Electronic Journal of Statistics
Vol. 10 (2016) 3310–3337
ISSN: 1935-7524
DOI: 10.1214/16-EJS1162

Bayesian inference for the extremal dependence

Giulia Marcon

Department of Political and Social Sciences, University of Pavia, Corso Strada Nuova 65, 27100, Italy
e-mail: giulia.marcon@unipv.it

Simone A. Padoan and Isadora Antoniano-Villalobos

Department of Decision Sciences, Bocconi University, Milan, via Roentgen 1, 20136, Italy
e-mail: simone.padoan@unibocconi.it; isadora.antoniano@unibocconi.it

Abstract: A simple approach for modeling 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 nonparametric. Although the analytical expression of the posterior is unknown, inference is possible via a trans-dimensional MCMC scheme. We show the efficiency of the proposed methodology by means of a simulation study. The extremal behaviour of log-returns of daily exchange rates between the Pound Sterling vs the U.S. Dollar and the Pound Sterling vs the Japanese Yen is analysed for illustrative purposes.

Keywords and phrases: Generalised extreme value distribution, extremal dependence, angular measure, max-stable distribution, Bernstein polynomials, Bayesian nonparametrics, trans-dimensional MCMC, exchange rates.
MSC 2010 subject classifications: 62G05, 62G07, 62G32.

Received December 2015.

Contents

1. Introduction

The estimation of future extreme episodes of a real process, such as heavy-rainfall, heat-waves or 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 random vector of so-called 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 nonparametric estimators 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. (2015), among others. Examples of Bayesian modelling of the extremal dependence are Boldi and Davison (2007), Guillotte and Perron (2008) and Sabourin and Naveau (2014) to cite a few.

In order to provide a comprehensive discussion of our approach, we restrict our attention to the bivariate case, that is to two-dimensional vectors of sample maxima. Specifically, we describe how Bernstein polynomials (Lorentz, 1986) can be used to model the extremal dependence within a Bayesian nonparametric framework. In recent years, Bernstein polynomials are attracting much attention in Bayesian nonparametric statistics, in that they are useful for constructing prior distributions of distribution functions (Petrone, 1999b), for density estimation (Petrone, 1999a) and they have nice properties (Ghosal, 2001, Petrone and Wasserman, 2002).

Our present proposal has the following key features that make it different from Marcon et al. (2015). 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 so-called angular (or spectral) measure. 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., 2015). Additionally, there is no need to choose between representing the dependence by means of the angular measure or the Pickands dependence function. Finally, the expression of approximate probabilities for simultaneous exceedances can be derived in closed-form and this implies that the predictive probability for such events is easy to calculate.

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 value theory, which we use for the development of the framework we propose. For more details see e.g. Chapters 4, 6 and 8 of Falk et al. (2010), de Haan and Ferreira (2006) and Beirlant et al. (2004), respectively.

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 exist 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 refer 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.

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

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,

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

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

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

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

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

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

(C1) The center of the mass of H must be at $1/2$, that is,

$$\int_{\mathcal{S}} w\, H(\mathrm{d}w) = \int_{\mathcal{S}} (1-w)\, H(\mathrm{d}w) = 1/2.$$

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).

More precisely, for any max-stable distribution G_0 there exists a finite measure, H^* on $\mathcal{S}$, satisfying the mean conditions $\int_{\mathcal{S}} w\, H^*(\mathrm{d}w) = \int_{\mathcal{S}} (1-w)\, H^*(\mathrm{d}w) = 1$, which implies $H^*(\mathcal{S}) = 2$, such that G can be represented by the general form (2.3), where the angular measure 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 measure with distribution function H satisfying (C1) generates a valid bivariate max-stable distribution (de Haan and Ferreira, 2006, Ch. 6). As usual practice, for simplicity we focus on a subset of all the possible angular measures (Beirlant et al., 2004, Ch. 8).

Assumption 2.1. *Let* $(\{0\}, \mathring{\mathcal{S}}, \{1\})$ *be a partition of* $\mathcal{S}$*, where* $\mathring{\mathcal{S}} = (0,1)$*. Consider angular measures of the form*

$$H([a,b]) = p_0 \delta_0([a,b]) + \Delta((a,b]) + p_1 \delta_1([a,b]),$$

for any $a, b \in \mathcal{S}$ *with* $a \leq b$ *and* $p_0, p_1 \in [0, 1/2]$*. Specifically,* $\delta_x(A)$ *is the Dirac measure for any* $x \in \mathbb{R}$ *and a measurable set* $A \subset \mathbb{R}$*,* $\Delta((a,b]) = \mathring{H}(b) - \mathring{H}(a)$ *is the Lebesgue-Stieltjes measure, where* $\mathring{H}(w) = \int_0^w h(t)\,\mathrm{d}t$ *and* $h(t) \geq 0$ *is a Lebesgue integrable function such that* $\int_0^1 h(w)\,\mathrm{d}w = 1 - p_0 - p_1$*.*

The role of the angular measure can be explained by means of its geometric interpretation. The more the dependence between variables increases (the more likely it is that they are similar in value), the more the mass of H tends to accumulate at the center of the simplex, i.e. $1/2$ by condition (C1). Conversely, the more the mass of H moves to the vertices of the simplex, the more the variables become independent. The distribution function of the angular measure is

$$H(w) = p_0 + \mathring{H}(w) + p_1 \mathbb{1}_{[0,w]}(1), \quad w \in \mathcal{S} \tag{2.5}$$

where $\mathbb{1}_A(x)$ is the indicator function of the set A. This means that H has atoms on the vertices $\{0\}$ and $\{1\}$, denoted by $p_0 = H(\{0\})$ and $p_1 = H(\{1\}) = H([0,1]) - H([0,1))$ respectively, and it is absolutely continuous on $\mathring{\mathcal{S}}$. Notice that, by the mean constraint (C1), the following two identities must be satisfied

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

We stress that although (2.5) excludes atoms in $\mathring{\mathcal{S}}$, it is already rich enough to describe the dependence of many practical applications. 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 $x_1, x_2 \geq 0$,

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

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

(C2) $A(t)$ is convex, i.e., $A(at + (1-a)t') \leq aA(t) + (1-a)A(t')$, for $a, t, t' \in \mathcal{S}$;
(C3) $A(t)$ has lower and upper bounds

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

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 et al., 2010, Ch. 4). Conversely, if a function $A \in \mathcal{A}$ has second derivatives on $\mathring{\mathcal{S}}$, then a valid angular measure H exists, such that

$$A(t) = 1 + 2\int_0^t H(w)\, \mathrm{d}w - t, \quad t \in \mathcal{S} \tag{2.8}$$

and therefore $A'(t) = -1 + 2H([0,t])$, where A' is seen as the right-hand derivative and $A''(t) = 2h(t)$, for $t \in \mathring{\mathcal{S}}$ (Beirlant et al., 2004, Ch. 8). From the above relation follows that the atoms on the vertices of the simplex can be expressed by the Pickands dependence function as $p_0 = \{1 + A'(0)\}/2$ and $p_1 = \{1 - A'(1)\}/2$, where $A'(1) = \sup_{t \in [0,1)} A'(t)$.

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

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

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

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

with $\chi \in [0,1]$. 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 polynomial representation

The basic idea behind our proposal is to define both the distribution function of the angular measure 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 (Lorentz, 1986) 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

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

Throughout the article, use will be made of the simple identities,

$$(k+1)\, b_j(x; k) = \mathrm{Be}(x | j+1, k-j+1), \quad x \in \mathcal{S}, \tag{3.1}$$

where $\mathrm{Be}(\cdot | a, b)$ denotes the beta density function with shape parameters $a > 0$ and $b > 0$, and for $k \geq 1$

$$b_j(0; k) = \delta_{j,0}, \quad b_j(1; k) = \delta_{j,k}, \tag{3.2}$$

where $\delta_{j,r}$ is the Kronecker delta function (e.g., Petrone, 1999b).

We start modeling the extremal dependence by representing the distribution function (2.5) through a polynomial of degree $k-1$ in Bernstein form, for some $k = 1, 2, \dots$. Specifically, we define

$$H_{k-1}(w) := \begin{cases} \sum_{j \le k-1} \eta_j \, b_j(w; k-1) & \text{if} \quad w \in [0,1) \\ 1 & \text{if} \quad w = 1 \end{cases} \tag{3.3}$$

and taking the first derivative of H_{k-1} with respect to w, we have that the density in the interior of $\mathcal{S}$ is equal to

$$H'_{k-1}(w) = \sum_{j=0}^{k-2} (\eta_{j+1} - \eta_j) \, \mathrm{Be}(w | j+1, k-j-1) =: h_{k-1}(w), \quad w \in \mathring{\mathcal{S}} \tag{3.4}$$

Proposition 3.1. *By forcing the coefficients $\eta_0, \dots, \eta_{k-1}$ in (3.3), for fixed polynomial degree k, to meet the restrictions:*

(R1) $0 \le p_0 = \eta_0 \le \eta_1 \le \dots \le \eta_{k-1} = 1 - p_1 \le 1$*;*
(R2) $\eta_0 + \dots + \eta_{k-1} = k/2$*;*

it is ensured that H_{k-1} is the distribution function of a valid angular measure satisfying Assumption 2.1.

Alternatively, we can also model the extremal dependence by representing the Pickands dependence function in (2.7) with a polynomial of degree $k = 0, 1, \dots$ in the Bernstein form. Specifically, let

$$A_k(t) := \sum_{j=0}^{k} \beta_j b_j(t; k), \qquad t \in \mathcal{S}, \tag{3.5}$$

then by forcing the coefficients $\beta_0, \dots, \beta_k$ in (3.5) to meet the restrictions:

(R3) $\beta_0 = \beta_k = 1 \ge \beta_j$, for all $j = 1, \dots, k-1$;
(R4) $\beta_1 = \frac{k-1+2p_0}{k}$ and $\beta_{k-1} = \frac{k-1+2p_1}{k}$;
(R5) $\beta_{j+2} - 2\beta_{j+1} + \beta_j \ge 0$, $j = 0, \dots, k-2$;

it is ensured that A_k satisfies conditions (C2)–(C3) and hence it is a proper Pickands dependence function (Marcon et al., 2015). This is easily explained by the following. First, by (3.2) we have that $A_k(0) = A_k(1) = 1$ if $\beta_0 = \beta_k = 1$ and it is immediate to check that $A_k(t) = 1$ for all $t \in \mathcal{S}$ when $\beta_0 = \dots = \beta_k = 1$. Because $b_j(t; k) \le 1$ for all $t \in \mathcal{S}$ then $A_k(t) \le 1$ by (R3). Second, $A_k(t) \ge \max(t, 1-t)$ for all $t \in \mathcal{S}$ if $A'_k(0) \ge -1$ and $A'_k(1) \le 1$, where

$$A'_k(t) = \sum_{j=0}^{k-1} (\beta_{j+1} - \beta_j) \, \mathrm{Be}(t | j+1, k-j), \quad t \in \mathcal{S}. \tag{3.6}$$

Since $A'_k(0) = k(\beta_1 - \beta_0)$, $A'_k(1) = k(\beta_k - \beta_{k-1})$ and, on the other hand, knowing also from (2.8) that $A'_k(0) = 2p_0 - 1$ and $A'_k(1) = 1 - 2p_1$, we obtain the conditions in (R4), which imply that $\beta_1 \geq 1 - 1/k$ and $\beta_{k-1} \geq 1 - 1/k$. Finally, $A_k(t)$ is convex if $A_k''(t) \geq 0$ for all $t \in \mathcal{S}$, where

$$A_k^{''}(t) = k \sum_{j=0}^{k-2} (\beta_{j+2} - 2\beta_{j+1} + \beta_j) \,\mathrm{Be}(t|j+1, k-j-1), \quad t \in \mathcal{S}. \tag{3.7}$$

Clearly the positivity of (3.7) is guaranteed by the conditions in (R5).

Under Assumption 2.1, the distribution function (3.3) and the Pickands dependence function (3.5) are linked, as described by the next result.

Proposition 3.2. *Let H_{k-1} be the distribution function of an angular measure with expression (3.3), and A_k be the Pickands dependence function given in (3.5). Then, the following are equivalent:*

i) Given A_k one may recover H_{k-1} by means of their coefficients' relationship:

$$\eta_j = \frac{k}{2}\left(\beta_{j+1} - \beta_j + \frac{1}{k}\right), \quad j = 0, \ldots, k-1. \tag{3.8}$$

Conversely, given H_{k-1}, one may recover A_k by means of their coefficients' relationship:

$$\beta_{j+1} = \frac{1}{k}\left(2\sum_{i=0}^{j} \eta_i + k - j - 1\right), \quad j = 0, \ldots, k-1, \tag{3.9}$$

with $\beta_0 = 1$.

ii) Restrictions (R1) and (R2) are satisfied and H_{k-1} meets condition (C1), if and only if restrictions (R3)–(R5) are verified and A_k meets conditions (C2) and (C3).

This result tells us that the one-to-one relationship between the angular measure and the Pickands dependence function, when these are represented with Bernstein polynomials, is simply expressed through a one-to-one relationship between their corresponding coefficients. Its implications are as follows. One can estimate the coefficients in (3.3) so that they meet the conditions (R1) and (R2) and then compute the coefficients in (3.5) by equation (3.9), which will automatically meet the conditions (R3)–(R5). Or vice versa, estimate the coefficients in (3.5) satisfying conditions (R3)–(R5) and derive the coefficients in (3.3) by (3.8), which will satisfy (R1) and (R2). As a consequence, for inference, there is no need to choose between one or the other way of representing the dependence structure.

Finally, H_{k-1} and A_k can provide accurate approximations of the true functions H and A.

Proposition 3.3. *Let*

$$\mathcal{H}_{k-1} = \{w \mapsto H_{k-1}(w) = \sum_{j \leq k-1} \eta_j \, b_j(w; k-1) :$$
$$\eta_0, \ldots, \eta_{k-1} \in [0,1] \textit{ and (R1)–(R2) are satisfied}\}$$

and

$$\mathcal{A}_k = \{t \mapsto A_k(t) = \sum_{j \le k} \beta_j b_j(t;k) : \beta_0, \dots, \beta_k \in [0,1] \textit{ and (R3)–(R5) are satisfied}\}.$$

Then, $\mathcal{A}_k$ and $\mathcal{H}_{k-1}$, $k = 1, 2, \dots$ are nested sequences in $\mathcal{A}$ and $\mathcal{H}$, respectively. Additionally, there are polynomials A_k and H_{k-1} such that

$$\lim_{k\to\infty} \sup_{t\in\mathcal{S}} |A_k(t) - A(t)| = 0 \tag{3.10}$$

and

$$\lim_{k\to\infty} \sup_{w\in\mathcal{S}} |H_{k-1}(w) - H(w)| = 0. \tag{3.11}$$

3.2. Bayesian inference

We provide details of the key ingredients of a Bayesian nonparametric model for the extremal dependence. This can be formulated through (3.3) or (3.5), indifferently since, as seen in Section 3.1, one expression can always be recovered from the other. We show the explicit forms in which the prior distribution and the likelihood function for one approach are linked to those of the other.

We start by constructing a prior probability on the space $\mathcal{H}$ of valid angular measures using the Bernstein polynomial representation (3.3), for some polynomial order k. Then, the prior on $\mathcal{H}$ is induced by a joint prior distribution on $(k, \boldsymbol{\eta}_k)$, where $\boldsymbol{\eta}_k = (\eta_0, \dots, \eta_{k-1})$. We conveniently express the prior distribution as

$$\Pi(k, \boldsymbol{\eta}_k) = \Pi(\boldsymbol{\eta}_k|k)\,\Pi(k). \tag{3.12}$$

Note that for $k < 3$, the resulting dependence structure is trivial, so we will only consider the case when $k \ge 3$. Some convenient choices for the prior distribution of the polynomial order are $\Pi(k) = \text{Pois}(k-3|\kappa_P)$ or $\Pi(k) = \text{nbin}(k-3|\kappa_{NB}, \sigma^2)$, where $\kappa_P, \kappa_{NB} > 0$ are the means of Poisson and negative binomial distributions, respectively. The latter, however, is more flexible through its variance σ^2. Specifically, the probability mass function for the negative binomial distribution is $\Gamma(x+s)/(\Gamma(s)x!)\ p^s\ (1-p)^x$, for $x = 0, 1, 2, \dots$, with target for number of successful trials $s > 0$ and probability of success in each trial $0 < p \le 1$. With this parametrization, the mean corresponds to $\kappa_{NB} = s(1-p)/p$ and the variance to $\sigma^2 = s(1-p)/p^2$. 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 $\mathscr{E} = \mathscr{E}(k) \subset \mathcal{S}^k$ of k-dimensional vectors satisfying (R1) and (R2). By (R1) we have that the atoms on the edges are represented by parameters $\eta_0 = p_0$ and $\eta_{k-1} = 1 - p_1$. Given the particular role that these quantities play in the model, and the relevance of their interpretation, we have decided to treat them separately when defining the prior. Furthermore, this choice seems empirically justified by the results obtained through simulation studies. Therefore, we define the conditional prior for the polynomial coefficients given the degree k in the following manner,

$$\Pi(\boldsymbol{\eta}_k|k) = \Pi(\eta_1, \dots, \eta_{k-2}|p_1, p_0, k)\, \Pi(p_1|k, p_0)\, \Pi(p_0).$$

Specifically, we let $\Pi(p_0) = \mathrm{Unif}(0, 1/2)$. Then,

$$(k-1)\, p_0 + (1-p_1) \leq \sum_{j=0}^{k-1} \eta_j = k/2 \leq p_0 + (k-1)(1-p_1), \tag{3.13}$$

where the identity follows from condition (R2), while the two inequalities stem from (R1). After simple manipulations, it follows that, in order for (R1) and (R2) to hold, a necessary condition is $(k-1)\, p_0 - k/2 + 1 \leq p_1 \leq (p_0 + k/2 - 1)/(k-1)$, so we set $\Pi(p_1|k, p_0) = \mathrm{Unif}(a, b)$, with interval limits given by $a = a(k, p_0) = \max\{0, (k-1)p_0 - k/2 + 1\}$ and $b = b(k, p_0) = (p_0 + k/2 - 1)/(k-1)$.

Now, conditional on k, η_0 and η_{k-1}, we set $X_0 = \eta_0$ and we extend the prior distribution to the remaining parameters $\eta_1, \dots, \eta_{k-1}$ by focusing on the differences $X_j = \eta_j - \eta_{j-1}$, $j = 1, \dots, k-1$, in order to guarantee that condition (R1) is satisfied. For simplicity, analogous to what we did with p_1, we make such differences conditionally uniformly distributed on appropriate intervals, specified below, in order to satisfy also condition (R2), that is

$$\sum_{j=0}^{k-1} \eta_j = \sum_{j=0}^{k-1} (k-j) X_j = k/2. \tag{3.14}$$

Notice that we can rewrite (3.14) as

$$(k-j)X_j + \sum_{l=j+1}^{k-1} (k-l)X_l = k/2 - \sum_{l=0}^{j-1} (k-l)X_l,$$

for $j = 1, \dots, k-2$. Thus, if we assume that $X_l = 0$ for $l = j+1, \dots, k-2$, so that $\eta_l = \eta_j$ for $l = j+1, \dots, k-2$, we attain the upper bound,

$$X_j \leq \frac{1}{k-j-1} \left(k/2 + p_1 - 1 - \sum_{l=0}^{j-1} (k-l-1)X_l \right),$$

for $j = 1, \dots, k-2$. On the other hand, if we assume that

$$\sum_{l=j+1}^{k-2} (k-l-1)X_l = (k-j-2)\left(1 - p_1 - \sum_{l=0}^{j} X_l\right),$$

corresponding to $\eta_l = 1 - p_1$ for $l = j+1, \dots, k-2$, then we attain, through few algebraic manipulations, the lower bound

$$X_j \geq \max\left\{0, k/2 + (j-k+1)(1-p_1) - \sum_{l=0}^{j-1} (j-l+1)X_l \right\},$$

for $j = 1, \dots, k-2$. Rewriting these inequalities in terms of η_j, we find that the widest valid range for the coefficients can be expressed in terms of intervals $\mathscr{E}_j = \mathscr{E}_j(k, \eta_0, \dots, \eta_{j-1}, \eta_{k-1})$, given by

$$\mathscr{E}_j = \Bigg[\max\left\{ \eta_{j-1}, \, \frac{k}{2} + (k-j-1)(p_1 - 1) - \sum_{l=0}^{j-1} \eta_l \right\}; \\ \min\left\{ 1 - p_1; \, \frac{1}{k-j-1}\Big(\frac{k}{2} + p_1 - 1 - \sum_{l=0}^{j-1} \eta_l \Big) \right\} \Bigg],$$

for $j = 1, \dots, k-2$. Finally, we let $\eta_1 | (k, \eta_0, \eta_{k-1})$ and $\eta_j | (k, \eta_0, \dots, \eta_{j-1}, \eta_{k-1})$ for $j = 2, \dots, k-2$ be uniformly distributed on such intervals, therefore arriving at the following conditional prior distribution

$$\Pi(\eta_1, \dots, \eta_{k-2} | k, p_1, p_0) = \prod_{j=1}^{k-2} \Pi(\eta_j | k, \eta_0, \dots, \eta_{j-1}, \eta_{k-1}) = \prod_{j=1}^{k-2} \mathrm{Unif}(\mathscr{E}_j). \tag{3.15}$$

A direct consequence of Proposition 3.2 is that a valid prior distribution is induced also on the space $\mathcal{A}$ of valid Pickands dependence functions, as expressed by the following result.

Corollary 3.4. *Let $\mathscr{B} = \mathscr{B}(k) \subset \mathcal{S}^{k+1}$ be the space of $(k+1)$-dimensional vectors satisfying restrictions (R3)–(R5). Then, for any fixed $k \geq 3$ the prior distribution (3.15) induces a prior distribution on the coefficients of A_k in (3.5). Precisely, $\beta_j | (\beta_0, \dots, \beta_{j-1}, \beta_{k-1}, \beta_k)$, for $j = 2, \dots, k-2$, turns out to be uniformly distributed on the intervals*

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

The prior distribution on $\boldsymbol{\beta}_k = (\beta_0, \dots, \beta_k)$ is then given by

$$\begin{aligned} \Pi(\boldsymbol{\beta}_k | p_1, p_0, k) &= \mathbb{1}_{\{1\}}(\beta_0)\, \mathbb{1}_{\{(k-1+2\,p_0)/k\}}(\beta_1) \prod_{j=2}^{k-2} \Pi(\beta_j | \beta_0, \dots, \beta_{j-1}, \beta_{k-1}) \\ &\times \mathbb{1}_{\{(k-1+2\,p_1)/k\}}(\beta_{k-1})\, \mathbb{1}_{\{1\}}(\beta_k) \\ &= \mathbb{1}_{\{1\}}(\beta_0)\, \mathbb{1}_{\{(k-1+2\,p_0)/k\}}(\beta_1)\, \mathbb{1}_{\{(k-1+2\,p_1)/k\}}(\beta_{k-1})\, \mathbb{1}_{\{1\}}(\beta_k) \\ &\times \prod_{j=2}^{k-2} \mathit{Unif}(\mathscr{B}_j) \, \left(\frac{k}{2}\right)^{k-3}. \end{aligned}$$

This result follows directly from the change of variable formula. In fact, letting $\beta(\eta;\, k)$ given by expression (3.9) denote the inverse transformation of $\eta(\beta;\, k)$, given by expression (3.8), the corresponding Jacobian is $(k/2)^{k-3}$. Notice that, in this representation, the point masses of H are given by $p_0 = 1/2 - k(1-\beta_1)/2$ and $p_1 = 1/2 - k(1-\beta_{k-1})/2$.

The prior thus constructed assigns positive probability to any subset of $(\{k\}\times \mathcal{S}^k)$, $k>1$ which is valid, in the sense of satisfying conditions (R1)–(R2) or, equivalently, to every subset of $(\{k\}\times\mathcal{S}^{k+1})$, $k>1$ which is valid, in the sense of satisfying conditions (R3)–(R5). It therefore follows from proposition 3.3 that the prior has a full support, in terms of the L_∞ norm, on the spaces $\mathcal{H}$ and $\mathcal{A}$.

We now derive the analytical expression of the likelihood function. To do so, we consider for simplicity the distribution (2.3) with stable tail dependence function represented by (2.7). Then, 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 $J(y_1,y_2) = (y_1y_2)^{-2}$. This is equal to

$$g(y_1,y_2) = G(y_1,y_2)\left[\frac{\{A(t)-t\,A'(t)\}\,\{A(t)+(1-t)\,A'(t)\}}{(y_1y_2)^2} + \frac{A''(t)}{(y_1+y_2)^3}\right].$$

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 p.d.f. $g(y_1,y_2)$. Assume that the Pickands dependence function is represented by (3.5), for some fixed k. Then, the log-likelihood function is equal to

$$\begin{aligned}
\ell(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = &-\sum_{i=1}^n \left(\frac{1}{y_{1,i}} + \frac{1}{y_{2,i}}\right)\sum_{j=0}^k \beta_j\; b_j(t_i;k)\\
&+\sum_{i=1}^n \log\Bigg\{\Big(\sum_{j=0}^k \beta_j\; b_j(t_i;k) - t_i\,k\sum_{j=0}^{k-1}(\beta_{j+1}-\beta_j)\; b_j(t_i;k-1)\Big)\\
&\times\frac{\sum_{j=0}^k \beta_j\; b_j(t_i;k) + (1-t_i)k\,\sum_{j=0}^{k-1}(\beta_{j+1}-\beta_j)\; b_j(t_i;k-1)}{(y_{1,i}y_{2,i})^2}\\
&+\frac{k\,(k-1)\,\sum_{j=0}^{k-2}(\beta_{j+2}-2\,\beta_{j+1}+\beta_j)\; b_j(t_i;k-2)}{(y_{1,i}+y_{2,i})^3}\Bigg\},
\end{aligned} \tag{3.16}$$

where $\boldsymbol{\theta} = (k,\beta_0,\ldots,\beta_k)\in\boldsymbol{\Theta}\subseteq(\mathbb{N}\times\mathcal{S}^{k+1})$, abusing notation. We denote by $\mathcal{L}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})$ the associated likelihood function. We may once again apply Proposition 3.2, to obtain the log-likelihood function in terms of $\boldsymbol{\theta} = (k,\eta_0,\ldots,\eta_{k-1})\in$

$\mathbf{\Theta} \subseteq (\mathbb{N} \times \mathcal{S}^k)$ which, abusing terminology, can be seen as a reparametrization. More formally, this corresponds to the representation of the distribution (2.5), in the stable-tail dependence function (2.4), by means of a polynomial angular distribution given by the expression (3.3).

There is no closed form for the posterior distribution $\Pi^n(\boldsymbol{\theta}|\boldsymbol{y}_{1:n})$, which is proportional to $\Pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta})$, 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 $\Pi(k, \boldsymbol{\eta}_k)$ to

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

where $\boldsymbol{\eta}_\infty = (\eta_0, \eta_1, \ldots)$ denotes an infinite sequence of which, given k, only the first k elements are relevant, and $\Pi(\eta_j)$ is any fully known distribution. In order to update the pair $(k^{(s)}, \boldsymbol{\eta}_\infty^{(s)})$ at 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 conditional prior distribution $\Pi(\boldsymbol{\eta}_k|k)$, 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 state s of the Markov chain and the proposal, indexed by $s+1$, the acceptance probability depends on the ratio

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

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

$$p\left(k^{(s+1)}, \boldsymbol{\eta}^{(s+1)}_{k^{(s+1)}}, k^{(s)}, \boldsymbol{\eta}^{(s)}_{k^{(s)}}\right) = \frac{\Pi(k^{(s+1)})}{\Pi(k^{(s)})}\, \frac{\mathcal{L}(\boldsymbol{y}_{1:n}|k^{(s+1)}, \boldsymbol{\eta}^{(s+1)}_{k^{(s+1)}})}{\mathcal{L}(\boldsymbol{y}_{1:n}|k^{(s)}, \boldsymbol{\eta}^{(s)}_{k^{(s)}})}.$$

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 3.5. *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* $\left(k^{(s)}, \boldsymbol{\eta}^{(s)}_{k^{(s)}} \in \mathscr{E}_{k^{(s)}}\right)$;
2. *Repeat* M *times the update of the parameters according to:*
   (a) *Draw the proposals:*
   $$k^{(s+1)} \sim q_k(k|k^{(s)}) \text{ and } \boldsymbol{\eta}^{(s+1)}_{k^{(s+1)}} \sim q_{\boldsymbol{\eta}}(\boldsymbol{\eta}_k|k^{(s+1)});$$
   (b) *Compute the acceptance probability:*
   $$p = \min\left(p\left(k^{(s+1)}, \boldsymbol{\eta}^{(s+1)}_{k^{(s+1)}}, k^{(s)}, \boldsymbol{\eta}^{(s)}_{k^{(s)}}\right),\ 1\right);$$
   (c) *Draw* $U \sim \mathit{Unif}(0,1)$ *and, if* $U > p$*, then set:*
   $$\left(k^{(s+1)},\ \boldsymbol{\eta}^{(s+1)}_{k^{(s+1)}}\right) = \left(k^{(s)},\ \boldsymbol{\eta}^{(s)}_{k^{(s)}}\right);$$
   (d) *Set* $s = s + 1$;

Thus, after an appropriate burn-in period of, say m iterations, the sequence $(k^{(s)}\ \boldsymbol{\eta}^{(s)}_{k^{(s)}})_{s=m+1}^{M}$ 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. A simple way to do so is to use formula (2.9). This task can be fully performed, within the Bayesian paradigm, through a Monte Carlo estimate of the posterior predictive distribution, i.e.

$$\mathbb{P}(Y_1 > y_1^*, Y_2 > y_2^*|\boldsymbol{y}_{1:n}) \approx \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{P}(Y_1 > y_1^*, Y_2 > y_2^*|\boldsymbol{\theta})\, \Pi^n(\boldsymbol{\theta}|\boldsymbol{y}_{1:n})\, \mathrm{d}\boldsymbol{\theta}, \quad (3.17)$$

where $y_1^*, y_2^* > 0$ are unobserved thresholds. For each element of the posterior sample, applying expressions (2.9) and (3.3), we have that

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

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 through a simulation study in which the extremal dependence of some well-known parametric models is inferred. In particular, we consider the symmetric logistic (SL) model (Coles, 2001, p. 146), the asymmetric logistic (AL) model (Tawn, 1990), the Hüsler-Reiss (HR) model (Hüsler and Reiss, 1989) and the Extremal-t (ET) model (Nikoloulopoulos, Joe, and Li, 2009).

For each model, a sample of $n = 100$ bivariate observations with common unit Fréchet marginal distributions is simulated. Using such datasets, MCMC posterior samples of the angular measure and the Pickands dependence function are simulated via Algorithm 3.5 and compared with the theoretical functions (see Figures 1 and 2). After a burn-in period of $m = 400$ thousand iterations, 100 thousand samples are considered. Figure 1 displays the results obtained using the AL model with a mild dependence structure. In particular, the dependence parameter is $\alpha = 0.6$ and the asymmetry parameters are $(\tau_1, \tau_2) = (0.3, 0.8)$. The four columns report the results attained using different prior distributions for the polynomial degree k, when modeling the distribution function H_{k-1}. Precisely, from left to right, a Poisson distribution with mean $\kappa_P = 7$ and negative binomials with parameters $(\kappa_{NB} = 0.57, \sigma^2 = 0.73)$, $(\kappa_{NB} = 12.40, \sigma^2 = 23.66)$, $(\kappa_{NB} = 3.2, \sigma^2 = 4.48)$ have been considered. The third row shows the prior and posterior distributions for k, in green and red, respectively. The posterior median values are equal to 9, 3, 13 and 5, respectively for the four cases, from left to right. In the fourth row the prior (green line) and posterior (red line) distributions for the atom p_0, in addition to its true value $p_0 = (1 - \tau_2)/2 = 0.35$ (black dashed line) are reported. For all the cases, we see that most of the mass of the posterior distribution is concentrated close to the true value. The corresponding median values of the posterior distributions for p_0 are 0.351, 0.342, 0.358 and 0.349, from left to right. For the atom $p_1 = (1-\tau_1)/2 = 0.10$ we obtain the median values 0.149, 0.155, 0.165 and 0.139. Then, we can conclude that the information about the point masses at the edges of the unit interval is well reproduced. The first and second rows report the point-wise mean (red line) and the point-wise 95% credibility bands (in grey) computed through the posterior samples of the angular density and the Pickands dependence function, respectively. The credibility bands are the point-wise 0.05- and 0.95-quantiles of the posterior samples. The solid black lines are the true functions. In the first row, the true point masses on the edges are represented by black dots and the means computed from the posterior distributions are represented by red dots. The grey points are 95% upper and lower limits of the credibility intervals for the point masses. The true functions (angular density and Pickands) and the point masses

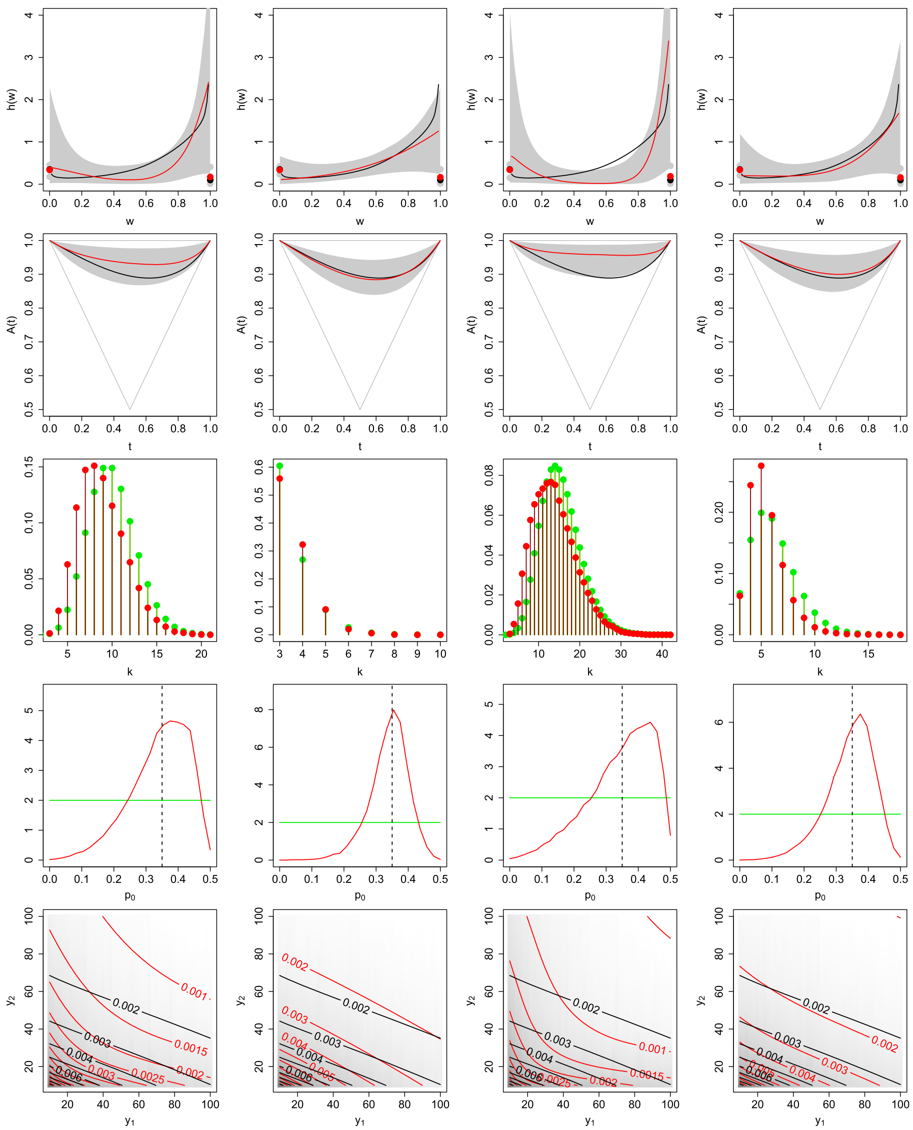

FIG 1. *Summary of the Bayesian nonparametric fitting of the extremal dependence. The true model is the Asymmetric Logistic model. Different prior distributions for the polynomial's degree k are considered, from left to right.*

fall within the point-wise 95% credibility bands in most of the cases, pointing out that our inferential method captures the dependence structure quite well. In the four cases, the results are quite similar; only in the third column (from the left), the 95% credibility bands do not include the true functions in a few points. So, it seems that our method is not too sensitive to the prior distribution for k. The fifth row reports the Monte Carlo predictive probabilities (red lines)

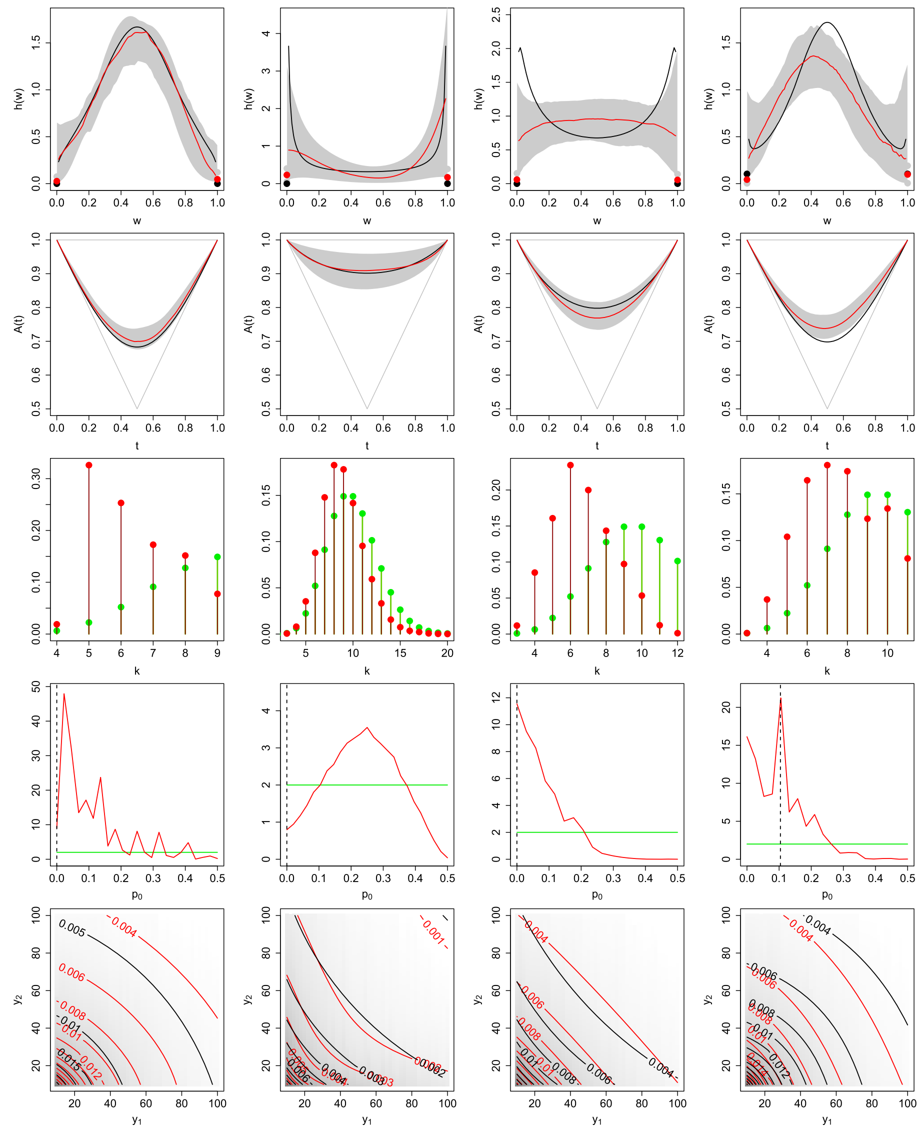

FIG 2. *Summary of the Bayesian nonparametric fitting for the extremal dependence models: Symmetric Logistic (mild and weak), Hüsler-Reiss and Extremal-t.*

of future simultaneous exceedances (3.17) for pairs of unobserved thresholds (y_1^*, y_2^*) ranging between 10 and 100. The black lines are the true probabilities. In the second and fourth cases (from the left) the estimates are very accurate, while they are less so for the first and the third.

Figure 2 reports (from left to right) the results obtained for data generated from the SL model with mild and weak dependence structures (denoted by SLm and SLw) given by dependence parameter values $\alpha = 0.45$ and 0.85; the HL and the ET models with mild dependence given by the dependence parameters $\lambda = 1.2$ and ($\omega = 0.8$, $\nu = 2$), respectively. The format of the graphs is the same as that of the previous figure. The results are obtained using the same prior distribution for the polynomial degree k, i.e. a Poisson distribution with mean $\kappa_P = 7$. Other prior settings can be considered (skipped here for brevity) and, as the previous study shows, the results do not change significantly. In the four cases, the posterior median values for k are 6, 9, 7 and 8. The only model that includes point masses on the edges is the ET, corresponding to $p_0 = p_1 = T_{\nu+1}(-\omega\{(\nu + 1)/(1 - \omega^2)\}^{1/2}) = 0.104$, where $T_{\nu+1}(\cdot)$ denotes a t distribution function with $\nu + 1$ degrees of freedom. The medians of the posterior distributions for the point masses are (0.018, 0.041), (0.235, 0.151), (0.047, 0.045) and (0.041, 0.097), respectively for the four cases. We see, from the first and second row-panels, that the posterior distributions adequately capture the different extremal dependence forms. Also, the plots in the last row show rather accurate predictions of the probabilities of joint exceedances, outlining the good performance of our inferential method.

Finally, going beyond visual checks, we measure the accuracy of our proposed method. To do so, we focus on the Pickands dependence function and we compute, for each element of the posterior MCMC sample, the integrated squared error:

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

where A is the true Pickands dependence function and $A^{(s)}$, $s = 1, \ldots, m$, is a Pickands dependence function sampled from the posterior. Table 1 reports, for different sample sizes and for each of the four models considered in Figures 1 and 2 (first column), the Monte Carlo posterior mean of the ISE (second column). Between parenthesis the 0.05- and 0.95- quantiles of the posterior distribution for the ISE are reported. For comparison purposes, the third and fourth columns report similar estimates obtained using the projection method discussed in Marcon et al. (2015) and focusing on the multivariate madogram (MD) and Capéraà-Fougères-Genest (CFG, Capéraà et al., 1997) estimators as pilot estimates, see Marcon et al. (2015) for details. In particular, for each dataset, 500 bootstrap replicates are produced and for each of these the ISE is computed, where in this case $A^{(s)}$ is the estimated Pickands dependence function obtained with the projection method. In the table, the mean and the 0.05- and 0.95- quantiles (in parenthesis) of the ISE computed over the 500 bootstrap replicates, are reported. Results in Table 1 show the slightly better performance of our proposed method with respect to the competitors, for the several examples considered. This supports our new proposal.

We have also compared our inferential approach with other proposals, for instance that in Einmahl et al. (2008) (see also, Klüppelberg et al., 2007, Krajina, 2012). They proposed a parametric method for estimating the tail of a bivariate distribution that is in the domain of attraction of a bivariate extreme value distribution (de Haan and Ferreira, 2006, Ch. 6). Instead, we directly model the extremal dependence of bivariate extreme value distributions. Thus, care must

TABLE 1

Mean, 95% credibility intervals (Bayesian method) and 95% bootstrap confidence intervals (projection method) of the ISE for the models in Figures 1 and 2, for increasing sample sizes.

Model	Inferential methods		
	Sample size 25		
	Bayesian	Projection-MD	Projection-CFG
AL	2.35×10^{-3}	5.10×10^{-3}	1.13×10^{-2}
	$(3.53 \times 10^{-4}; 5.65 \times 10^{-3})$	$(1.81 \times 10^{-4}; 1.90 \times 10^{-2})$	$(8.02 \times 10^{-4}; 2.45 \times 10^{-2})$
SLm	7.64×10^{-3}	6.63×10^{-3}	1.47×10^{-3}
	$(8.33 \times 10^{-4}; 2.09 \times 10^{-2})$	$(8.57 \times 10^{-5}; 3.01 \times 10^{-2})$	$(6.15 \times 10^{-5}; 6.98 \times 10^{-3})$
SLw	1.75×10^{-3}	3.81×10^{-3}	4.36×10^{-3}
	$(1.23 \times 10^{-4}; 4.21 \times 10^{-3})$	$(3.26 \times 10^{-4}; 6.10 \times 10^{-3})$	$(3.95 \times 10^{-4}; 1.31 \times 10^{-2})$
HR	8.75×10^{-3}	4.58×10^{-3}	6.75×10^{-3}
	$(4.95 \times 10^{-4}; 1.75 \times 10^{-2})$	$(3.10 \times 10^{-4}; 9.94 \times 10^{-3})$	$(1.51 \times 10^{-3}; 9.92 \times 10^{-3})$
ET	3.43×10^{-2}	7.00×10^{-2}	6.55×10^{-2}
	$(2.35 \times 10^{-2}; 5.18 \times 10^{-2})$	$(6.17 \times 10^{-2}; 8.63 \times 10^{-2})$	$(6.18 \times 10^{-2}; 7.31 \times 10^{-2})$
	Sample size 50		
	Bayesian	Projection-MD	Projection-CFG
AL	1.23×10^{-3}	2.04×10^{-3}	1.96×10^{-3}
	$(4.73 \times 10^{-5}; 4.09 \times 10^{-3})$	$(1.10 \times 10^{-4}; 6.48 \times 10^{-3})$	$(8.67 \times 10^{-5}; 6.67 \times 10^{-3})$
SLm	1.76×10^{-3}	6.52×10^{-4}	4.17×10^{-4}
	$(1.16 \times 10^{-4}; 5.04 \times 10^{-3})$	$(2.53 \times 10^{-5}; 2.35 \times 10^{-3})$	$(1.87 \times 10^{-5}; 1.18 \times 10^{-3})$
SLw	1.47×10^{-3}	2.14×10^{-3}	2.33×10^{-3}
	$(9.18 \times 10^{-5}; 3.89 \times 10^{-3})$	$(3.74 \times 10^{-4}; 5.59 \times 10^{-3})$	$(2.39 \times 10^{-4}; 7.08 \times 10^{-3})$
HR	8.87×10^{-4}	2.71×10^{-3}	4.38×10^{-3}
	$(4.53 \times 10^{-5}; 3.26 \times 10^{-3})$	$(2.47 \times 10^{-4}; 6.82 \times 10^{-3})$	$(9.63 \times 10^{-4}; 8.29 \times 10^{-3})$
ET	3.20×10^{-2}	7.46×10^{-2}	7.08×10^{-2}
	$(2.42 \times 10^{-2}; 4.51 \times 10^{-2})$	$(6.68 \times 10^{-2}; 8.52 \times 10^{-2})$	$(6.53 \times 10^{-2}; 7.69 \times 10^{-2})$
	Sample size 100		
	Bayesian	Projection-MD	Projection-CFG
AL	5.71×10^{-4}	9.48×10^{-4}	6.51×10^{-4}
	$(1.60 \times 10^{-5}; 2.02 \times 10^{-3})$	$(7.10 \times 10^{-5}; 6.47 \times 10^{-3})$	$(2.98 \times 10^{-5}; 2.30 \times 10^{-3})$
SLm	3.58×10^{-4}	1.85×10^{-4}	1.91×10^{-4}
	$(7.67 \times 10^{-6}; 1.13 \times 10^{-3})$	$(1.53 \times 10^{-5}; 2.85 \times 10^{-4})$	$(2.10 \times 10^{-5}; 2.84 \times 10^{-4})$
SLw	8.44×10^{-4}	1.21×10^{-3}	1.17×10^{-3}
	$(4.77 \times 10^{-5}; 2.74 \times 10^{-3})$	$(9.67 \times 10^{-5}; 3.88 \times 10^{-3})$	$(1.23 \times 10^{-4}; 4.02 \times 10^{-3})$
HR	5.61×10^{-4}	2.16×10^{-3}	2.37×10^{-3}
	$(3.89 \times 10^{-5}; 1.67 \times 10^{-3})$	$(2.32 \times 10^{-4}; 4.71 \times 10^{-3})$	$(5.38 \times 10^{-4}; 4.24 \times 10^{-3})$
ET	2.49×10^{-2}	6.66×10^{-2}	6.68×10^{-2}
	$(2.14 \times 10^{-2}; 2.99 \times 10^{-2})$	$(6.49 \times 10^{-2}; 7.23 \times 10^{-2})$	$(6.49 \times 10^{-2}; 7.09 \times 10^{-2})$
	Sample size 200		
	Bayesian	Projection-MD	Projection-CFG
AL	3.76×10^{-4}	6.09×10^{-4}	4.95×10^{-4}
	$(1.87 \times 10^{-5}; 1.22 \times 10^{-3})$	$(4.50 \times 10^{-5}; 1.92 \times 10^{-3})$	$(3.31 \times 10^{-5}; 1.63 \times 10^{-3})$
SLm	5.62×10^{-5}	4.52×10^{-4}	4.84×10^{-4}
	$(6.45 \times 10^{-6}; 1.50 \times 10^{-4})$	$(4.69 \times 10^{-5}; 1.03 \times 10^{-3})$	$(1.01 \times 10^{-4}; 9.65 \times 10^{-4})$
SLw	5.16×10^{-4}	8.10×10^{-4}	1.19×10^{-3}
	$(2.87 \times 10^{-5}; 1.72 \times 10^{-3})$	$(4.59 \times 10^{-5}; 2.54 \times 10^{-3})$	$(6.41 \times 10^{-5}; 3.39 \times 10^{-3})$
HR	2.53×10^{-4}	3.91×10^{-4}	3.62×10^{-4}
	$(1.73 \times 10^{-5}; 8.55 \times 10^{-4})$	$(2.22 \times 10^{-5}; 1.20 \times 10^{-3})$	$(2.59 \times 10^{-5}; 1.09 \times 10^{-3})$
ET	2.28×10^{-2}	6.25×10^{-2}	6.16×10^{-2}
	$(2.09 \times 10^{-2}; 2.56 \times 10^{-2})$	$(6.10 \times 10^{-2}; 6.93 \times 10^{-2})$	$(6.15 \times 10^{-2}; 6.93 \times 10^{-2})$

be taken when interpreting the results, which for brevity are not presented here. For specific parametric families of dependence models, their parametric method outperforms our nonparametric proposal. However, the integrated squared error (used for comparison) is of the same order in both techniques, suggesting that our model-free proposal is equally appealing, in addition to providing wide applicability.

In conclusion, we stress that the computational cost of running our proposed Bayesian model is moderately low. For example, to run $M = 500$ thousand iterations of the MCMC algorithm, it takes only 114.03 seconds, with an intel Core i7 processor at 2.2 GHz. The code for the model fitting will soon be available with the `R`-package `ExtremalDep`. The data simulation was performed using the `R`-package `evd` (Stephenson, 2004).

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 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 $\mu_2 = 0.0068$,

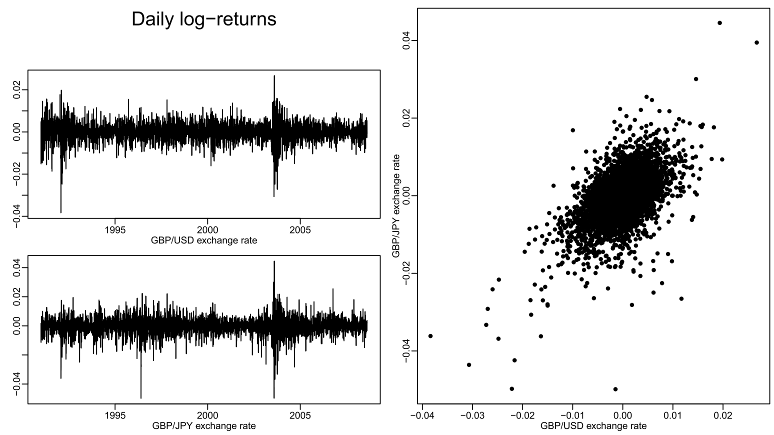

FIG 3. *Daily log-returns of GBP/USD and GBP/JPY exchange rates.*

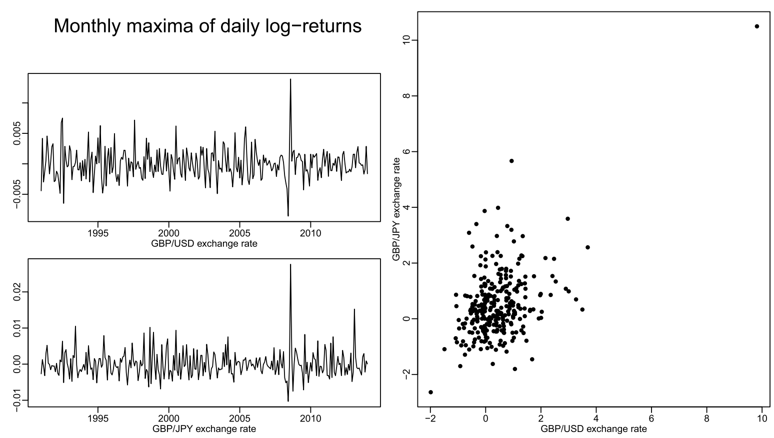

FIG 4. *Monthly-maxima of log-returns of GBP/USD and GBP/JPY exchange rates.*

$\sigma_2 = 0.0030$, $\xi_2 = 0.1199$, 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. Secondly, we transform the data to obtain unit Fréchet margins, by means of transformation (2.2) and using the estimated marginal parameters. The data transformed in this way can be assumed to be a sample coming approximately from a bivariate max-stable distribution of the type (2.3). The extremal dependence of monthly-maxima of log-returns is then inferred by using the method described in Section 3.

The set-up for computing the approximate posterior distributions is the same as that considered for the models illustrated in Figure 2 of Section 4. The sum-

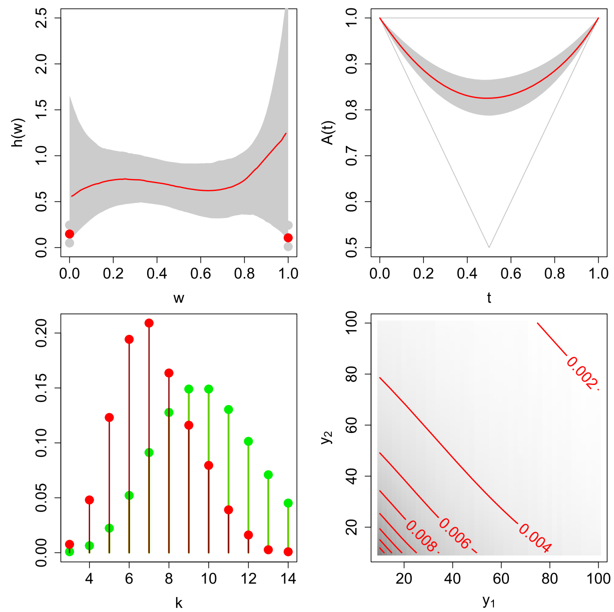

FIG 5. *Summary of the Bayesian nonparametric fitting of the extremal dependence for the monthly-maxima of GBP/USD and GBP/JPY log-returns of exchange rates.*

maries of results obtained from the posterior distribution are displayed in Figure 5. The first row reports the point-wise posterior means (red lines) and 95% credibility bands (in grey) of the angular density (left panel) and the Pickands dependence function (right panel). The results regarding the Pickands dependence function suggest that the dependence structure is symmetric. Results about both the Pickands dependence function and the angular density suggest a mild dependence, with posterior median values of the point masses p_0 and p_1 equal to 0.149 and 0.093, respectively. The bottom-left panel of Figure 5 displays the prior (green dots) and posterior distributions (red dots) for the polynomial degree k, with posterior median value 7. The bottom-right panel displays the predicted probabilities of joint exceedances, given by (3.17), for combinations of values ranging between 10 and 100. These results highlight that the probability of joint exceedances is also symmetric and hence the two variables can be considered exchangeable. However, as we have previously discussed, the marginal distribution of monthly-maxima of GBP/USD log-returns is different from that of GBP/USD. Therefore, 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 the 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, for q_1 we obtain the thresholds $y_{2,1}^{*} = 14.12$ and $y_{1,1}^{*} = 57.25$ and the joint predictive probability (3.17) is equal to 0.0050. Therefore, we obtain the final result $\mathbb{P}(\text{GBP/JPY} > q_1 \,|\, \text{GBP/USD} > q_1,) \approx \mathbb{P}(Y_2 > y_{2,1}^{*} \,|\, Y_1 > y_{1,1}^{*}) = 0.2880$. Similarly, for q_2 we obtain the thresholds $y_{1,2}^{*} = 450.23$ and $y_{2,2}^{*} = 52.32$ and the joint predictive probability (3.17) is equal to 0.0007. Therefore, we obtain the final result $\mathbb{P}(\text{GBP/USD} > q_2 \,|\, \text{GBP/JPY} > q_2) \approx \mathbb{P}(Y_1 > y_{1,2}^{*} \,|\, Y_2 > y_{2,2}^{*}) = 0.0386$.

In conclusion, in contrast to the case of the joint exceedances, since the GBP/JPY tends to assume larger values than GBP/USD, the conditional probability of the log-returns of GBP/USD given elevated values of log-returns of GBP/JPY is quite high.

Appendix: Proofs

Proof of Proposition 3.1. Using the identities in (3.2) we have that $H_{k-1}(0) = p_0$, $H_{k-1}(1) = 1 - p_1$, where the former is the mass at $\{0\}$ and $H_{k-1}(1) = \sup_{w \in [0,1)} H_{k-1}(w)$. As a result $H_{k-1}([0,1]) - H_{k-1}(1) = p_1$, which is the mass at $\{1\}$.

Second, for any $w_1 \leq w_2 \in [0,1)$ we have

$$H_{k-1}([0,w_2]) - H_{k-1}([0,w_1]) = \sum_{j=0}^{k-1} (\eta_{j+1} - \eta_j) \int_{w_1}^{w_2} \mathrm{Be}(v|j+1, k-j)\mathrm{d}v \geq 0,$$

where the inequality holds because by (R1) we have that $\eta_{j+1} - \eta_j \geq 0$, for $j = 0, \ldots, k-1$ and therefore $H_{k-1}([0,w_1]) \leq H_{k-1}([0,w_2])$.

Third, note that

$$\int_0^1 w\, h_{k-1}(w)\, \mathrm{d}w + p_1 = p_1 + \frac{1}{k} \sum_{j=0}^{k-2} (\eta_{j+1} - \eta_j)(j+1), \tag{A.1}$$

and

$$p_0 + \int_0^1 (1-w)\, h_{k-1}(w)\, \mathrm{d}w = p_0 + \frac{1}{k} \sum_{j=0}^{k-2} (\eta_{j+1} - \eta_j)(k-j-1). \tag{A.2}$$

Equating (A.1) and (A.2) to 1/2 we attain the condition in (R2). Then H_{k-1} satisfies the mean constraint (C1) by applying (R2) to its coefficients. $\square$

Proof of Proposition 3.2. By (2.8) we have that $H([0,w]) = (A'(w)+1)/2$ for $w \in [0,1)$. Applying such a relationship between H_{k-1} and A_k we attain

$$\begin{aligned} H_{k-1}([0,w]) &= \frac{1}{2}\left\{k\sum_{j=0}^{k-1}(\beta_{j+1}-\beta_j)b_j(w;k-1)+1\right\} \\ &= \sum_{j=0}^{k-1}\frac{1}{2}\left\{k(\beta_{j+1}-\beta_j)+1\right\}b_j(w;k-1) \\ &= \sum_{j=0}^{k-1}\eta_j b_j(w;k-1), \end{aligned}$$

where we have used the identity $\sum_{j\leq k-1} b_j(w;k-1) = 1$. From the above formula the result in (3.8) follows. On the other hand we have

$$\begin{aligned} A'_k(t) &= 2H_{k-1}([0,t])-1 \\ k\sum_{j=0}^{k-1}(\beta_{j+1}-\beta_j)b_j(t;k-1) &= \sum_{j=0}^{k-1}(2\eta_j-1)b_j(t;k-1) \end{aligned}$$

where the last identity holds if and only if $k(\beta_{j+1}-\beta_j) = 2\eta_j - 1$ for all $j = 0,\dots k-1$. Resolving for β_{j+1} we attain the formula $\beta_{j+1} = \beta_j + (2\eta_j - 1)/k$. From this we get $\beta_1 = (2\eta_0 + k - 1)/k$, for $j = 0$, since $\beta_0 = 1$. Applying it recursively we get $\beta_2 = (2(\eta_0+\eta_1)+k-2)/k$, for $j = 1$. Repeating this reasoning for $j = 2, 3, \dots$ we attain the general recursive formula in (3.9). Thus, statement $i)$ is shown.

Consider A_k in (3.5) and assume it fulfills (R3)–(R5). Then, we must check that H_{k-1} in (3.3) with coefficients given by (3.8) fulfill (R1) and (R2).

By (3.8) we have $\eta_0 = k(\beta_1 - 1 + 1/k)/2$ and $\eta_{k-1} = k(1-\beta_{k-1}+1/k)/2$ for $j = 0$ and $j = k-1$. By (R4) we have therefore that $\eta_0 = p_0$ and $\eta_{k-1} = 1 - p_1$. Next, it needs to be shown that $\eta_j \leq \eta_{j+1}$ for $j = 0,\dots,k-2$. By (3.8) this inequality is equal to $k(\beta_{j+1} - \beta_j + 1/k)/2 \leq k(\beta_{j+2} - \beta_{j+1} + 1/k)/2$ for $j = 0,\dots,k-2$. This holds if and only if $\beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0$ and this is true by (R5). Thus H_{k-1} fulfills (R1).

It remains to show that $\eta_0 + \cdots + \eta_{k-1} = k/2$. By (3.8) with a few steps we attain

$$\begin{aligned} \frac{k}{2} + p_1 - 1 - p_0 &= \sum_{j=1}^{k-2}\eta_j \\ &= \sum_{j=1}^{k-2}\left(\frac{k(\beta_{j+1}-\beta_j+1/k)}{2}\right) \\ &= \frac{k}{2} - 1 + \frac{k}{2}\sum_{j=1}^{k-2}(\beta_{j+1}-\beta_j) \end{aligned}$$

and from the last identity we obtain $2(p_1 - p_0) = \beta_{k-1} - \beta_1$. By (R4) it is straightforward to check that the last equation holds. Therefore H_{k-1} fulfills also (R2) and it is the distribution of a valid angular measure.

Now, consider H_{k-1} in (3.3) and assume it fulfills (R1) and (R2). Then, we must check that A_k in (3.5) with coefficients given by (3.9) fulfills (R3)–(R5).

Applying (3.9) with $j = k - 1$ we have that $\beta_k = 1$ and this is attained using the condition (R2). Next, it needs to be shown that $\beta_{j+1} \leq 1$ for any $j = 0, \dots, k-1$. Applying (3.9) to check that such inequalities hold is equivalent to checking that $\sum_{i \leq j} \eta_i \leq (j+1)/2$ for any $j = 0, \dots, k-1$. Thus, when $j = 0$ we have $\eta_0 \leq 1/2$ and this holds since that $\eta_0 = p_0$ by (R1) and $p_0 \in [0, 1/2]$ by Assumption 2.1. For any $j = 1, \dots, k-2$ suppose, on the contrary, that $(\eta_0 + \cdots + \eta_j) > (j+1)/k$. From this and taking into account (R1) and that $p_1 \in [0, 1/2]$ by Assumption 2.1, it follows the contradiction that (R2) is not valid. As a consequence, the opposite inequalities hold. Since $\beta_0 = 1$ by definition, then A_k fulfills (R3).

By (3.9), for $j = 0$ and $j = k - 2$, we derive with some manipulations $\beta_1 = (2p_0 + k - 1)/k$ and $\beta_{k-1} = (2p_1 + k - 1)/k$. These results are attained by using (R1) and (R2), respectively. Therefore A_k fulfills (R4).

It remains to show that $\beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0$ for all $j = 0, \dots k-2$. Applying (3.9) and with some manipulations we have

$$
\begin{aligned}
0 &\leq \beta_{j+2} - 2\beta_{j+1} + \beta_j \\
&\leq \frac{1}{k}\left(2\sum_{i=0}^{j+1}\eta_i + k - j - 2\right) - \frac{2}{k}\left(2\sum_{i=0}^{j}\eta_i + k - j - 1\right) + \frac{1}{k}\left(2\sum_{i=0}^{j-1}\eta_i + k - j\right) \\
&\leq \frac{2}{k}(\eta_{j+1} - \eta_j)
\end{aligned}
$$

for $j = 0, \dots k-2$. This result holds since $\eta_j \leq \eta_{j+1}$ for $j = 0, \dots k-2$ by (R1). Therefore A_k fulfills also (R5) and is a valid Pickands dependence function. Then the proof is concluded. □

Proof of Proposition 3.3. The fact that $\mathcal{A}_k$, $k = 1, 2, \dots$ is nested in $\mathcal{A}$ has been shown by Proposition 3.3 in Marcon et al. (2015). Here we only need to show that $A_{k+1}(t)$ satisfies the conditions in (R4), where

$$
A_{k+1}(t) = \sum_{j=0}^{k+1} \beta_j^* \, b_j(t; k+1), \quad \beta_j^* = \left(\beta_j \frac{k+1-j}{k+1} + \beta_{j-1}\frac{j}{k+1}\right).
$$

Applying the above formula we have $\beta_1^* = (k\beta_1 + \beta_0)/(k+1)$ and $\beta_k^* = (\beta_k + k\beta_{k-1})/(k+1)$. Substituting with $\beta_0 = \beta_k = 1$, $\beta_1 = (2p_0 + k - 1)/k$ and $\beta_{k-1} = (2p_1 + k - 1)/k$, we obtain $\beta_1^* = (2p_0 + k)/(k+1)$ and $\beta_k^* = (2p_1 + k)/(k+1)$. Therefore, the result is shown. We now show that also $\mathcal{H}_k$, $k = 1, 2, \dots$ is nested in $\mathcal{H}$. Let

$$
H_k(w) = \sum_{j=0}^{k} \eta_j^* \, b_j(w; k), \quad \eta_j^* = \eta_j \frac{k-j}{k} + \eta_{j-1}\frac{j}{k}.
$$

We can verify that $\eta_j^* \leq \eta_{j+1}^*$, for $j = 0, \ldots, k-1$. Using the definition of η_j^* we obtain

$$-\frac{j}{k}\left(\eta_j - \eta_{j-1}\right) \quad \leq \quad \frac{k-j-1}{k}\left(\eta_{j+1} - \eta_j\right)$$

and the left-hand and right-hand side of the above inequality are always negative and positive, respectively, by (R1). Therefore, also H_k satisfies condition (R1). Furthermore, we have

$$\begin{aligned} \frac{k+1}{2} &= \sum_{j=0}^{k} \eta_j^* \\ &= \sum_{j=0}^{k} \left(\eta_j \frac{k-j}{k} + \eta_{j-1}\frac{j}{k}\right) \\ &= \sum_{j=0}^{k-1} \eta_j + \frac{1}{2}, \end{aligned}$$

where the last equation holds by (R2). As a consequence also H_k satisfies conditions in (R2) and hence $\mathcal{H}_k$, $k = 1, 2, \ldots$, is nested in $\mathcal{H}$.

Now, let

$$B_A(w;k) = \sum_{j=0}^{k} A\left(\frac{j}{k}\right) b_j(w;k), \quad k = 1, 2, \ldots,$$

then, by Proposition 3.1 in Marcon et al. (2015) we have

$$\sup_{w\in[0,1]} |B_A(w;k) - A(w)| \leq \frac{1}{2\sqrt{k}}.$$

Therefore, by Proposition 3.3 in Marcon et al. (2015) the result in (3.10) follows. Next, consider H_{k-1} as in (3.3), where $H_{k-1}(w) = (A_k'(w) + 1)/2$ for $w \in [0, 1)$ and A_k as in (3.5), satisfying (R3)–(R5). Then, $H_{k-1} \in \mathcal{H}$ by Proposition 3.2 and $\mathcal{H}_k$ is nested in $\mathcal{H}$ as has been shown above. Furthermore, let $\tilde{B}_H(w; k-1) = (B_A'(w;k) + 1)/2$ for $w \in \mathcal{S}$, then

$$|\tilde{B}_H(w;k-1) - H(w)| = |B_A'(w;k) - A'(w)|, \quad w \in \mathcal{S}.$$

As a consequence the result (3.10) implies that also result (3.11) holds, by the uniform convergence of the first derivative of convex functions (see Theorem 25.7 in Rockafellar, 2015). $\square$

Acknowledgements

We thank a referee, an associate editor and the editor for useful suggestions and comments.

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. MR3173743

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd., Chichester. MR2108013

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. MR3089619

Boldi, M. O. and A. C. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B 69*(2), 217–229. MR2325273

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. MR2893858

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

Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer, London. MR1932132

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer. MR2234156

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

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. MR2456011

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics 37*(5B), 2990–3022. MR2541453

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics 29*(5), 1264–1280. MR1873330

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. MR1939699

Guillotte, S. and F. Perron (2008). A Bayesian estimator for the dependence function of a bivariate extreme-value distribution. *Canadian Journal of Statistics 36*(3), 383–396. MR2456012

Hüsler, J. and R. Reiss (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters 7*, 283–286. MR0980699

Klüppelberg, C., G. Kuhn, L. Peng, et al. (2007). Estimating the tail dependence function of an elliptical distribution. *Bernoulli 13*(1), 229–251. MR2307405

Krajina, A. (2012). A method of moments estimator of tail dependence in meta-elliptical models. *Journal of Statistical Planning and Inference 142*(7), 1811–1823. MR2903392

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. A. Padoan, P. Naveau, and P. Muliere (2015). Nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference, Under revision*.

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. MR2515644

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics 27*(1), 105–126. MR1703623

Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics 26*, 373–393. MR1712051

Petrone, S. and L. Wasserman (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(1), 79–100. MR1881846

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. MR0820979

Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press. MR1451876

Sabourin, A. and P. Naveau (2014). Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics & Data Analysis 71*, 542–567. MR3131989

Stephenson, A. (2004). A user's guide to the 'evd' package (version 2.1). *Department of Statistics. Macquarie University. Australia*.

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