Bernoulli **22**(1), 2016, 213–241
DOI: 10.3150/14-BEJ656

Polynomial Pickands functions

SIMON GUILLOTTE[1] and FRANÇOIS PERRON[2]

[1]*Département de Mathématiques, Université du Québec à Montréal, 201 Ave du Président-Kennedy, Montréal, Québec H2X 3Y7, Canada. E-mail: guillotte.simon@uqam.ca*
[2]*Département de Mathématiques et Statistique, Université de Montréal, C.P. 6128, Succursale Centreville, Montréal, Québec H3C 3J7, Canada. E-mail: perronf@dms.umontreal.ca*

Pickands dependence functions characterize bivariate extreme value copulas. In this paper, we study the class of polynomial Pickands functions. We provide a solution for the characterization of such polynomials of degree at most $m+2$, $m \geq 0$, and show that these can be parameterized by a vector in $\mathbb{R}^{m+1}$ belonging to the intersection of two ellipsoids. We also study the class of Bernstein approximations of order $m+2$ of Pickands functions which are shown to be (polynomial) Pickands functions and parameterized by a vector in $\mathbb{R}^{m+1}$ belonging to a polytope. We give necessary and sufficient conditions for which a polynomial Pickands function is in fact a Bernstein approximation of some Pickands function. Approximation results of Pickands dependence functions by polynomials are given. Finally, inferential methodology is discussed and comparisons based on simulated data are provided.

Keywords: Bernstein's theorem; extreme value copulas; Lorentz degree; Pickands dependence function; polynomials; spectral measure

1. Introduction

Bivariate extreme value copulas are characterized by the Pickands dependence functions, these are functions $A:[0,1] \to \mathbb{R}$ which satisfy the conditions:

1. *Boundary conditions*: $(1-t) \vee t \leq A(t) \leq 1$, $t \in [0,1]$.
2. *Convexity condition*: A is convex.

In the presence of the *convexity condition*, the above *boundary conditions* are simply saying that the lines $\ell_1(t) = 1-t$ and $\ell_2(t) = t$, $t \in [0,1]$, are both *support lines* of A at $t=0$ and $t=1$ respectively, and it follows that they can be equivalently replaced by

1. *Endpoint conditions*: $A(0) = 1 = A(1)$ and $-1 \leq A'(0)$, $A'(1) \leq 1$ (see, e.g., Roberts and Varberg [24], page 14, problem K). Note also that in the case where A is twice differentiable the *convexity condition* can be replaced by $A''(t) \geq 0$, for all $t \in [0,1]$.

Let $\mathcal{A}$ be the space of Pickands dependence functions. For $A \in \mathcal{A}$, the copula is

$$C_A(u,v) = \exp\left\{\log(uv) A\left(\frac{\log v}{\log uv}\right)\right\}, \qquad 0 < u, v \leq 1, \tag{1.1}$$

and C_A has a density if A' is absolutely continuous on $[0,1]$. Bivariate extreme value copulas may also be parameterized by the so-called spectral measure H, that is, a positive measure on $[0,1]$ such that $\int_{[0,1]} w H(\mathrm{d}w) = 1 = \int_{[0,1]} (1-w) H(\mathrm{d}w)$. Let $\mathcal{H}$ denote the space of spectral

1350-7265

measures. The one-to-one correspondence between $\mathcal{A}$ and $\mathcal{H}$, is given by

$$A(t) = \int_{[0,1]} \{(1-t)w\} \vee \{t(1-w)\} H(\mathrm{d}w), \qquad t \in [0, 1], \tag{1.2}$$

and $H([0, w]) = 1 + A'_{+}(w)$, for all $w \in [0, 1)$, where A'_{+} is the right derivative of A, see Beirlant *et al.* [1].

In the past literature, parametric models for the function A have been studied for instance in Tawn [30], Coles and Tawn [5], Joe *et al.* [17], Ledford and Tawn [21] and Dupuis and Tawn [8]. The modeling of H has also been considered, as it offers some advantages, especially when extensions to the multivariate case are desired. Inference within parametric families of spectral measures can be found in Boldi and Davison [2], Coles and Tawn [5], Coles and Tawn [6], de Haan *et al.* [7], Einmahl *et al.* [9], Joe *et al.* [17], Ledford and Tawn [21] and Smith [28]. Polynomial splines models have also been proposed by Guillotte *et al.* [14] and Fougères *et al.* [11].

In view of the infinite-dimensional nature of the space of Pickands functions, inference is often done nonparametrically for maximum flexibility, see for instance Capéraà *et al.* [4], Hall and Tajvidi [15], Fils-Villetard *et al.* [10] and recently Bücher *et al.* [3], to cite only a few. However, the nonparametric estimators do not usually satisfy the properties of Pickands functions for finite samples and are often modified to do so. This can be cumbersome, see, for instance, Fils-Villetard *et al.* [10] and the references therein. One alternative to the nonparametric approaches is based on having a series of nested models indexed by m, $m \geq 0$. Each model is parametric but m is not bounded. Within each model, a genuine estimator of the Pickands dependence function is proposed. In this spirit, we consider modeling the Pickands function A using polynomials on $[0, 1]$. The latter are natural candidates here, the quadratic case being known as the symmetric mixed model, while the cubic case corresponds to the asymmetric mixed model, see Beirlant *et al.* [1], pages 308 and 309, and the references therein. An attempt at characterizing the space of Pickands polynomials for higher degrees was made in Klüppelberg and May [20], page 1472, Theorem 2.5. It is stated there that a polynomial (in the power basis) $A(t) = 1 - (\sum_{k=2}^{m} a_k)t + \sum_{k=2}^{m} a_k t^k$, $t \in [0, 1]$, is a Pickands dependence function if and only if its coefficients satisfy the four following conditions:

$$0 \leq a_2, \qquad 0 \leq \sum_{k=2}^{m} a_k, \qquad 0 \leq \sum_{k=2}^{m} (k-1)a_k \leq 1 \quad \text{and} \quad 0 \leq \sum_{k=2}^{m} k(k-1)a_k.$$

It turns out that these conditions fail to be sufficient for $m = 4$. In fact, a counterexample dates back to Beirlant *et al.* [1], page 308: the polynomial $A(t) = 1 - t^3 + t^4$ satisfies the above conditions but is not convex. We will return to this example in Section 3.

First, Section 2 is a short preliminary section giving the essential definitions and general mathematical results concerning the Bernstein basis used throughout the paper. The reason for this choice of basis here is, essentially, its ability to reflect geometric and algebraic particularities of this problem on the coefficients. This will be reinforced, with an example and pointers to the text in the concluding remarks of Section 7. In Section 3, the aim is to characterize, for each polynomial having degree at most m, $m \geq 0$, the domain in which the coefficients of the polynomial (in the Bernstein basis) must belong. We are looking for a constructive solution leading ultimately to a parameterization, which could be directly applicable in the search of the maximum likelihood estimate, for instance. We approach the problem by exploiting the one-to-one correspondence between the Pickands function and the related spectral measure mentioned above. For polynomials, this boils down to the link between the Pickands function and its second derivative (a nonnegative polynomial). The solution therefore exploits the characterization of nonnegative polynomials given by Lukács, see Karlin and Shapley [19]. We subsequently refer to this solution as being the solution to the *full model*. In Section 4, the focus is on characterizing, for each degree $m \geq 0$, the space of polynomials that can be obtained from a Bernstein approximation of a Pickands dependence function. We call this model the *submodel*. We try to give an answer to the question: what is the gap between the *full model* and the *submodel*? Now in Section 5, we use tools from approximation theory and probability for obtaining accurate bounds for measuring the closeness between the space of Pickands dependence functions and the one obtained from the *submodel*. Note that Bernstein approximations of copulas have appeared in the past literature and some of their properties have been studied in Sancetta and Satchell [25] and Sancetta and Satchell [26]. Finally, in Section 6, we present the results of a simulation comparing the maximum likelihood estimator from the full model with the one from the submodel and a version of the popular nonparametric "CFG" estimator in Capéraà *et al.* [4]. Our estimators should be easily implemented, and could become part of standard extreme value statistics packages, such as the EVD package in R. The existing packages offer only polynomial models of degree at most three, this may be explained by the absence of a successful characterization for higher degrees in the previous literature.

2. Bernstein polynomials

Let $\mathcal{P}_m$ be the space of polynomials on the interval [0, 1], with degree at most m, $m \geq 0$. We shall represent polynomials in $\mathcal{P}_m$ using the commonly called Bernstein basis. For self-containedness of the exposition, we begin with a short (preliminary) section describing some of the key properties of this basis that will serve in the following developments. The connection between this basis and the binomial distribution provides a very powerful tool in some of the proofs. We are also going to exploit the binomial identities given below.

2.1. The basis

Henceforth, let S_n be a Binomial(n, x) random variable, $n \geq 0$, where the value of $x \in [0, 1]$ will always be clearly indicated by the context. The members of the Bernstein basis of degree m, $m \geq 0$, are the polynomials

$$b_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k} = \mathrm{P}_x(S_m = k), \qquad x \in [0, 1],$$

for $0 \le k \le m$. The coefficients of a polynomial $P \in \mathcal{P}_m$ in the Bernstein basis of degree m, will be denoted by $\{c(k,m;\,P) : k = 0, \dots, m\}$, and so for $P \in \mathcal{P}_m$ we have

$$P(x) = \sum_{k=0}^{m} c(k,m;\,P) b_{k,m}(x) = \mathrm{E}_x\big[c(S_m, m;\,P)\big], \qquad x \in [0,1].$$

Let Δ be the forward difference operator. Here, when applied to a function f of two arguments: $(k,m) \mapsto f(k,m)$, it is understood that Δ operates on the first argument: $\Delta f(k,m) = f(k+1,m) - f(k,m)$ and $\Delta^2 f(k,m) = f(k+2,m) - 2f(k+1,m) + f(k,m)$, for all (k,m). Here are some basic results related to the Bernstein basis.

Proposition 2.1. *For $0 \le k \le m$, we have*

(i) $b'_{k,m} = m(b_{k-1,m-1} - b_{k,m-1}) = -m\Delta b_{k-1,m-1}$, *for $m \ge 1$, with the convention $b_{j,m} = 0$ for $j \notin \{0,\dots,m\}$,*
(ii) *for $0 \le \ell \le n$, $b_{k,m} b_{\ell,n} = \frac{\binom{m}{k}\binom{n}{\ell}}{\binom{m+n}{k+\ell}} b_{k+\ell,m+n}$,*
(iii) $\int_0^t b_{k,m}(w)\,\mathrm{d}w = \frac{1}{m+1}\sum_{j=k+1}^{m+1} b_{j,m+1}(t) = \frac{1}{m+1} - \int_t^1 b_{k,m}(w)\,\mathrm{d}w$, $t \in [0,1]$.

The following are direct but useful consequences of the above proposition. Properties (i) and (ii) (obtained using the summation by parts formula) show how derivatives of polynomials represented in the Bernstein basis act on the coefficients.

Proposition 2.2. *Let $P \in \mathcal{P}_m$, with P', P'', its first and second derivative respectively.*

(i) $c(k,m-1;\,P') = m\Delta c(k,m;\,P)$, $0 \le k \le m-1$, $m \ge 1$, *so that*

$$P'(t) = m\mathrm{E}_t\big\{\Delta c(S_{m-1}, m;\,P)\big\}, \qquad t \in [0,1],$$

(ii) $c(k,m-2;\,P'') = m(m-1)\Delta^2 c(k,m;\,P)$, $0 \le k \le m-2$, $m \ge 2$, *so that*

$$P''(t) = m(m-1)\mathrm{E}_t\big\{\Delta^2 c(S_{m-2}, m;\,P)\big\}, \qquad t \in [0,1],$$

(iii) *for $0 \le i \le m$ and $0 \le j \le n$, and $t \in [0,1]$,*

$$\int_0^t b_{i,m}(x) b_{j,n}(x)\,\mathrm{d}x = \frac{\binom{m}{i}\binom{n}{j}}{\binom{m+n}{i+j}} \frac{1}{m+n+1} \mathrm{E}_t\big\{\mathbb{1}(i+j < S_{m+n+1})\big\},$$

$$\int_t^1 b_{i,m}(x) b_{j,n}(x)\,\mathrm{d}x = \frac{\binom{m}{i}\binom{n}{j}}{\binom{m+n}{i+j}} \frac{1}{m+n+1} \mathrm{E}_t\big\{\mathbb{1}(i+j \ge S_{m+n+1})\big\}.$$

2.2. Bernstein approximations

In Section 4, we will construct a submodel based on Bernstein approximations of Pickands functions. In general, the mth-order Bernstein approximation of any function $f : [0,1] \to \mathbb{R}$ is given by

$$B_m(f,t)=\sum_{k=0}^{m} f(k/m)b_{k,m}(t)=\mathrm{E}_t\big[f(S_m/m)\big],\qquad t\in[0,1].$$

It transforms f into the polynomial $B_m(f,\cdot)\in\mathcal{P}_m$, with coefficients equal to the function values on the uniformly spaced grid $\{k/m : k=0,\dots,m\}$. Notice that $B_m(f,0)=f(0)$ and $B_m(f,1)=f(1)$. However, $B_m(f,\cdot)$ does not necessarily interpolate f on the grid. As a first implication of Proposition 2.2 above, we get that the convexity of f implies that of $B_m(f,\cdot)$.

2.3. Binomial identities

We will make extensive use of the following identities. Let $X_1,\dots,X_m$ be independent Bernoulli(t) random variables, $t\in[0,1]$, and let $f:\{0,\dots,m\}\to\mathbb{R}$ be a function. Here $S_m=\sum_{k=1}^m X_k\sim$ Binomial(m,t), and

$$\begin{aligned}\mathrm{E}_t\big\{f(S_m)\big\}&=\mathrm{E}_t\big\{f(S_{m-1}+X_m)\big\}=\mathrm{E}_t\big[\mathrm{E}_t\big\{f(S_{m-1}+X_m)\mid S_{m-1}\big\}\big]\\&=t\mathrm{E}_t\big\{f(S_{m-1}+1)\big\}+(1-t)\mathrm{E}_t\big\{f(S_{m-1})\big\},\end{aligned}\tag{2.1}$$

$$\begin{aligned}\mathrm{E}_t\big\{S_m f(S_m)\big\}&=m\mathrm{E}_t\big\{X_m f(S_{m-1}+X_m)\big\}=m\mathrm{E}_t\big[\mathrm{E}_t\big\{X_m f(S_{m-1}+X_m)\mid S_{m-1}\big\}\big]\\&=mt\mathrm{E}_t\big\{f(S_{m-1}+1)\big\},\end{aligned}\tag{2.2}$$

so that from (2.1) and (2.2), we have $\mathrm{E}_t\{(m-S_m)f(S_m)\}=m(1-t)\mathrm{E}_t\{f(S_{m-1})\}$. These identities are special cases of more general identities developed for the exponential families, see Hudson [16].

3. The characterization

The problem of characterizing the polynomial Pickands functions is solved in Section 3.2 below. To do so, we will take further advantage of the one-to-one correspondence between the spectral measure H and the related Pickands function A. In fact, under absolute continuity of A', we will show that the problem of modeling A (or H) boils down to modeling some nonnegative function h on $(0,1)$ satisfying the condition

$$\left\{\int_0^1(1-w)h(w)\,\mathrm{d}w\right\}\vee\left\{\int_0^1 wh(w)\,\mathrm{d}w\right\}\le 1.\tag{3.1}$$

Although this is (briefly) mentioned in Beirlant *et al.* [1], page 269, we provide the details in Section 3.1 because the use of this special function h is a key element in the paper. Note that this framework is slightly more general then what is really needed for obtaining the characterization later on, because we will be working with polynomials rather than merely absolutely continuous functions. However, it is effortless to do so here.

3.1. A representation theorem

Let $A \in \mathcal{A}$, let H on $[0, 1]$ be the spectral measure associated with A and let μ be the Lebesgue measure on $[0, 1]$. Assume that A' is absolutely continuous on $[0, 1]$, then, in particular, the copula C_A in equation (1.1) has a density with respect to the Lebesgue measure on $[0, 1]^2$. In this case, we derive a convenient integral representation for $A(t)$, $t \in [0, 1]$, in terms of the Radon–Nikodým derivative h of the restriction of H on $(0, 1)$. Note that H may still have point masses at 0 and 1, which can be seen by $H(\{0\}) = 1 + A'(0)$ and $H(\{1\}) = 1 - A'(1)$. Under the above regularity condition on A', we will show that the knowledge (almost everywhere) of A'' alone is enough to evaluate A, and similarly, the knowledge of the Radon–Nikodým derivative h of the restriction of H on $(0, 1)$ is enough to know the measure H. In particular, we have $h = A''$ almost everywhere on $(0, 1)$.

Theorem 3.1 (Integral representation).

(i) *Let $A \in \mathcal{A}$, let H be the spectral measure associated with A. Assume that A' is absolutely continuous and let $h = A''$ almost everywhere on* $(0, 1)$. *We have*

$$A(t) = 1 - \int_0^1 \bigl[\bigl\{(1-t)w\bigr\} \wedge \bigl\{t(1-w)\bigr\}\bigr] h(w)\,\mathrm{d}w, \qquad t \in [0, 1], \tag{3.2}$$

and the function h satisfies the condition (3.1). *Let μ be the Lebesgue measure and δ_x denote a Dirac measure at x, the spectral measure is given by $H = h_0\delta_0 + \mathring{H} + h_1\delta_1$, with*

$$\begin{aligned} \mathrm{d}\mathring{H}/\mathrm{d}\mu &= h, \\ h_0 = H\bigl(\{0\}\bigr) &= 1 - \int_0^1 (1-w)h(w)\,\mathrm{d}w, \\ h_1 = H\bigl(\{1\}\bigr) &= 1 - \int_0^1 wh(w)\,\mathrm{d}w. \end{aligned} \tag{3.3}$$

(ii) *Conversely, let $h : (0, 1) \to \mathbb{R}$ be a nonnegative function satisfying condition* (3.1). *The measure H given by* (3.3) *is a spectral measure, its corresponding Pickands function A is given by* (3.2), *and we have $A'' = h$ almost everywhere on* $(0, 1)$.

Proof. (i) Integration by parts yields

$$\int_0^t wh(w)\,\mathrm{d}w = tA'(t) + \bigl\{1 - A(t)\bigr\}, \qquad t \in [0, 1], \tag{3.4}$$

and

$$\int_t^1 (1-w)h(w)\,\mathrm{d}w = \bigl\{1 - A(t)\bigr\} - (1-t)A'(t), \qquad t \in [0, 1], \tag{3.5}$$

and by combining these two equations, we obtain

$$\begin{aligned}\{1 - A(t)\} &= (1-t)\int_0^t wh(w)\,\mathrm{d}w + t\int_t^1 (1-w)h(w)\,\mathrm{d}w \\ &= \int_0^1 \{(1-t)w\} \wedge \{t(1-w)\} h(w)\,\mathrm{d}w, \qquad t \in [0,1].\end{aligned}$$

Letting $t = 1$ in (3.4) and $t = 0$ in (3.5) shows that the condition (3.1) is satisfied:

$$\int_0^1 wh(w) = A'(1) \leq 1 \quad \text{and} \quad \int_0^1 (1-w)h(w) = -A'(0) \leq 1.$$

From the one-to-one correspondence between Pickands functions and spectral measures, the only thing left to show is that if H is the measure given by (3.3), then it is the spectral measure associated to A, that is, equation (1.2) holds:

$$\begin{aligned}A(t) &= 1 - \int_0^1 \big[\{(1-t)w\} \wedge \{t(1-w)\}\big] h(w)\,\mathrm{d}w \\ &= (1-t)\left(1 - \int_0^t wh(w)\,\mathrm{d}w\right) + t\left(1 - \int_t^1 (1-w)h(w)\,\mathrm{d}w\right) \\ &= (1-t)\left(h_1 + \int_t^1 wh(w)\,\mathrm{d}w\right) + t\left(h_0 + \int_0^t (1-w)h(w)\,\mathrm{d}w\right) \\ &= (1-t)h_1 + th_0 + \int_0^1 \big[\{(1-t)w\} \vee \{t(1-w)\}\big] h(w)\,\mathrm{d}w \\ &= \int_{[0,1]} \{(1-t)w\} \vee \{t(1-w)\} H(\mathrm{d}w), \qquad t \in [0,1],\end{aligned} \tag{3.6}$$

as claimed.

(ii) For the converse, if $h : (0,1) \to [0,\infty)$ satisfies the condition (3.1), then the measure H given by (3.3) is a spectral measure: it is clearly positive, and

$$\int_{[0,1]} wH(\mathrm{d}w) = h_1 + \int_0^1 wh(w)\,\mathrm{d}w = 1 = h_0 + \int_0^1 (1-w)h(w)\,\mathrm{d}w = \int_{[0,1]} (1-w)H(\mathrm{d}w).$$

Let A be its corresponding Pickands function given by equation (1.2). The above equalities that lead to (3.6) show that A is also given by (3.2). Finally, a direct calculation shows that $A'' = h$ almost everywhere on $(0,1)$. □

3.2. The space of polynomial Pickands functions

The space $\mathcal{A}_m = \mathcal{P}_m \cap \mathcal{A}$ corresponds to polynomial Pickands functions with degree at most m, $m \geq 0$. The reader should refer to Section 2 for notations used in the following. It is immediate that $\mathcal{A}_0 = \mathcal{A}_1$, and $A \in \mathcal{A}_0$ if and only if $A(t) = 1$, for all $t \in [0, 1]$. The nontrivial cases start with Pickands polynomials of degree at least two. We will characterize $\mathcal{A}_m$, for all $m \geq 2$. Note that in the Bernstein basis, the *endpoint conditions* are quite simple to verify because they are directly related to the first and last two coefficients only.

Proposition 3.2. *Let $A \in \mathcal{P}_{m+2}$, $m \geq 0$. We have $A \in \mathcal{A}_{m+2}$ if and only if the two following conditions are verified*:

1. *Endpoint conditions*:

$$c(0, m+2; A) = 1 = c(m+2, m+2; A),$$

and

$$c(1, m+2; A) \wedge c(m+1, m+2; A) \geq (m+1)/(m+2).$$

2. *Convexity condition*: $A''(t) \geq 0$, $t \in [0, 1]$.

Proof. We need to verify the *endpoint conditions*, that is, $A(0), A(1) = 1, -A'(0), A'(1) \leq 1$. In the Bernstein basis, by Proposition 2.2, we have

$$\begin{aligned} c(0, m+2; A) = A(0) = 1 \quad &\text{and} \quad c(m+2, m+2; A) = A(1) = 1, \\ 1 \geq -A'(0) &= -c\big(0, m+1; A'\big) = -(m+2)\Delta c(0, m+2; A) \\ &= (m+2)\big\{1 - c(1, m+2; A)\big\}, \end{aligned}$$

and

$$\begin{aligned} 1 \geq A'(1) &= c\big(m+1, m+1; A'\big) = (m+2)\Delta c(m+1, m+2; A) \\ &= (m+2)\big\{1 - c(m+1, m+2; A)\big\}. \end{aligned}$$

□

It is therefore the *convexity condition* above that is more delicate to express in terms of the coefficients of A. However, the spaces $\mathcal{A}_2$ and $\mathcal{A}_3$ are still easy to characterize, as can be seen in the following example.

Example 1 ($\mathcal{A}_2$ and $\mathcal{A}_3$). We have $A \in \mathcal{A}_2$ if and only if $c(0, 2; A) = 1 = c(2, 2; A)$ and $1/2 \leq c(1, 2; A) \leq 1$. To see this, let $A \in \mathcal{P}_2$. The polynomial A satisfies the *endpoint conditions* of Proposition 3.2 if and only if $c(0, 2; A) = 1 = c(2, 2; A)$ and $1/2 \leq c(1, 2; A)$. Moreover $A''(t) = A''(0)$, for all $t \in [0, 1]$, and by Proposition 2.2, $A''(0) = 2\Delta^2 c(0, 2; A) = 4\{1 - c(1, 2; A)\}$. Therefore, A is convex if and only if $c(1, 2; A) \leq 1$.

We have $A \in \mathcal{A}_3$ if and only if $c(0, 3; A) = 1 = c(3, 3; A)$, and the couple $(c(1, 3; A), c(2, 3; A))$ belongs to the polytope derived from the four linear inequalities:

$$c(1, 3; A) \wedge c(2, 3; A) \geq 2/3$$

and

$$\big\{2c(1, 3; A) - c(2, 3; A)\big\} \vee \big\{2c(2, 3; A) - c(1, 3; A)\big\} \leq 1.$$

To show this, let $A \in \mathcal{P}_3$. The polynomial A satisfies the *endpoint conditions* of Proposition 3.2 if and only if $c(0,3;A) = 1 = c(3,3;A)$, and $c(1,3;A) \wedge c(2,3;A) \geq 2/3$. Also, since $A''(t) = (1-t)A''(0) + tA''(1)$, $t \in [0,1]$, we have $A'' \geq 0$, in $[0,1]$ if and only if $A''(0) \wedge A''(1) \geq 0$. By evaluating $A''(k) = 6\Delta^2 c(k,3;A)$ for $k = 0,1$, we get that A is convex if and only if $\{2c(1,3;A) - c(2,3;A)\} \vee \{2c(2,3;A) - c(1,3;A)\} \leq 1$.

For $m \in \{0,1\}$, finding $\mathcal{A}_{m+2}$ explicitly (in Example 1 above) is easy mostly because $A'' \in \mathcal{P}_1$ is nonnegative if and only if $A''(0) \wedge A''(1) \geq 0$. When $A \in \mathcal{A}_{m+2}$ with $m > 1$ things become more complicated, as the following example illustrates.

Example 2 (The counterexample polynomial in $\mathcal{P}_4$). Consider the polynomial $A(t) = 1 - t^3 + t^4$, which served as the counterexample provided by Beirlant *et al.* [1] discussed in the introduction. Here, $A''(t) = 12t(t - 1/2)$, so $A''(0) \wedge A''(1) = 0$ but $\{t : A''(t) \geq 0\} = \{0\} \cup [1/2, 1] \neq [0,1]$, and therefore $A \in \mathcal{P}_4 \setminus \mathcal{A}_4$.

Essentially, it all boils down to finding the coefficients making A'' nonnegative, and the problem is now reoriented towards the characterization of such A''. We will essentially be using the *Integral representation* Theorem 3.1 and a characterization of the nonnegative polynomials on $[0,1]$ due to Lukács. The main result is given in Theorem 3.5, while a parametric representation of the polynomial Pickands functions and the geometric shape of the corresponding parameter space is provided via Theorem 3.6. To get to these results, we first need two technical lemmas, namely Lemma 3.3 and Lemma 3.4.

Because of the new orientation of the problem as mentioned above, it is convenient to introduce new spaces $\mathcal{H}_m \subset \mathcal{P}_m$, $m \geq 0$. We say that a polynomial h belongs to $\mathcal{H}_m$, $m \geq 0$, when the following conditions are satisfied:

1. *Endpoint derivatives conditions:*

$$\frac{1}{m+1}\sum_{j=0}^{m}\left(1 - \frac{j+1}{m+2}\right)c(j,m;h) \leq 1 \quad \text{and} \quad \frac{1}{m+1}\sum_{j=0}^{m}\frac{j+1}{m+2}c(j,m;h) \leq 1. \tag{3.7}$$

2. *Nonnegativity condition: $h(t) \geq 0$, $t \in [0,1]$.*

Remark. These are the same conditions as those given at the very beginning of this section when restricted to polynomials, in particular, the *endpoint derivatives conditions* (3.7) come directly from the condition (3.1), by using Proposition 2.1. When a polynomial A on $[0,1]$ has an integral reprentation (3.2), the condition (3.1) corresponds to $-1 \leq A'(0), A'(1) \leq 1$, hence the name *endpoint derivatives conditions*. From the expression (3.2) we get the following result (notice the resemblance between the polynomial A and its coefficients in the Bernstein basis…).

Lemma 3.3. *Let $m \geq 0$.*

(i) *If $h \in \mathcal{P}_m$ and*

$$A(t) = 1 - \int_0^1 \left[\{(1-t)w\} \wedge \{t(1-w)\}\right] h(w)\,\mathrm{d}w, \qquad t \in [0,1], \tag{3.8}$$

then $A \in \mathcal{P}_{m+2}$ and its kth coefficient, $k = 0, 1, \ldots, m+2$, in the Bernstein basis is

$$
\begin{aligned}
&c(k, m+2; A) \\
&= 1 - \frac{1}{m+1} \sum_{j=0}^{m} \left[\left\{ \left(1 - \frac{k}{m+2}\right) \frac{j+1}{m+2} \right\} \wedge \left\{ \frac{k}{m+2} \left(1 - \frac{j+1}{m+2}\right) \right\} \right] c(j, m; h).
\end{aligned}
\tag{3.9}
$$

(ii) *If $A \in \mathcal{P}_{m+2}$ and $h = A''$, then $h \in \mathcal{P}_m$ and*

$$
c(k, m; h) = (m+2)(m+1)\Delta^2 c(k, m+2; A), \qquad k = 0, \ldots, m.
$$

Proof. (i) We have the equality

$$
\begin{aligned}
&\int_0^1 \left[\{(1-t)w\} \wedge \{t(1-w)\}\right] h(w)\,\mathrm{d}w \\
&\quad = \int_0^t w h(w)\,\mathrm{d}w + t \int_t^1 h(w)\,\mathrm{d}w - t \int_0^1 w h(w)\,\mathrm{d}w.
\end{aligned}
\tag{3.10}
$$

Using Proposition 2.2(iii) and the binomial identities (2.2), we obtain:

$$
\begin{aligned}
\int_0^t w h(w)\,\mathrm{d}w &= \frac{1}{m+1} \sum_{j=0}^{m} \frac{j+1}{m+2} \mathrm{E}_t\left\{\mathbb{1}(j+1 < S_{m+2})\right\} c(j, m; h), \\
t \int_t^1 h(w)\,\mathrm{d}w &= \frac{t}{m+1} \sum_{j=0}^{m} \mathrm{E}_t\left\{\mathbb{1}(j \geq S_{m+1})\right\} c(j, m; h) \\
&= \frac{1}{m+1} \sum_{j=0}^{m} \mathrm{E}_t\left\{ \frac{S_{m+2}}{m+2} \mathbb{1}(j+1 \geq S_{m+2}) \right\} c(j, m; h), \\
t \int_0^1 w h(w)\,\mathrm{d}w &= \frac{1}{m+1} \sum_{j=0}^{m} \frac{j+1}{m+2} \mathrm{E}_t\left(\frac{S_{m+2}}{m+2} \right) c(j, m; h),
\end{aligned}
$$

and putting the three pieces together shows that (3.10) equals

$$
\mathrm{E}_t\left(\frac{1}{m+1} \sum_{j=0}^{m} \left[\left\{ \left(1 - \frac{S_{m+2}}{m+2}\right) \frac{j+1}{m+2} \right\} \wedge \left\{ \frac{S_{m+2}}{m+2} \left(1 - \frac{j+1}{m+2}\right) \right\} \right] c(j, m; h) \right),
$$

which in turn equals $\mathrm{E}_t\{1 - c(S_{m+2}, m+2; A)\}$, and this gives (3.9).

(ii) This is directly obtained from Proposition 2.2. □

Example 3 ($\mathcal{A}_4$). We can show directly that $h \in \mathcal{P}_2$ is nonnegative on [0, 1] if and only if

$$
c(0, 2; h) \wedge c(2, 2; h) \geq 0 \quad \text{and} \quad c(1, 2; h) \geq -\sqrt{c(0, 2; h) c(2, 2; h)}.
$$

Indeed, if $h \in \mathcal{P}_2$ is nonnegative on $[0, 1]$, then necessarily $c(0, 2; h) = h(0) \geq 0$ and $c(2, 2; h) = h(1) \geq 0$. Since the latter conditions need to be satisfied, lets assume them and look at the remaining coefficient $c(1, 2; h)$. If $c(1, 2; h) \geq 0$, then h is nonnegative on $[0, 1]$. If however $c(1, 2; h) < 0$, then we have $\Delta^2 c(0, 2; h) > 0$ and $-\Delta c(0, 2; h)/\Delta^2 c(0, 2; h) \in (0, 1)$. By expressing

$$h(t) = \frac{c(0, 2; h)c(2, 2; h) - c(1, 2; h)^2}{\Delta^2 c(0, 2; h)} + \Delta^2 c(0, 2; h)\left(t + \frac{\Delta c(0, 2; h)}{\Delta^2 c(0, 2; h)}\right)^2, \qquad t \in [0, 1],$$

we see that h is convex and attains its minimum on $(0, 1)$. The result follows from the sign of the minimum. The above conditions on $c(0, 2; h)$, $c(1, 2; h)$ and $c(2, 2; h)$, together with the *endpoint derivatives conditions* and a direct application of Lemma 3.3 gives the characterization of $\mathcal{A}_4$.

Notice that, so far (for $m \leq 4$), we have characterized $\mathcal{A}_m$ using only elementary mathematics. For higher degrees however, we use a finer result by Lukács, concerning the characterization of nonnegative polynomials, which can be found in for instance Karlin and Shapley [19], Szegő [29] and Pólya and Szegő [22]. It says that a polynomial $h \in \mathcal{P}_m$ of degree at most m, is nonnegative if and only if there are polynomials P and Q with $\deg(P) \leq \lfloor m/2 \rfloor$ and $\deg(Q) \leq \lfloor (m-1)/2 \rfloor$, such that

$$h(t) = \begin{cases} P^2(t) + t(1-t)Q^2(t), & \text{if } m \text{ is even}, \\ tP^2(t) + (1-t)Q^2(t), & \text{if } m \text{ is odd}, \end{cases} \tag{3.11}$$

for all $t \in [0, 1]$. The following lemma exploits Lukács' result and provides expressions for the coefficients, in the Bernstein basis, of nonnegative $h \in \mathcal{P}_m$, $m \geq 1$, using the Hypergeometric distribution. Here $Y \sim \text{Hypergeo}(n, M, N)$ if

$$P(Y = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad (n + M - N) \vee 0 \leq k \leq n \wedge M, 0 \leq n, M \leq N.$$

Lemma 3.4. *A polynomial $h \in \mathcal{P}_m$, $m \geq 1$, is nonnegative on* $[0, 1]$ *if and only if there exist polynomials P and Q such that, the coefficients of h are related to the coefficients of P and Q* (*in the Bernstein basis*) *in the following way*: (*for notational simplicity here, let $h(k, m) = c(k, m; h)$, $p(k, m) = c(k, m; P)$ and $q(k, m) = c(k, m; Q)$ denote the coefficients*)

(i) *when m is even, $P \in \mathcal{P}_{m/2}$, $Q \in \mathcal{P}_{(m-2)/2}$ and*

$$\begin{aligned} h(k, m) = \mathrm{E}\bigg\{ & p\left(Y_1, \frac{m}{2}\right) p\left(k - Y_1, \frac{m}{2}\right) \\ & + \frac{k(m-k)}{m(m-1)} q\left(Y_2, \frac{m-2}{2}\right) q\left(k - Y_2 - 1, \frac{m-2}{2}\right)\bigg\}, \end{aligned} \tag{3.12}$$

with $Y_1 \sim$ Hypergeo$(k, m/2, m)$, *$k = 0, \ldots, m$, and $Y_2 \sim$* Hypergeo$(k - 1, (m-2)/2, m - 2)$, *$k = 1, \ldots, m - 1$,*

(ii) *when m is odd*, $P, Q \in \mathcal{P}_{(m-1)/2}$ *and*

$$h(k,m) = \frac{1}{m}\mathrm{E}\left\{kp\left(Y_1, \frac{m-1}{2}\right)p\left(k - Y_1 - 1, \frac{m-1}{2}\right) + (m-k)q\left(Y_2, \frac{m-1}{2}\right)q\left(k - Y_2, \frac{m-1}{2}\right)\right\}, \tag{3.13}$$

with $Y_1 \sim \mathrm{Hypergeo}(k-1, (m-1)/2, m-1)$, $k = 1, \ldots, m$, *and* $Y_2 \sim \mathrm{Hypergeo}(k, (m-1)/2, m-1)$, $k = 0, \ldots, m-1$.

Proof. The proof is a bit technical, and can be skipped without affecting the readability of what follows. See the Appendix. □

By putting everything together, we get our main theorem.

Theorem 3.5 (Characterization).

(i) *For $m \geq 1$, we have $h \in \mathcal{H}_m$ if and only if h satisfies the endpoint derivatives conditions* (3.7) *and there exist polynomials P and Q such that the coefficients of h* (*in the Bernstein basis*) *are given by* (3.12) *when m is even or by* (3.13) *when m is odd.*
(ii) *If $h \in \mathcal{H}_m$, $m \geq 0$, then A given by* (3.8) *is in $\mathcal{A}_{m+2}$. Lemma* 3.3 *gives the coefficients of A in terms of the ones of h.*
(iii) *If $A \in \mathcal{A}_{m+2}$, $m \geq 0$, then $A'' \in \mathcal{H}_m$. Lemma* 3.3 *gives the coefficients of h in terms of the ones of A.*

A parameterization of $\mathcal{H}_m$ (and therefore of the polynomial Pickands functions) can therefore be made via the coefficients of the polynomials P and Q above. The following theorem describes the corresponding parameter space.

Theorem 3.6 (Parameterization). *Let $m \geq 0$, for polynomials $P \in \mathcal{P}_{\lfloor m/2 \rfloor}$ and when $m \geq 1$ $Q \in \mathcal{P}_{\lfloor (m-1)/2 \rfloor}$, let θ be the concatenation of the coefficients of P and the ones of Q in the Bernstein basis, that is,*

$$\theta(k) = \begin{cases} c\big(k-1, \lfloor m/2 \rfloor; P\big), & \text{if } 1 \leq k \leq \lfloor m/2 \rfloor + 1, \\ c\big(k - \lfloor m/2 \rfloor - 2, \big\lfloor (m-1)/2 \big\rfloor; Q\big), & \text{if } \lfloor m/2 \rfloor + 2 \leq k \leq m+1, m \geq 1. \end{cases}$$

Let $h_\theta \in \mathcal{P}_m$ be constructed using polynomials P and Q in formula (3.11). *When $m \geq 1$, the coefficients of h_θ are given in Lemma* 3.4. *The parameter space $\Theta_m = \{\theta : h_\theta \in \mathcal{H}_m\}$ is given by*

$$\Theta_m = E_0 \cap E_1, \tag{3.14}$$

where E_0 and E_1 are two ellipsoids in $\mathbb{R}^{m+1}$.

Proof. For every fixed value of $t \in [0, 1]$, the function $\theta \mapsto h_\theta(t)$ is a positive semidefinite quadratic form. In fact, using the Bernstein basis and (3.11), we find that $h_\theta(t)$ is given by

$$\begin{cases} \left(\sum_{k=1}^{(m/2)+1} b_{k-1,m/2}(t)\theta_k\right)^2 \\ \quad + \left(\sum_{k=1}^{m/2} \sqrt{t(1-t)}\, b_{k-1,(m/2)-1}(t)\theta_{k+(m/2)+1}\right)^2, & \text{if } m \text{ is even,} \\ \left(\sum_{k=1}^{(m+1)/2} \sqrt{t}\, b_{k-1,(m-1)/2}(t)\theta_k\right)^2 \\ \quad + \left(\sum_{k=1}^{(m+1)/2} \sqrt{1-t}\, b_{k-1,(m-1)/2}(t)\theta_{k+(m+1)/2}\right)^2, & \text{if } m \text{ is odd.} \end{cases}$$

Note that h_θ is the zero polynomial if and only if $\theta = 0$, and so if $\|\theta\| > 0$, then the function $t \mapsto h_\theta(t)$ is positive except perhaps at a finite number of points (roots). Now let A_θ be the polynomial obtained from formula (3.9) with h being replaced by h_θ. Since

$$-A'_\theta(0) = \int_0^1 (1-w)h_\theta(w)\,\mathrm{d}w \quad \text{and} \quad A'_\theta(1) = \int_0^1 w h_\theta(w)\,\mathrm{d}w,$$

if $\|\theta\| > 0$, it follows that $-A'_\theta(0)$ and $A'_\theta(1)$ are both positive definite quadratic forms. Finally, Theorem 3.5 says that $h_\theta \in \mathcal{H}_m$ if and only if the *endpoint derivatives conditions* are verified:

$$\left\{\int_0^1 (1-w)h_\theta(w)\,\mathrm{d}w\right\} \vee \left\{\int_0^1 w h_\theta(w)\,\mathrm{d}w\right\} \le 1,$$

and so the sets

$$E_0 = \left\{\theta \in \mathbb{R}^{m+1} : -A'_\theta(0) \le 1\right\} \quad \text{and} \quad E_1 = \left\{\theta \in \mathbb{R}^{m+1} : A'_\theta(1) \le 1\right\},$$

are therefore both ellipsoids in $\mathbb{R}^{m+1}$ and $\Theta_m = E_0 \cap E_1$. □

4. The Bernstein approximations submodel

Let

$$\mathcal{A}_m^+ = \left\{B_m(A, \cdot) : A \in \mathcal{A}\right\}, \qquad m \ge 0,$$

be the set of Bernstein's approximations of Pickands functions. The second part of the following lemma gives a useful necessary and sufficient geometric condition on the coefficients (in the Bernstein basis) for a polynomial $A \in \mathcal{P}_m$ to belong to $\mathcal{A}_m^+$. The first part is used for this characterization and is also useful in the next propositions.

Lemma 4.1. *For any polynomial $A \in \mathcal{P}_m$, $m \geq 1$, represented in the Bernstein basis by $A(t) = \mathrm{E}_t\{c(S_m, m; A)\}$, $t \in [0,1]$, let A^* be the piecewise linear function interpolating the points $\{(k/m, c(k,m;A))\}_{k=0}^m$,*

$$A^*(t) = \big(\lfloor mt \rfloor + 1 - mt\big)c\big(\lfloor mt \rfloor, m; A\big) + \big(mt - \lfloor mt \rfloor\big)c\big(\lfloor mt \rfloor + 1, m; A\big), \qquad t \in [0,1].$$

(i) *If $\Delta^2 c(k,m;A) \geq 0$, $k = 0, \dots, m-2$, $m \geq 2$, then A^* is convex,*
(ii) (*characterization*) *$A \in \mathcal{A}_m^+$ if and only if $A^* \in \mathcal{A}$.*

Proof. (i) Consider the step function $\varphi(t) = m\Delta c(\lfloor mt \rfloor \wedge (m-1), m; A)$, $t \in [0,1]$. Geometrically, φ corresponds to the right derivative of A^* on [0,1). Now, $\Delta^2 c(k,m;A) \geq 0$, $k = 0, \dots, m-2$, implies that $k \mapsto \Delta c(k,m;A)$, $k = 0, \dots, m-1$, is nondecreasing, and therefore φ is nondecreasing on $[0,1]$. Since $A^*(t) = 1 + \int_0^t \varphi(x)\,\mathrm{d}x$, for all $t \in [0,1]$, it follows that A^* is convex.

(ii) We have $A(\cdot) = B_m(A^*, \cdot)$. If $A^* \in \mathcal{A}$, then $B_m(A^*, \cdot) \in \mathcal{A}_m^+$ and therefore $A \in \mathcal{A}_m^+$. For the converse, if $A \in \mathcal{A}_m^+$, then there is some $A_0 \in \mathcal{A}$ such that $A(\cdot) = B_m(A_0, \cdot)$. For $m = 1$, the fact that $A^* \in \mathcal{A}$ follows trivially. For $m \geq 2$, the convexity of A_0 implies

$$\Delta^2 c(k,m;A) = A_0\left(\frac{k+2}{m}\right) - 2A_0\left(\frac{k+1}{m}\right) + A_0\left(\frac{k}{m}\right) \geq 0, \qquad k = 0, \dots, m-2,$$

and (i) above shows that A^* is convex. Also,

$$A^*(0) = c(0,m;A) = A_0(0) = 1 = A_0(1) = c(m,m;A) = A^*(1) = 1,$$

and finally

$$\big(A^*\big)'(0) = m\Delta c(0,m;A) = \frac{A_0(1/m) - 1}{1/m} \geq A_0'(0) \geq -1$$

and

$$\big(A^*\big)'(1) = m\Delta c(m-1,m;A) = \frac{1 - A_0(1-1/m)}{1/m} \leq A_0'(1) \leq 1.$$

This shows that $A^* \in \mathcal{A}$. □

The following proposition says that Bernstein approximations of Pickands functions are themselves Pickands functions.

Proposition 4.2. *We have $\mathcal{A}_m^+ \subset \mathcal{A}_m$, for every $m \geq 1$. Moreover, $\mathcal{A}_m^+ = \mathcal{A}_m$ for $m = 1,2,3$, while $\mathcal{A}_4^+ \neq \mathcal{A}_4$.*

Proof. Let $V(t) = t \vee (1-t)$, $t \in [0,1]$. For the first statement, let $A \in \mathcal{A}$ and $B_m(A,\cdot) \in \mathcal{A}_m^+$, $m \geq 1$. We have $B_m(A,0) = A(0) = 1 = A(1) = B_m(A,1)$, and $B_m(A,\cdot)$ is convex, see Section 2.2. Using Jensen's inequality, we obtain

$$V(t) \leq A(t) = A\big(\mathrm{E}_t(S_m/m)\big) \leq \mathrm{E}_t\big(A(S_m/m)\big) = B_m(A,t), \qquad t \in [0,1],$$

and this shows that the *boundary conditions* of a Pickands function are verified and so $B_m(A, \cdot) \in \mathcal{A}_m$.

For the second statement, recall that $\mathcal{A}_1 = \mathcal{A}_0 = \{1\} = \mathcal{A}_1^+$. If $A \in \mathcal{A}_m$ for either $m = 2$ or $m = 3$, then the arguments presented in Example 1 can serve to verify that the piecewise linear function A^* interpolating the points $\{(k/m, c(k, m; A))\}_{k=0}^m$ belongs to $\mathcal{A}$. Lemma 4.1 then implies that $A \in \mathcal{A}_m^+$, when $m = 2, 3$. The fact that A^* belongs to $\mathcal{A}$ is not necessarily true for $m = 4$ as the following shows: take for instance

$$A(t) = 1 - t(1-t)\big\{1 - 2t(1-t)\big\}, \qquad t \in [0, 1]. \tag{4.1}$$

It can be easily verified that $A \in \mathcal{A}_4$, and we have $A(t) = \mathrm{E}_t\{c(S_4, 4; A)\}$, with $c(k, 4; A) = 1$, for $k = 0, 2, 4$ and $c(1, 4; A) = 3/4 = c(3, 4; A)$. Suppose that $A(\cdot) = B_4(A_0, \cdot)$, for some $A_0 \in \mathcal{A}$. Then $1 = c(2, 4; A) = c(2, 4; B_4(A_0, \cdot)) = A_0(1/2)$. This implies that $A_0(t) = 1$, for all $t \in [0, 1]$, which in turn implies that $1 = B_4(A_0, t) = A(t)$, for all $t \in [0, 1]$. Therefore, $A \in \mathcal{A}_4 \setminus \mathcal{A}_4^+$. $\square$

Remark. The above result says that $\mathcal{A}_4^+ \neq \mathcal{A}_4$, but we can say even more than this: $\mathcal{A}_m^+ \neq \mathcal{A}_m$ for every $m \geq 4$. This follows as a direct application of *The gap* theorem coming up (notice that for A given by in (4.1) we have $A''(1/2) = 0$).

For $m \geq 0$, we call $\mathcal{A}_m$ the *full model* and $\mathcal{A}_m^+$ the *submodel*. In order to give a parameterization of the submodel, let

$$\mathcal{H}_m^+ = \big\{A'' : A \in \mathcal{A}_{m+2}^+\big\} \subset \mathcal{H}_m.$$

The next theorem shows that the (parametric) space of the coefficients

$$C_m^+ = \big\{\big(c(0, m; h), \dots, c(m, m; h)\big) : h \in \mathcal{H}_m^+\big\},$$

is a polytope by showing that $\mathcal{H}_m^+ = \{h \in \mathcal{H}_m : c(k, m; h) \geq 0, k = 0, \dots, m\}$, $m \geq 0$.

Theorem 4.3 (Parameterization). *We have $h \in \mathcal{H}_m^+$, $m \geq 0$, if and only if*

$$\frac{1}{m+1}\sum_{j=0}^{m}\left(1 - \frac{j+1}{m+2}\right)c(j, m; h) \leq 1, \qquad \frac{1}{m+1}\sum_{j=0}^{m}\frac{j+1}{m+2}c(j, m; h) \leq 1 \tag{4.2}$$

and

$$c(k, m; h) \geq 0, \qquad k = 0, \dots, m. \tag{4.3}$$

Proof. Let $m \geq 0$. If $h \in \mathcal{H}_m^+$ then $h \in \mathcal{H}_m$ and, by definition of $\mathcal{H}_m$, the conditions

$$\frac{1}{m+1}\sum_{j=0}^{m}\left(1 - \frac{j+1}{m+2}\right)c(j, m; h) \leq 1, \qquad \frac{1}{m+1}\sum_{j=0}^{m}\frac{j+1}{m+2}c(j, m; h) \leq 1,$$

hold. When $h \in \mathcal{H}_m^+$, there exists $A \in \mathcal{A}_{m+2}$ such that $h(\cdot) = B''_{m+2}(A, \cdot)$. Therefore, by Lemma 3.3 and convexity of A,

$$c(k, m; h) = (m+2)(m+1)\left\{A\left(\frac{k+2}{m+2}\right) - 2A\left(\frac{k+1}{m+2}\right) + A\left(\frac{k}{m+2}\right)\right\} \geq 0, \qquad k = 0, \ldots, m.$$

Conversely, suppose that a polynomial $h \in \mathcal{P}_m$ satisfies (4.2) and (4.3), then $h \in \mathcal{H}_m$, and from Theorem 3.1, there exists an $A \in \mathcal{A}_{m+2}$ such that $h = A''$. Now let $A^* : [0, 1] \to \mathbb{R}$ be the piecewise linear interpolation of $\{(k/(m+2), c(k, m+2; A)\}_{k=0}^{m+2}$. The *endpoint conditions* on the coefficients of A given by Proposition 3.2 directly imply that A^* satisfies the *endpoint conditions* of a Pickands function. The convexity of A^* is obtained by the nonnegativity of the coefficients of h: Lemma 3.3 gives $(m+2)(m+1)\Delta^2 c(k, m+2; A) = c(k, m; h) \geq 0$, $k = 0, \ldots, m$, and the first part of Lemma 4.1 gives the convexity of A^*. Therefore, $A^* \in \mathcal{A}$, and by the second part of Theorem 4.2, $A \in \mathcal{A}_{m+2}^+$, which means that $h \in \mathcal{H}_m^+$. □

A useful property of the submodel $\mathcal{A}_m^+$, $m \geq 1$, is that it is nested. We show this in the next theorem, and we find the gap between the full model and the submodel. The proof relies on the Lorentz degree of a positive polynomial on (0, 1), see for instance Powers and Reznick [23]. In the Bernstein basis, it is easy to see that $P \in \mathcal{P}_m$ is nonnegative if $c(k, m; P) \geq 0$, $k = 0, \ldots, m$. This sufficient condition is not a necessary condition, see Karlin and Shapley [19], Szegő [29] and Pólya and Szegő [22]. It may happen that for some $M > m$,

$$\min\{c(k, m; P) : k = 0, \ldots, m\} < 0, \qquad \text{while } \min\{c(k, M; P) : k = 0, \ldots, M\} \geq 0.$$

The question is when does this happen? For the (interesting) case: $\deg(P) > 0$, the necessary and sufficient condition for having $\min\{c(k, M; P) : k = 0, \ldots, M\} \geq 0$ for some $M \geq \deg(P)$ is that P be positive on (0, 1). This is known in the literature as Bernstein's theorem, and it motivates the definition of the Lorentz degree of a polynomial $P \in \mathcal{P}_m$ which is positive on (0, 1):

$$r(P) = \min\{M \geq m : c(k, M; P) \geq 0, k = 0, \ldots, M\}.$$

To illustrate this notion, here is a very instructive example that will lead us naturally to *The gap* theorem. The symmetric polynomials in $\mathcal{A}_4 \setminus \mathcal{A}_2$ have the following form:

$$A_{\alpha,\beta}(t) = 1 - \alpha t(1-t)\{1 - \beta t(1-t)\},$$
$$t \in [0, 1], \text{ with } \alpha \in (0, 1] \text{ and } \beta \in [-1, 2] \setminus \{0\}.$$

As can be seen in Figure 1 below, α (which is the absolute value of the first derivative at the endpoints) sets the triangle in which the curves belong, and then β controls their curvature. Let $h_{\alpha,\beta} = A''_{\alpha,\beta} \in \mathcal{H}_2$, and lets find the values of $m \geq 2$, such that $h_{\alpha,\beta} \in \mathcal{H}_m^+$. The smallest such m, if it exists, is the Lorentz degree of $h_{\alpha,\beta}$, and as we will see, it depends on β here. The function

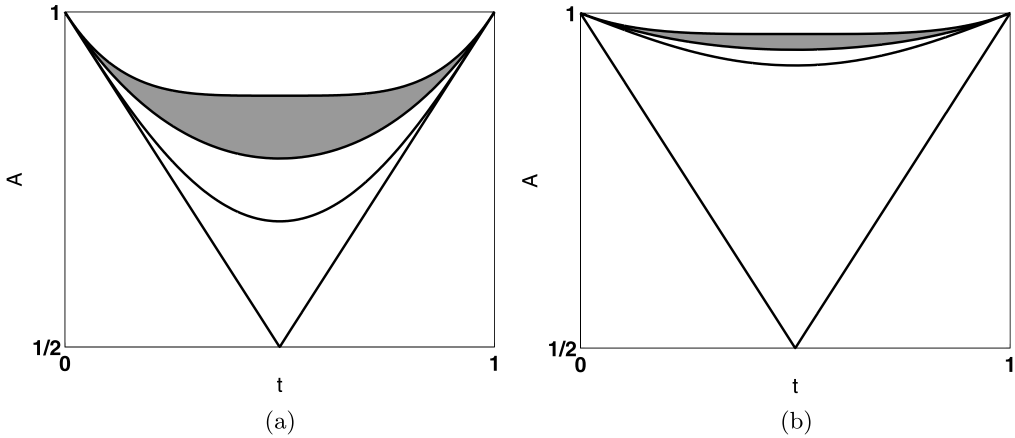


Figure 1. Illustration of the gap between $\mathcal{A}_4^+$ and $\mathcal{A}_4$ in the symmetric case. Here $A = A_{\alpha,\beta}$ with $A_{\alpha,\beta}(t) = 1 - \alpha t(1-t)[1 - \beta t(1-t)]$, $t \in [0, 1]$. In both plots, the three curves correspond to $\beta = -1, 1/2$ and 2 respectively (from bottom to top). The shaded regions hold the curves in $\mathcal{A}_4 \setminus \mathcal{A}_4^+$. (a) $\alpha = 1$. (b) $\alpha = 1/4$.

$h_{\alpha,\beta}$ is given, for $t \in [0, 1]$, by

$$\begin{aligned} h_{\alpha,\beta}(t) &= 2\alpha\bigl\{(1+\beta) - 6\beta t(1-t)\bigr\} \\ &= 2\alpha\bigl[(1+\beta)\bigl\{(1-t)+t\bigr\}^m - 6\beta t(1-t)\bigl\{(1-t)+t\bigr\}^{m-2}\bigr], \qquad m \geq 2. \end{aligned}$$

Using Newton's binomial formula, we get

$$c(k, m; h_{\alpha,\beta}) = 2\alpha\left\{(1+\beta) - 6\beta\frac{k(m-k)}{m(m-1)}\right\}, \qquad k = 0, \ldots, m, m \geq 2. \tag{4.4}$$

Fix $\alpha \in (0, 1]$. From (4.4), we see that when $\beta \in [-1, 0)$, we have $c(k, m; h_{\alpha,\beta}) \geq 0$, for $k = 0, \ldots, m$, and this holds for every $m \geq 2$. Therefore, the Lorentz degree in this case is $r(h_{\alpha,\beta}) = 2$, and $h_{\alpha,\beta} \in \mathcal{H}_m^+$ if and only if $m \geq 2$. Next, when $\beta \in (0, 2)$ the smallest coefficient occurs when $k = \lfloor m/2 \rfloor$, and we get

$$\begin{aligned} &\min\bigl\{c(k, m; h_{\alpha,\beta}) : k = 0, \ldots, m\bigr\} \\ &\quad = 2\alpha\left\{(1+\beta) - 6\beta\frac{\lfloor m/2\rfloor(m - \lfloor m/2\rfloor)}{m(m-1)}\right\}, \qquad m \geq 2. \end{aligned} \tag{4.5}$$

By looking at the sign of (4.5) and solving the inequality for m we obtain $c(\lfloor m/2\rfloor, m; h_{\alpha,\beta}) \geq 0$ if and only if $m \geq 2\lceil(1+\beta)/(2-\beta)\rceil$. This tells us that the Lorentz degree in this case is

$$r(h_{\alpha,\beta}) = 2\bigl\lceil(1+\beta)/(2-\beta)\bigr\rceil, \qquad \beta \in (0, 2),$$

and also that $h_{\alpha,\beta} \in \mathcal{H}_m^+$ if and only if $m \geq r(h_{\alpha,\beta})$. The Lorentz degree increases without bound as β increases to 2. Finally when $\beta = 2$, we get

$$c\big(\lfloor m/2 \rfloor, m; h_{\alpha,2}\big) = \frac{-6\alpha}{2\lceil m/2 \rceil - 1} < 0, \qquad \text{for every } m \geq 2,$$

so $h_{\alpha,2} \in \mathcal{H}_2 \setminus \bigcup_{m\geq 2} \mathcal{H}_m^+$ or equivalently $A_{\alpha,2} \in \mathcal{A}_4 \setminus \bigcup_{m\geq 4} \mathcal{A}_m^+$. As the following theorem shows, this is true because $h_{\alpha,2}(1/2) = 0$, that is, $h_{\alpha,2}$ is nonnegative but fails to be positive on $(0,1)$. Finally, for this example, we note that $A_{\alpha,\beta} \in \mathcal{A}_4^+$ if and only if $(\alpha,\beta) \in (0,1] \times [-1,0) \cup (0,1/2]$, see Figure 1.

Theorem 4.4 (The gap). *For all $m \geq 1$, we have $\mathcal{A}_m^+ \subset \mathcal{A}_{m+1}^+$. Moreover, for $h \in \bigcup_{m=0}^\infty \mathcal{H}_m$ we have $h \notin \bigcup_{m=0}^\infty \mathcal{H}_m^+$ if and only if $\deg(h) > 0$ and $h(t) = 0$ for some $t \in (0,1)$.*

Proof. First, since $\mathcal{A}_m^+ = \mathcal{A}_m$, for $m = 1,2,3$, we already have $\mathcal{A}_1^+ \subset \mathcal{A}_2^+ \subset \mathcal{A}_3^+$. Let $h \in \mathcal{H}_m^+$, for some $m \geq 1$. We want to show that $h \in \mathcal{H}_{m+1}^+$. Since $\mathcal{H}_m^+ \subset \mathcal{H}_m \subset \mathcal{H}_{m+1}$ we just need to show that $c(k, m+1; h) \geq 0$, $k = 0, \dots, m+1$. Using the binomial identities from Section 2.3, we have

$$\begin{aligned}
\mathrm{E}_t\big\{c(S_{m+1}, m+1; h)\big\} &= \mathrm{E}_t\big\{c(S_m, m; h)\big\} \\
&= \mathrm{E}_t\big\{(1-t)c(S_m, m; h) + tc(S_m, m; h)\big\} \\
&= \frac{1}{m+1}\mathrm{E}_t\big\{(m+1-S_{m+1})c(S_{m+1}, m; h) + S_{m+1}c(S_{m+1}-1, m; h)\big\},
\end{aligned}$$

leading to the identity

$$c(j, m+1; h) = \frac{j}{m+1}c(j-1, m; h) + \left(1 - \frac{j}{m+1}\right)c(j, m; h), \qquad j = 0, \dots, m+1.$$

From Theorem 4.3 it follows that $h \in \mathcal{H}_{m+1}^+$, which proves the first assertion.

Now let $h \in \bigcup_{m=0}^\infty \mathcal{H}_m$. To prove the second assertion, we show that we have $h \in \bigcup_{m=0}^\infty \mathcal{H}_m^+$ if and only if $\deg(h) = 0$ or $h(t) > 0$ for all $t \in (0,1)$.

First, assume that $h \in \bigcup_{m=0}^\infty \mathcal{H}_m^+$. If $h \in \mathcal{H}_0^+$, then $\deg(h) = 0$. Otherwise, $h \in \mathcal{H}_m^+ \setminus \mathcal{H}_0^+$, so $\deg(h) > 0$ and this implies that $\mathrm{P}_t\{c(S_m, m; h) \geq 0\} = 1$ and $\mathrm{P}_t\{c(S_m, m; h) > 0\} > 0$ for all $t \in (0,1)$. It follows that $h(t) = \mathrm{E}_t\{c(S_m, m; h)\} > 0$, for all $t \in (0,1)$.

Conversely, if $\deg(h) = 0$ then $h \in \mathcal{H}_0^+$, while if h is positive on $(0,1)$ then $h \in \mathcal{H}_{r(h)}^+$, where $r(h)$ is the Lorentz degree of h. □

5. On the quality of the Bernstein approximations

We are now concerned with the flexibility of the submodel $\mathcal{A}_m^+$ consisting of the Bernstein approximations of Pickands functions. More precisely, we provide an answer to how well a Pickands function $A \in \mathcal{A}$ can be approached in the space $\mathcal{A}_m^+$. We also work out the range $\tau(\mathcal{A}_m^+)$, as m varies, of some dependence measures τ on $\mathcal{A}$.

Theorem 5.1 (Approximation). *For $m \geq 1$, $A \in \mathcal{A}$, $t \in [0, 1]$, we have*

$$A(t) \leq B_m(A, t) \leq A(t) + 2t(1-t)\mathrm{P}_t\big(S_{m-1} = \lfloor mt \rfloor\big), \tag{5.1}$$

where

$$2t(1-t)\mathrm{P}_t\big(S_{m-1} = \lfloor mt \rfloor\big) = \left\{\frac{2t(1-t)}{m\pi}\right\}^{1/2} + \mathrm{O}\big(m^{-3/2}\big), \qquad (m \to \infty).$$

Moreover, when $A = V$, with $V(t) = (1-t) \vee t$, $t \in [0, 1]$, we get the finer approximation

$$B_m(V, t) - V(t) \leq \big\{1 - V(t)\big\}\mathrm{P}_t\big(S_{m-1} = \lfloor m/2 \rfloor\big), \tag{5.2}$$

where $\{1 - V(t)\}\mathrm{P}_t(S_{m-1} = \lfloor m/2 \rfloor) = 0$, for $t \in \{0, 1\}$, while

$$\big\{1 - V(t)\big\}\mathrm{P}_t\big(S_{m-1} = \lfloor m/2 \rfloor\big) \leq \frac{1}{2}\left\{\frac{1}{t} \wedge \frac{1}{(1-t)}\right\}\left\{\frac{2t(1-t)}{m\pi}\right\}^{1/2} + \mathrm{O}\big(m^{-3/2}\big), \qquad (m \to \infty),$$

for $t \in (0, 1)$, with equalities in the last two inequalities when $t = 1/2$.

Proof. The inequality $B_m(A, \cdot) \geq A$ follows from Jensen's inequality

$$B_m(A, t) = \mathrm{E}_t\big\{A(S_m/m)\big\} \geq A(t), \qquad t \in [0, 1].$$

The approximation formula (5.1) is obtained as follows. First, the convexity and the fact that $V(t) \leq A(t) \leq 1$, $t \in [0, 1]$ imply that $|A(t_1) - A(t_2)| \leq |t_1 - t_2|$, for all $t_1, t_2 \in [0, 1]$. From that, for $t \in [0, 1]$,

$$\begin{aligned}
B_m(A, t) - A(t) &= \big|\mathrm{E}_t\big\{A(S_m/m) - A(t)\big\}\big| \\
&\leq \mathrm{E}_t\big\{\big|A(S_m/m) - A(t)\big|\big\} \\
&\leq \mathrm{E}_t\big\{|S_m/m - t|\big\} \\
&= 2t(1-t)\mathrm{P}_t\big(S_{m-1} = \lfloor mt \rfloor\big) \\
&= \left\{\frac{2t(1-t)}{m\pi}\right\}^{1/2} + \mathrm{O}\big(m^{-3/2}\big), \qquad (m \to \infty),
\end{aligned}$$

the last two equalities can be found in, for instance, Johnson [18]. In particular, the last equality is obtained using Stirling's formula, $n! = (2\pi n)^{1/2}(n/\mathrm{e})^n\{1 + \mathrm{O}(1/n)\}$, $(n \to \infty)$. Here we should point out that the Cauchy inequality gives the weaker result

$$B_m(A, t) - A(t) \leq \mathrm{E}_t\big(|S_m/m - t|\big) \leq \big\{t(1-t)/m\big\}^{1/2},$$

although the rate of convergence remains of the same order.

To show (5.2), making use of the binomial identities (once again), we have, for $t \in [0, 1]$,

$$\begin{aligned} B_m(V,t) &= \mathrm{E}_t\big\{V(S_m/m)\big\} \\ &= \mathrm{E}_t\big\{(1 - S_m/m)\mathbb{1}(S_m \le m/2) + (S_m/m)\mathbb{1}(S_m > m/2)\big\} \\ &= (1-t)\mathrm{P}_t(S_{m-1} \le m/2) + t\mathrm{P}_t(S_{m-1} > m/2 - 1). \end{aligned}$$

Thus,

$$\begin{aligned} B_m(V,t) - V(t) &= \big\{1 - V(t)\big\}\mathrm{P}_t(m/2 - 1 < S_{m-1} \le m/2) \\ &\quad + \big\{1 - t - V(t)\big\}\mathrm{P}_t(S_{m-1} \le m/2 - 1) + \big\{t - V(t)\big\}\mathrm{P}_t(S_{m-1} > m/2) \\ &= \big\{1 - V(t)\big\}\mathrm{P}_t(m/2 - 1 < S_{m-1} \le m/2) \\ &\quad - |2t-1|\big\{\mathrm{P}_t(S_{m-1} \le m/2 - 1)\mathbb{1}(t > 1/2) + \mathrm{P}_t(S_{m-1} > m/2)\mathbb{1}(t \le 1/2)\big\} \\ &\le \big\{1 - V(t)\big\}\mathrm{P}_t(m/2 - 1 < S_{m-1} \le m/2) \\ &= \big\{1 - V(t)\big\}\mathrm{P}_t\big(S_{m-1} = \lfloor m/2 \rfloor\big). \end{aligned}$$

The latter inequality being an equality at $t = 1/2$.

Finally, $\mathrm{P}_t(S_{m-1} = \lfloor m/2 \rfloor) \le \mathrm{P}_t(S_{m-1} = \lfloor mt \rfloor)$, and the rate of convergence follows again by Stirling's formula. □

Dependence measures for bivariate extremes have been studied in the literature, see for instance Tawn [30] and Weissman [31]. In particular, for $A \in \mathcal{A}$ the following measures were proposed

$$\tau_1(A) = 2\big\{1 - A(1/2)\big\}, \qquad \tau_2(A) = \mathrm{E}\big[4\big\{1 - A(U)\big\}\big], \qquad U \sim \mathcal{U}(0,1).$$

Let $A \in \mathcal{A}$ and consider the regions $R_i \subset \mathbb{R}^2$, $i = 1, 2, 3$, given by

$$\begin{aligned} R_1 &= \big\{(t, y)\colon 0 \le t \le 1,\, A(1/2) + \big(1 - A(1/2)\big) \mid 2t - 1 \mid \le y \le 1\big\}, \\ R_2 &= \big\{(t, y)\colon 0 \le t \le 1,\, A(t) \le y \le 1\big\}, \\ R_3 &= \big\{(t, y)\colon 0 \le t \le 1,\, V(t) \le y \le 1\big\}. \end{aligned}$$

Since $A \in \mathcal{A}$ we have $R_1 \subset R_2 \subset R_3$ and $0 \le \tau_1 \le \tau_2 \le 1$ because,

$$\tau_i = 4 \times \mathrm{area}(R_i), \qquad i = 1, 2 \quad \text{and} \quad 0 \le \mathrm{area}(R_1) \le \mathrm{area}(R_2) \le \mathrm{area}(R_3) = 1/4.$$

In particular,

$$\tau_1\big\{B_m(A,\cdot)\big\} = \mathrm{E}_{1/2}\big[2\big\{1 - A(S_m/m)\big\}\big], \qquad \tau_2\big\{B_m(A,\cdot)\big\} = \mathrm{E}\big[4\big\{1 - A(U_m/m)\big\}\big],$$

where $U_m \sim \mathcal{U}\{0, \ldots, m\}$. Note that since S_m/m and U_m/m converge in distribution to t and $U \sim \mathcal{U}(0,1)$ respectively, and since A is continuous and bounded, we obtain that $\tau_i\{B_m(A,\cdot)\} \to \tau_i(A)$, as $m \to \infty$, $i \in \{1, 2\}$. The next proposition gives the range of these measures for the Bernstein approximations.

Proposition 5.2 (Dependence measures). *Let* $V(t) = (1-t) \vee t$, $t \in [0,1]$ (*the Pickands function corresponding to complete monotone dependence*). *For* $m \geq 1$, $\tau_i(\mathcal{A}_m^+) = [0, \tau_i\{B_m(V,\cdot)\}]$, $i \in \{1,2\}$, *with*

$$\tau_1\big\{B_m(V,\cdot)\big\} = 1 - \mathrm{P}_t\big(S_{m-1} = \lfloor m/2 \rfloor\big), \qquad \tau_2\big\{B_m(V,\cdot)\big\} = \lfloor m/2 \rfloor / \big(\lfloor m/2 \rfloor + 1/2\big).$$

Proof. First, if $A_1 \leq A_2$, $A_1, A_2 \in \mathcal{A}$, then $B_m(A_1,\cdot) \leq B_m(A_2,\cdot)$, so that $\tau_i\{B_m(A_1,\cdot)\} \geq \tau_i\{B_m(A_2,\cdot)\}$, $i = 1,2$, and this implies that $\tau_i(\mathcal{A}_m^+) \subset [0, \tau_i\{B_m(V,\cdot)\}]$. The reverse inclusion follows from the convexity of the space $\mathcal{A}_m^+$, the fact that $B_m(V,\cdot)$ is a Pickands function and the linearity (under convex combinations) of the functionals τ_1 and τ_2. □

6. Simulation experiment

We now compare the maximum likelihood estimator from the full model and the one from the submodel through simulated data. In both cases, the maximum likelihood estimator (MLE) is computed numerically. For a fixed $m \geq 0$, when using the full model, the estimator is the polynomial $A_{\hat{\theta}}$, where $\hat{\theta}$ is the solution of the nonlinear constrained maximization problem

$$\hat{\theta} = \arg\max_{\theta \in \Theta_m} L_F(\theta \mid u, v), \tag{6.1}$$

where Θ_m is given in (3.14), for $m \geq 1$, $\Theta_0 = [0,2]$, $(u,v) = \{(u_1,v_1),\ldots,(u_n,v_n)\}$ is the data and the likelihood L_F is obtained by computing the coefficients of h via Lemma 3.4, by deriving its associated Pickands function A via Lemma 3.3, and then by using the mixed partial derivative of (1.1). The estimator for the Bernstein approximations submodel is $A_{\hat{h}}$, the problem to solve is

$$\hat{c} = \arg\max_{c \in C_m^+} L_S(c \mid u, v), \tag{6.2}$$

where C_m^+ is the polytope given in Theorem 4.3 and $\hat{h} \in \mathcal{H}_m^+$ is such that $\hat{c} = (c(k,m;\hat{h}))_{k=0}^m$. Again the likelihood L_S is obtained by Lemma 3.3, and (1.1). Both problems have solutions by continuity of the likelihood over their (compact) feasible regions Θ_m and C_m^+, respectively. These are easily solved by using the MATLAB global search algorithm, for instance.

We consider the three following models for the simulation: the asymmetric logistic model

$$A(t) = (1-\psi_1)t + (1-\psi_2)(1-t) + \big[(\psi_1 t)^{1/\alpha} + \big\{\psi_2(1-t)\big\}^{1/\alpha}\big]^\alpha, \qquad t \in [0,1], \tag{6.3}$$

with $\alpha \in (0,1]$, $0 \leq \psi_1, \psi_2 \leq 1$, the symmetric mixed model

$$A(t) = 1 - \psi t + \psi t^2, \qquad t \in [0,1], \tag{6.4}$$

$\psi \in [0,1]$, and the polynomial A obtained via Lemma 3.3 with $c(0,2;h) = 2$, $c(1,2;h) = -1/3$ and $c(2,2;h) = 1/5$, leading to

$$A(t) = 1 - (83/180)t + t^2 - (7/9)t^3 + (43/180)t^4, \qquad t \in [0,1]. \tag{6.5}$$

While it is clear from the characterization of $\mathcal{H}_2$ given after the proof of Lemma 3.3 that $A \in \mathcal{A}_4 \setminus \mathcal{A}_4^+$, it turns out that it has Lorentz degree $r(h) = 6$, where $h = A''$, so that $A \in \mathcal{A}_8^+ \setminus \mathcal{A}_7^+$. The first two models have also been considered for simulations studies in Fils-Villetard *et al.* [10] and Bücher *et al.* [3]. Here we take the parameter values $\alpha = 1/2$, $\psi_1 = 1/10$, $\psi_2 = 1/2$ in the first model, and $\psi = 9/10$ in the second model. Note that for this particular choice of asymmetric logistic model, $A \notin \bigcup_{m\geq 0} \mathcal{A}_m$, and for the symmetric mixed model, $A \in \mathcal{A}_2^+ = \mathcal{A}_2$.

A general algorithm for drawing independent couples from the copulas associated to these models is provided by Ghoudi *et al.* [13]. Here we draw 1000 samples of sizes $n = 100$ from the copulas corresponding to each of the above three models. For every sample, the maximum likelihood estimate is obtained using the full model by solving (6.1) and the submodel by solving (6.2) both with $m = 5$. This gives polynomial estimates of degree at most 7, that is $A_{\hat{\theta}} \in \mathcal{A}_7$ for the full model and $A_{\hat{h}} \in \mathcal{A}_7^+$ for the submodel. For comparisons, we also consider the popular A_{CFG} estimator from Capéraà *et al.* [4]. The latter can lead to estimates which are not genuine Pickands functions because they do not satisfy either the *boundary conditions* or the *convexity condition*. A modification which leads to Pickands functions as estimates is $\hat{A}_{\text{CFG}} =$ greatest convex minorant of $1 \wedge \{A_{\text{CFG}} \vee V\}$, where $V(t) = (1-t) \vee t$, $t \in [0, 1]$. Another good estimator, according to Genest and Segers [12], is the optimally corrected CFG estimator A_{CFGopt} proposed initially by Segers [27]. The estimator that we use for the comparisons is

$$\hat{A}_{\text{CFGopt}} = \text{greatest convex minorant of } 1 \wedge \{A_{\text{CFGopt}} \vee V\}. \tag{6.6}$$

We show 95% point-wise confidence intervals for model (6.5) in Figure 2. To compare the performance of the various estimators in the cases considered here, we look at estimates of the mean squared error

$$\text{MSE}_A\big\{\hat{A}(t)\big\} = \text{E}\big\{\hat{A}(t) - A(t)\big\}^2, \qquad t \in [0, 1],$$

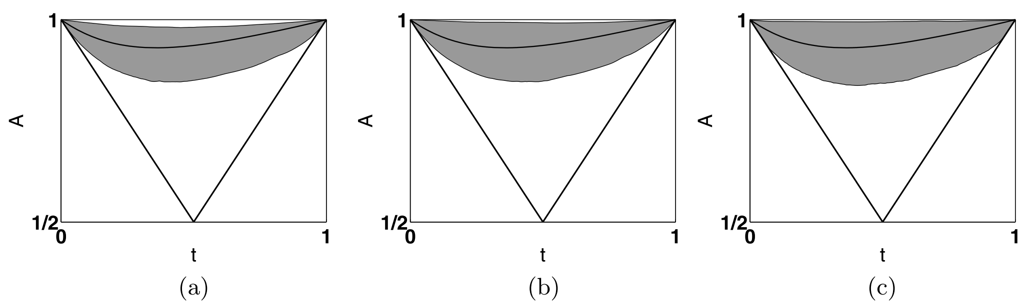


Figure 2. Shaded region corresponds to point-wise 95% confidence intervals for the true Pickands function (black curve) of model (6.5). On part (a) are the results using the maximum likelihood estimator from the full model computed via (6.1), in part (b) using the maximum likelihood estimator from the submodel computed via (6.2) and on part (c) using the optimal CFG estimator (6.6). Here, $m = 5$ for the polynomials. (a) Model (6.5), full MLE $A_{\hat{\theta}}$. (b) Model (6.5), sub MLE $A_{\hat{h}}$. (c) Model (6.6), $\hat{A}_{\text{CFGopt}}$.

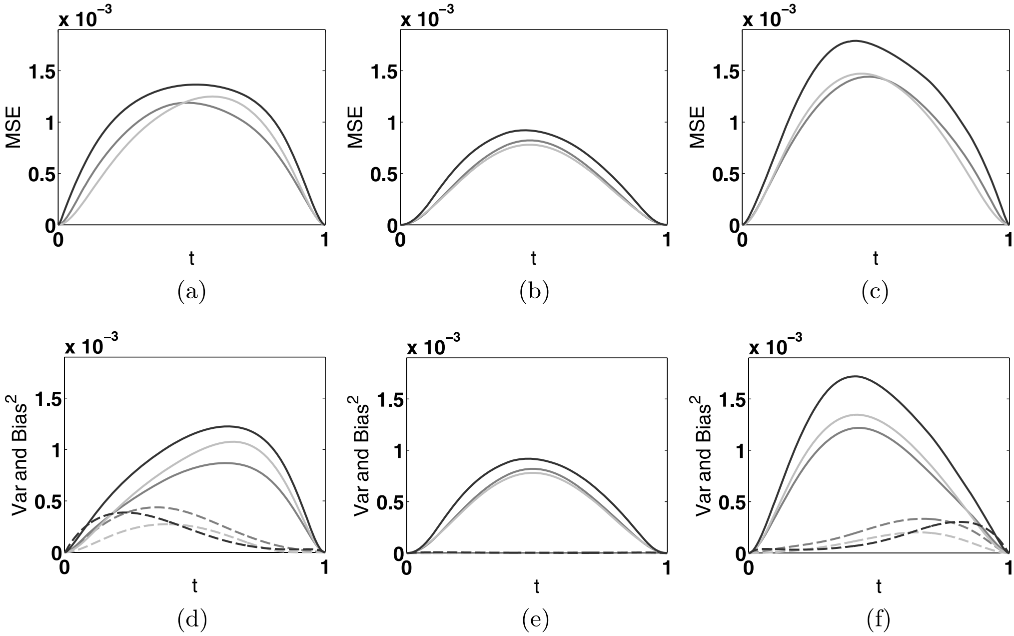


Figure 3. In all the above illustrations, different shades of grey represent (from dark to light grey): the optimal CFG estimator, the MLE from the full model, and the MLE from the submodel respectively. In parts (a)–(c), the curves represent the estimated mean squared error of the estimators. Parts (d)–(f) show the variances (thick curves) and squared bias (dashed curves). Here, $m = 5$ for the polynomials. (a) Model (6.3). (b) Model (6.4). (c) Model (6.5). (d) Model (6.3). (e) Model (6.4). (f) Model (6.5).

where $\hat{A}$ is an estimator and A is the true Pickands function. We also look at the variance and bias separately. These results are plotted in Figure 3. In the simulation results, the MLE of the submodel (uniformly) outperforms the optimal CFG estimator (6.6) in terms of mean squared error and also variance. The MLE of the full model has uniformly smaller variance than (6.6). The bias of all three estimators is much smaller in the symmetric case than in the other two cases. When $m = 5$, the results do not clearly indicate an overall winner between the two maximum likelihood estimators for sample sizes $n = 100$. This seems to change, however (according to further simulations we have done), when m is greater with respect to the data size n. This will be noticed in the following.

We run the entire simulation again, using the same data set, but this time with $m = 8$. This gives polynomial estimates of degree at most 10, that is $A_{\hat{\theta}} \in \mathcal{A}_{10}$ for the full model and $A_{\hat{h}} \in \mathcal{A}_{10}^{+}$ for the submodel. The results are plotted in Figure 4. We can observe that the performance of the estimator from the submodel is roughly the same as in the case $m = 5$, but the performance of the estimator from the full model has somehow worsened in the two nonsymmetric cases in part due to an increase of bias.

As a final numerical investigation, we compare the performance of the MLE from the full model and the one from the submodel, this time, when the sample size n increases and $m = 5$ is

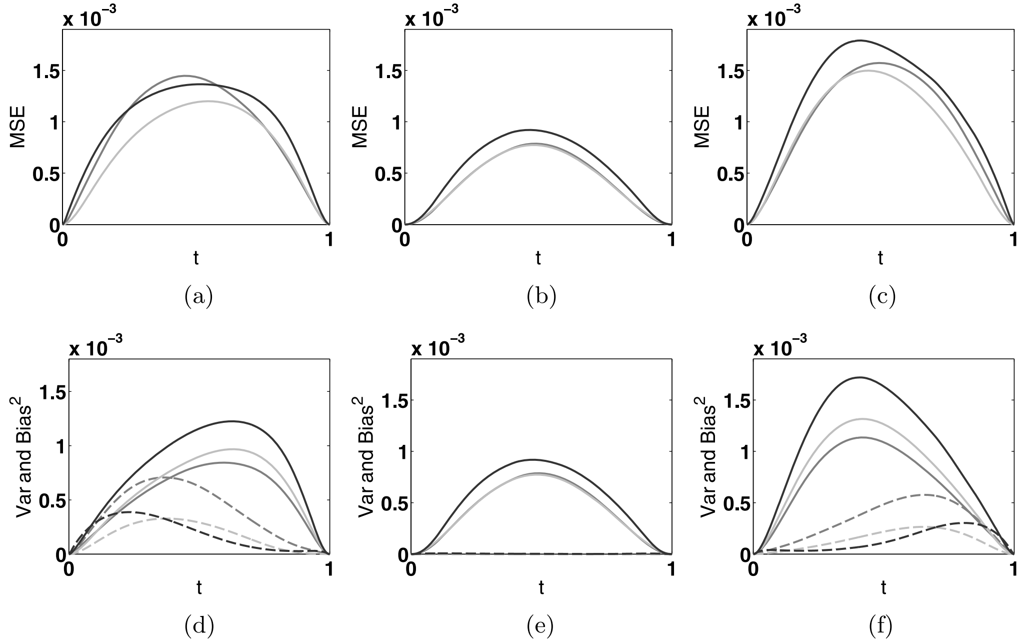


Figure 4. In all the above illustrations, different shades of grey represent (from dark to light grey): the optimal CFG estimator, the MLE from the full model, and the MLE from the submodel respectively. In parts (a)–(c), the curves represent the estimated mean squared error of the estimators. Parts (d)–(f) show the variances (thick curves) and squared bias (dashed curves). Here, $m = 8$ for the polynomials. (a) Model (6.3). (b) Model (6.4). (c) Model (6.5). (d) Model (6.3). (e) Model (6.4). (f) Model (6.5).

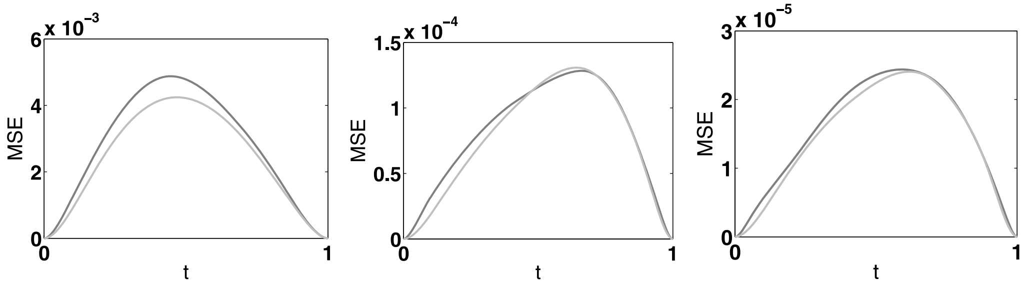
held fixed. Here, in every case we have tried, the mean squared error of the MLE of the submodel was uniformly smaller than that of the full model when the polynomial degree m was large in comparison to the sample size n. Then, when n increases and m is held fixed, the two estimators' mean squared error curves are close and cross eachother. Figure 5 gives a typical illustration.

7. Concluding remarks and comments

The choice of basis

First, the results obtained in Section 3 could be developed in any other polynomial basis, and we took the popularity of the power basis into consideration. However, it appears that some of the results have an easy interpretation when the coefficients are expressed in the Bernstein basis but seem meaningless when the coefficients are expressed in the power basis. For instance, the link between $A \in \mathcal{A}_{m+2}$ and $h = A''$, together with the *endpoint derivatives conditions* can be written

$$A(t) = 1 - \mathrm{E}\Big(\Big[\big\{(1-t)U\big\} \wedge \big\{t(1-U)\big\}\Big]h(U)\Big),$$



(a) (b) (c)

Figure 5. Dark grey curve is the MSE of the MLE from the full model and light grey curve is the MSE of the MLE from the submodel. Here, the true Pickands function is (6.3) and $m = 5$ for the polynomials. (a) $n = 30$. (b) $n = 1000$. (c) $n = 5000$.

with $t \in [0, 1]$, and

$$\mathrm{E}\{(1-U)h(U)\} \le 1, \qquad \mathrm{E}\{Uh(U)\} \le 1,$$

with $U \sim \mathcal{U}(0, 1)$. When we express these expressions in terms of the coefficients in the Bernstein basis, see Lemma 3.3 and inequalities (3.7) respectively, we get a striking similarity

$$c(k, m+2; A) = 1 - \mathrm{E}\big(\big[\{(1-t)U_m\} \wedge \{t(1-U_m)\}\big]c\big((m+2)U_m - 1, m; h\big)\big),$$

with $t = k/(m+2)$, $k = 0, \dots, m+2$, and

$$\mathrm{E}\{(1-U_m)c\big((m+2)U_m - 1, m; h\big)\} \le 1, \qquad \mathrm{E}\{U_m c\big((m+2)U_m - 1, m; h\big)\} \le 1,$$

with $U_m \sim \mathcal{U}\{1/(m+2), \dots, (m+1)/(m+2)\}$. Now let

$$A(t) = \sum_{k=0}^{m+2} \alpha_k t^k, \qquad h(t) = \sum_{k=0}^{m} \eta_k t^k, \qquad t \in [0, 1].$$

If we do the same exercise in the power basis this time, we obtain

$$\alpha_0 = 1, \qquad \alpha_1 = -\sum_{k=0}^{m} \frac{1}{(k+1)(k+2)} \eta_k,$$

$$\alpha_k = \frac{1}{k(k-1)} \eta_{k-2}, \qquad 2 \le k \le m+2$$

and

$$\sum_{k=0}^{m} \frac{1}{(k+1)(k+2)} \eta_k \le 1, \qquad \sum_{k=0}^{m} \frac{1}{(k+2)} \eta_k \le 1,$$

these are definitely difficult to grasp.

Secondly, in view of the model constructed in Section 4, the Bernstein basis is clearly the right basis for approximating continuous functions on [0, 1] by polynomials, think of the Weierstrass theorem for instance. It turns out that the Bernstein basis is very appealing for other reasons as well. For example, the geometry of the Pickands functions is directly reflected in the coefficients, this is made clear by Lemma 4.1. It is also shown that $\mathcal{A}_m^+ = \mathcal{A}_m$ for $m = 1, 2, 3$, but $\mathcal{A}_4^+ \neq \mathcal{A}_4$. Working with the Bernstein basis throughout made this finding easier.

Full model vs submodel

The convexity condition on the full model in Section 3 is obtained via intermediate polynomials (P and Q), so that the coefficients of the resulting Pickands function A are parameterized by those of P and Q. In this parameterization, there is an identifiability problem with the parameters. To see this, in (3.11), take for example the four couples (P, Q), $(-P, Q)$, $(Q, -P)$ and $(-P, -Q)$, these will all produce the same function h. However, for inferential purposes, the parameters of interest remain the coefficients of h (or A), not the ones of P and Q. The submodel is not concerned with this issue at all, the parameters of the submodel are the coefficients. The simplicity of this model, its polytopal parameter space, the gap theorem and its flexibility makes it more appealing, for us at least, than the full model in Section 3. A practical consequence of the simplicity of the submodel is in finding the maximum likelihood estimator; it is more than twice as fast, numerically, than for the complete model. Finally, we mention that although the likelihood is not concave (and rather complicated), local maxima could indeed make difficult the search for a global maximum. However, using a state-of-the-art MATLAB global search algorithm, we have not encountered problematic situations.

Appendix

Proof of Lemma 3.4. Consider $X_1, X_2, \ldots,$ a sequence of independent Bernoulli(t) random variables. Let $S_1, S_2, \ldots,$ be the sequence of the cumulative sums, $S_n = \sum_{k=1}^n X_k$. Suppose that m is even, $m > 0$, $P \in \mathcal{P}_{m/2}$, $Q \in \mathcal{P}_{(m-2)/2}$. We can write, by equation (3.11),

$$h(t) = P^2(t) + t(1-t)Q^2(t), \qquad t \in [0, 1].$$

Now,

$$\begin{aligned} P^2(t) &= \mathrm{E}_t\bigl\{p(S_{m/2}, m/2)p(S_m - S_{m/2}, m/2)\bigr\} \\ &= \mathrm{E}_t\bigl[\mathrm{E}\bigl\{p(Y, m/2)p(S_m - Y, m/2) \mid S_m\bigr\}\bigr], \end{aligned}$$

with $\mathcal{L}(Y \mid S_m = k) = \text{Hypergeo}(k, m/2, m)$, $k = 0, 1, \ldots, m$. Similarly,

$$\begin{aligned} Q^2(t) &= \mathrm{E}_t\bigl\{q\bigl(S_{(m-2)/2}, (m-2)/2\bigr)q\bigl(S_{m-2} - S_{(m-2)/2}, (m-2)/2\bigr)\bigr\} \\ &= \mathrm{E}_t\bigl[\mathrm{E}\bigl\{q\bigl(Y, (m-2)/2\bigr)q\bigl(S_{m-2} - Y, (m-2)/2\bigr) \mid S_{m-2}\bigr\}\bigr], \end{aligned} \tag{A.1}$$

with $\mathcal{L}(Y \mid S_{m-2} = k) = \text{Hypergeo}(k, (m-2)/2, m-2)$, $k = 0, 1, \dots, m-2$. From (A.1) and the binomial identities, we get

$$t(1-t)Q^2(t) = \mathrm{E}_t\left[\frac{S_m(m-S_m)}{m(m-1)}\mathrm{E}\{q\big(Y_2, (m-2)/2\big)q\big(S_m - Y_2 - 1, (m-2)/2\big) \mid S_m\}\right],$$

with $\mathcal{L}(Y_2 \mid S_m = k) = \text{Hypergeo}(k-1, (m-2)/2, m-2)$, $k = 1, \dots, m-1$.

Now, when m is odd, $P, Q \in \mathcal{P}_{(m-1)/2}$, and we can write

$$h(t) = tP^2(t) + (1-t)Q^2(t), \qquad t \in [0,1].$$

Here,

$$\begin{aligned} P^2(t) &= \mathrm{E}_t\{p\big(S_{(m-2)/2}, (m-1)/2\big)p\big(S_{m-1} - S_{(m-1)/2}, (m-1)/2\big)\} \\ &= \mathrm{E}_t\big[\mathrm{E}\{p\big(Y, (m-1)/2\big)p\big(S_{m-1} - Y, (m-1)/2\big) \mid S_{m-1}\}\big], \end{aligned} \tag{A.2}$$

with $\mathcal{L}(Y \mid S_{m-1} = k) = \text{Hypergeo}(k, (m-1)/2, m-1)$, $k = 0, 1, \dots, m-1$. Moreover, from (A.2) and the binomial identities,

$$tP^2(t) = \mathrm{E}_t\left[\frac{S_m}{m}\mathrm{E}\{p\big(Y_1, (m-1)/2\big)p\big(S_m - Y_1 - 1, (m-1)/2\big) \mid S_m\}\right],$$

with $\mathcal{L}(Y_1 \mid S_m = k) = \text{Hypergeo}(k-1, (m-1)/2, m-1)$, $k = 1, \dots, m$.

Similarly,

$$Q^2(t) = \mathrm{E}_t\big[\mathrm{E}\{q\big(Y, (m-1)/2\big)q\big(S_{m-1} - Y, (m-1)/2\big) \mid S_{m-1}\}\big],$$

with $\mathcal{L}(Y \mid S_{m-1} = k) = \text{Hypergeo}(k, (m-1)/2, m-1)$, $k = 0, 1, \dots, m-1$. Finally, again from the binomial identities,

$$(1-t)Q^2(t) = \mathrm{E}_t\left[\frac{m - S_m}{m}\{q\big(Y_2, (m-1)/2\big)p\big(S_m - Y_2, (m-1)/2\big) \mid S_m\}\right],$$

with $\mathcal{L}(Y_2 \mid S_m = k) = \text{Hypergeo}(k, (m-1)/2, m-1)$, $k = 0, \dots, m-1$. □

Acknowledgements

We are particularly grateful to the two referees and especially to the Associate Editor, the comments and suggestions have led to a very much improved version of this paper. We wish to thank Libasse Coly for bringing the problem to our attention and Johan Segers (Université Catholique de Louvain) for great advice and interesting discussions. We also acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes*: *Theory and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. With contributions from Daniel De Waal and Chris Ferro. MR2108013

[2] Boldi, M.-O. and Davison, A.C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 217–229. MR2325273

[3] Bücher, A., Dette, H. and Volgushev, S. (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist.* **39** 1963–2006. MR2893858

[4] Capéraà, P., Fougères, A.-L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985

[5] Coles, S.G. and Tawn, J.A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **53** 377–392. MR1108334

[6] Coles, S.G. and Tawn, J.A. (1994). Statistical methods for multivariate extremes: An application to structural design. *Appl. Statist.* **43** 1–48.

[7] de Haan, L., Neves, C. and Peng, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99** 1260–1275. MR2419346

[8] Dupuis, D.J. and Tawn, J.A. (2001). Effects of mis-specification in bivariate extreme value problems. *Extremes* **4** 315–330 (2002). MR1924233

[9] Einmahl, J.H.J., Krajina, A. and Segers, J. (2008). A method of moments estimator of tail dependence. *Bernoulli* **14** 1003–1026. MR2543584

[10] Fils-Villetard, A., Guillou, A. and Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canad. J. Statist.* **36** 369–382. MR2456011

[11] Fougères, A.-L., Mercadier, C. and Nolan, J.P. (2013). Dense classes of multivariate extreme value distributions. *J. Multivariate Anal.* **116** 109–129. MR3049895

[12] Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453

[13] Ghoudi, K., Khoudraji, A. and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.* **26** 187–197. MR1624413

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

[15] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6** 835–844. MR1791904

[16] Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484. MR0467991

[17] Joe, H., Smith, R.L. and Weissman, I. (1992). Bivariate threshold methods for extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **54** 171–183. MR1157718

[18] Johnson, N.L. (1957). A note on the mean deviation of the binomial distribution. *Biometrika* **44** 532–533. MR0093069

[19] Karlin, S. and Shapley, L.S. (1953). Geometry of moment spaces. *Mem. Amer. Math. Soc.* **1953** 93. MR0059329

[20] Klüppelberg, C. and May, A. (2006). Bivariate extreme value distributions based on polynomial dependence functions. *Math. Methods Appl. Sci.* **29** 1467–1480. MR2247312

[21] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163

[22] Pólya, G. and Szegő, G. (1998). *Problems and Theorems in Analysis*. *II*. *Theory of Functions*, *Zeros*, *Polynomials*, *Determinants*, *Number Theory*, *Geometry*. *Classics in Mathematics*. Berlin: Springer. Translated from the German by C.E. Billigheimer. Reprint of the 1976 English translation. MR1492448

[23] Powers, V. and Reznick, B. (2000). Polynomials that are positive on an interval. *Trans*. *Amer. Math*. *Soc*. **352** 4677–4692. MR1707203
[24] Roberts, A.W. and Varberg, D.E. (1973). *Convex Functions*. *Pure and Applied Mathematics* **57**. New York–London: Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers]. MR0442824
[25] Sancetta, A. and Satchell, S. (2001). Bernstein approximation to copula function and portfolio optimization. DAE working paper, Univ. Cambridge.
[26] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory* **20** 535–562. MR2061727
[27] Segers, J. (2007). Nonparametric inference for bivariate extreme-value copulas. In *Topics in Extreme Values* (M. Ahsanullah and S.N.U.A. Kirmani, eds.) 181–203. New York: Nova Science Publishers.
[28] Smith, R.L. (1994). Multivariate threshold methods. In *Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simiu, eds.) 225–248. Dordrecht: Kluwer.
[29] Szegő, G. (1975). *Orthogonal Polynomials*, 4th ed. *Colloquium Publications* **XXIII**. Providence, RI: Amer. Math. Soc. MR0372517
[30] Tawn, J.A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415. MR0967580
[31] Weissman, I. (2008). On some dependence measures for multivariate extreme value distributions. In *Advances in Mathematical and Statistical Modeling*. *Stat*. *Ind*. *Technol*. 171–180. Boston, MA: Birkhäuser. MR2790117

Received May 2013 and revised June 2014