frac{1}{d}\sum_{i=1}^{d} \int_0^1 \mathbb{D}_n(1, \ldots, 1, x^{w_i}, 1, \ldots, 1)\mathrm{d}x - \int_0^1 \mathbb{D}_n(x^{w_1}, \ldots, x^{w_d})\mathrm{d}x$$

and clearly a similar expression is obtained for $\sqrt{n}\,\{\hat{v}_n(\boldsymbol{w}) - v(\boldsymbol{w})\}$ but replacing $\mathbb{D}_n$ with $\hat{\mathbb{D}}_n$. Now, two related results: in the space $l^\infty([0, 1]^d)$ equipped with the supremum norm, $\mathbb{D}_n \rightsquigarrow \mathbb{D}$, as $n \to \infty$, where $\mathbb{D}$ is a C-Brownian bridge, and if Condition 2.3 holds, then $\hat{\mathbb{D}}_n \rightsquigarrow \hat{\mathbb{D}}$, as $n \to \infty$, where $\hat{\mathbb{D}}$ is the Gaussian process defined in (2.7). The map

$$\varphi : l^\infty([0, 1]^d) \to l^\infty(S_{d-1}) : f \mapsto \varphi(f)$$

defined by

$$(\varphi(f))(\boldsymbol{w}) = \frac{1}{d}\sum_{i=1}^{d} \int_0^1 f(1, \ldots, 1, x^{w_i}, 1, \ldots, 1)\,\mathrm{d}x - \int_0^1 f(x^{w_1}, \ldots, x^{w_d})\,\mathrm{d}x$$

is linear and bounded, and therefore continuous. The continuous mapping theorem then implies

$$\sqrt{n}(v_n - v) = \varphi(\mathbb{D}_n) \rightsquigarrow \varphi(\mathbb{D}), \qquad \sqrt{n}(\hat{v}_n - v) = \varphi(\hat{\mathbb{D}}_n) \rightsquigarrow \varphi(\hat{\mathbb{D}}), \qquad n \to \infty,$$

in $l^\infty(S_{d-1})$. The Gaussian process $\hat{\mathbb{D}}$ satisfies

$$\mathrm{P}\{\forall\, i = 1, \ldots, d : \forall\, u \in [0, 1] : \hat{\mathbb{D}}(1, \ldots, 1, u, 1, \ldots, 1) = 0\} = 1.$$

This property follows from the continuity of its sample paths and by the form of the covariance function (2.6). We find, for $\boldsymbol{w} \in S_{d-1}$,

$$(\varphi(\hat{\mathbb{D}}))(\boldsymbol{w}) = -\int_0^1 \hat{\mathbb{D}}(x^{w_1}, \ldots, x^{w_d})\,\mathrm{d}x.$$

Finally, apply the functional delta method (van der Vaart 2000, Ch. 20) to arrive at the conclusion. □

Proof of Proposition 3.1. We have $|B_A(\boldsymbol{w}; k) - A(\boldsymbol{w})| \le \mathrm{E}|A(\mathbf{Y}_k/k) - A(\boldsymbol{w})|$, where $\mathbf{Y}_k = (Y_{k,i};\, i = 1, \ldots, d)$ is a multinomial random vector with k trials, d possible outcomes, and success probabilities $w_1, \ldots, w_d$. Any function $A \in \mathsf{A}$ is Lipschitz-1, so that $|B_A(\boldsymbol{w}; k) - A(\boldsymbol{w})| \le \sum_{i=1}^{d} \mathrm{E}|Y_{k,i}/k - w_i|$. By the Cauchy–Schwarz inequality and the fact that the $Y_{i,k}$ are binomially distributed, it follows that $|B_A(\boldsymbol{w}; k) - A(\boldsymbol{w})| \le \sum_{i=1}^{d} (\mathrm{E}(Y_{k,i}/k - w_i)^2)^{1/2} \le d/(2\sqrt{k})$. □

Proof of Proposition 3.2. On the one hand we have that if $B_A(\boldsymbol{w}; k) \ge \max(w_1, \ldots, w_d)$, then $D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B_A(\boldsymbol{v}_j; k) \ge -1$. Indeed, $\max(w_1, \ldots, w_d)$ is the intersection of the planes $z_0 = 1 - w_1 - w_2 - \cdots - w_{d-1}$, $z_1 = w_1$, …, $z_{d-1} = w_{d-1}$, then by the assumption

$$B_A(\boldsymbol{v}_j; k) \ge z_j, \qquad j = 0, 1, \ldots, d-1.$$

The directional derivatives of B_A calculate for $\boldsymbol{v}_j$, $j = 0, 1, \ldots, d-1$, are equal to

$$D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B(\boldsymbol{v}_j; k) = \begin{cases} D_{\boldsymbol{v}_i - \boldsymbol{v}_0} B(\boldsymbol{v}_0; k) & \text{if } i \ne 0 = j \\ -D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) & \text{if } i = 0 \ne j \\ D_{\boldsymbol{v}_i - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) - D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) & \text{if } i \ne 0 \ne j, \; i \ne j \end{cases} \tag{A.4}$$

which are bounded from below by -1. Then, considering the directional derivatives on both sides of the above inequality we obtain

$$D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B_A(\boldsymbol{v}_j; k) \ge -1, \qquad \forall\, i, j = 0, 1, \ldots, d-1, \; i \ne j,$$

and hence the result.

On the other hand if $D_{\boldsymbol{v}_i - \boldsymbol{v}_j} B_A(\boldsymbol{v}_j; k) \ge -1$, $j = 0, \ldots, d-1$ then $B_A(\boldsymbol{w}; k) \ge \max(w_1, \ldots, w_d)$. Since B_A lies above the tangent plane

$$B_A(\boldsymbol{w}; k) \ge B_A(\boldsymbol{w}^j; k) + (\boldsymbol{w}^j - \boldsymbol{w})^{\mathrm{T}} \nabla B_A(\boldsymbol{w}^j; k), \qquad \forall\, \boldsymbol{w}, \boldsymbol{w}^j \in \mathsf{S}_{d-1}. \tag{A.5}$$

by the convexity assumption, then evaluating this inequality for $\boldsymbol{w}^j = \boldsymbol{v}_j$ for $j \in \{0, 1, \ldots, d-1\}$ we obtain the desired result $B_A(\boldsymbol{w}; k) \ge w_j$ for all $\boldsymbol{w} \in \mathsf{S}_{d-1}$. Indeed, considering (A.5) at $\boldsymbol{w}^j = \boldsymbol{v}_0$ we find, for $\boldsymbol{w} \in \mathsf{S}_{d-1}$,

$$B_A(\boldsymbol{w}; k) \ge 1 + \boldsymbol{w}^{\mathrm{T}} \nabla B_A(\boldsymbol{v}_0; k) = 1 + \sum_{i=1}^{d-1} w_i \, D_{\boldsymbol{v}_i - \boldsymbol{v}_0} B(\boldsymbol{v}_0; k) \ge 1 + \sum_{i=1}^{d-1} w_i \, (-1) = w_d$$

where $w_d = 1 - w_1 - \cdots - w_{d-1}$, as required. Furthermore, considering (A.5) at $\boldsymbol{w}^j = \boldsymbol{v}_j$ for

$j \in \{1, \ldots, d-1\}$ we find for $\boldsymbol{w} \in S_{d-1}$,

$$\begin{aligned} B_A(\boldsymbol{w};\, k) &\geq 1 + (\boldsymbol{w} - \boldsymbol{v}_j)^{\mathrm{T}} \nabla B_A(\boldsymbol{v}_j;\, k) \\ &= 1 + (w_j - 1) D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) + \sum_{\substack{i=1 \\ i \neq j}}^{d} w_i\, D_{\boldsymbol{v}_i - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) \\ &\geq 1 + (w_j - 1) D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) + \sum_{\substack{i=1 \\ i \neq j}}^{d-1} w_i \quad D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) - 1 \\ &= 1 + (w_j - 1) D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) + (1 - w_j - w_d) \quad D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) - 1 \\ &= w_j + w_d \quad 1 - D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) \quad \geq w_j \end{aligned}$$

given that $D_{\boldsymbol{v}_i - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) \geq D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) - 1$ and $1 - D_{\boldsymbol{v}_j - \boldsymbol{v}_0} B(\boldsymbol{v}_j; k) \geq 0$. □

Proof of Proposition 3.3. Firstly, the polynomials in $\mathbb{A}^k$ are nested (e.g., Wang and Ghosh, 2012; Farin, 1986). By the degree-raising property we have

$$B(\boldsymbol{w}; k) = \sum_{\boldsymbol{\alpha} \in \Gamma_k} \beta_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w};\, k) = \sum_{\boldsymbol{\alpha} \in \Gamma_{k+1}} \tilde{\beta}_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\boldsymbol{w};\, k+1) = \tilde{B}(\boldsymbol{w};\, k+1)$$

where

$$\tilde{\beta}_{\boldsymbol{\alpha}} = \sum_{h=1}^{d} \frac{\alpha_h}{k+1} \beta_{\boldsymbol{\alpha} - \boldsymbol{v}_{h-1}}. \tag{A.6}$$

We need to show that the coefficients $\tilde{\beta}_{\boldsymbol{\alpha}}$ satisfy the constraints R1)-R2)-R3). For the case R1) we need to check that

$$\Delta^2_{i,0} \tilde{\beta}_{\boldsymbol{\alpha}} - \sum_{j \neq i} |\Delta_{i,0} \Delta_{j,0} \tilde{\beta}_{\boldsymbol{\alpha}}| \geq 0, \qquad \forall \boldsymbol{\alpha} \in \Gamma_{(k+1)-2}, \quad i = 1, \ldots, d-1.$$

This can be rewritten as

$$\Delta^2_{i,0} \tilde{\beta}_{\boldsymbol{\alpha}} - \sum_{j \quad i} (-1)^{\prime_{s,t}} \Delta_{i,0} \Delta_{j,0} \tilde{\beta}_{\boldsymbol{\alpha}} \geq 0,$$

where $I_{s,t}$ is the set of all the possible combinations with repetition of the set $\{1, 2\}$ in sequences of d 2 terms, $s = 1, \ldots, d$ 2 and $t = 1, \ldots, 2^{d-2}$. Using the relation in (A.6) we have

$$\tilde{\beta}_{\boldsymbol{\alpha}} = \sum_{h=1} \frac{\alpha_h}{k+1} \beta_{\boldsymbol{\alpha} - \boldsymbol{v}_{h-1}} \quad = \sum_{h=1}^{d} \frac{\alpha_h}{k+1} \beta_{\boldsymbol{\alpha} - \boldsymbol{e}_h} + \frac{\alpha_d}{k+1} \beta_{\boldsymbol{\alpha}}.$$

Then, we obtain

$$\begin{aligned}
\Delta^2_{i,0}\tilde{b}_{\boldsymbol{\alpha}} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0}\Delta_{j,0}\tilde{b}_{\boldsymbol{\alpha}} &= \Delta^2_{i,0}\left(\sum_{h=1}^{d} \frac{\alpha_h}{k+1} b_{\boldsymbol{\alpha}-\boldsymbol{e}_h} + \frac{\alpha_d}{k+1} b_{\boldsymbol{\alpha}}\right) \\
&\quad - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0}\Delta_{j,0}\left(\sum_{h=1}^{d} \frac{\alpha_h}{k+1} b_{\boldsymbol{\alpha}-\boldsymbol{e}_h} + \frac{\alpha_d}{k+1} b_{\boldsymbol{\alpha}}\right) \\
&= \sum_{h=1}^{d-1} \frac{h}{k+1}\left(\Delta^2_{i,0} b_{\boldsymbol{\alpha}-\boldsymbol{e}_h} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0}\Delta_{j,0} b_{\boldsymbol{\alpha}-\boldsymbol{e}_h}\right) \\
&\quad + \sum_{h=1}^{d-1} \frac{\alpha_d}{k+1}\left(\Delta^2_{i,0} b_{\boldsymbol{\alpha}} - \sum_{j \neq i} (-1)^{I_{s,t}} \Delta_{i,0}\Delta_{j,0} b_{\boldsymbol{\alpha}}\right) \geq 0,
\end{aligned}$$

and hence the result. For the case R2), using (A.6), it is immediate to verify for the set $\{\tilde{b}_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \Gamma_{k+1} : \boldsymbol{\alpha} = \mathbf{0}$ or $\boldsymbol{\alpha} = (k+1)\boldsymbol{e}_i,\ \forall i = 1, \ldots, d-1\}$ that $\tilde{b}_{\boldsymbol{\alpha}} = b_{\boldsymbol{\alpha}} = 1$. Finally, for the case R3) we need to check that $1 - 1/(k+1) < \tilde{b}_{\boldsymbol{\alpha}}$, where $\{\tilde{b}_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \Gamma_{k+1} : \boldsymbol{\alpha} = \boldsymbol{e}_i$ or $\boldsymbol{\alpha} = k\boldsymbol{e}_i$ or $\boldsymbol{\alpha}_I = k\boldsymbol{e}_i + \boldsymbol{e}_j,\ \forall j \neq i = 1, \ldots, d-1\}$. By definition we have

$$\tilde{b}_{\boldsymbol{e}_i} = \frac{k}{k+1} b_{\boldsymbol{e}_i} + \frac{1}{k+1},\quad \tilde{b}_{k\boldsymbol{e}_i} = \frac{k}{k+1} b_{(k-1)\boldsymbol{e}_i} + \frac{1}{k+1},\quad \tilde{b}_{k\boldsymbol{e}_i+\boldsymbol{e}_j} = \frac{k}{k+1} b_{(k-1)\boldsymbol{e}_i+\boldsymbol{e}_j} + \frac{1}{k+1}.$$

Substituting $\tilde{b}_{\boldsymbol{\alpha}}$, with $\boldsymbol{\alpha} = \boldsymbol{e}_i$, $\boldsymbol{\alpha} = k\boldsymbol{e}_i$, $\boldsymbol{\alpha} = k\boldsymbol{e}_i + \boldsymbol{e}_j$, in the previous inequality we obtain

$$\begin{aligned}
\tilde{b}_{\boldsymbol{\alpha}} &\geq 1 - \frac{1}{k+1} \\
\frac{1}{k+1} 1 + k b_{\boldsymbol{\alpha}-\boldsymbol{v}_{i-1}} &\geq \frac{1}{k+1} 1 + k\left(1 - \frac{1}{k}\right) \\
b_{\boldsymbol{\alpha}-\boldsymbol{v}_{i-1}} &\geq 1 - \frac{1}{k}
\end{aligned}$$

for $i = 1, \ldots, d-1$, and hence the result. Thus the first statement is proven.

Secondly, let A be a Pickands dependence function and consider the Bernstein polynomial

$$A_k(\boldsymbol{w}) = \sum_{\boldsymbol{\alpha} \in \Gamma_k} A(\boldsymbol{\alpha}/k)\ b_{\boldsymbol{\alpha}}(\boldsymbol{w};\ k),$$

that is, $A_k = B_A(\ \cdot\ ;\ k)$ as in (3.3). Constraint R1) holds by assumption (3.8). Since $\max(w_1, \ldots, w_d) \leq A(\boldsymbol{w}) \leq 1$ for all $\boldsymbol{w} \in S_{d-1}$, the constraints in R2) and R3) are satisfied too. Finally, we have uniform convergence $A_k \to A$ by Proposition 3.1. □

Proof of Proposition 3.4. Consider the projection of the madogram estimator on the full space A (rather than on the subspace A_k):

$$\tilde{A}^{\mathrm{MD}}_n = \arg\min_{B \in \mathsf{A}} \|\hat{A}^{\mathrm{MD}}_n - B\|_2.$$

From Theorem 2.4 it follows that $\sqrt{n}(\hat{A}_n^{\mathrm{MD}} - A) \rightsquigarrow Z$ in $L^2(S_{d-1})$ as $n \to \infty$ where Z is a Gaussian process. Theorem 1 in Fils-Villetard et al. (2008) then implies that

$$\sqrt{n}(\tilde{A}_n^{\mathrm{MD}} - A) \rightsquigarrow \underset{Z' \in T_A(A)}{\arg\min} \|Z' - Z\|_2, \qquad n \to \infty.$$

It remains to show that we can replace $\tilde{A}_n^{\mathrm{MD}}$ by $\tilde{A}_{n,k_n}^{\mathrm{MD}}$. It suffices to show that

$$\|\tilde{A}_{n,k_n}^{\mathrm{MD}} - \tilde{A}_n^{\mathrm{MD}}\|_2 = o_p(n^{-1/2}), \qquad n \to \infty.$$

By the first inequality in Lemma 1 in Fils-Villetard et al. (2008) with, in their notation, $\mathsf{F} = \mathsf{A}$ and $\mathsf{G} = \mathsf{A}_k$, we find that

$$\|\tilde{A}_{n,k_n}^{\mathrm{MD}} - \tilde{A}_n^{\mathrm{MD}}\|_2 \le [\delta_{k_n}(2\|\hat{A}_n^{\mathrm{MD}} - \tilde{A}_n^{\mathrm{MD}}\|_2 + \delta_{k_n})]^{1/2},$$

where δ_{k_n} is bounded by the L_2 Hausdorff distance between A and A_{k_n}. Proposition 3.1 yields $\delta_{k_n} = O(k_n^{-1/2})$, which is $o(n^{-1/2})$ by the assumption on k_n. Furthermore, since $A \in \mathsf{A}$, we find, by definition of the projection estimator,

$$\|\hat{A}_n^{\mathrm{MD}} - \tilde{A}_n^{\mathrm{MD}}\|_2 \le \|\hat{A}_n^{\mathrm{MD}} - A\|_2 = O_p(n^{-1/2}), \qquad n \to \infty.$$

Combine these relations to complete the proof. □

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd.,Chichester.

Berghaus, B., A. Bücher, and H. Dette (2013). Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. *Journal de la Société Française de Statistique 154* (1), 116–137.

Bernard, E., P. Naveau, M. Vrac, and O. Mestre (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate 26*, 7929–7937.

Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics 39* (4), 1963–2006.

Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika 84*, 567–577.

Carnicer, J. and J. Peña (1993). Shape preserving representations and optimality of the Bernstein basis. *Advances in Computational Mathematics 1*(2), 173–196.

Chang, I. S., C. A. Hsiung, Y. J. Wu, and C. C. Yang (2005). Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics 32*, 447–466.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters 12*, 429–439.

Ditzian, Z. (1986). Inverse theorems for multidimensional Bernstein operators. *Pacific Journal of Mathematics 121* (2), 293–319.

Falk, M., J. Hüsler, and R. D. Reiss (2010). *Laws of Small Numbers: Extremes and Rare Events* (Third ed.). Birkhäuser Boston.

Farin, G. (1986). Triangular Bernstein–Bézier patches. *Computer Aided Geometric Design 3* (2), 83–127.

Fils-Villetard, A., A. Guillou, and J. Segers (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics 36* (3), 369–382.

Fonseca, C., L. Perreira, H. Ferreira, and A. Martins (2015). Generalized madogram and pairwise dependence of maxima over two regions of a random field. *Kybernetika 51* (2), 193–211.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics 37* (5B), 2990–3022.

Goldfarb, D. and A. Idnani (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming 27* (1), 1–33.

Gudendorf, G. and J. Segers (2011). Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis 102*, 37–47.

Gudendorf, G. and J. Segers (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference 142*, 3073–3085.

Hall, P. and N. Tajvidi (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli 6* (5), 835–844.

Hüsler, J. and R. Reiss (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters 7*, 283–286.

Kotz, S. and S. Nadarajah (2000). *Extreme value distributions theory and applications*. Imperial College Press.

Lai, M.-J. (1993). Some sufficient conditions for convexity of multivariate Bernstein-Bézier polynomials and box spline surfaces. *Studia Scientiarum Mathematicarum Hungarica 28* (3), 363–374.

Li, F. J. (2011). Interpolation and convergence of Bernstein-Bézier coefficients. *Acta Mathematica Sinica, English Series 27* (9), 1769–1782.

Lorentz, G. G. (1986). *Bernstein Polynominals* (Second ed.). Chelsea Publishing Company, New York.

Naveau, P., A. Guillou, D. Cooley, and J. Diebolt (2009). Modelling pairwise dependence of maxima in space. *Biometrika 96* (1), 1–17.

Padoan, S. A. (2011). Multivariate extreme models based on underlying skew-t and skew-normal distributions. *Journal of Multivariate Analysis 102* (5), 977–991.

Padoan, S. A. (2013). Extreme value analysis. *Encyclopedia of Environmetrics*.

Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics 26*, 373–393.

Petrone, S. (2004). On the role of mixtures in Bayesian nonparametrics. *Atti della XLII Riunione Scientifica della Societa Italiana di Statistica 1*, 257–268.

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

Ressel, P. (2013). Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis 117*, 246–256.

Sauer, T. (1991). Multivariate Bernstein polynomials and convexity. *Computer Aided Geometric Design 8* (6), 465–478.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli 18* (3), 764–782.

Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript. University of North Carolina, Chapel Hill, U. S.

Smith, R. L., J. A. Tawn, and H. K. Yuen (1990). Statistics of multivariate extremes. *Int. Statist. Rev. 58*, 47–58.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika 77* (2), 245–253.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Wang, J. and S. Ghosh (2012). Shape restricted nonparametric regression with Bernstein polynomials. *Computational Statistics and Data Analysis 56* (9), 2729–2741.

Zhang, D., M. T. Wells, and L. Peng (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis 99* (4), 577–588.