Bayesian Inference for the Extremal Dependence

I. Antoniano-Villalobos, G. Marcon and S. A. Padoan *

January 8, 2016

Abstract

A simple approach for modelling multivariate extremes is to consider the vector of component-wise maxima and their max-stable distributions. The extremal dependence can be inferred by estimating the angular measure or, alternatively, the Pickands dependence function. We propose a nonparametric Bayesian model that allows, in the bivariate case, the simultaneous estimation of both functional representations through the use of polynomials in the Bernstein form. The constraints required to provide a valid extremal dependence are addressed in a straightforward manner, by placing a prior on the coefficients of the Bernstein polynomials which gives probability one to the set of valid functions. The prior is extended to the polynomial degree, making our approach fully nonparametric. Although the analytical expression of the posterior is unknown, inference is possible via a trans-dimensional MCMC scheme. We show the efficacy of the proposed methodology by means of a simulation study. The extremal behaviour of log-returns of daily exchange rates between Great Britain pound vs USA dollar and Great Britain pound vs Japan yen is analysed for illustrative purposes.

Keywords: Generalised extreme value distribution, Extremal dependence, Angular measure, Max-stable distribution, Bernstein polynomials, Bayesian nonparametrics, Trans-dimensional MCMC, Exchange rate.

*Antoniano-Villalobos, Marcon and Padoan work at the Department of Decision Sciences, Bocconi University of Milan, via Roentgen 1, 20136 Milano, Italy, E-mail: isadora.antoniano@unibocconi.it, giulia.marcon@phd.unibocconi.it, simone.padoan@unibocconi.it.

1 Introduction

The estimation of future extreme episodes of a real process, such as for instance heavy-rainfall, heat-waves and simultaneous losses in the financial market, is of crucial importance for risk management. In most applications, an accurate assessment of such types of risks requires an appropriate modelling and inference of the dependence structure of multiple extreme values.

A simple definition of multiple extremes is obtained by applying the definition of block (or partial)-maximum (Coles 2001, Ch. 3) to each of the variables considered. Then, the probabilistic modelling concerns the joint distribution of the so-called random vector of component-wise (block) maxima, in short *sample maxima*, whose joint distribution is named a *multivariate extreme value distribution* (de Haan and Ferreira 2006, Ch. 6). Within this approach parametric models for the dependence structure have been widely discussed and applied in the literature (e.g. Coles 2001, Beranger and Padoan 2015), but a major downside is that a model which may be useful for a specific application is often too restrictive for many others. As a consequence, more recently, much attention has been devoted to the study of nonparametric estimators or estimation methods for assessing the extremal dependence (see e.g. de Haan and Ferreira 2006, Ch. 7). Some examples focused on the estimation of the Pickands dependence function (Pickands 1981) are provided in Capéraà et al. (1997), Genest and Segers (2009), Bücher et al. (2011), Berghaus et al. (2013) and Marcon et al. (2014), among others.

In order to provide a comprehensive discussion of our approach, we restrict our attention to the bivariate case, that is a two-dimensional vector of sample maxima. Specifically, we describe how Bernstein polynomials can be used to model the extremal dependence within a Bayesian nonparametric framework. The proposal has two key features. Firstly, the use of this particular polynomial expansion makes it possible to accommodate different representations of the dependence structure, such as the Pickands dependence function and the angular (or spectral) distribution. This ensures that in each case, there is the fulfillment of some specific constraints which guarantee that a proper extreme value distribution is defined. Secondly, model fitting, inference and model assessment can be achieved via MCMC methods, preserving the relation between both extremal dependence forms. Information about the polynomial degree is yielded from the data as part of the inferential procedure, and there is no need for a preliminary estimate as is often the case when regularization methods are applied, (e.g. Fils-Villetard et al. 2008, Marcon et al. 2014). Additionally, there is no need to choose between representing the dependence by means of the angular distribution or the Pickands dependence function.

The paper is organised as follows. In Section 2 we briefly describe some basic concepts regarding the extremal dependence structure. In Section 3 we propose a Bayesian nonparametric model for the extremal dependence along with an MCMC approach for posterior simulation. Section 4 illustrates the flexibility of the proposed approach by estimating the dependence structure of data simulated from some popular parametric dependence models. Section 5 provides a real data application, in which we analyse the exchange rates of the Pound Sterling against the U.S. Dollar and Japanese Yen, jointly, at extremal levels during the past few decades.

2 Extremal Dependence

In this section, we present some main ideas regarding multivariate extreme theory, which we use for the development of the framework we propose. For more details see e.g. Chapter 6 of de Haan and Ferreira (2006).

Assume that $\boldsymbol{Z} = (Z_1, Z_2)$ is a bivariate random vector of sample maxima with an extreme value distribution G. A distribution as such has the attractive feature of being *max-stable*, that is for all $n = 1, 2, \ldots$, there exists sequences of constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that $G^n(a_n z_1 + b_n, c_n z_2 + d_n) = G(z_1, z_2)$, for all $z_1, z_2 \in \mathbb{R}$. Hereafter, we refere to G as a bivariate max-stable distribution. In particular, the margins of G, denoted by $G_i(z) = \mathbb{P}(Z_i \leq z)$, for all $z \in \mathbb{R}$ and $i = 1, 2$, are members of the Generalised Extreme Value (GEV) distribution (Coles 2001, Ch. 3), i.e.

$$G_i(z_i; \mu_i, \sigma_i, \xi_i) = \exp\left\{ - \left(1 + \xi_i \frac{z_i - \mu_i}{\sigma_i}\right)_+^{-1/\xi_i} \right\}, \qquad (2.1)$$

where $z_i, \mu_i, \xi_i \in \mathbb{R}$, $\sigma_i > 0$ for $i = 1, 2$ and $(x)_+ = \max(0, x)$ and, hence, are univariate max-stable distributions. Taking the transformation, with the marginal parameters assumed to be known,

$$Y_i = \left(1 + \xi_i \frac{Z_i - \mu_i}{\sigma_i}\right)_+^{-1/\xi_i}, \quad i = 1, 2, \qquad (2.2)$$

then, the marginal distributions of $\boldsymbol{Y} = (Y_1, Y_2)$ are unit Fréchet, i.e. $\mathbb{P}(Y_i \leq y) = e^{-1/y}$, for all $y > 0$ with $i = 1, 2$, and the bivariate max-stable distribution takes the form

$$G_0(y_1, y_2) = \exp\{-L(1/y_1, 1/y_2)\}, \quad y_1, y_2 > 0, \tag{2.3}$$

where $L : [0, \infty)^2 \to [0, \infty)$, named the stable-tail dependence function (de Haan and Ferreira 2006, pp. 221–226) is given by

$$L(x_1, x_2) = 2 \int_{\mathcal{S}} \max\{x_1 u, x_2(1-u)\} H(\mathrm{d}u), \quad x_1, x_2 \geq 0. \tag{2.4}$$

$\mathcal{S} = [0, 1]$ denotes the one-dimensional simplex and H, named the angular (or spectral) distribution, is a probability distribution supported on $\mathcal{S}$ and satisfying the following condition

$$\int_{\mathcal{S}} u \, H(\mathrm{d}u) = \int_{\mathcal{S}} (1-u) \, H(\mathrm{d}u) = 1/2, \tag{C1}$$

that is, the center of the mass must be at 1/2. We stress that marginal parameters can always be estimated separately using some standard methods (e.g. de Haan and Ferreira 2006, Ch. 3, Coles 2001, Ch. 3,9) and hence be used to achieve the representation (2.3).

Furthermore, for any max-stable distribution G_0 there exists a finite measure, H^* on $\mathcal{S}$, satisfying the mean conditions $\int_{\mathcal{S}} u \, H^*(\mathrm{d}u) = \int_{\mathcal{S}} (1-u) \, H^*(\mathrm{d}u) = 1$, which implies $H^*(\mathcal{S}) = 2$, such that G_0 can be represented by the general form (2.3), where the angular distribution is given by the normalization $H := H^*/H^*(\mathcal{S})$. We will use H to denote both the probability measure and its distribution function, since the difference can be derived from the context. Conversely any probability distribution H, satisfying (C1), generates a valid bivariate max-stable distribution (de Haan and Ferreira 2006, Ch. 6).

Consider the partition $(\{0\}, \mathring{\mathcal{S}}, \{1\})$ of $\mathcal{S}$, where $\mathring{\mathcal{S}} = (0, 1)$ is the interior of the simplex. The angular distribution H can place mass on $\mathring{\mathcal{S}}$ as well as on the vertices $\{0\}$ and $\{1\}$ and, if H is absolutely continuous on $\mathring{\mathcal{S}}$, we can write

$$H(u) = p_0 + \mathbb{1}_{\mathring{\mathcal{S}}}(u) \mathring{H}(u) + p_1 \mathbb{1}_{\{1\}}(u), \quad u \in \mathcal{S} \tag{2.5}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A, $p_0, p_1 \in [0, 1/2]$ denotes point masses at the vertices of the simplex, and $\mathring{H}(u) = \int_0^u h(t) \, \mathrm{d}t$ for some continuous

$h : \mathring{\mathcal{S}} \to [0,1]$ such that $\int_0^1 h(u)\,\mathrm{d}u = 1 - p_0 - p_1$. Notice that, by the mean constraint (C1), the following two identities must be satisfied

$$p_1 = 1/2 - \int_0^1 u\,h(u)\,\mathrm{d}u, \quad p_0 = 1/2 - \int_0^1 (1-u)\,h(u)\,\mathrm{d}u. \tag{2.6}$$

In the following sections, we will denote by $\mathcal{H}$ the space of angular distributions defined in this way, so that each $H \in \mathcal{H}$ is defined by a valid triplet $(p_0, p_1, \mathring{H})$. The properties of the stable-tail dependence function are: a) it is homogeneous of order 1, that is $L(vx_1, vx_2) = vL(x_1, x_2)$ for all $v, x_1, x_2 > 0$; b) $L(x, 0) = L(0, x) = x$ for all $x > 0$; c) it is continuous and convex, i.e. $L(v(x_1, x_2) + (1-v)(x_1', x_2')) \leq vL(x_1, x_2) + (1-v)L(x_1', x_2')$ for all $x_1, x_2, x_1', x_2' \geq 0$ and $v \in \mathcal{S}$; d) $\max(x_1, x_2) \leq L(x_1, x_2) \leq x_1 + x_2$ for all $x_1, x_2 \geq 0$. The lower and upper bounds of the last condition represent the cases of complete dependence and independence, respectively. By the homogeneity of L we have that, for all $x_1, x_2 \geq 0$,

$$L(x_1, x_2) = (x_1 + x_2)A(w), \qquad A(w) = 2\int_{\mathcal{S}} \max\{w(1-u), (1-w)u\}H(\mathrm{d}u), \tag{2.7}$$

where $w = x_2/(x_1 + x_2) \in \mathcal{S}$. The function A is called the Pickands dependence function and, by the properties of L, it satisfies the following conditions:

(C2) $A(w)$ is convex, i.e., $A(aw + (1-a)w') \leq aA(w) + (1-a)A(w')$, for $a, w, w' \in \mathcal{S}$;

(C3) $A(w)$ has lower and upper bounds

$$1/2 \leq \max(w, 1-w) \leq A(w) \leq 1; \quad w \in \mathcal{S}.$$

In condition (C3), the lower and upper bounds represent the cases of complete dependence and independence, respectively. In other words, any Pickands dependence function belongs to the class $\mathcal{A}$ of functions $A : \mathcal{S} \to [1/2, 1]$ satisfying the above conditions (Falk, Hüsler, and Reiss, 2010, Ch. 4). Conversely, if a function $A \in \mathcal{A}$ has second derivatives on $\mathring{\mathcal{S}}$, then a proper angular distribution H exists, such that

$$A(w) = 1 + 2\int_0^w H(u)\mathrm{d}u - w$$

and therefore $A'(w) = -1 + 2H(w)$ and $A''(w) = 2h(w)$. Finally, it follows from (2.6) that the masses placed by H on the vertices of the simplex can be expressed in terms of the Pickands dependence function as $p_0 = \{1 + A'(0)\}/2$ and $p_1 = \{1 - A'(1)\}/2$.

The angular distribution is also used to define another important tail dependence function, R, given by

$$R(x_1, x_2) = 2\int_{\mathcal{S}} \min\{x_1 u, x_2(1-u)\} H(\mathrm{d}u), \quad x_1, x_2 \geq 0, \tag{2.8}$$

or equivalently, by $R(x_1, x_2) = x_1 + x_2 - L(x_1, x_2)$. This function can be used to approximate the probability of simultaneous exceedances, i.e.

$$\mathbb{P}(Y_1 > y_1, Y_2 > y_2) \approx R(1/y_1, 1/y_2), \tag{2.9}$$

for high enough thresholds $y_1, y_2 > 0$ (see e.g. Beranger and Padoan 2015), as well as to compute the coefficient of upper tail dependence (see e.g. Coles 2001, p.163), i.e.

$$\chi = \lim_{y\to+\infty} \mathbb{P}(Y_1 > y | Y_2 > y) = \lim_{y\to+\infty} \mathbb{P}(Y_2 > y | Y_1 > y) \equiv R(1,1) \in [0,1]. \tag{2.10}$$

This is an important summary measure of the extremal dependence between two random variables. Y_1 and Y_2 are independent in the upper tail when $\chi = 0$, whereas they are completely dependent when $\chi = 1$.

3 Bayesian nonparametric modeling of H and A

3.1 Bernstein Polynomials Representation

The basic idea behind our proposal is to define both the angular distribution function and the Pickands dependence function as polynomials, restricted to $\mathcal{S}$, of the form $\sum_{j=0}^{k} a_j b_j(x)$, where each a_j is a real-valued coefficient and the $b_j(\cdot)$, $j = 1, 2, \ldots$ form an adequate polynomial basis. Denote by $\mathcal{P}_k$ the space of polynomials of degree k, and let $\mathcal{H}$ and $\mathcal{A}$ be the sets of angular distributions and Pickands dependence functions, respectively, as in the previous section. Since $\bigcup_{k=0}^{\infty} \mathcal{P}_k$ is dense in the spaces $\mathcal{H}$ and $\mathcal{A}$, we know that any angular distribution function in $\mathcal{H}$ as well as any Pickands dependence function in $\mathcal{A}$, can be arbitrarily well approximated by a polynomial in $\mathcal{P}_k$ for some k. Due to their shape preserving properties, it is convenient to use a Bernstein polynomial basis that, when restricted to $\mathcal{S}$, will allow us to construct proper functions on $\mathcal{H}$ and $\mathcal{A}$ by identifying valid sets of coefficients.

For each $k = 1, 2, \ldots$, the Bernstein basis polynomials of degree k are defined as

$$b_j(x; k) = \binom{k}{j} x^j (1-x)^{k-j}, \quad j = 0, \ldots, k. \tag{3.1}$$

where

$$\binom{k}{j} = \frac{k!}{j!(k-j)!}.$$

is the binomial coefficient. For $x \in \mathcal{S}$, a simple identity relates the polynomial basis and Beta density functions, namely $(k+1)b_j(x;k) = \mathrm{Be}(x|j+1, k-j+1)$, where $\mathrm{Be}(\cdot|a,b)$, denotes the beta density function with shape parameters $a, b > 0$.

We start to develop the model by defining the continuous component $\mathring{H}$ of an angular distribution in expression (2.5) as polynomial of degree k in the Bernstein form (see e.g., Lorentz 1986), for some $k = 0, 1, \ldots$ i.e.

$$\mathring{H}_k(u) := \sum_{j=0}^{k} \eta_j b_j(u;k). \tag{3.2}$$

Notice that, in order to construct an angular distribution, we only require the restriction of $\mathring{H}$ to $\mathring{\mathcal{S}}$, but the polynomial itself is well defined and infinitely differentiable on the whole real line. It is straightforward to show that the first derivative of $\mathring{H}_k$ with respect to u can be expressed as a finite linear combination of beta densities.

$$\mathring{H}_k'(u) = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j)\mathrm{Be}(u|j+1, k-j) =: h_{k-1}(u), \quad u \in \mathcal{S}. \tag{3.3}$$

It can be verified that

$$\int_0^1 h_{k-1}(u)\,\mathrm{d}u = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) = \eta_k - \eta_0 \tag{3.4}$$

and

$$\int_0^1 u\, h_{k-1}(u)\,\mathrm{d}u = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j)\frac{j+1}{k+1}, \tag{3.5}$$

$$\int_0^1 (1-u)\, h_{k-1}(u)\,\mathrm{d}u = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j)\frac{k-j}{k+1}. \tag{3.6}$$

In order for $\mathring{H}$ to define a valid angular distribution H, the coefficients, $\eta_0, \ldots, \eta_k$, must be such that the integrals (3.4)-(3.6) take the values $1 - p_0 - p_1$, $1/2 - p_1$ and $1/2 - p_0$, respectively. Furthermore, recall from the previous section that, $p_0 = H(\{0\})$ and $p_1 = H(\{1\})$, so we must have that $\mathring{H}_k(1) = 1 - p_0 - p_1$ and $\mathring{H}_k(0) = 0$. All of these conditions hold under the following restrictions

(R1) $0 = \eta_0 \leq \eta_1 \leq \ldots \leq \eta_k = 1 - p_0 - p_1$;

(R2) $\eta_1 + \cdots + \eta_k = (k+1)(1/2 - p_0)$;

(R3) $0 < p_0 < 1/2$ and $\max\{0, kp_0 - (k-1)/2\} < p_1 < p_0/k + (k-1)/2k$

Alternatively, we can start by defining a Pickands dependence function as polynomial of degree r in the Bernstein form, for some $r = 0, 1, \ldots$, i.e.

$$A_r(w) := \sum_{j=0}^{r} \beta_j b_j(w; r), \qquad w \in \mathcal{S}, \tag{3.7}$$

A_r must satisfiy conditions (C2)-(C3) in order to be a proper Pickands dependence function, and this happens if and only if the following restrictions on the coefficients hold

(R4) $\beta_0 = \beta_r = 1 \geq \beta_j$, for all $j = 1, \ldots, r-1$;

(R5) for some $0 < p_0 < 1/2$ and $\max\{0, (r-1)p_0 - (r-2)/2\} < p_1 < p_0/(r-1) + (r-2)/2(r-1)$, $\beta_1 = \frac{r-1+2p_0}{r}$ and $\beta_{r-1} = \frac{r-1+2p_1}{r}$;

(R6) $\beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0$, $j = 0, \ldots, r-2$.

Taking the second derivative of A_r with respect to w we obtain

$$A_r''(w) = r \sum_{j=0}^{r-2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) \operatorname{Be}(w|j+1, r-j-1) =: h_{r-2}(w), \quad w \in \mathcal{S}. \tag{3.8}$$

It can be checked that

$$\int_0^1 h_{r-2}(w)\,\mathrm{d}w = r(2 - \beta_1 - \beta_{r-1}) \equiv 2 - \{1 + A_r'(0)\} - \{1 - A_r'(1)\}, \tag{3.9}$$

where A_r' is the first derivative of A_r with respect to w, and

$$\int_0^1 w\, h_{r-2}(w)\,\mathrm{d}w = r(1 - \beta_{r-1}) \equiv 1 - \{1 - A_{r-1}'(1)\}, \tag{3.10}$$

$$\int_0^1 (1-w)\, h_{r-2}(w)\,\mathrm{d}w = r(1 - \beta_1) \equiv 1 - \{1 + A_{r-1}'(0)\} \tag{3.11}$$

and these results are consistent with the theoretical arguments discussed in Section 2. The following result provides a major benefit of defining the angular distribution function and the Pickands dependence function through polynomials.

Proposition 3.1. *Let H be defined by (2.5) with point masses p_0 and p_1, and $\mathring{H}$ as in (3.2). And let A be defined by (3.7). Then, the following are equivalent:*

i) Starting from H, given by $\mathring{H}_k$, one may recover $A = A_r$ by setting $r = k+1$ and the coefficients:

$$\beta_j = \frac{2}{k+1}\left(\sum_{i=0}^{j-1}\eta_i + j\,p_0 + \frac{k+1-j}{2}\right), \quad j = 0,\dots,k+1 \tag{3.12}$$

Conversely, starting from $A = A_r$ one may recover H from the point masses p_0 and p_1 and the polynomial $\mathring{H}_k$ given by $k = r-1$ and the coefficients:

$$\eta_j = \frac{k+1}{2}\left(\beta_{j+1} - \beta_j + \frac{1-2p_0}{k+1}\right), \quad j = 0,\dots,k. \tag{3.13}$$

ii) Restrictions (R1)-(R3) are satisfied and H meets condition (C1), if and only if restrictions (R4)-(R6) are verified and A meets conditions (C2)-(C3).

This result tells us that, by using Bernstein polynomials, the two representations of the extremal dependence namely, the angular distribution and the Pickands dependence function can be recovered simultaneously, through a one-to-one relationship between their corresponding coefficients. In other words the η_j coefficients in equation (3.2) can be calculated from the β_j coefficients of equation (3.7) and vice versa. And when one set of coefficients satisfies restrictions (R1)-(R3) the corresponding transformation also satisfies conditions (R4)-(R6). As a consequence, for inference, there is no need to choose between different characterizations of the dependence structure.

3.2 Prior, likelihood and posterior

We provide details of the key components of a Bayesian nonparametric model for the extremal dependence, which can be formulated in terms of H or A, indifferently since, as seen in Section 3.1, one expression can always be recovered from the other. We show that the explicit forms of the prior distribution and the likelihood function for one approach are linked to those of the other.

We construct a prior distribution on the space $\mathcal{H}$ of valid angular distributions through the expression $H(u) = p_0 + \mathbb{1}_{\mathring{\mathcal{S}}}(u)\mathring{H}(u) + \mathbb{1}_{\{1\}}(u)p_1$ for all $u \in \mathcal{S}$, by writing $\mathring{H} := \mathring{H}_k$ as in (3.2), for some polynomial order k. For simplicity we assume in the present work that p_0 and p_1 are known constants. Thus, the prior on $\mathcal{H}$ is induced by a joint prior on $(k, \eta_0, \ldots, \eta_k)$. which we denote by

$$\Pi(\boldsymbol{\eta}_k, k) = \Pi(\boldsymbol{\eta}_k | k) \cdot \Pi(k). \tag{3.14}$$

Note that for $k < 3$, the resulting dependence structure is trivial, so we will only consider the case when $k \geq 3$. A convenient choice for the prior distribution of the polynomial order is to use $\Pi(k) = \text{Pois}(k + 3|\kappa)$, where $\kappa > 0$ is the rate of Poisson distribution. We propose to set the hyperparameter equal to $\kappa = \exp(-2\chi + 3.5)$, where χ is as in (2.10). In this way the range of κ is between around 4 and 33, favoring low values in the case of strong dependence and high values in the case of weak dependence. In order to define a valid prior on $\mathcal{H}$, $\Pi(\boldsymbol{\eta}_k|k)$ must assign for each $k \in \mathbb{N}$, probability one to the set $\mathscr{E} = \mathscr{E}(k) \subset \mathcal{S}^{k+1}$ of $(k+1)$-dimensional vectors satisfying (R1)-(R3). Recalling that η_0 and η_k are fixed values depending only on p_0 and p_1, we construct such a prior by focusing on the differences $X_j = \eta_j - \eta_{j-1}$, $j = 1, \ldots, k-1$, and assuming that they are conditionally uniformly distributed on appropriate intervals that we will now specify. It follows from the restrictions that

$$\sum_{j=1}^{k-1} \eta_j = \sum_{j=1}^{k-1} (k - j) X_j = (k-1)/2 + p_1 - k p_0. \tag{3.15}$$

Thus, letting $X_1 = \eta_1$, it follows, with few manipulations, that

$$\max(0, (3-k)/2 + p_1 - k\, p_0) \leq X_1 \leq 1/2 + (p_1 - k\, p_0)/(k-1).$$

We repeat this reasoning sequentially, for $j = 1, \ldots, k-1$. Since $\eta_{j-1} \leq \eta_j \leq 1$ we have that $0 \leq X_j \leq 1 - (X_1 + \cdots + X_{j-1})$ and from (3.15) with few algebraic steps we obtain the inequalities

$$X_j \geq \max\left(0, \frac{2j+1-k}{2} + p_1 - k p_0 - \sum_{i=1}^{j-1} (j+1-i) X_i\right)$$

and

$$X_j \leq \min\left(1 - \sum_{i=1}^{j-1} X_i, \frac{(k-1)/2 + p_1 - k p_0 - \sum_{i=1}^{j-1} (j+1-i) X_i}{k-j}\right).$$

Re-expressing the inequalities we have that, for $j = 1, \ldots, k-1$, $\eta_j \in \mathscr{E}_j$, given by

$$\mathscr{E}_j = \left\{ \max\left(\eta_{j-1}, \frac{2j+1-k}{2} + p_1 - k p_0 - \sum_{i=1}^{j-1} \eta_i\right), \frac{(k-1)/2 + p_1 - k p_0 - \sum_{i=1}^{j-1} \eta_i}{k-j} \right\}.$$

We therefore let $\eta_j|(\eta_1,\ldots,\eta_{j-1}) \sim \mathrm{unif}(\eta_j|\mathscr{E}_j)$, for $j=1,\ldots,k-1$, i.e., conditional on $(\eta_1,\ldots,\eta_{j-1})$, η_j is chosen, a priori, to be uniform in $\mathscr{E}_j$. Thus, given k, we obtain a valid prior

$$\begin{aligned}\Pi(\boldsymbol{\eta}_k|k) &= \mathbb{1}_{\{0\}}(\eta_0)\,\Pi(\eta_1)\prod_{j=2}^{k-1}\Pi(\eta_j|\eta_1,\ldots,\eta_{j-1})\mathbb{1}_{\{1-p_0-p_1\}}(\eta_k)\\ &= \mathbb{1}_{\{0\}}(\eta_0)\prod_{j=1}^{k-1}\mathrm{unif}(\eta_j|\mathscr{E}_j)\mathbb{1}_{\mathscr{E}_j}(\eta_j)\,\mathbb{1}_{\{1-p_0-p_1\}}(\eta_k). \end{aligned} \tag{3.16}$$

A direct consequence of Proposition 3.1 is that a valid prior distribution is induced also on the space $\mathcal{A}$ of Pickands dependence functions, as expressed by the following result.

Corollary 3.2. *Let $\mathscr{B}=\mathscr{B}(k)\subset\mathcal{S}^{k+2}$ be the space of $(k+2)$-dimensional vectors satisfying restrictions (R4)-(R6). Then, for any fixed $k\geq 3$ the prior distribution* (3.16) *induces a prior on the beta coefficients of A given by*

$$\begin{aligned}\Pi(\boldsymbol{\beta}_k|k) = \mathbb{1}_{\{1\}}(\beta_0)\,\mathbb{1}_{\{(k+2p_0)/(k+1)\}}(\beta_1)\left(\frac{k+1}{2}\right)^{k-2}\prod_{j=2}^{k-1}\mathit{unif}(\beta_j|\mathscr{B}_j)\mathbb{1}_{\mathscr{B}_j}(\beta_j)\\ \times\,\mathbb{1}_{\{(k+2p_1)/(k+1)\}}(\beta_k)\,\mathbb{1}_{\{1\}}(\beta_{k+1}).\end{aligned}$$

In other words, $\beta_j|(\beta_2,\ldots,\beta_{j-1})\sim \mathit{unif}(\beta_j|\mathscr{B}_j)$, where

$$\mathscr{B}_j = \Bigg[\max\left\{2\,\beta_{j-1}-\beta_{j-2},\ \frac{1}{k+1}\Big(j+2p_1-2p_0(k-j)\Big)\right\},$$

$$\frac{1}{k-j+1}\left(\frac{k+2p_1}{k+1}+(k-j)\beta_{j-1}\right)\Bigg].$$

This result follows from the multivariate change of variables. Let $\beta(\eta|\,k)$ denotes the inverse transformation, with expression (3.12), of the transformation $\eta(\beta|\,k)$, with expression (3.13). The Jacobian of the inverse transformation is $\{(k+1)/2\}^{k-2}$. Then, by applying the change of variable formula we obtained the conditional probability density function reported above.

We derive the analytical expression of the likelihood function and we focus on the joint distribution for the bivariate observations based on the Pickands dependence function (2.7), therefore avoiding the need for computing an integral that would be required by the alternative representation in (2.8). With this, the joint probability density function (p.d.f.) is given by

$$g(y_1, y_2) = |J(y_1, y_2)| \, \frac{\partial^2}{\partial x_1 \partial x_2} G(1/x_1, 1/x_2)\Big|_{x_1=1/y_1, x_2=1/y_2},$$

for all $y_1, y_2 > 0$, where $J(y_1, y_2) = (y_1 y_2)^{-2}$. Therefore, we obtain

$$g(y_1, y_2) = G(y_1, y_2) \left[\frac{\{A(w) - w\, A'(w)\}\, \{A(w) + (1-w)\, A'(w)\}}{(y_1 y_2)^2} + \frac{A''(w)}{(y_1 + y_2)^3} \right].$$

Let $\boldsymbol{Y}_{1:n} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ be i.i.d. copies of a bivariate max-stable random vector, with joint density $g(y_1, y_2)$ and polynomial Pickands dependence function given by expression (3.7), for fixed $r = k+1$ (following Proposition 3.1). Then, the log-likelihood function for the data is

$$\begin{aligned}
\ell(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = &- \sum_{i=1}^{n} \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}} \right) \sum_{j=0}^{k+1} \beta_j \; b_j(w_i; k+1) \\
&+ \sum_{i=1}^{n} \log \left\{ \frac{\sum_{j=0}^{k+1} \beta_j \; b_j(w_i; k+1) - w_i\, (k+1) \, \sum_{j=0}^{k} (\beta_{j+1} - \beta_j) \; b_j(w_i; k)}{(y_{1,i} y_{2,i})^2} \right. \\
&\times \frac{\sum_{j=0}^{k+1} \beta_j \; b_j(w_i; k+1) + (1 - w_i)\, (k+1) \, \sum_{j=0}^{k} (\beta_{j+1} - \beta_j) \; b_j(w_i; k)}{(y_{1,i} y_{2,i})^2} \\
&\left. + \frac{k\,(k+1) \, \sum_{j=0}^{k-1} (\beta_{j+2} - 2\,\beta_{j+1} + \beta_j) \; b_j(w_i; k-1)}{(y_{1,i} + y_{2,i})^3} \right\},
\end{aligned} \tag{3.17}$$

where $\boldsymbol{\theta} = (k, \beta_0, \ldots, \beta_{k+1}) \in \boldsymbol{\Theta} \subseteq (\mathbb{N} \times \mathcal{S}^{k+2})$. We denote by $\mathcal{L}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})$ the associated likelihood function. We may once again apply Proposition 3.1, to obtain the log-likelihood function in terms of $\boldsymbol{\theta} = (k, \eta_0, \ldots, \eta_k) \in \boldsymbol{\Theta} \subseteq (\mathbb{N} \times \mathcal{S}^{k+1})$ which, abusing terminology, can be seen as a reparametrization. More formally, this corresponds to the representation of the angular distribution (2.5), in the joint distribution (2.3), by means of a polynomial angular distribution given by the expression (3.2).

There is no closed form for the posterior distribution $\Pi^n(\boldsymbol{\theta}|\boldsymbol{Y}_{1:n}) \propto \Pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})$, regardless of the representation considered. For this reason we base the model inference on a complex MCMC posterior simulation scheme and, to be concise, we only describe the estimation procedure of the polynomial angular distribution, since it has been established that the Pickands dependence function can be obtained through a transformation. The main difficulty stems from the fact that, at each MCMC iteration, the dimension of the vector of coefficients $\boldsymbol{\eta}_k$ changes with k. We therefore resort to a trans-dimensional MCMC scheme proposed by Godsill (2001) and, in the infinite-dimensional case, applied by Antoniano-Villalobos and Walker (2013). Thus, we extend $\Pi(\boldsymbol{\eta}_k, k)$ to

$$\Pi(\boldsymbol{\eta}_\infty, k) = \Pi(\boldsymbol{\eta}_k|k)\ \Pi(k)\ \prod_{j>k} \Pi(\eta_j),$$

where $\boldsymbol{\eta}_\infty = (\eta_0, \eta_1, \ldots)$ denotes an infinite sequence of which, given k only the first $k+1$ elements are relevant, and $\Pi(\eta_j)$ is any fully known distribution. In order to update the pair $(k^{(s)}, \boldsymbol{\eta}_\infty^{(s)})$ to the current state s of the Markov chain, we propose a Metropolis-Hastings step with the following proposal distribution,

$$q(k, \boldsymbol{\eta}_\infty|k^{(s)}, \boldsymbol{\eta}_\infty^{(s)}) = q_k(k|k^{(s)}) \cdot q_\eta(\boldsymbol{\eta}_k|k) \cdot \prod_{j>k} \Pi(\eta_j)$$

where $q_\eta(\boldsymbol{\eta}_k|k)$ coincides with the prior $\Pi(\boldsymbol{\eta}_k|k)$, $\Pi(\eta_j)$ is a fully specified density on $\mathcal{S}$ and

$$q_k\left(k = k^{(s)} + 1|k^{(s)}\right) = \begin{cases} 1 & \text{if } k^{(s)} = 3 \\ 1/2 & \text{if } k^{(s)} > 3 \end{cases}$$

and

$$q_k\left(k = k^{(s)} - 1|k^{(s)}\right) = \begin{cases} 0 & \text{if } k^{(s)} = 3 \\ 1/2 & \text{if } k^{(s)} > 3. \end{cases}$$

Thus, given the current states s of the Markov chain and the proposals indexed by $s+1$, the acceptance probability depends on the ratio

$$p\left(k^{(s+1)}, \boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)}, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}^{(s)}\right) = \frac{\Pi^n(k^{(s+1)}, \boldsymbol{\eta}_\infty^{(s+1)}|\boldsymbol{Y}_{1:n})\, q(k^{(s)}, \boldsymbol{\eta}_\infty^{(s)}|k^{(s+1)}, \boldsymbol{\eta}_\infty^{(s+1)})}{\Pi^n(k^{(s)}, \boldsymbol{\eta}_\infty^{(s)}|\boldsymbol{Y}_{1:n})\, q(k^{(s+1)}, \boldsymbol{\eta}_\infty^{(s+1)}|k^{(s)}, \boldsymbol{\eta}_\infty^{(s)})}$$

which, for any $k^{(s)} > 3$, simplifies to

$$p\left(k^{(s+1)}, \boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)}, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}^{(s)}\right) = \frac{(k^{(s)} - 3)!\, \kappa^{k^{(s+1)} - k^{(s)}}}{(k^{(s+1)} - 3)!} \, \frac{\mathcal{L}(\boldsymbol{y}_{1:n}; k^{(s+1)}, \boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)})}{\mathcal{L}(\boldsymbol{y}_{1:n}; k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}^{(s)})},$$

For $k^{(s)} = 3$, we have $k^{(s+1)} = k^{(s)} + 1$ with probability one, so there is a 1/2 factor multiplying the ratio.

This leads to the following algorithm

Algorithm 3.3. *MCMC scheme to draw samples from the posterior distribution* $\Pi^n(k, \boldsymbol{\eta}_k|\boldsymbol{Y}_{1:n})$ *of the polynomial order and coefficients.*

1. *Set* $s = 0$ *and some starting values for the parameters* $\left(k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}^{(s)} \in \mathscr{E}_{k^{(s)}}\right)$*;*

2. *Repeat* M *times the update of the parameters according to:*

   (a) *Draw the proposals* $k^{(s+1)} \sim q_k(k|k^{(s)})$ *and* $\boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)} \sim q_{\boldsymbol{\eta}}(\boldsymbol{\eta}_k|k^{(s+1)}, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}})$*;*

   (b) *Compute the acceptance probability:*

$$p = \min\left(p\left(k^{(s+1)}, \boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)}, k^{(s)}, \boldsymbol{\eta}_{k^{(s)}}^{(s)}\right),\ 1\right);$$

   (c) *Draw* $U \sim unif(0,1)$ *and if* $U > p$ *then set:*

$$\left(k^{(s+1)},\ \boldsymbol{\eta}_{k^{(s+1)}}^{(s+1)}\right) = \left(k^{(s)},\ \boldsymbol{\eta}_{k^{(s)}}^{(s)}\right);$$

   (d) *Set* $s = s + 1$*;*

Thus, after an appropriate burning period of say m iterations, the sequence $(k^{(s)}\, \boldsymbol{\eta}_{k^{(s)}}^{(s)})_{s=m+1}^{M+1}$ provides a sample from the posterior distribution $\Pi^n(k, \boldsymbol{\eta}_k|\boldsymbol{Y}_{1:n})$.

An important goal of an extreme value analysis is to predict the probability of future simultaneous exceedances, as in expression (2.9). This task can be fully performed, within the Bayesian paradigm, through a Monte Carlo estimate of the posterior predictive distribution, i.e.

$$\mathbb{P}(Y_1 > y_1^*, Y_2 > y_2^*|\boldsymbol{Y}_{1:n}) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{P}(Y_1 > y_1^*, Y_2 > y_2^*|\boldsymbol{\theta})\Pi^n(\boldsymbol{\theta}|\boldsymbol{Y}_{1:n})\mathrm{d}\boldsymbol{\theta}, \qquad (3.18)$$

where $y_1^*, y_2^* > 0$ are unobserved thresholds. For each element of the posterior sample, applying expressions (2.9), (2.5), (3.2), we have that

$$\begin{aligned}\mathbb{P}(Y_1 > y_1^*, Y_2 > y_2^*|\boldsymbol{\theta}) &= \frac{1}{k+1}\sum_{j=0}^{k-1}(\eta_{j+1} - \eta_j)\left(\frac{(j+1)\mathrm{B}(y_1^*/(y_1^*+y_2^*)|j+2, k-j)}{y_1}\right. \\ &+ \left.\frac{(k-j)\mathrm{B}(y_2^*/(y_1^*+y_2^*)|k-j+1, j+1)}{y_2}\right),\end{aligned}$$

where $B(x|a,b)$, for $x \in \mathcal{S}$, denotes the cumulative distribution function of a Beta random variable with shape parameters $a, b > 0$. Therefore, an estimate can be obtained by averaging these quantities over the complete posterior sample.

The efficacy of our proposed model and inference methodology is numerically illustrated in the next section.

4 Numerical Examples

We illustrate the performance and flexibility of our methodology with a simulation study where the extremal dependence of some well-known parametric models is estimated. For each example, a data sample of size $n = 100$, with common unit Fréchet marginal distributions is simulated and the angular measure and Pickands dependence function are estimated. Estimation results are compared with the the corresponding theoretical functions, see Figures 1 and 2. Furthermore, we consider $M + 1 = 200$ thousand iterations for the MCMC Algorithm 3.3 and a burning period of $m = 80$ thousand iterations. Throughout the estimation procedure we fixed $p_0 = p_1 = 0$.

Firstly, the symmetric logistic model is considered (e.g., Coles 2001, p. 146), with strong, mild and weak dependence structures obtained setting the parameter values $\alpha = 0.45, 0.60$ and 0.85, respectively. Using the criteria described in the previous section leads to values of the prior parameter for the polynomial orders equal to $\kappa = 10.10, 10.74$ and 16.85, respectively, for decreasing dependence strength. The corresponding prior distributions are represented by the green line in the fourth-row panels of Figure 1, together with the corresponding posteriors in red. Notice that the stronger the dependence is within maxima, the more concentrated the posterior is around its median value. The top panels in Figure 1 illustrate the estimated angular probability density functions (red line) with the theoretical ones (black line). Specifically, a reasonable functional estimator of $h(w)$ is obtained through the polynomial representation (3.3), with polynomial orders $k^* = 12, 11, 18$, which are the median values of each posterior distribution. The corresponding coefficient vectors, $\boldsymbol{\eta}^*_{k^*}$ are obtained as the component-wise mean of all $\boldsymbol{\eta}^{(s)}_{k^*}$, $s = m+1, \dots, M+1$, corresponding to MCMC samples with polynomial order k^*. This is not the usual Monte Carlo estimate, but it is justified by the high concentration of the posterior density around k^* and the need to guarantee that the curve preserves the required properties (see Section 2), which would not be the case for a point-wise estimate. The grey shades represent credibility bands obtained through functional boxplots (Sun and Genton 2011) of the realizations from the posterior distribution of $h(w)$ induced by the $(k^{(s)}, \boldsymbol{\eta}^{(s)}_k)$ pairs with k ranging between the 2.5% and 97.5% percentile of its posterior distribution. The angular density is well estimated, in particular for strong and weak dependence structures. Generally, the estimate approximates well the true function in the central region of the simplex, while it may not reach the true total mass accumulated at the vertices of the simplex, zero and one, for the case of weaker dependence. The corresponding estimation of the Pickands dependence function, obtained by means of equation (3.12), is illustrated in the second-row panels. Additionally, in order to give a measure of the variability of the posterior, credibility bands are computed based on the same approach as for the angular distribution, say by considering the 2.5% and 97.5% quantiles of the posterior distribution of the model (grey shade). Overall, good results are obtained even for different strengths of dependence. Going beyond visual checks, we measure the discrepancy between the true curve A and a realization $A^{(s)}$, obtained from the Bernstein polynomial representation at each iteration of the MCMC, through the integrated squared error:

$$\mathrm{ISE}(A^{(s)}, A) = \int_0^1 \left(A^{(s)}(w) - A(w)\right)^2 \mathrm{d}w.$$

In the case of strong dependence the ISE drops at the starting iterations, whereas for the weak dependence structure the value is close to zero from the beginning as depicted from the panels in the third row. This illustrates the fast convergence of algorithm 3.3, for an arbitrary starting value $\left(k^{(0)}, \boldsymbol{\eta}^{(0)}_{k^{(0)}}\right)$ of the chain. The same behaviour is observed on the fourth-row panels where the prior (green line) and posterior (red line) distributions for the degree of the polynomial are shown. In the left two panels the posterior distribution concentrates more on smaller values with respect to the prior, while in the case of weak dependence, the two distributions almost coincide. Once the estimation of the dependence is assessed, we focus on the prediction of the probabilities of future simultaneous exceedances (3.18). The resulting estimated probabilities are displayed in the bottom-panels of Figure 1 (red line) and compared to the theoretical ones (black line) for a grid of thresholds (y_1^*, y_2^*) ranging between 10 and 100. The point predictions of the curves match the true values, especially for the strong and mild dependence structures. In the case of weak dependence a slight underestimation of the probabilities is observed, with the estimated curves decreasing a little faster then the true ones, but the symmetry in the dependence is adequately recovered.

Analogous considerations apply to Figure 2, concerning dependence estimation for data generated from different parametric extremal models, with a comparable mild dependence structure. The panels in the first column correspond to the asymmetric logistic model (Tawn 1990) with dependence parameter $\alpha = 0.6$ and asymmetry moderated by $(\tau_1, \tau_2) = (0.3, 0.8)$. Overall, most of the dependence and, in particular the asymmetry, is recovered by the model. An exception concerns the behaviour of the estimated angular probability distribution close to the vertices of the simplex and the estimated Pickands dependence function close to end-point one. Such a result was expected, since the asymmetric logistic model places point masses $1 - \tau_1$ and $1 - \tau_2$ on the corners zero and one respectively, and this stresses the need to incorporate the point masses $p_0, p_1 > 0$. This has a consequence on the estimated predictive probabilities. Indeed, the red contour lines of the bottom panel correctly describe the joint probabilities of exceedance for any threshold y_2^* between 10 and 100 and a moderately low threshold y_1^* for the first component, but show higher exceedance probabilities than the true for increasing thresholds y_1^*, due to the excessive mass that the estimated angular distribution concentrates close to one. The prior and posterior distributions for the degree of the polynomial seem to coincide having the rate of the poisson distribution equal to $\kappa = 15.27$, therefore an approximated median $\kappa + 1/3 - 0.02/\kappa \approx 16$, and median of the posterior $k = 16$.

The second-column panels of Figure 2 refer to the Hüsler-Reiss model (Hüsler and Reiss 1989) with dependence parameter $\lambda = 1.2$. We can see that in terms of the Pickands dependence function, the structure is well captured even by the median posterior curve. The median posterior curve of $h(w)$ does not recover the smoothness of the true curve, but it still approximates the true behaviour well enough. In fact, the pattern of the angular density is completely covered by the credibility bands. In this case, there is a significant difference between the prior and posterior distributions of the polynomial order, where the former coincides with a Poisson with rate $\kappa = 4.60$ and approximated median equal to 5 and the latter to a distribution with median $k = 9$. It can be interpreted as a good performance of our algorithm in terms of incorporating the information contained in the data regarding the set of polynomials which properly represent the dependence structure.

Finally, the last-column panels of Figure 2 illustrate the estimation of the dependence structure of the Extremal-t model (Nikoloulopoulos, Joe, and Li 2009) with scale parameter $\omega = 0.8$ and degrees of freedom $\nu = 2$. This extremal model places a point mass equal to $T_{\nu+1}(-\omega\{(\nu + 1)/(1 - \omega^2)\}^{1/2})$ on both vertices of the simplex, where $T_{\nu+1}(\cdot)$ denotes a t distribution with $\nu+1$ degrees of freedom. We fixed the rate for the prior on the polynomial degree to $\kappa = 8.79$, corresponding to an approximated median of 9, and the corresponding posterior median is $k = 11$. In this case, although there are positive point masses on the corners, the estimation results are surprisingly good, as the first and the second panels show. Finally, for all three models, the ISE is very close to zero, although it does not seem to converge to zero, stressing a good performance of our methodology.

In conclusion, estimates of the angular distribution or the Pickands dependence function are achieved with low computational cost. Indeed, for instance to run $M + 1 = 200$ thousands iterations of the MCMC algorithm, takes only 82.50 seconds, with an intel Core i7 processor at 2.2 GHz.

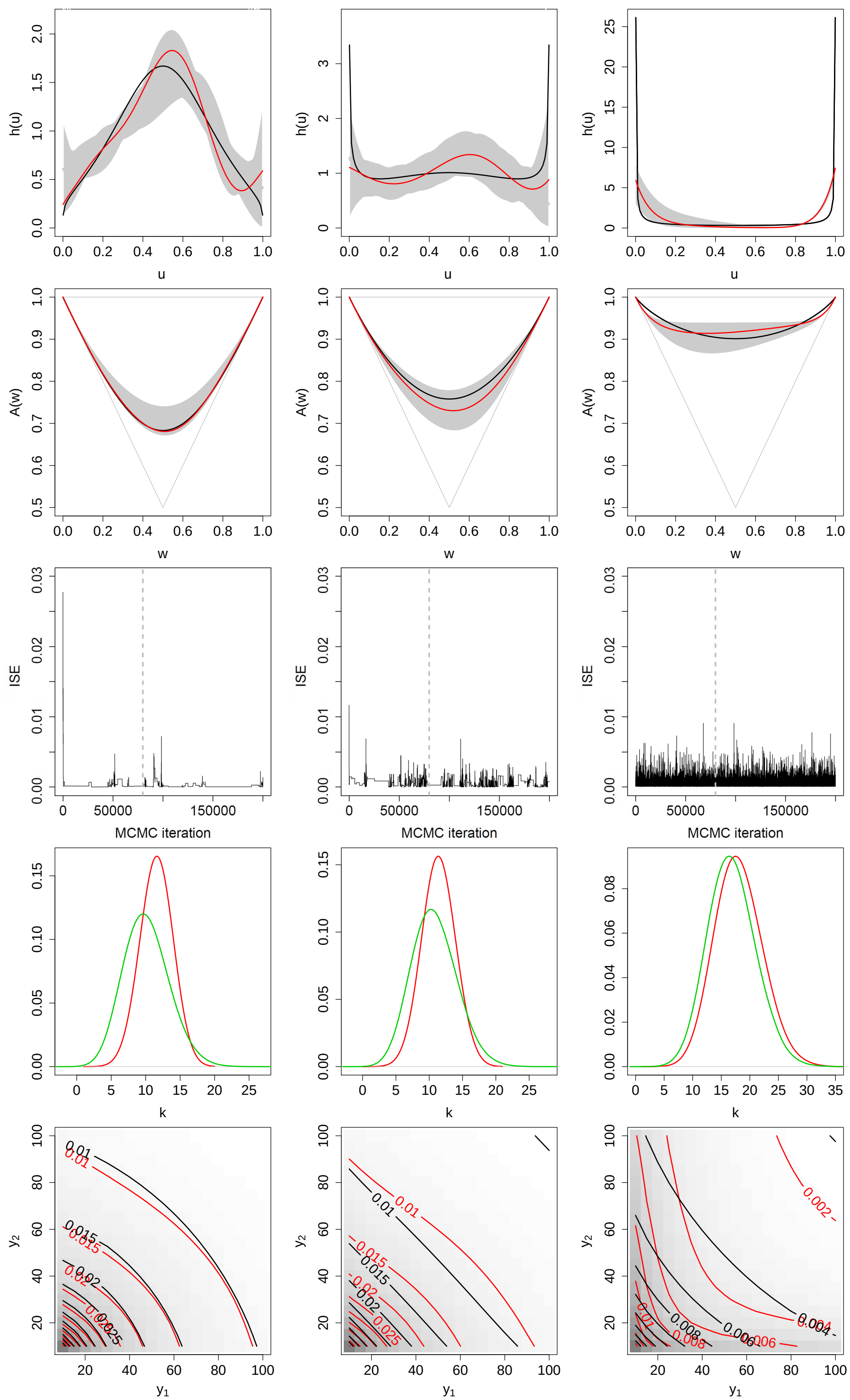

Figure 1: Bayesian nonparametric estimation of the extremal dependence for the Symmetric Logistic model with strong, mild and weak dependence ($\alpha = 0.45, 0.6, 0.85$).

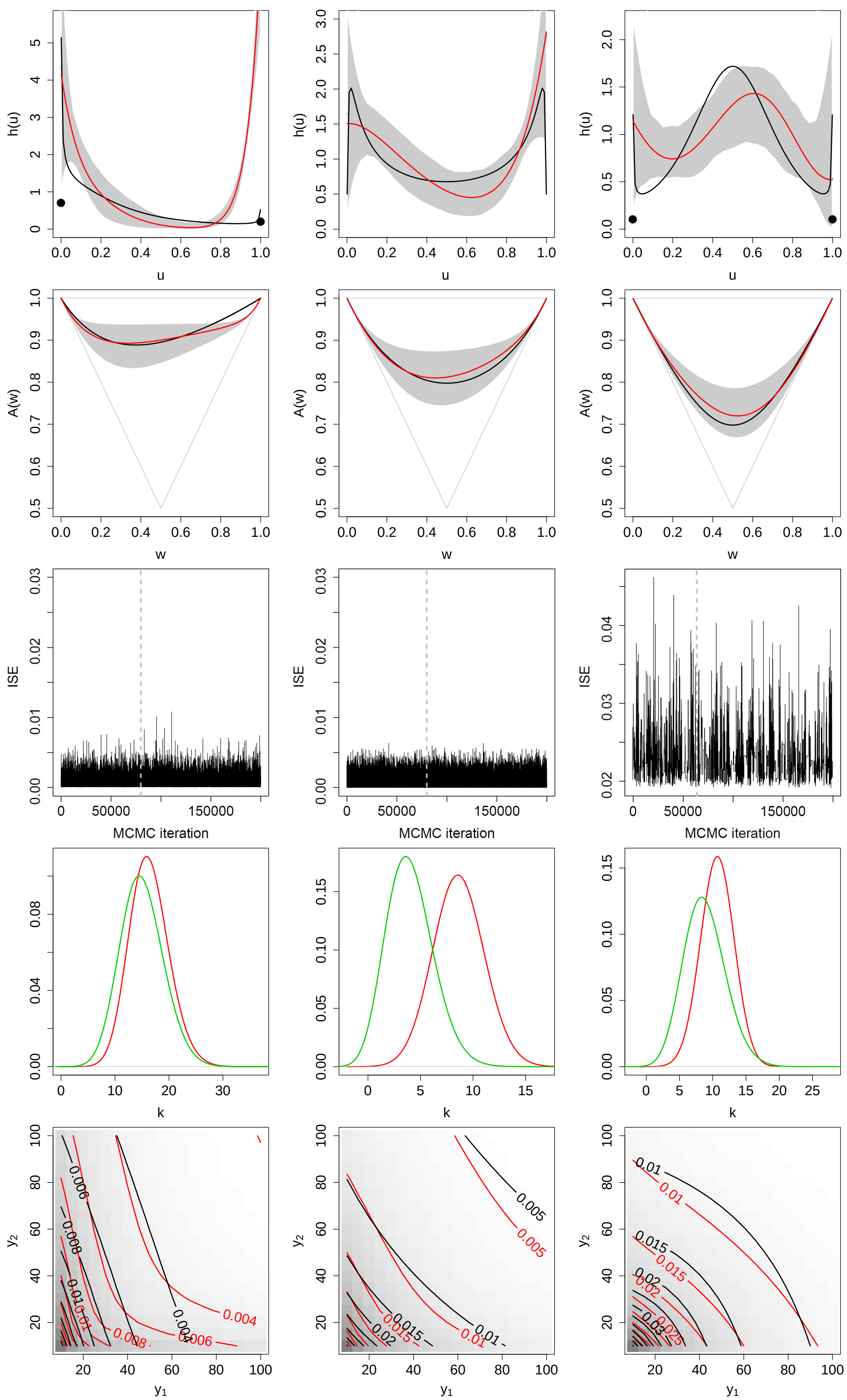

Figure 2: Bayesian nonparametric estimation of the extremal dependence for the Asymmetric Logistic ($\alpha = 0.6, \tau_1 = 0.3, \tau_2 = 0.8$), Hüsler-Reiss ($\lambda = 1.2$) and Extremal-$t$ models ($\nu = 2, \omega = 0.8$).

5 Analysis of Extreme Log-return Exchange Rates

Predicting exchange rates is one of the most challenging tasks in economics. A seminal paper by Meese and Rogoff (1983) showed that predictions of exchange rates based on macroeconomic models are unable to outperform those derived from a random walk. However, recent literature (e.g. Engel and West 2005) has established a link between exchange rates and fundamental economic principles. The modern asset market approach relies on a supply-and-demand analysis of the exchange rate viewed as the price of domestic assets in terms of foreign assets (Madura 2014). In the short-term, the exchange rate is influenced by a positive interest rate differential, which causes an appreciation of the home currency. In the long-term, a rise in the home country's price level causes the depreciation of its currency, while higher productivity or an increased demand for exports cause the appreciation of the currency (the opposite holds true for an increased demand for imports).

The United States and Japan share some common features, such as the presence of titanic enterprises and a similar monetary policy, so a strong dependence between the exchange rates of the Pound Sterling against the US dollar (GBP/USD) and the Japanese yen (GBP/JPY) is to be expected. In fact, Figure 3 shows a remarkable relation between the daily log-returns for this pair of exchange rates from March 1991 to October 2015. Our interest is in estimating extremely high (or low) joint levels of the exchange rates, thus we focus on monthly-maxima of log-returns. An

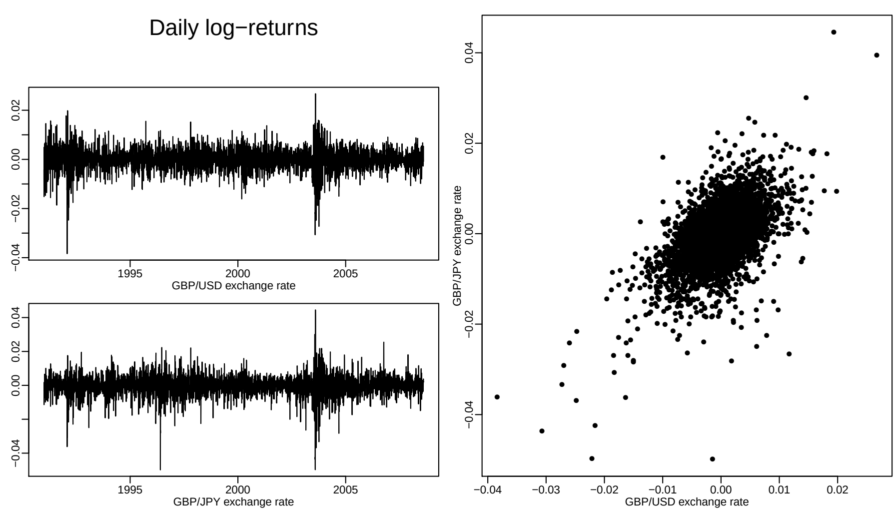

Figure 3: Daily log-returns of GBP/USD and GBP/JPY exchange rates.

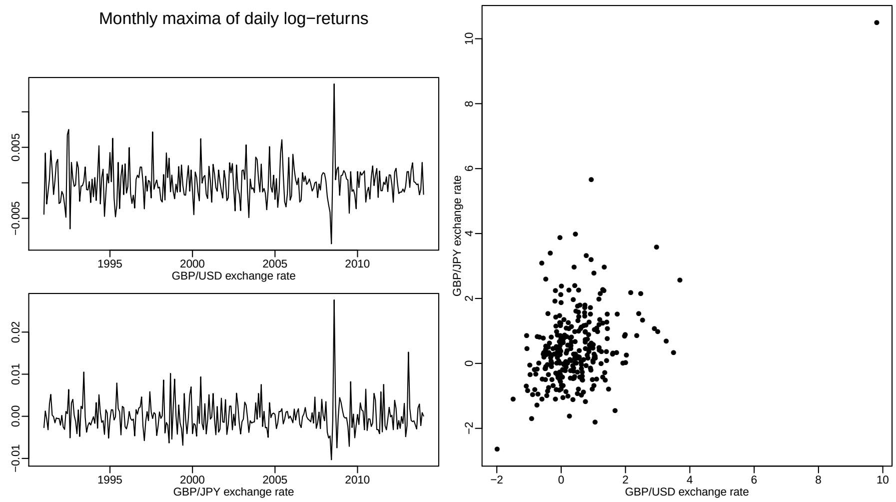

Figure 4: Monthly-maxima of log-returns of GBP/USD and GBP/JPY exchange rates.

inspection of the data shows, for instance, that monthly-maxima often occur on the same day of the month. An adequate quantification of the dependence of the bivariate maxima is crucial for predicting future extremely high exchange rates of GBP/JPY based on occurrences of extremely high exchange rates of GBP/USD, and vice versa. Figure 4 shows that an important degree of extremal dependence persists, even after removing the trend and seasonality from each of the monthly-maxima series. Firstly, we estimate the marginal GEV parameters of each series of residuals, by the maximum likelihood method. The parameter estimates for GBP/USD and GBP/JPY are $\mu_1 = 0.0055$, $\sigma_1 = 0.0025$, $\xi_1 = 0.0249$ and $\mu_2 = 0.0068$, $\sigma_2 = 0.0030$, $\xi_2 = 0.1199$, respectively. Note that ξ_2 is higher than ξ_1. Since the shape parameter drives the heaviness of the tail, the larger it is, the heavier the tail is, therefore the higher the marginal probability of observing extreme values is for GBP/JPY as opposed to GBP/USD. Then, we transform the data to obtain unit Fréchet margins, by means of transformation (2.2) and using the estimated marginal parameters. The data thus transformed can be modelled as a sample from a bivariate max-stable distribution of the type characterized by expression (2.3) and the extremal dependence of monthly-maxima of log-returns residuals can be estimated through the method we propose and describe in Section 3.

We set the parameter for the polynomial order to $\kappa = 13.60$, according to the criteria previously discussed and consider an MCMC posterior sample size of 120

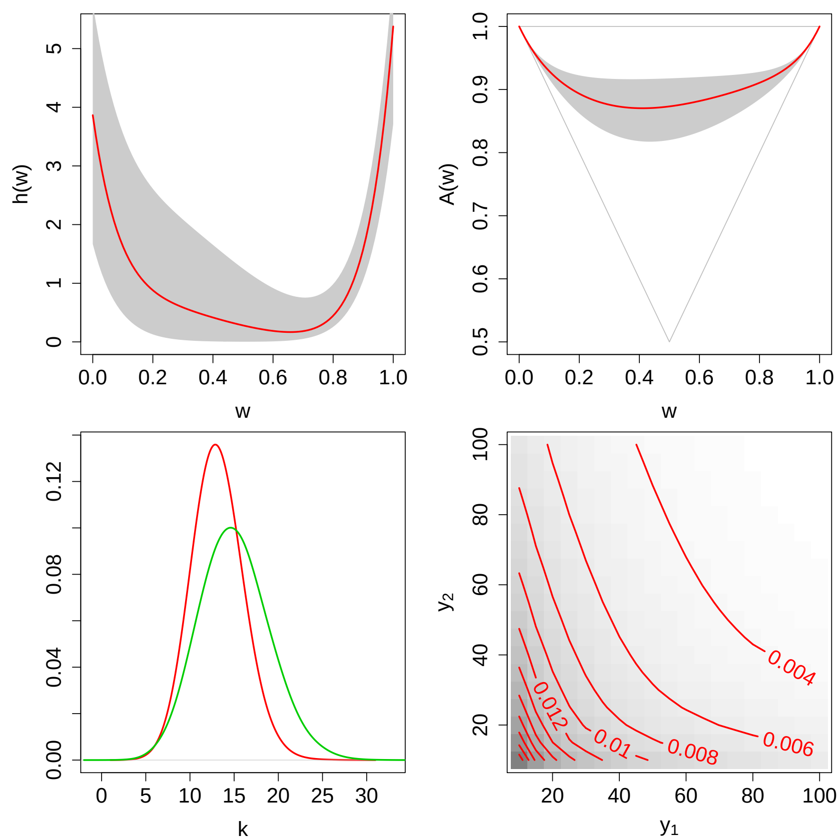

Figure 5: Bayesian nonparametric estimation of the extremal dependence for the monthly-maxima of GBP/USD and GBP/JPY log-returns of exchange rates.

thousand iterations after a burn-in of 80 thousand. The results of the analysis are summarized in Figure 5. The first row shows the estimated angular density (left) and the Pickands dependence function (right). The red lines correspond to the polynomials with coefficients equal to the mean of the posterior sample of corresponding coefficients, given a polynomial degree equal to the median of the posterior sample. With this method, it is ensured that the curves considered preserve all the necessary conditions, which would be violated by point-wise estimation methods. The use of the median value for the polynomial degree is justified by the shape of the posterior distribution for k (red line of the bottom-left panel) which concentrates its mass on a relatively small number of values. The credibility bands are computed similarly to those in the simulation study of Section 4. The green line of the bottom-left panel represents the prior distribution of the polynomial order, so we can see that the information contained in the data is reflected by the more concentrated posterior distribution. The bottom-right panel reports an estimate of the probability of joint exceedances given by expression (2.9) for combinations of the thresholds y_1^* and y_2^* ranging between 10 and 100. These have been calculated as the Monte Carlo average based on the posterior sample of the exceedance probabilities given by expression (3.18) which, as stated before, depend only on the polynomial degree and the values of the coefficients. We compared this estimate with the exceedance probability obtained by considering only the curve given by the posterior mean of the coefficients, conditional to a polynomial degree equal to the posterior median of k, as explained previously, noting no significant differences.

The estimated angular density and Pickands dependence function are asymmetric with GBP/JPY tending to assume larger values than GBP/USD. The two variables are not interchangeable and the probability that GBP/JPY exceeds a high threshold, given that GBP/USD has already exceeded such a threshold, is greater than the probability of the vice versa occurring. The predictive probabilities depicted in the bottom-right panel of Figure 5 reveal this feature. Bringing this small case study to a close, we compute both conditional probabilities when the conditioning variable exceeds its 99% percentile, i.e. $\mathbb{P}(\text{GBP/JPY} > q_1 \,|\, \text{GBP/USD} > q_1)$ and $\mathbb{P}(\text{GBP/USD} > q_2 \,|\, \text{GBP/JPY} > q_2)$. To do so we proceed as follows. We calculate q_1 and q_2 as the 99% percentiles of the marginal GEV distributions of log-returns of exchange rates GBP/USD and GBP/JPY, respectively, using the estimated marginal parameters. These are equal to $q_1 = 0.0162$ and $q_2 = 0.0221$. We transform these thresholds in order to represent them in unit-Fréchet scale by

$$y_{i,j}^* = \left\{ 1 + \xi_i \left(\frac{q_j - \mu_i}{\sigma_i} \right) \right\}_+^{(1/\xi_i)}, \qquad i, j = 1, 2.$$

Now, with q_1 we obtain the thresholds $y_{2,1}^* = 14.12$ and $y_{1,1}^* = 57.25$ and the joint predictive probability (3.18) is equal to 0.0078. Therefore, we obtain the final result $\mathbb{P}(\text{GBP/JPY} > q_1 \,|\, \text{GBP/USD} > q_1, \boldsymbol{\theta}) \approx \mathbb{P}(Y_2 > y_{2,1}^* \,|\, Y_1 > y_{1,1}^*, \boldsymbol{\theta}) = 0.4513$, Similarly, with q_2 we obtain the thresholds $y_{1,2}^* = 450.23$ and $y_{2,2}^* = 52.32$ and the joint predictive probability (3.18) is equal to 0.0014. Therefore, we obtain the final result $\mathbb{P}(\text{GBP/USD} > q_2 \,|\, \text{GBP/JPY} > q_2, \boldsymbol{\theta}) \approx \mathbb{P}(Y_1 > y_{1,2}^* \,|\, Y_2 > y_{2,2}^*, \boldsymbol{\theta}) = 0.0719$.

These results are consistent with our conclusion that GBP/JPY tends to assume larger values than GBP/USD, and the conditional probability of the log-returns of GBP/USD given high values of log-returns of GBP/JPY is quite high.

Appendix A: Proofs

Proof of Proposition 3.1. Expressions (3.3) and (3.8) provide two alternative representations of the angular probability density, which must be equivalent. As a consequence we must have that

$$k\sum_{j=0}^{k-1}\left(\eta_{j+1}-\eta_j\right)b_j(w;k-1)=\frac{r(r-1)}{2}\sum_{i=0}^{r-2}\left(\beta_{i+2}-2\beta_{i+1}+\beta_i\right)b_i(w;r-2),\quad w\in[0,1].$$

Since the $b_j(\cdot;\cdot)$ form a polynomial basis, the equality is obtained by setting $r=k+1$, the right-hand side of the above equality can be rephrased as a sum for $j=0,\dots,k-1$ and observing that each of the coefficients on the right-hand expression must be equal to the corresponding coefficient on the left side, i.e.

$$\eta_{j+1}-\eta_j=\frac{k+1}{2}\left(\beta_{j+2}-2\beta_{j+1}+\beta_j\right),\quad j=0,\dots,k-1. \tag{5.1}$$

For the proof of claim i), we first, consider expression (3.2), assuming that (R1)-(R3) are verified by the η_j coefficients. In particular, $\eta_0=0$, so applying (5.1) recursively and solving with respect to β_{j+2} we obtain,

$$\begin{aligned}
\beta_2 &= \frac{2}{k+1}\left(\eta_1+(k+1)\beta_1-\frac{k+1}{2}\beta_0\right)\\
\beta_3 &= \frac{2}{k+1}\left(\eta_1+\eta_2+\frac{3(k+1)}{2}\beta_1-(k+1)\beta_0\right)\\
\beta_4 &= \frac{2}{k+1}\left(\eta_1+\eta_2+\eta_3+2(k+1)\beta_1-\frac{3(k+1)}{2}\beta_0\right)\\
\beta_5 &= \frac{2}{k+1}\left(\eta_1+\eta_2+\eta_3+\eta_4+\frac{5(k+1)}{2}\beta_1-2(k+1)\beta_0\right)\\
&\vdots\\
\beta_{j+2} &= \frac{2}{k+1}\left(\sum_{i=0}^{j+1}\eta_i+\frac{(j+2)(k+1)}{2}\beta_1-\frac{(j+1)(k+1)}{2}\beta_0\right),
\end{aligned} \tag{5.2}$$

And $\beta_0=\beta_{k+1}=1$. In addition, we have that

$$p_0=\frac{1+A'(0)}{2}=\frac{1+(k+1)(\beta_1-\beta_0)}{2} \tag{5.3}$$

and

$$p_1=\frac{1-A'(1)}{2}=\frac{1-(k+1)(\beta_{k+1}-\beta_k)}{2} \tag{5.4}$$

from which we obtain

$$\beta_1 = \frac{2\,p_0 + k}{k+1}, \qquad \beta_k = \frac{2\,p_1 + k}{k+1}. \tag{5.5}$$

Substituting β_0 and β_1 in (5.2) and re-indexing, i.e. substituting $j+2$ with j, the general expression (3.12) is obtained.

Conversely, if we consider expression (3.7), assuming that (R4)-(R6) are verified for the β_j coefficients, applying (5.1) and resolving with respect to η_{j+1} leads to

$$\begin{aligned}
\eta_1 &= \eta_0 + \frac{k+1}{2}(\beta_2 - 2\beta_1 + \beta_0)\\
\eta_2 &= \eta_0 + \frac{k+1}{2}(\beta_3 - \beta_2 - \beta_1 + \beta_0)\\
\eta_3 &= \eta_0 + \frac{k+1}{2}(\beta_4 - \beta_3 - \beta_1 + \beta_0)\\
\eta_4 &= \eta_0 + \frac{k+1}{2}(\beta_5 - \beta_4 - \beta_1 + \beta_0)\\
&\vdots\\
\eta_{j+1} &= \eta_0 + \frac{k+1}{2}(\beta_{j+2} - \beta_{j+1} - \beta_1 + \beta_0),
\end{aligned} \tag{5.6}$$

which has a unique solution for $\eta_0 = 0$. In addition, substituting β_0 and β_1 in (5.6) and re-indexing, i.e. substituting $j+1$ with j, the general formula (3.13) is obtained.

Now, to prove claim ii), first we consider expression (3.12). When $j = 0$, under the convention that $\eta_{-1} = 0$ we obtain $\beta_0 = 1$. When $j = k+1$ we obtain $\beta_{k+1} = 2(\eta_1 + \cdots + \eta_k + (k+1)p_0)/(k+1)$ and then using (R2) this becomes $\beta_{k+1} = 1$. Thus (R4) is verified. When $j = 1$ we obtain the following result

$$\beta_1 = \frac{2p_0 + k}{k+1} \geq 1 - \frac{1}{k+1},$$

and when $j = k$, after some algebraic manipulation we obtain

$$\beta_k = \frac{2p_1 + k}{k+1} \geq 1 - \frac{1}{k+1}.$$

Thus (R5) is verified and the above inequalities ensure that the lower bound condition in (C3) is satisfied. Finally, from (3.12), again with some algebraic manipulation it can be checked that

$$\beta_{j+2} - 2\beta_{j+1} + \beta_t = \frac{2}{k+1}(\eta_{j+1} - \eta_j), \quad j = 0, \ldots, k-1.$$

Since $\eta_{j+1} \geq \eta_j$ for all $j = 0, \ldots, k-1$ by (R1), so (R6) holds.

Starting now from (3.12), by (R6), the right hand side of (5.1) is non-negative. Furthermore, taking $j = 0$, using $\beta_0 = 1$ and β_1 in (R5) we obtain

$$\begin{aligned} \eta_0 &= \frac{k+1}{2}\left(\beta_1 - \beta_0 + \frac{1-2p_0}{k+1}\right) \\ &= \frac{k+1}{2}\left(\frac{2p_0+k}{k+1} - 1 + \frac{1-2p_0}{k+1} -\right) = 0. \end{aligned}$$

Similarly, for $j = k$, using $\beta_{k+1} = 1$ and β_k in (R5) we obtain

$$\begin{aligned} \eta_k &= \frac{k+1}{2}\left(\beta_{k+1} - \beta_k + \frac{1-2p_0}{k+1}\right) \\ &= \frac{k+1}{2}\left(1 - \frac{2p_1+k}{k+1} + \frac{1-2p_0}{k+1}\right) = 1 - p_0 - p_1. \end{aligned}$$

Therefore, (R1) is verified. Finally, applying (3.13) repeatedly and with a few algebraic manipulations we obtain

$$\sum_{t=0}^{k} \eta_t = \frac{k+1}{2}\left(\beta_{k+1} - \beta_0 + \frac{(k+1)(1-2p_0)}{k+1}\right) = \frac{(k+1)}{2}\left(1-2p_0\right),$$

where we used the fact that $\beta_{k+1} = \beta_0 = 1$. Hence also (R2) is verified. Condition (R3) follows from condition (R5), and vice versa.

□

References

Antoniano-Villalobos, I. and S. G. Walker (2013). Bayesian nonparametric inference for the power likelihood. *Journal of Computational and Graphical Statistics 22*(4), 801–813.

Beranger, B. and S. A. Padoan (2015). Extreme dependence models. In D. Dey and J. Yan (Eds.), *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman and Hall/CRC.

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.

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.

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values.* Springer, London.

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

Engel, C. and K. D. West (2005). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517.

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

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

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

Godsill, S. J. (2001). On the relationship between markov chain monte carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics 10*(2), 230–248.

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

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

Madura, J. (2014). *Financial markets and institutions.* Cengage learning.

Marcon, G., S. Padoan, P. Naveau, and P. Muliere (2014). Multivariate nonparametric estimation of the pickands dependence function using bernstein polynomials. *arXiv preprint arXiv:1405.5228*.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics 14*(1), 3–24.

Nikoloulopoulos, A. K., H. Joe, and H. Li (2009). Extreme value properties of multivariate t copulas. *Extremes 12*(2), 129–148.

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.

Sun, Y. and M. G. Genton (2011). Functional boxplots. *Journal of Computational and Graphical Statistics 20*(2).

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