



# Regression-type analysis for multivariate extreme values

Miguel de Carvalho<sup>1</sup> · Alina Kumukova<sup>2</sup> · Gonçalo dos Reis<sup>1,3</sup>

Received: 25 July 2021 / Revised: 6 May 2022 / Accepted: 7 August 2022 /

Published online: 21 October 2022

© The Author(s) 2022, corrected publication 2022

## Abstract

This paper devises a regression-type model for the situation where both the response and covariates are extreme. The proposed approach is designed for the setting where the response and covariates are modeled as multivariate extreme values, and thus contrarily to standard regression methods it takes into account the key fact that the limiting distribution of suitably standardized componentwise maxima is an extreme value copula. An important target in the proposed framework is the regression manifold, which consists of a family of regression lines obeying the latter asymptotic result. To learn about the proposed model from data, we employ a Bernstein polynomial prior on the space of angular densities which leads to an induced prior on the space of regression manifolds. Numerical studies suggest a good performance of the proposed methods, and a finance real-data illustration reveals interesting aspects on the conditional risk of extreme losses in two leading international stock markets.

**Keywords** Angular measure · Bernstein polynomials · Extreme value copula · Joint extremes · Multivariate extreme value distribution · Quantile regression · Statistics of extremes

## 1 Introduction

Whereas classical statistical modeling is mostly concerned with inferences surrounding the bulk of a distribution, the field of statistics of extremes deals with the rather challenging situation of conducting inferences about the tail of a

---

✉ Miguel de Carvalho  
Miguel.deCarvalho@ed.ac.uk

<sup>1</sup> School of Mathematics, University of Edinburgh, Edinburgh, UK

<sup>2</sup> Maxwell Institute for Mathematical Sciences, School of Mathematics, University of Edinburgh, Edinburgh, UK

<sup>3</sup> Centro de Matemática e Aplicações (CMA), FCT, UNL, Caparica, Portugal

distribution. The behavior of extreme values in large samples is often mathematically tractable, and this tractability is often used to build sound statistical methods for modeling risk and extreme values. As an example of this asymptotic tractability, it is well known that if  $Y_1, \dots, Y_n$  is a random sample with sample maximum  $M_n = \max(Y_1, \dots, Y_n)$  and if there exist sequences  $\{a_n > 0\}$  and  $\{b_n\}$  such that  $(M_n - b_n)/a_n \xrightarrow{d} Z$ , then  $Z$  follows a GEV (Generalized Extreme Value) distribution with location, scale, and shape parameters  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , and  $\xi \in \mathbb{R}$  respectively; see, for instance, Embrechts et al. (1997, Theorem 3.2.3). Details on the paradigm of statistics of extremes can be found in monographs (e.g. Coles 2001; Beirlant et al. 2004; de Haan and Ferreira 2006; Resnick 2007) as well as review papers (e.g. Davison and Huser 2015).

In this paper, we devise a regression-type method for the situation where both the response and the covariates are themselves extreme. Here and below, the expression “regression-type” is used to refer to the class of statistical models that relate the conditional quantiles of a response with covariates via a joint distribution—rather than by specifying a functional relation between response and covariate as, for example, in quantile regression (Koenker and Bassett 1978). An important result in the field of statistics of extremes—that will be fundamental for our developments—is that the properly standardized vector of block maxima converges in distribution to a so-called extreme value copula (Gudendorf and Segers 2010). Thus, a key target in the proposed framework is what we will refer to below as the regression manifold, that is, a family of regression lines that obeys the latter large sample result. Our methods thus take on board information on the dependence structure between the extreme values so to assess what effects an extreme value covariate can have on an extreme value response. To learn about the proposed model from data, we develop a prior in the space of regression manifolds by resorting to a flexible Bernstein polynomial prior on the space of angular densities as recently proposed by Hanson et al. (2017).

Our approach contributes to the literature on conditional modeling given large observed values (e.g. Wang and Stoev 2011; Cooley et al. 2012), nonetheless, our focus differs from the latter papers in a number of important ways as we describe next. The main difference is that, as anticipated above, here the focus is on devising a regression framework for extreme covariates and extreme responses, whereas the latter papers focus mainly on using the conditional density as a way to make probabilistic statements about the likelihood of an extreme given the occurrence of another extreme. Conditional quantile estimation within the framework of multivariate extreme value distributions was first looked into by Cooley et al. (2012). Here, among other things, we aim to characterize the entire regression manifold, its cross sections as well as how extremal dependence affects its shape; additionally, we aim to tackle the nontrivial task of defining a prior on the space of regression manifolds. Since our main target of analysis is regression, our method has some links with statistical approaches for nonstationary extremes (e.g. Coles 2001, Sect. 6; Yee and Stephenson 2007; Eastoe and Tawn 2009; Wang and Tsai 2009; Katz 2013); the most elementary version of approaches for nonstationary extremes aims to learn about how the limiting law of a suitably standardized block maxima response ( $S_{\mathbf{x}}$ ) changes according to a covariate  $\mathbf{x} = (x_1, \dots, x_p)^T$ , via the specification

$$(S \mid \mathbf{X} = \mathbf{x}) \sim \text{GEV}(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}, \xi_{\mathbf{x}}). \quad (1)$$

Since the approach in (1) is built from the univariate theory of extremes it is not tailored for conditioning on another variable being extreme as it fails to take on board information from the dependence structure between the extremes.

Additionally, the method proposed in this work is loosely related to quantile regression (Koenker and Bassett 1978), whose original version consists in modeling the conditional quantile of a response  $Y$  given a covariate  $\mathbf{X} = (X_1, \dots, X_p)^\top$  in a linear fashion, that is

$$F^{-1}(q \mid \mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}_q, \quad 0 < q < 1, \quad (2)$$

where  $F^{-1}(q \mid \mathbf{x}) = \inf\{y : F(y \mid \mathbf{x}) \geq q\}$  and  $F(y \mid \mathbf{x})$  is the distribution function of  $Y \mid \mathbf{X} = \mathbf{x}$ . Versions of quantile regression that aim to equip (2) with the ability to extrapolate into the tail of  $Y$  are often known as extremal quantile regression methods (e.g. Chernozhukov 2005). While flexible and sturdy, such quantile regression-based approaches do not take into account information on the fact that the limiting joint distribution of suitably standardized componentwise maxima is an extreme value copula, and thus fail to be equipped with the ability to extrapolate into the joint tail. The approach proposed in this paper will take such knowledge on the limiting joint distribution into consideration and will assume a conditional law that stems from such knowledge—rather than imposing a linear specification as in (2); yet, the proposed approach is not to be seen as a competitor to quantile regression but rather as a method based on some loosely related principles and specific to the context where we have a block maxima response and a block maxima covariate.

The remainder of the paper unfolds as follows. In Sect. 2 we introduce the proposed model and Sect. 3 devises an approach for learning about it from data. Section 4 reports the main findings of a Monte Carlo simulation study. We showcase the proposed methodology in a real data application to stock market data in Sect. 5. Finally, in Sect. 6 we present closing remarks. Supporting technical details can be found in the Appendix, and further numerical experiments and other details are presented in the supplementary material.

## 2 Modelling conditional multivariate extreme value distributions

### 2.1 Background on multivariate extremes

Prior to introducing a regression of block maxima on block maxima we need to lay groundwork on multivariate extremes. Let  $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$  be a sequence of independent random vectors with unit Fréchet marginal distributions, i.e.  $\exp(-1/z)$ , for  $z > 0$ . In our setup,  $Y_i$  should be understood as a response, whereas  $\mathbf{X}_i = (X_{1,i}, \dots, X_{p,i})$  should be understood as a  $p$ -dimensional covariate. Let the componentwise block maxima be  $\mathbf{M}_n = (M_{n,x_1}, \dots, M_{n,x_p}, M_{n,y})$  with  $M_{n,y} = \max\{Y_1, \dots, Y_n\}$  and  $M_{n,x_j} = \max(X_{j,1}, \dots, X_{j,n})$ , for  $j = 1, \dots, p$ . Under this setup, it is well-known that the vector of normalized componentwise maxima

$\mathbf{M}_n/n$  converges in distribution to a random vector  $(\mathbf{X}, Y)$  which follows a multivariate extreme value distribution with the joint distribution function

$$G(\mathbf{x}, y) = \exp\{-V(\mathbf{x}, y)\}, \quad (\mathbf{x}, y) \in (0, \infty)^{p+1}. \tag{3}$$

Here,

$$V(\mathbf{x}, y) = d \int_{\Delta_d} \max\left(\frac{w_1}{x_1}, \dots, \frac{w_p}{x_p}, \frac{w_{p+1}}{y}\right) H(d\mathbf{w}),$$

is the exponent measure and  $d = p + 1$ ; see, for instance, de Haan and Resnick (1977), Pickands (1981), and Coles (2001, Theorem 8.1). In addition,  $H$  is a parameter of the multivariate extreme value distribution  $G$  known as angular measure, which controls the dependence between the extreme values; specifically,  $H$  is a probability measure on the unit simplex  $\Delta_d = \{(w_1, \dots, w_d) \in [0, 1]^d, \sum_{i=1}^d w_i = 1\} \subset \mathbb{R}^d$ , and obeying the mean constraint

$$\int_{\Delta_d} \mathbf{w} H(d\mathbf{w}) = \frac{1}{d} \mathbf{1}_d, \tag{4}$$

where  $\mathbf{1}_d$  is a vector of ones in  $\mathbb{R}^d$ . If  $H$  is absolutely continuous with respect to the Lebesgue measure then its density is given by the Radon–Nikodym derivative  $h = dH/d\mathbf{w}$ , for  $\mathbf{w} \in \Delta_d$ .

### 2.2 Regression manifold for conditional multivariate extreme value distributions

We are now ready to introduce our regression method. We define the regression manifold as the family of regression lines,

$$\mathcal{L} = \{L_q : 0 < q < 1\} \quad \text{with} \quad L_q = \{y_{q|\mathbf{x}} : \mathbf{x} \in (0, \infty)^p\}, \tag{5}$$

where

$$y_{q|\mathbf{x}} = \inf\{y > 0 : G_{Y|\mathbf{X}}(y | \mathbf{x}) \geq q\}, \tag{6}$$

is a conditional quantile of a multivariate extreme value distribution, with  $q \in (0, 1)$  and  $\mathbf{x} \in (0, \infty)^p$ , and  $G_{Y|\mathbf{X}}(y | \mathbf{x}) = \mathbb{P}(Y \leq y | \mathbf{X} = \mathbf{x})$  is a conditional multivariate extreme value distribution function. The next proposition sheds light on sufficient conditions in order for the regression manifold in (5) to be smooth.

**Proposition 1** *Let  $G_{Y|\mathbf{X}}(y | x) = \mathbb{P}(Y \leq y | X = x)$  be a conditional multivariate extreme value distribution function. Suppose that  $y \mapsto G_{Y|\mathbf{X}}(y | \mathbf{x})$  is continuous and strictly increasing over  $[a, b] \subset (0, \infty)$ , and that  $\mathbf{x} \mapsto G_{Y|\mathbf{X}}(y | \mathbf{x})$  is continuous over  $\mathbb{R}^p$ . Then,  $y_{q|\mathbf{x}}$  is continuous with respect to the sup norm for all  $(\mathbf{x}, q) \in \mathbb{R}^p \times [c_{\mathbf{x}}, d_{\mathbf{x}}]$ , where  $c_{\mathbf{x}} = G_{Y|\mathbf{X}}(a | \mathbf{x})$  and  $d_{\mathbf{x}} = G_{Y|\mathbf{X}}(b | \mathbf{x})$ .*

**Proof** Let  $(\mathbf{x}_0, z_0)$  be a point in  $\mathbb{R}^p \times [c_{\mathbf{x}_0}, d_{\mathbf{x}_0}]$ . We aim to show that for every  $\varepsilon > 0$ ,



$$y_0 - \varepsilon < y_{q|\mathbf{x}} < y_0 + \varepsilon, \quad \text{whenever } \|(\mathbf{x}, q) - (\mathbf{x}_0, q_0)\|_\infty < \delta, \tag{7}$$

where  $y_0 := y_{q_0|\mathbf{x}_0}$  and  $\|(\mathbf{x}, z)\|_\infty = \max\{|x_1|, \dots, |x_p|, |z|\}$  is the sup norm. In order to find  $\delta > 0$  that works for (7) we start by recalling that since by assumption  $y \mapsto G_{Y|\mathbf{X}}(y | \mathbf{x})$  is continuous and strictly increasing, then it can be shown after some calculations by following a standard argument by Apostol (1980, Theorem 3.10) that there exists  $\delta_1 > 0$ , such that whenever  $q_0 - \delta_1 < q < q_0 + \delta_1$  it holds that

$$G_{Y|\mathbf{X}}(y_0 - \varepsilon | \mathbf{x}_0) < q < G_{Y|\mathbf{X}}(y_0 + \varepsilon | \mathbf{x}_0), \tag{8}$$

Indeed, to obtain (8) it suffices to take

$$\delta_1 = \min\{G_{Y|\mathbf{X}}(y_0 + \varepsilon | \mathbf{x}_0) - G_{Y|\mathbf{X}}(y_0 | \mathbf{x}_0), G_{Y|\mathbf{X}}(y_0 | \mathbf{x}_0) - G_{Y|\mathbf{X}}(y_0 - \varepsilon | \mathbf{x})\}.$$

Now, by assumption  $\mathbf{x} \mapsto G_{Y|\mathbf{X}}(y | \mathbf{x})$  is continuous and hence there exists a neighborhood of  $\mathbf{x}_0$  with radius  $\delta_2 > 0$  such that for every  $\mathbf{x}$  in that neighborhood it holds that  $G_{Y|\mathbf{X}}(y_0 - \varepsilon | \mathbf{x}) < q < G_{Y|\mathbf{X}}(y_0 + \varepsilon | \mathbf{x})$ , as a consequence of (8). This then implies that  $y_0 - \varepsilon < y_{q|\mathbf{x}} < y_0 + \varepsilon$ , and hence (7) holds with  $\delta = \min(\delta_1, \delta_2)$ , from where the final result follows.

In higher dimensions  $G_{Y|\mathbf{X}}$  can be expressed with the help of a joint multivariate extreme value density  $g_{\mathbf{X},Y}$  and its expression has been derived by Stephenson and Tawn (2005). By applying Bayes' theorem, we deduce  $G_{Y|\mathbf{X}}(y | \mathbf{x}) = \int_0^y g_{Y|\mathbf{X}}(z | \mathbf{x}) dz$  from  $g_{\mathbf{X},Y}$  with  $g_{Y|\mathbf{X}}$  given as follows:

$$g_{Y|\mathbf{X}}(y | \mathbf{x}) = \frac{\exp\{-V(\mathbf{x}, y)\} \sum_{i=1}^d \sum_{j=1}^{n_i} (-1)^i \prod_{\Lambda \in r_{ij}} V(\mathbf{x}, y)}{\sum_{i=1}^d \sum_{j=1}^{n_i} \int_0^\infty \exp\{-V(\mathbf{x}, y)\} \prod_{\Lambda \in r_{ij}} V_\Lambda(\mathbf{x}, y) dy}, \quad y, \mathbf{x} > 0, \tag{9}$$

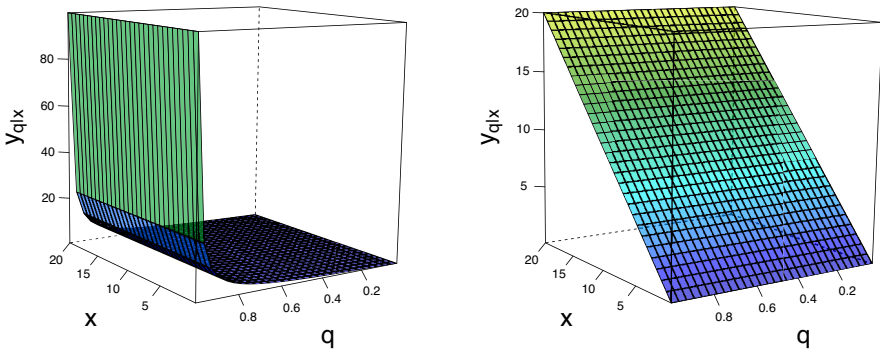
where  $V_\Lambda(\mathbf{x}, y)$  corresponds to mixed partial derivative of the exponent measure  $V(\mathbf{x}, y)$  with respect to the  $l$ th components of  $(\mathbf{x}, y)$  such that  $l \in \Lambda$ ,  $n_i$  is the number of partitions of  $\{1, \dots, d\}$  of size  $i = 1, \dots, d$ , and  $r_{ij}$  is the  $j$ th partition of  $\{1, \dots, d\}$  of size  $i$ , with  $1 \leq j \leq n_i$ .

In the particular case where we have a single covariate ( $p = 1$ ), the regression manifold  $\mathcal{L}$  in (5) can be derived using properties of bivariate copulas; see Appendix A. Accordingly, for an absolutely continuous angular measure  $H$  (with density  $h$ ), it follows that

$$G_{Y|\mathbf{X}}(y | x) = 2 \exp(1/x)G(x, y) \int_{\omega(x,y)}^1 wh(w) dw, \quad x, y > 0, \tag{10}$$

where  $\omega(x, y) = x/(x + y)$ , and  $y_{q|\mathbf{x}}$  is then calculated via (6).

We now derive regression manifolds  $\mathcal{L}$  in (5) for the cases of independent and perfectly dependent extremes, which are depicted in Fig. 1. When extremes are



**Fig. 1** Regression manifolds for cases of complete independence (left) and perfect dependence (right)

independent,  $H$  assigns equal mass to the boundaries of the simplex, which also corresponds to asymptotic independence of  $\mathbf{X}$  and  $Y$  (Hüsler and Li 2009), resulting in

$$L_q = \{-1/\log q : \mathbf{x} \in (0, \infty)^p\}, \tag{11}$$

with

$$G_{Y|\mathbf{X}}(y | \mathbf{x}) = \int_0^y \frac{\exp(-z^{-1} - x_1^{-1} - \dots - x_p^{-1})(-z^{-2}) \prod_{j=1}^p (-x_j^{-2})}{\exp(-x_1^{-1} - \dots - x_p^{-1}) \prod_{j=1}^p (-x_j^{-2})} dz = \exp(-1/y), \quad y > 0.$$

When extremes are perfectly dependent, the angular measure  $H$  assigns all its mass to the barycenter of the simplex,  $d^{-1}\mathbf{1}_d$ , leading to  $G(\mathbf{x}, y) = \exp\{-\max(x_1^{-1}, \dots, x_p^{-1}, y^{-1})\}$ . Taking derivatives of  $G(\mathbf{x}, y)$  in this case is non-trivial, and we replace the maximum function with a soft maximum (Cook 2011) so to obtain an approximation for the shape of the regression lines for perfectly dependent extremes. Thus, the soft maximum approximation for the regression lines for perfectly dependent extremes is

$$\tilde{L}_q = \{\min(x_1, \dots, x_p) : \mathbf{x} \in (0, \infty)^p\}. \tag{12}$$

That is, regression lines for the case of perfectly dependent extremes do not depend on  $q$ . See Appendix B for the derivation, and Fig. 1 for a chart of its regression manifold.

We end this section with comments on properties of regression manifolds. Trivially, regression lines obey the standard properties of quantile functions (van der Vaart 1998, Chap. 21). Less trivial is however the fact that monotone regression dependence of bivariate extremes (Guillem 2000, Theorem 1) implies that regression lines  $y_{q|x}$  in (6) are non-decreasing in  $x$ , for  $p = 1$ , under some mild assumptions.

**Proposition 2** *Let  $G_{Y|X}(y | x) = \mathbb{P}(Y \leq y | X = x)$  be a conditional bivariate extreme value distribution function, which we assume to be jointly continuously differentiable and strictly increasing in  $y$  for any fixed  $x \in (0, \infty)$ . Then, the regression lines for bivariate extremes  $(0, \infty) \ni x \mapsto y_{q|x}$  are non-decreasing for all  $q \in (0, 1)$ .*

**Proof** Since  $y \mapsto G_{Y|X}(y | x)$  is continuous (strictly increasing) for all  $x \in (0, \infty)$ ,  $y_{q|x}$  given by (6) is the solution to  $G_{Y|X}(y | x) = q$  for a fixed  $q \in (0, 1)$ . Then  $y$  satisfying  $G_{Y|X}(y | x) = q$  is an implicit function of  $x$  parametrized by  $q$ . Under our assumptions we apply the implicit function theorem and calculate the derivative of  $y_{q|x}$  with respect to  $x$  via

$$\frac{\partial}{\partial x} y_{q|x} = - \frac{\frac{\partial}{\partial x} G_{Y|X}(y | x)}{\frac{\partial}{\partial y} G_{Y|X}(y | x)}. \tag{13}$$

Equation (13) combined with the monotone regression dependence property, i.e.  $x \mapsto G_{Y|X}(y | x)$  is non-increasing for all  $y \in (0, \infty)$  (Guillem 2000, Theorem 1), and the strict monotonicity of  $y \mapsto G_{Y|X}(y | x)$  (increasing) for all  $x \in (0, \infty)$  gives

$$\frac{\partial}{\partial x} y_{q|x} \geq 0.$$

This completes the proof.

The assumption of  $y \mapsto G_{Y|X}(y | x)$  being strictly increasing for all  $x \in (0, \infty)$  is a mild assumption, which is satisfied by any bivariate GEV distribution that possesses an absolutely continuous angular measure with a positive density on  $(0, 1)$ . This can be easily seen by noting that under those circumstances, it follows from (10) that

$$\frac{\partial G_{Y|X}}{\partial y} = 2 \exp(1/x)G(x, y) \left( \frac{2}{y^2} \int_{\omega(x,y)}^1 wh(w) dw \int_0^{\omega(x,y)} (1-w)h(w) dw + \omega^3(x, y)h(\omega(x, y))/x \right) > 0.$$

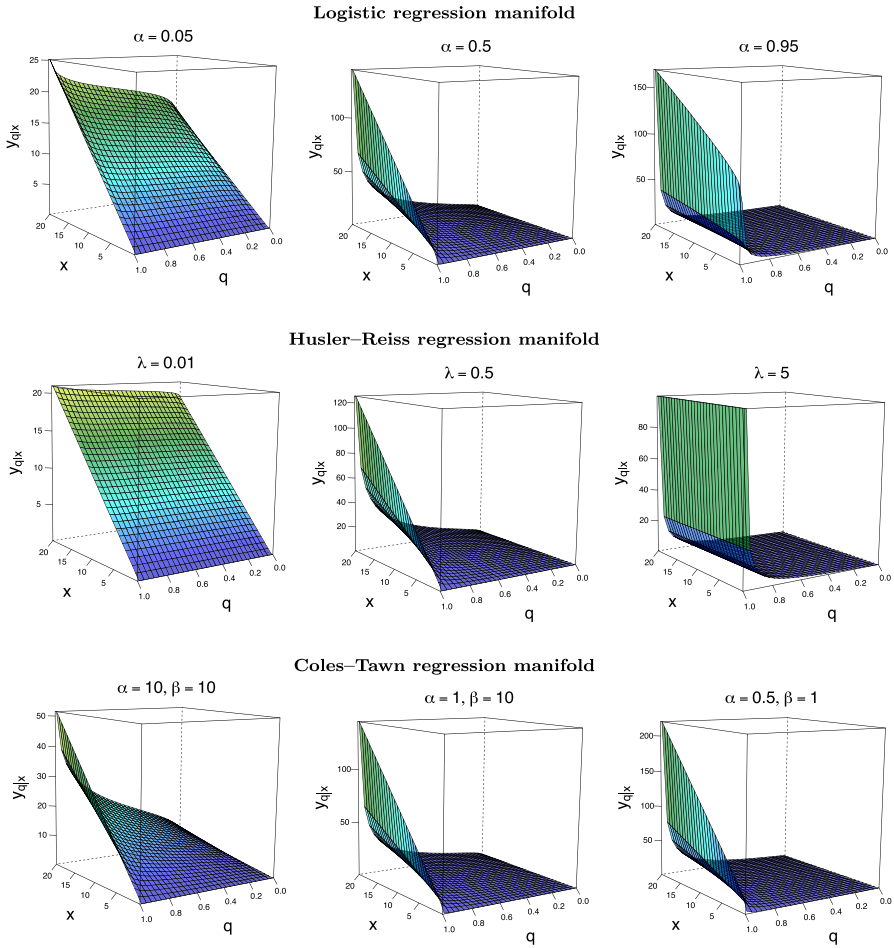
Thus, monotonicity of regression manifold holds for Examples 1–3 below.

### 2.3 Parametric instances of regression manifolds

We now consider some parametric instances of regression manifolds as defined in (5). Charts of regression manifolds for these parametric examples are depicted in Fig. 2. In Appendix C, we show that for sufficiently large  $x$ , the following linear approximation holds for the regression manifold,  $L_q = \{y_{q|x} : x \in (0, \infty)\}$ , of the Logistic model from Example 1 with

$$y_{q|x} = \gamma_q + \beta_q x + o(x). \tag{14}$$

Here,  $\gamma_q$  and  $\beta_q$  are functions of both  $\alpha$  and  $q$  (see Eqs. (24) and (25)), and  $o(x)$  is little- $o$  of  $x$  in Bachmann–Landau notation; the numerical accuracy of this approximation is illustrated in the supplementary material.



**Fig. 2** Regression manifold  $\mathcal{L}$ , as defined in (5), for bivariate Logistic, Husler–Reiss, and Coles–Tawn models (top to bottom) with strong dependence, intermediate and weak extremal dependence (left to right)

**Example 1 (Logistic)** An instance of the Logistic regression manifold can be found in Fig. 2 (top). It stems from the Logistic bivariate extreme value distribution function given by

$$G(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\}, \quad x, y > 0,$$

where  $\alpha \in (0, 1]$  characterizes the dependence between extremes: the closer  $\alpha$  to 0, the stronger the dependence, with the limit  $\alpha \rightarrow 0$  corresponding to the case of perfect dependence. The conditional distribution of  $Y$  given  $X$  is

$$G_{Y|X}(y | x) = G(x, y)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1} x^{1-1/\alpha} \exp(1/x), \quad x, y > 0,$$

thus leading to the following family of regression lines  $L_q$  in (5) where

$$y_{q|x} = \left[ \left\{ \frac{1-\alpha}{\alpha} x W \left( \frac{\alpha}{1-\alpha} x^{-1} e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)} \right) \right\}^{1/\alpha} - 1 \right]^{-\alpha} x, \tag{15}$$

and  $x > 0$ . Here,  $W$  is the so-called Lambert  $W$  function, that is, the multivalued analytic inverse of  $f(z) = z \exp(z)$  with  $z$  denoting a real or complex number (Borwein and Lindstrom 2016); see the supplementary material for further details. As it can be seen from Fig. 2 (top), the regression lines obey what is claimed in Proposition 2 in the sense that  $(0, \infty) \ni x \mapsto y_{q|x}$  are non-decreasing for all  $q \in (0, 1)$ .

**Example 2 (Husler–Reiss)** An instance of the Husler–Reiss regression manifold is depicted in Fig. 2 (middle). It follows from the Husler–Reiss bivariate extreme value distribution function which has the following form:

$$G(x, y) = \exp \left\{ -x^{-1} \Phi \left( \lambda + \frac{1}{2\lambda} \log \frac{y}{x} \right) - y^{-1} \Phi \left( \lambda + \frac{1}{2\lambda} \log \frac{x}{y} \right) \right\}, \quad x, y > 0,$$

where  $\Phi$  is the standard Normal distribution function and  $\lambda \in (0, \infty]$  is the parameter regulating the dependence between extremes:  $\lambda \rightarrow 0$  corresponds to perfect dependence and the limit case  $\lambda \rightarrow \infty$  corresponds to complete independence. The family of regression lines  $L_q$  in (5) for this model does not have explicit representations and is obtained using (6) with

$$G_{Y|X}(y | x) = \left[ \Phi \left( \lambda + \frac{1}{2\lambda} \log \frac{y}{x} \right) + \frac{1}{2\lambda} \phi \left( \lambda + \frac{1}{2\lambda} \log \frac{y}{x} \right) - \frac{xy^{-1}}{2\lambda} \phi \left( \lambda + \frac{1}{2\lambda} \log \frac{x}{y} \right) \right] \times G(x, y) \exp(1/x), \quad x, y > 0,$$

where  $\phi$  is the standard Normal density function.

**Example 3 (Coles–Tawn)** An instance of the Coles–Tawn regression manifold is depicted in Fig. 2 (bottom). It follows from the Coles–Tawn bivariate extreme value distribution function which has the following form:

$$G(x, y) = \exp[-x^{-1} \{1 - \text{Be}(q; \alpha + 1, \beta)\} - y^{-1} \text{Be}(q; \alpha, \beta + 1)], \quad x, y > 0,$$

where  $\text{Be}(q; a, b)$  is the cumulative distribution function of a Beta distribution with parameters  $a, b > 0$ ,  $q = \alpha y^{-1} / (\alpha y^{-1} + \beta x^{-1})$  and  $\alpha, \beta > 0$  are the parameters regulating dependence between extremes; the case  $\alpha = \beta = 0$  corresponds to complete independence, whereas  $\alpha = \beta \rightarrow \infty$  corresponds to perfect dependence. For fixed  $\alpha$  ( $\beta$ ) the strength of dependence increases with  $\beta$  ( $\alpha$ ). The family of regression lines  $L_q$  in (5) for this model does not have an explicit representation and is calculated using (6), for  $x, y > 0$ , with

$$G_{Y|X}(y | x) = \left[ 1 - \text{Be}(q; \alpha + 1, \beta) + \frac{(\alpha + 1)\beta}{\gamma} \text{be}(q; \alpha + 2, \beta + 1) - \frac{x}{y} \frac{\alpha(\beta + 1)}{\gamma} \text{be}(q; \alpha + 1, \beta + 2) \right] G(x, y) \exp(1/x),$$

where  $be(q;a, b)$  is the density function of the Beta distribution with parameters  $a, b > 0$  and  $\gamma = (\alpha + \beta + 2)(\alpha + \beta + 1)$ .

Section 2 introduced our key parameter of interest—regression manifolds for multivariate extreme values, i.e.  $\mathcal{L}$  as in (5)—, it commented on some of its properties, and gave examples of parametric instances. Next, we discuss Bayesian inference for  $\mathcal{L}$ .

### 3 Learning about regression manifolds via Bernstein polynomials

#### 3.1 Induced prior on the space of regression manifolds for $p = 1$

In this section we discuss how to learn about regression manifolds from data. To achieve this, we resort to the Bayesian paradigm and will define an induced prior on the space of regression manifolds by resorting to a flexible prior on the space of all angular measures that was recently proposed by Hanson et al. (2017). To lay the groundwork, we start by defining the setup of interest. Let  $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$  be a sequence of independent random vectors with unit Fréchet marginal distributions; define  $R_i = Y_i + \sum_{j=1}^p X_{j,i}$  and  $\mathbf{W}_i = (\mathbf{X}_i, Y_i)/R_i$ , known as the pseudo-angular decomposition of the observations. de Haan and Resnick (1977) showed the equivalence of the convergence of normalized componentwise maxima to  $G$  to the following weak convergence of measures

$$\mathbb{P}(\mathbf{W} \in \cdot \mid R > u) \xrightarrow{d} H(\cdot), \quad \text{as } u \rightarrow \infty.$$

This means that when the radius  $R$  is sufficiently large, the pseudo-angles  $\mathbf{W}$  are nearly independent of  $R$  and follow approximately a distribution associated with the angular measure  $H$ . Thus, to learn about  $L_q$  in (5), we first learn about  $H$  based on  $k = |\{\mathbf{W}_i : R_i > u, i = 1, \dots, n\}|$  exceedances above a large threshold  $u$ , where  $k = o(n)$ , with a methodology that we describe next.

Following Hanson et al. (2017), we model the angular density  $h$  via a Bernstein polynomial defined on the unit simplex  $\Delta_d$ , and hence basis polynomials are Dirichlet densities. More precisely, our specification for the angular density is

$$h(\mathbf{w}) = b(\mathbf{w}, J, \boldsymbol{\pi}_J) := \sum_{|\boldsymbol{\alpha}|=J} \pi_{\boldsymbol{\alpha}} \text{dir}_d(\mathbf{w}; \boldsymbol{\alpha}), \tag{16}$$

with  $\mathbf{w} \in \Delta_d$  and  $\boldsymbol{\pi}_J = (\pi_{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}| = J)$ . Here,  $\text{dir}_d$  is the density of a Dirichlet distribution supported on  $\Delta_d$ , that is,

$$\text{dir}_d(\mathbf{w}; \boldsymbol{\alpha}) = \frac{\Gamma(|\boldsymbol{\alpha}|)}{d \prod_{i=1}^d \Gamma(\alpha_i)} \prod_{i=1}^d w_i^{\alpha_i-1},$$

where  $\alpha \in \mathbb{N}^d$  (with  $\mathbb{N} := \{1, 2, 3, \dots\}$ ),  $|\alpha| = \sum_{j=1}^d \alpha_j$ , and  $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$  is the gamma function; finally in (16) the  $\pi_\alpha > 0$  are weights and  $J \in \mathbb{N}$  controls the order of the resulting polynomial.

To ensure that the resulting  $h(\mathbf{w})$  is a valid angular density (i.e. an actual density satisfying the moment constraint (4)), the weights must obey

$$\sum_{|\alpha|=J} \pi_\alpha = 1, \quad \sum_{i=1}^{J-d+1} i \sum_{|\alpha|=J, \alpha_j=i} \pi_\alpha = \frac{J}{d}, \tag{17}$$

for  $j = 1, \dots, d$ . The normalization and mean constraints in (17) imply that there are  $m - d$  parameters, where  $m = \binom{J-1}{d-1}$  is the number of basis functions in (16); denote such free weights as  $\{\pi_\alpha : \alpha \in \mathcal{F}\}$ , where  $\mathcal{F} = \{\alpha \in \mathbb{N}^d, |\alpha| = J, \text{ and } \alpha \notin \{\mathbf{a}_1, \dots, \mathbf{a}_d\}\}$  with  $\mathbf{a}_i$  being a  $J$ -vector of ones except element  $i$  is  $J - d + 1$ . Similarly to Hanson et al. (2017), we parametrize the free weights via a generalized logit transformation that implicitly defines the auxiliary parameters  $\pi'_\alpha$ , that is,

$$\pi_\alpha = \frac{\exp(\pi'_\alpha)}{d + \sum_{\tilde{\alpha} \in \mathcal{F}} \exp(\pi'_\alpha)}. \tag{18}$$

Now, to induce a prior in the space of regression manifolds we plug-in the angular density in (16) into (9); subsequent integration with respect to  $y$  and inversion of  $G_{Y|X}(y|\mathbf{x})$  leads to an induced prior on the space of regression lines  $L_q$ . In detail, to define a prior on the space of regression manifolds we proceed as follows. The Bernstein polynomial prior in (16) induces a prior on the space of regression lines  $L_q = \{y_{q|x} : x \in (0, \infty)\}$ , where  $y_{q|x}$  is a solution to equation,  $G_{Y|X}(y|x) = q$ , for  $q \in (0, 1)$ , where

$$\begin{aligned} &G_{Y|X}(y|x) \\ &= \frac{2}{J} \exp \left\{ -\frac{2}{J} \sum_{|\alpha|=J} \pi_\alpha [\alpha_1 x^{-1} \{1 - \text{Be}(\omega(x, y); \alpha_1 + 1, \alpha_2)\} + \alpha_2 y^{-1} \text{Be}(\omega(x, y); \alpha_1, \alpha_2 + 1)] \right\} \\ &\times \sum_{|\alpha|=J} \pi_\alpha \alpha_1 \{1 - \text{Be}(\omega(x, y); \alpha_1 + 1, \alpha_2)\} \exp(1/x), \end{aligned} \tag{19}$$

where  $\omega(x, y) = x/(x + y)$ , for  $x, y > 0$ . Finally, to complete the model specification we set the following Dirichlet prior on the free parameters

$$p(\pi_\alpha : \alpha \in \mathcal{F}) \propto \text{dir}_d(\mathbf{w} | c \mathbf{1}_m) \prod_{j=1}^d I \left\{ \sum_{i=1}^{J-d+1} i \sum_{|\alpha|=J, \alpha_j=i} \pi_\alpha = \frac{J}{d} \right\},$$

where  $I$  is the indicator function, which accordingly induces a prior on the auxiliary parameters  $\pi'_\alpha$  in (18).

### 3.2 Induced prior on the space of regression manifolds for $p \geq 1$

This section will show how an approximation due to Cooley et al. (2012) can be used for conducting inference in the case  $p \geq 2$ ; while the main focus will be on  $p \geq 2$  the provided approximation works as well in the case  $p = 1$ , and thus it can be regarded as an alternative to the exact approach from Sect. 3.1.

When  $p \geq 2$  we proceed as in Sect. 3.1, that is our induced prior in the space of regression lines is again induced by the Bernstein polynomial prior for the angular density in (16), and it follows by solving  $G_{Y|X}(y | \mathbf{x}) = q$ , with  $h(\mathbf{w})$  as in (16). The expression for the conditional multivariate extreme value distribution  $G_{Y|X}(y | \mathbf{x})$  for  $p \geq 2$  is however not as manageable as the one in (19). We thus propose an approach for learning about the regression manifold  $\mathcal{L}$ , as defined in (5), via an approximation to the conditional multivariate GEV density. Let  $\mathbf{u} = (\mathbf{x}, y) \in (0, \infty)^d$  and  $\mathbf{u}(t) = (\mathbf{x}, t) \in (0, \infty)^d$  with  $\|\mathbf{u}\| = y + \sum_{i=1}^p x_i > u$  for a large threshold  $u$ .

Then, following Cooley et al. (2012, Proposition 1) the conditional density of a multivariate extreme value distribution can be approximated, via a point process representation for extremes, as follows

$$g_{Y|X}(y | \mathbf{x}) \approx \frac{\|\mathbf{u}\|^{-d-1} h(\mathbf{u}/\|\mathbf{u}\|)}{\int_0^\infty \|\mathbf{u}(t)\|^{-d-1} h(\mathbf{u}(t)/\|\mathbf{u}(t)\|) dt}. \tag{20}$$

An induced prior for  $g$  can be devised by plugging the approximation in (20) with the specification from Sect. 3.1, which leads to the following prior for the conditional multivariate extreme value distribution function,

$$G_{Y|X}(y | \mathbf{x}) \approx \frac{\sum_{|\alpha|=J} \pi_\alpha / B_\alpha \prod_{i=1}^p x_i^{\alpha_i-1} \int_0^y t^{\alpha_d-1} \|\mathbf{u}(t)\|^{-J-1} dt}{\sum_{|\alpha|=J} \pi_\alpha / B_\alpha \prod_{i=1}^p x_i^{\alpha_i-1} \int_0^\infty t^{\alpha_d-1} \|\mathbf{u}(t)\|^{-J-1} dt}, \tag{21}$$

where  $B_\alpha = \prod_{i=1}^d \Gamma(\alpha_i) / \Gamma(|\alpha|)$  is the multivariate beta function. Hence, we can learn about the regression manifold  $\mathcal{L}$  by estimating  $\pi_\alpha$  as described in Sect. 3.1, that is, by plugging in the Bernstein polynomial estimates (16) into the approximation (20) and numerically inverting (21) with respect to  $y$ . This strategy is illustrated numerically in the supplementary material.

## 4 Simulation study

### 4.1 Preliminary experiments

We study the finite sample performance of the proposed methods under three data generating scenarios that were introduced in Sect. 2; see Examples 1–3. Specifically, we simulate data as follows:



- **Scenario 1**—strongly dependent extremes: Husler–Reiss model with  $\lambda = 0.1$ .
- **Scenario 2**—weakly dependent extremes: Logistic model with  $\alpha = 0.9$ .
- **Scenario 3**—asymmetric intermediate dependence: Coles–Tawn model with  $\alpha = 0.5$ ,  $\beta = 100$ .

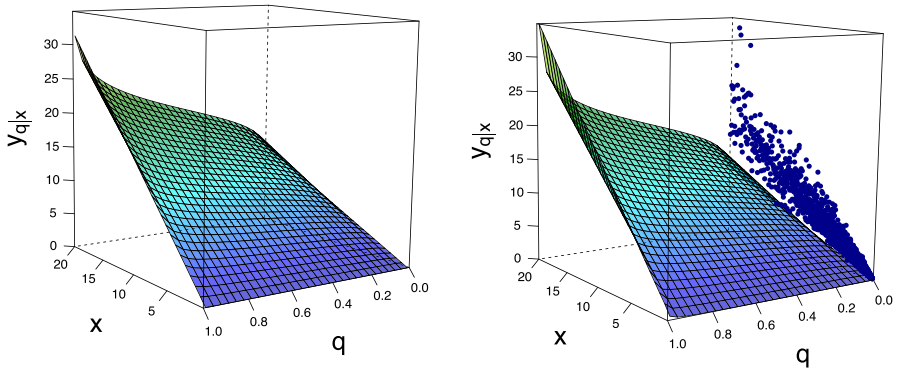
For now we focus on illustrating the methods in a single-run experiment; a Monte Carlo simulation study will be reported in Sect. 4.2. To illustrate how the resulting estimates compare with the true regression lines on a one-shot experiment, in each scenario we generate  $n = 5\,000$  samples  $\{(\mathcal{X}_i, \mathcal{Y}_i)\}_{i=1}^n$ . For the analysis we use observations for which  $\hat{X}_i + \hat{Y}_i > u$ , where  $u$  is the 98% quantile; here, the raw data are transformed to unit Fréchet margins via the transformation  $(\hat{X}_i, \hat{Y}_i) = (-1/\log\{\hat{F}_X(\mathcal{X}_i)\}, -1/\log\{\hat{F}_Y(\mathcal{Y}_i)\})$ , where  $\hat{F}_X$  and  $\hat{F}_Y$  respectively denote the empirical distribution functions (normalized by  $n + 1$  rather than by  $n$  to avoid division by zero). To learn about regression lines from data, we employ a standard componentwise adaptive Markov Chain Monte Carlo (MCMC) (Haario et al. 2005) with a Dirichlet prior,  $\text{Dirichlet}(10^{-4}\mathbf{1}_k)$ , defined on a generalized logit transformation of weights  $\pi_\alpha$ . The length of each MCMC chain is 10 000 with a burn-in period of 4 000.

In Fig. 3 we plot true and estimated regression manifolds under the three scenarios above over the range  $(x, y) \in (0, 20] \times (0, 20]$ , where 20 corresponds to the 95% quantile of the unit Fréchet marginal distributions. Figure 3 shows that, for these one shot experiments, the proposed methods recover well the shape of  $\mathcal{L}$  for all three cases, although as expected the fits are slightly less accurate for  $q$  closer to 0 and 1. To have a closer look into the outputs from these numerical experiments, we depict in Fig. 4 cross sections of the angular manifold, over  $q$  and over  $x$ , thus leading to regression lines and conditional quantiles for the Husler–Reiss (top), Logistic (central), and Coles–Tawn (bottom) models. Once more, we see that the fits are fairly reasonable overall although a bit more of bias is visible for  $q$  closer to 0 and 1, as can be seen in the charts of the conditional quantile curves. Interestingly, it can also be noticed in Fig. 4 that regression lines are approximately linear for the Logistic model, and we prove that this indeed is the case for large  $x$ ; see Appendix C.

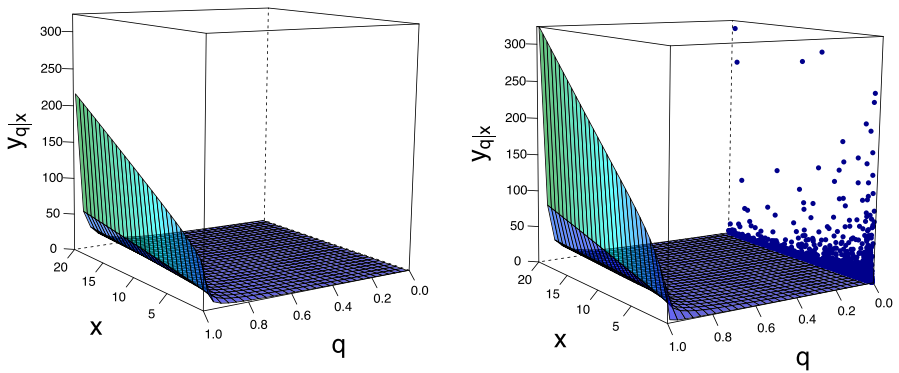
Figure 4 also anticipates a feature that we will revisit in Sect. 4.2, i.e. that the more the regression lines change with  $q$ , the larger is the uncertainty; this explains why the bands of the regression lines for Scenario 3 are wide, while those of Scenario 1 are narrow. Similarly, the more abruptly the regression lines change with  $q$ , the larger is the uncertainty. For example, for Scenario 2 it can be noticed that the true regression line for  $q = 0.9$  attains values much larger than the cross sections  $q = 0.1, 0.45, 0.55$ , and indeed the bands are wider for  $q = 0.9$ .

Motivated by the computational experience of one of the authors on a recent paper (Galasso et al. 2022), we have opted not to set a prior on  $J$ . This allows for any user to replicate our Monte Carlo simulation studies, and not just those who have access to Google Cloud, Amazon Web Services or a University server, as our methods are easy to code and can be readily implemented in a standard desktop machine; for example, the full Monte Carlo simulation study reported in Sect. 4.2 takes around 5 hours to run on a 3,2 GHz 6-Core Intel Core i7 (64 GB) machine.

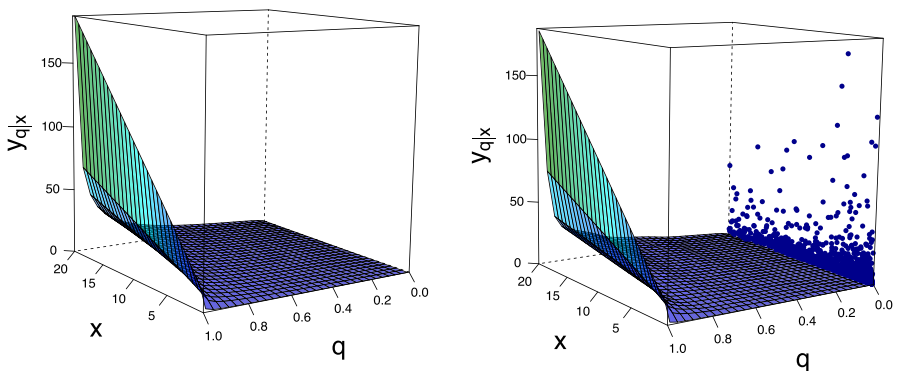
**Scenario 1**—strongly dependent extremes: Husler–Reiss model



**Scenario 2**—weakly dependent extremes: Logistic model

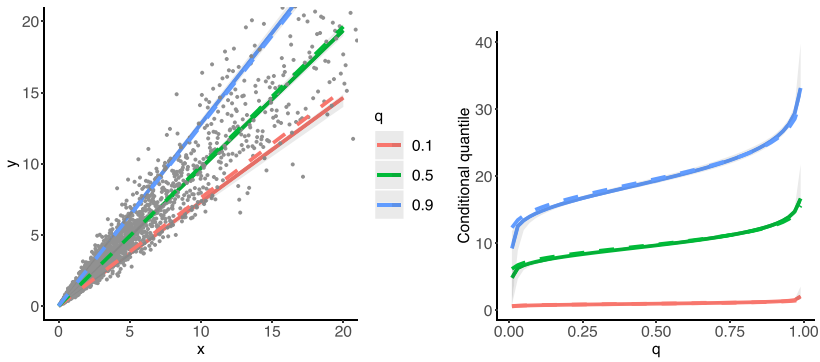


**Scenario 3**—asymmetric intermediate dependence: Coles–Tawn model

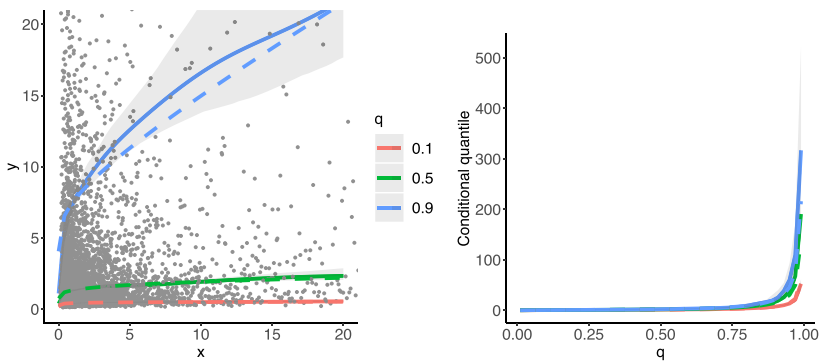


**Fig. 3** True regression manifold  $\mathcal{L}$ , as defined in (5), along with its posterior mean estimate obtained using the methods from Sect. 3 for Husler–Reiss, Logistic, and Coles–Tawn bivariate extreme value models (top to bottom) on a single-run experiment. Simulated data are overlaid on one of the faces of the box

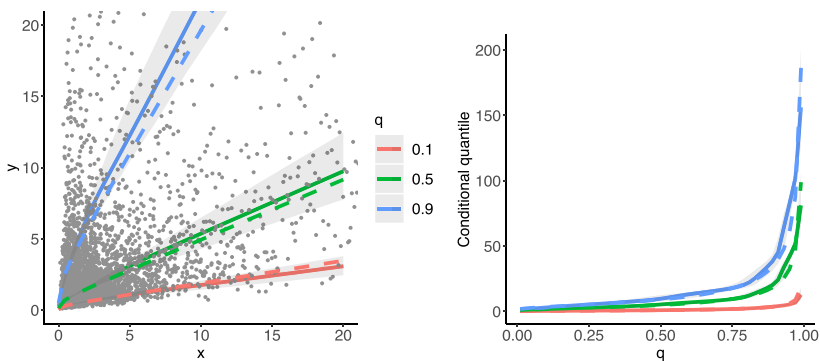
**Scenario 1**—strongly dependent extremes: Husler–Reiss model



**Scenario 2**—weakly dependent extremes: Logistic model



**Scenario 3**—asymmetric intermediate dependence: Coles–Tawn model



**Fig. 4** Posterior mean regression lines  $L_q$  for  $q \in \{0.1, 0.5, 0.9\}$  and  $x \in (0, 20]$  (left) and conditional quantile curves  $\{y_{q|x} : q \in (0, 1)\}$  along with credible bands, for  $x = \{1, 10, 20\}$  (right) for Husler–Reiss, Logistic, and Coles–Tawn bivariate extreme value models (top to bottom) on a single-run experiment. The dashed lines represent the true

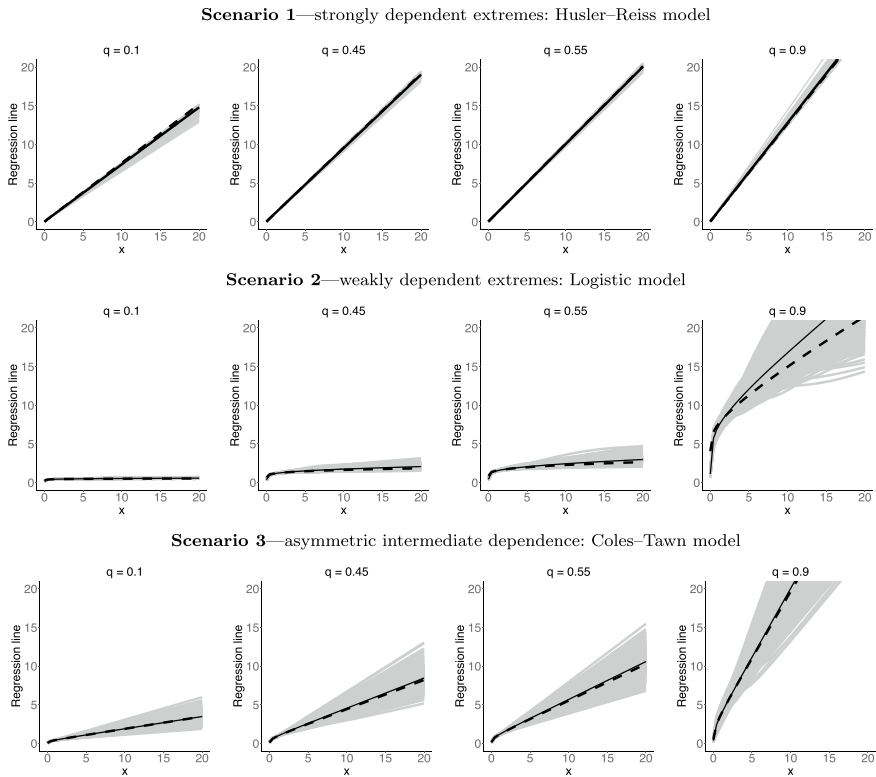
Hence, rather than setting a prior on  $J$ , we have opted for the computationally appealing and effective approach of Hanson et al. (2017) where  $J$  is chosen to be maximized subject to the number of basis functions being smaller or equal to the number of pseudo-angles. Part of the rationale for setting a large  $J$  is the well-known fact that, as  $J \rightarrow \infty$ , Bernstein polynomials can uniformly approximate any continuous and bounded density on the unit interval (Altomare and Campiti 1994, p. 333) and multivariate Bernstein polynomials obey a similar property (Barrientos et al. 2015, Sect. 4.1). As shown in Sect. 3 of the supplementary material, this strategy of choosing  $J$  of Hanson et al. (2017) leads to comparable performance, if not superior, with respect to the likelihood ratio changepoint approach of Guan (2016).

## 4.2 Monte Carlo simulations

To conduct a simulation study we generate 500 Monte Carlo samples of size  $n = 5000$  thresholded at the 98% quantile for the three scenarios described in Sect. 4.1. We use the MCMC algorithm as described in Sect. 4.1 with the same prior specification. The performance of our methods will be visualized via a comparison of posterior mean estimates of the regression lines with the true regression lines  $L_q$  for a few fixed  $q \in (0, 1)$ . We focus on the region  $x \in (0, 20]$  as the bivariate extreme value concentrates most of its mass (at least 90%) in the set  $(0, 20] \times (0, 20]$ .

The regression lines corresponding to the described scenarios are shown in Fig. 5. Figure 5 shows that the model fits the data for all scenarios reasonably well but again it can be noticed that the fit is slightly less accurate in Scenario 3 for  $q = 0.9$ . As mentioned in Sect. 4.1, our computational experience with the method leads us to believe that this is due to how suddenly the cross section changes for higher  $q$ , as indeed the true regression line for  $q = 0.9$  attains values much larger than those of  $q = 0.1, 0.45$ , and  $0.55$ . Interestingly, while the Monte Carlo means for Scenario 1 are in line with the true there is some asymmetry visible in the rare outlying fits in Fig. 5 though their frequency is not sufficient to shift the Monte Carlo means away from the true.

In addition to the numerical experiments reported above we have conducted a battery of other studies. Section 3 of the supplementary material includes a sensitivity analysis on the selected threshold as well as a similar Monte Carlo study for  $n = 10000$ . The results in the supplementary material suggest similar findings when the 95% quantile is used for thresholding the data, but with the asymmetry effect mentioned above being slightly attenuated; one can also notice in that additional study a moderate improvement in the fits when  $n = 10000$ . Beyond the simulation scenarios examined above, we also report in the supplementary material an additional scenario with a discrete angular measure; despite the fact that our prior is defined on the space of continuous angular measures the results available from the supplementary material indicate a satisfactory performance even on the latter case.



**Fig. 5** Posterior mean regression lines  $L_q$  for  $q \in \{0.1, 0.45, 0.55, 0.9\}$  and  $x \in (0, 20]$  for each of the 500 Monte Carlo samples (gray lines) plotted against the true (dashed line) for Husler–Reiss, Logistic, and Coles–Tawn bivariate extreme value models (top to bottom). The solid black line represents the Monte Carlo mean

## 5 Real data illustration

### 5.1 Data, preprocessing, and applied rationale for the analysis

#### 5.1.1 Description and scope

We now apply the proposed method to two of the world’s biggest stock markets—the NASDAQ (National Association of Securities Dealers Automated Quotations) and the NYSE (New York Stock Exchange). According to the Statistics Portal of the World Federation of Exchanges (<https://statistics.world-exchanges.org>), the total equity market capitalization of NASDAQ and NYSE are respectively 20.99 and 24.67 trillion US\$, as of April 2021, thus illustrating well the scale of these players in the worldwide stock-exchange industry. The data, available from the R package DATAstudio, were gathered from Yahoo Finance, and consist of daily closing prices of the NASDAQ and NYSE composite indices over the period from February

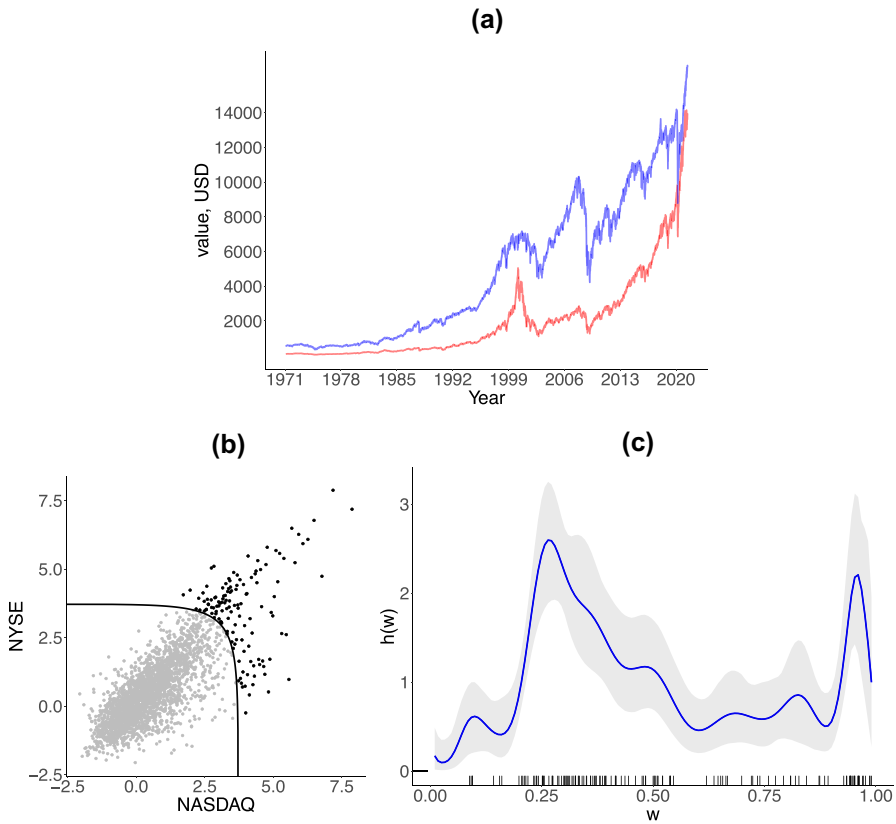
5, 1971 to June 9, 2021. A key goal of the analysis to be reported in Sect. 5.2 is to showcase how our model can be used to learn about the conditional risk of an extreme loss in the NYSE, given an extreme loss in the NASDAQ. Before we present the bulk of the conditional risk analysis via our methods we first offer below some remarks on the data.

We will focus on negative log returns, which can be regarded as a proxy for losses and which consist of first differences of prices on a log-scale; the resulting sequence of  $n$  componentwise weekly maxima losses for NASDAQ and NYSE is denoted below as  $\{(\mathcal{X}_i, \mathcal{Y}_i)\}_{i=1}^n$ . The sample period under analysis is sufficiently broad to cover a variety of major downturns and selloffs including, for example, those related with the 2007–2010 subprime mortgage crisis, the ongoing China–US trade war, and with the 2020 COVID-19 pandemic. Similarly to Sect. 4, we convert negative log returns to unit Fréchet margins via the transformation  $(\hat{X}_i, \hat{Y}_i) = (-1/\log\{\hat{F}_{\mathcal{X}}(\mathcal{X}_i)\}, -1/\log\{\hat{F}_{\mathcal{Y}}(\mathcal{Y}_i)\})$ , where  $\hat{F}_{\mathcal{X}}$  and  $\hat{F}_{\mathcal{Y}}$  respectively denote the empirical distribution functions (normalized by  $n + 1$  rather than by  $n$  to avoid division by zero) of negative log returns for NASDAQ ( $\mathcal{X}$ ) and NYSE ( $\mathcal{Y}$ ); the supplementary material includes the reverse analysis that swaps the roles of NASDAQ and NYSE (i.e. NASDAQ becomes  $\mathcal{Y}$  and NYSE becomes  $\mathcal{X}$ ). To validate the use of our model we further test negative log returns on NYSE and NASDAQ for multivariate regular variation (MRV) following methodology introduced by Einmahl et al. (2021). The results of the conducted analysis are presented in Sect. 4 of the supplementary material and suggest that at a 5% significance level there is no evidence to reject that the pair (NYSE, NASDAQ) follows a MRV distribution for a broad range of thresholds.

### 5.1.2 Visualization

The raw data and resulting preprocessed data are depicted in Fig. 6. As can be seen from the latter figure the composite indices exhibit a similar dynamics reacting to different economic shocks (9/11 attacks, 2001; 2008 financial crisis; China-US trade war started in 2018) alike. Also, as can be seen from Fig. 6, the shape of the scatterplot of the negative log returns brought to unit Fréchet margins in log-log scale above the boundary threshold evidences intermediate level of extremal dependence between negative log returns.

Before we focus on conditional risk (Sect. 5.2), we start by fitting the angular density via the Bernstein polynomial-based approach from Sect. 3.1. We follow essentially the same settings for prior specification and MCMC as in Sect. 4, threshold the unit Fréchet data at its 95% quantile, set the number of basis functions using once more the approach of Hanson et al. (2017), and run an MCMC chain of length 25 000 with a burn-in period of 10 000. The specified chain has the multivariate effective sample size of 1 726 657. The obtained fit for the angular density is reported in Fig. 6 (right). As is illustrated by this plot most of the observed pseudo-angles lie closer to the middle of the interval (0, 1) and the estimate resembles a bell-shaped right-skewed density which suggests there is an asymmetric intermediate dependence between extremal losses on NASDAQ and NYSE composite indices. Such asymmetry suggests a tendency for the losses of the NYSE to be more extreme than those of NASDAQ, when both are extreme.



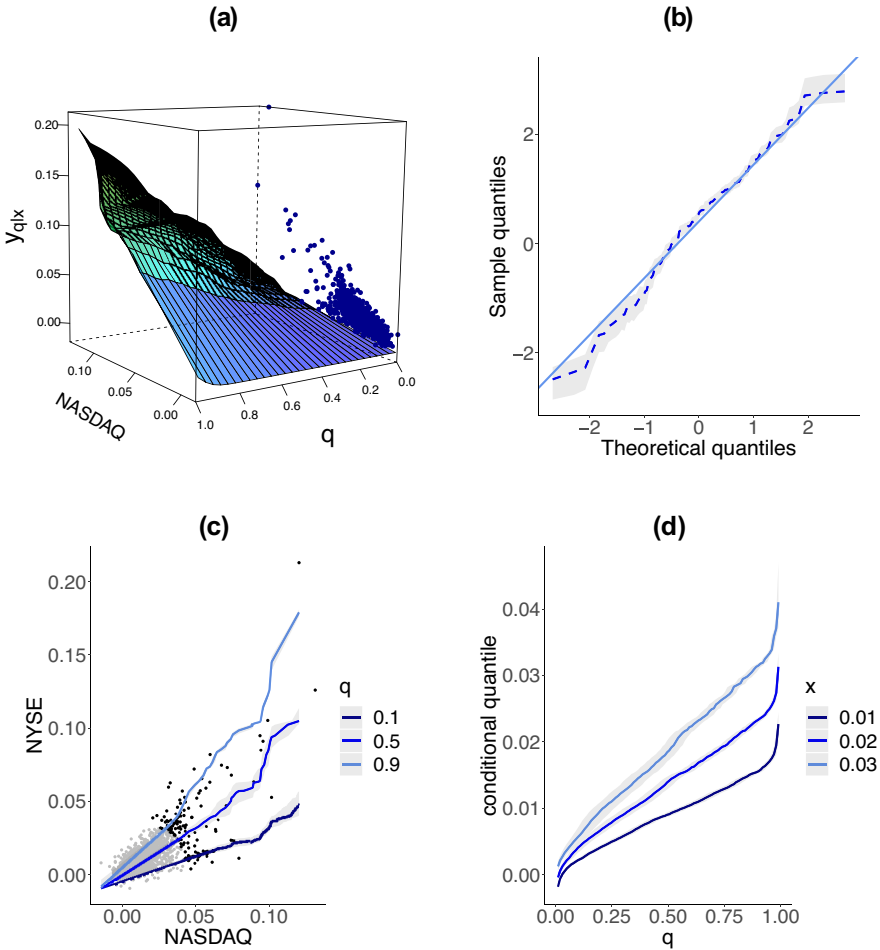
**Fig. 6** (a) NASDAQ (red) and NYSE (blue) composite indices. (b) Scatterplot of negative log returns of NASDAQ and NYSE composite indices converted to unit Fréchet margins; the solid line corresponds to the boundary threshold in the log-log scale, with both axes being logarithmic. (c) Angular density estimate with 95% credible band along with a rug of pseudo-angles

### 5.2 Modeling conditional risk via regression manifolds

This section presents the bulk of the conditional risk analysis. Specifically, we will show how the regression manifold can be used to model the conditional risk of an extreme loss on the NYSE, given an extreme loss on the NASDAQ. Figure 7a represents the resulting estimates of the regression manifold on the original scale together with cross-sections in  $q$  and  $x$ , respectively in (b) and (d) for negative log-returns on NASDAQ and NYSE composite indices in the original margins. Having evaluated the regression manifold, we are now ready to extract from its cross sections information on by how much the NYSE can plummet, conditionally on the NASDAQ plummeting. To examine this, we report in Table 1 predicted 75%, 90% and 95% quantiles of losses on NYSE evaluated for 1%, 2% and 3% weekly maxima losses on NASDAQ. This table follows from the regression manifold, and its interpretation is as follows. First, from a qualitative

**Table 1** Predicted 75%, 90% and 95% quantiles of losses on NYSE evaluated for 1%, 2% and 3% weekly maxima losses on NASDAQ, with 95% credible intervals in brackets; negative log-returns used as proxy for losses

NYSE	NASDAQ		
	0.01	0.02	0.03
75%	0.0129 (0.0125, 0.0134)	0.0194 (0.0187, 0.0201)	0.0266 (0.0259, 0.0274)
90%	0.0156 (0.0153, 0.0159)	0.0233 (0.0229, 0.0237)	0.0315 (0.0309, 0.0319)
95%	0.0171 (0.0166, 0.0177)	0.0249 (0.0245, 0.0255)	0.0333 (0.0328, 0.0339)



**Fig. 7** (a) Posterior mean regression manifold on the original scale for NYSE given NASDAQ along with joint negative log returns overlaid on one of the faces of the box. (b) QQ-plot of randomized quantile residuals; the dashed line represents the posterior mean plotted along with credible bands. (c) Posterior mean regression lines  $L_q$  for  $q \in \{0.1, 0.5, 0.9\}$  for NYSE given NASDAQ along with 95% credible bands and plotted against joint negative log returns. (d) Posterior mean conditional quantile curves  $\{y_{q|x} : q \in (0, 1)\}$  of negative log returns on NYSE for  $x \in \{0.01, 0.02, 0.03\}$ , along with 95% credible bands, corresponding to negative log returns on NASDAQ in the original margins



viewpoint, Table 1 indicates that conditionally on the NASDAQ plummeting, the NYSE tends to plummet reasonably by the same amount. Second—and more interesting from a financial outlook—are the quantitative claims that can be made from the analysis. For example, Table 1 indicates that whenever there is a 1% weekly maximum loss on the NASDAQ, only in 5% of the times we expect to suffer a loss in the NYSE above 1.71%. As another example, Table 1 indicates that whenever there is a 3% weekly loss on the NASDAQ, only in 5% of the times we expect to suffer a loss in the NYSE above 3.33%.

As can be seen from Fig. 7 the regression manifold is highly non-linear and the regression lines on the middle graph substantially differ from those corresponding to independence and tend to be closer to the identity line. Moreover, the cross-sections for different values of  $x$  reveal considerable variation in quantiles of  $y$  supporting the conclusion about presence of the dependence between negative log-returns. To assess the quality of the fitted regression manifold we depict in Fig. 6b a QQ-plot of a version of Dunn and Smyth (1996) randomized quantile residuals adapted to our model, defined as  $\varepsilon_i = \Phi^{-1}(G_H(Y_i | X_i))$ , for  $Y_i + X_i > u$ , with  $u$  denoting the 95% quantile of the pseudo-radius. The latter chart depicts randomized quantile residuals against the theoretical standard Normal quantiles, and it suggests an acceptably good fit of the proposed model. Beyond the results shown above we have conducted a variety of other empirical analysis which are available from the supplementary material (Sect. 5). In particular, we report in the supplementary material the fitted regression manifold that results from combining our random Bernstein model with the approximation from Sect. 3.2. As expected, the results are tantamount to the ones presented herein.

## 6 Closing remarks

We propose a regression-type model for the setup where both the response and the covariate are extreme. The modeling starting point is the result that the limiting behavior of the vector of properly standardized componentwise maxima is given by a multivariate extreme value distribution. Conceptually, the model is then constructed in a similar fashion as in quantile regression, that is, by assessing how the conditional quantile of the response reacts to changes in the covariate while it takes into account the latter asymptotic result. An important target in the proposed framework is the regression manifold, which consists of a family of regression lines obeying the proviso of multivariate extreme value theory. A Bernstein polynomial prior on the space of angular densities is used to learn about the model from data, with numerical studies showcasing its flexibility. While it is not too difficult to show that a variant of our model for the angular density is consistent under the same proviso as that of Sabourin and Naveau (2014, Sect. 3.3), as can be seen from Sect. 1 of the supplementary material, the large sample behavior of the proposed methods under a more realistic asymptotic framework remains an open question for future analysis. A comprehensive asymptotic analysis related with the current framework can be found in Padoan and Rizzelli (2022).

One could wonder why not to resort to statistical models for nonstationary extremes (e.g. Coles 2001, Sect. 6) as an alternative to methods proposed herein, as these can be used for assessing the effect of covariates on an extreme-valued response, by indexing the parameters of the GEV distribution with a covariate. Yet, since the latter models are built from the univariate theory of extremes they are not tailored for conditioning on another variable being extreme, as they fail to take on board information from the dependence structure between the extremes. Other related approaches include extremal quantile regression methods (Chernozhukov 2005), which similarly to the statistical models for nonstationary extremes, have not been designed for conditioning on another variable being extreme, as they do not take into account the dependence structure between the extremes.

While not explored here, the comparison of the fitted models for both  $Y | X = x$  and  $X | Y = y$ , would look natural for some applied settings of interest so to get an idea of cause and effects, and indeed related ideas are analyzed by Mhalla et al. (2020). We close the paper with some comments on future research. For regressions with many predictors, it is likely that most covariates will have little effect on the response and thus one could wonder how to devise a version of the proposed method that shrinks towards zero the effect of such irrelevant covariates; the development of a Lasso (Tibshirani 1996) version of the proposed model would thus seem natural for such situation, and is left as an open problem for future research. Another natural avenue for future research would be to devise regression-type methods for exceedances on exceedances by resorting to the so-called multivariate generalized Pareto distribution (Kiriliouk et al. 2019), rather than with the multivariate extreme value distribution as herein. Additionally, the development of a version of the model that could take into account asymptotic independence by resorting to the hidden angular measure (Ramos and Ledford 2009), rather than the standard angular measure as herein, would seem natural as well. Finally, while the proposed inference methods apply to both the block maxima and threshold exceedance frameworks for multivariate extremes, if only block maxima are available there is still room for further improving the inferences of regression manifolds, and such topic might be worth further investigation and future analysis; a natural starting point would be to work directly with the likelihood of the multivariate extreme value distribution along the lines of what is discussed, for instance, in Coles (2001, Sect. 8.2.2) and Dombry et al. (2017).

## Appendix

### Appendix A. Conditional bivariate extreme value distribution

Here we derive the expression for the conditional bivariate extreme value distribution function in (10). Sklar's theorem (Nelsen 2006, Theorem 2.3.3), implies that a joint bivariate distribution function  $G : \mathbb{R}^2 \rightarrow [0, 1]$  with continuous marginal distributions  $G_X : \mathbb{R} \rightarrow [0, 1]$  and  $G_Y : \mathbb{R} \rightarrow [0, 1]$  can be uniquely represented through a copula function  $C$  for  $(x, y) \in \mathbb{R}^2$   $G(x, y) = C(G_X(x), G_Y(y))$ , or, equivalently,

$C(u, v) = G(G_X^{-1}(u), G_Y^{-1}(v))$ , for  $(u, v) \in [0, 1]^2$ , where  $G_X^{-1}(q) = \inf \{x : G_X(x) \geq q\}$ . Using the following well-known property of copulas,

$$C_{V|U}(v | u) := \mathbb{P}(V \leq v | U = u) = \frac{\partial C(u, v)}{\partial u}, \quad (u, v) \in [0, 1]^2,$$

we calculate the conditional distribution  $(Y | X)$  as

$$G_{Y|X}(y | x) = C_{V|U}(e^{-1/y} | e^{-1/x}).$$

In our setting

$$C(u, v) = \exp \left[ -2 \int_0^1 \max \{-w \log u, -(1-w) \log v\} H(dw) \right].$$

Assuming  $H$  is absolutely continuous with density  $h$ , we have

$$\begin{aligned} & \frac{\partial}{\partial u} \int_0^1 \max \{-w \log u, -(1-w) \log v\} H(dw) \\ &= -\frac{\partial}{\partial u} \left( \log u \int_{\omega(u,v)}^1 w H(dw) + \log v \int_0^{\omega(u,v)} (1-w) H(dw) \right) \\ &= -u^{-1} \int_{\omega(u,v)}^1 wh(w) dw - (\log u \log^2 v - \log^2 v \log u) \frac{u^{-1}}{\log^3 uv} h(\omega(u, v)) \\ &= -u^{-1} \int_{\omega(u,v)}^1 wh(w) dw, \end{aligned}$$

where  $\omega(u, v) = \log v / \log uv$ . Then, the conditional copula has the following form

$$C_{V|U}(v | u) = 2u^{-1} C(u, v) \int_{w(u,v)}^1 wh(w) dw,$$

which in turn yields

$$G_{Y|X}(y | x) = 2 \exp \left\{ -2 \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) h(w) dw + x^{-1} \right\} \int_{w(x,y)}^1 wh(w) dw, \quad x, y > 0,$$

where  $w(x, y) = x / (x + y)$ .

**Appendix B. Soft-maximum approximation for regression manifold of perfectly dependent extremes**

Here we give details on the soft maximum approximation for the regression lines for perfectly dependent extremes claimed in (12). We use a smooth approximation of a maximum function called *soft-maximum*,

$$f(z_1, \dots, z_d; N) = \frac{1}{N} \log(e^{Nz_1} + \dots + e^{Nz_d}),$$

which is infinitely differentiable everywhere and converges to the maximum function as  $N \rightarrow \infty$  (Cook 2011). Then, the approximation of a multivariate GEV distribution function for the case of perfect dependent extremes,  $G(\mathbf{x}, y) = \max\{x_1^{-1}, \dots, x_p^{-1}, y^{-1}\}$  is

$$\begin{aligned} \tilde{G}(\mathbf{x}, y; N) &= \exp \left\{ -\frac{1}{N} \log(e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}}) \right\} \\ &= (e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N}, \end{aligned}$$

and its partial derivative of order  $d$  is

$$\begin{aligned} \tilde{g}(\mathbf{x}, y; N) &= \frac{\partial^d}{\partial x_1 \dots \partial x_p \partial y} \tilde{G}(\mathbf{x}, y; N) \\ &= y^{-2} \prod_{i=1}^p (1 + iN)x_i^{-2} \exp \left( Ny^{-1} + \sum_{i=1}^p x_i^{-1} \right) (e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-d}. \end{aligned}$$

This yields the following approximation of the conditional multivariate GEV density for perfectly dependent extremes,

$$\begin{aligned} \tilde{g}_{Y|\mathbf{X}}(y | \mathbf{x}; N) &= \frac{y^{-2} \prod_{i=1}^p (1 + iN)x_i^{-2} \exp \left( Ny^{-1} + \sum_{i=1}^p x_i^{-1} \right) (e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-d}}{\prod_{i=1}^{p-1} (1 + iN) \prod_{i=1}^p x_i^{-2} \exp \left( \sum_{i=1}^p x_i^{-1} \right) (e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-p}} \\ &= (1 + pN)y^{-2} e^{Ny^{-1}} \frac{(e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-d}}{(e^{Nx_1^{-1}} \dots + e^{Nx_p^{-1}})^{-1/N-p}}, \end{aligned}$$

and the following approximation for the corresponding conditional cumulative distribution function

$$\begin{aligned} \tilde{G}_{Y|\mathbf{X}}(y | \mathbf{x}; N) &= \int_0^y \tilde{g}_{Y|\mathbf{X}}(z | \mathbf{x}) dz \\ &= (1 + pN)(e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{1/N+p} \\ &\quad \int_0^y z^{-2} e^{Nz^{-1}} (e^{Nz^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-d} dz \\ &= \frac{(1 + pN)(-1/N)}{-1/N - d + 1} (e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{1/N+p} \\ &\quad (e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}})^{-1/N-d+1} \\ &= \left( \frac{e^{Ny^{-1}} + e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}}}{e^{Nx_1^{-1}} + \dots + e^{Nx_p^{-1}}} \right)^{-1/N-p}. \end{aligned}$$

Passing the last expression to the limit as  $N \rightarrow \infty$  provides an ansatz for the true conditional distribution function

$$\tilde{G}_{Y|X}(y | \mathbf{x}) = \begin{cases} 1, & y \geq \min(x_1, \dots, x_p) \\ 0, & y < \min(x_1, \dots, x_p) \end{cases}$$

with  $x_1, \dots, x_p, y > 0$ , from where (12) follows.

### Appendix C. Exact and limiting regression manifolds for logistic model

Here we give details on how the exact (15) and approximated (14) regression manifolds for the logistic model can be derived. The derivations below require the use of Lampert  $W$  function (Borwein and Lindstrom 2016) on which some properties and details can be found in the supplementary material.

**C.1. Exact regression manifold** Here we compute the conditional quantiles for bivariate extreme value distribution and their linear approximation for large  $x$ . Using (10), we calculate the conditional distribution function for the logistic model; for  $(x, y) \in (0, \infty)^2$ , it follows that

$$\begin{aligned} G_{Y|X}(y | x) &= G(x, y)(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1} x^{-1/\alpha-1} x^2 \exp(x^{-1}) \\ &= \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha + x^{-1}\} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1} x^{1-1/\alpha}. \end{aligned}$$

The conditional quantiles behave differently depending on the strength of dependence between extremes. The special case of the logistic model is for  $\alpha = 1$ , corresponding to independence between extremes for which the family of regression lines are known to be given by (11). We now derive conditional quantiles for bivariate dependent extremes, i.e. when  $\alpha \in [0, 1)$ . Since  $y \mapsto G_{Y|X}$  is continuous, a conditional quantile is a solution to

$$\exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha-1} x^{1-1/\alpha} \exp(x^{-1}) = q, \tag{22}$$

which can be written in terms of the Lambert  $W$  function. Rewriting (22) as

$$\begin{aligned} &\exp\left\{-\frac{\alpha}{\alpha-1}(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\right\} (x^{-1/\alpha} + y^{-1/\alpha})^{(\alpha-1)\alpha/(\alpha-1)} x^{(\alpha-1)/\alpha/(\alpha-1)} \\ &\times \exp\left\{\frac{\alpha}{\alpha-1}x^{-1}\right\} = q^{\alpha/(\alpha-1)} \\ \Leftrightarrow \exp\left\{\frac{\alpha}{1-\alpha}(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\right\} \frac{\alpha}{1-\alpha}(x^{-1/\alpha} + y^{-1/\alpha})^\alpha &= \frac{\alpha}{1-\alpha}x^{-1} \exp\left\{\frac{\alpha}{1-\alpha}x^{-1}\right\} q^{\alpha/(\alpha-1)} \\ \Leftrightarrow \frac{\alpha}{1-\alpha}(x^{-1/\alpha} + y^{-1/\alpha})^\alpha &= W\left(\frac{\alpha}{1-\alpha}x^{-1} e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)}\right), \end{aligned}$$

gives

$$y_{q|x} = \left[ \left\{ \frac{1-\alpha}{\alpha} x W\left(\frac{\alpha}{1-\alpha} x^{-1} e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)}\right) \right\}^{1/\alpha} - 1 \right]^{-\alpha} x, \quad x > 0. \tag{23}$$

From properties of the Lambert  $W$  function (see supplementary material for details) it follows that

$$\lim_{x \rightarrow \infty} x W\left(\frac{\alpha}{1-\alpha} x^{-1} e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)}\right) = \frac{\alpha}{1-\alpha} q^{\alpha/(\alpha-1)},$$

and that the conditional quantiles tend to infinity as  $x \rightarrow \infty$ .

**C.2. Limiting regression manifold** Below we show that (14) holds. To find  $\gamma_q$  and  $\beta_q$  we use the expansion of the principal branch of the Lambert  $W$  function, a solution to  $z = we^w$  when  $z > 0$ , around 0, that is

$$W(z) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} z^n = z + O(z^2),$$

where  $O(z)$  is big- $O$  of  $z$ . Substituting in (23) the expansion of  $W$  leads to

$$\begin{aligned} y_{q|x} &= \left[ \left\{ \frac{1-\alpha}{\alpha} x \left( \frac{\alpha}{1-\alpha} x^{-1} e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)} + O(x^{-2} e^{2\alpha/(1-\alpha)x^{-1}}) \right) \right\}^{1/\alpha} - 1 \right]^{-\alpha} x \\ &= \left[ \left\{ e^{\alpha/(1-\alpha)x^{-1}} q^{\alpha/(\alpha-1)} + O(x^{-1} e^{2\alpha/(1-\alpha)x^{-1}}) \right\}^{1/\alpha} - 1 \right]^{-\alpha} x, \end{aligned}$$

and taking  $x \rightarrow \infty$  we find the linear asymptote of  $x \mapsto y_{q|x}$  which is based on

$$\beta_q = \lim_{x \rightarrow \infty} \frac{y_{q|x}}{x} = \{q^{-1/(1-\alpha)} - 1\}^{-\alpha}, \tag{24}$$

and, the more involved calculation of  $\gamma_q$ . The derivative of the asymptotic expansion of a function does not necessarily correspond to the asymptotic expansion of the derivative of the function; hence, we use L'Hospital with (23) and only then inject the asymptotic expansions. We have then after some tedious derivations that

$$\gamma_q = \lim_{x \rightarrow \infty} (y_{q|x} - \beta_q x) = \frac{\alpha}{1-\alpha} \{q^{1/(\alpha-1)} - 1\}^{-\alpha-1} \{q^{\alpha/(1-\alpha)} - 1\} q^{1/(\alpha-1)}. \tag{25}$$

The resulting linear approximation is as follows, for  $q \in (0, 1)$  and  $x \gg 1$ :

$$\tilde{y}_{q|x} = \frac{\alpha}{1-\alpha} \{q^{1/(\alpha-1)} - 1\}^{-\alpha-1} \{q^{\alpha/(1-\alpha)} - 1\} q^{1/(\alpha-1)} + \{q^{-1/(1-\alpha)} - 1\}^{-\alpha} x = \alpha_q + \beta_q x.$$

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10687-022-00446-6>.

**Acknowledgements** The authors thank the Editor, Associate Editor, and two Referees for their constructive remarks and fruitful recommendations. In addition, the authors thank, without implicating, Johan Segers (Université catholique de Louvain) and Raphaël Huser (King Abdullah University of Science and Technology) for insightful discussions, suggestions, and comments. M. de Carvalho acknowledges support from the *Fundação para a Ciência e a Tecnologia* (Portuguese NSF) through the projects PTDC/MAT-STA/28649/2017 and UID/MAT/00006/2020. G. dos Reis acknowledges support from the *Fundação para a Ciência e a Tecnologia* (Portuguese NSF) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações CMA/FCT/UNL). A. Kumukova was supported by The Maxwell

Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot–Watt University, and the University of Edinburgh.

**Data availability** The datasets analysed during the current study are available from the R package DATAstudio (<https://cran.r-project.org/web/packages/DATAstudio/>).

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Altomare, F., Campiti, M.: Korovkin-Type Approximation Theory and Its Applications. de Gruyter, Berlin (1994)
- Apostol, T.M.: Calculus, vol. II. Wiley, New York (1980)
- Barrientos, A.F., Jara, A., Quintana, F.A.: Bayesian density estimation for compositional data using random Bernstein polynomials. *J. Stat. Plann. Infer.* **166**, 116–125 (2015)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of extremes: Theory and applications. Wiley, Hoboken, NJ (2004)
- Borwein, J.M., Lindstrom, S.B.: Meetings with Lambert  $\mathcal{W}$  and other special functions in optimization and analysis. *Pure Appl. Funct. Anal.* **1**(3), 361–396 (2016)
- Chernozhukov, V.: Extremal quantile regression. *Ann. Stat.* **33**(2), 806–839 (2005)
- Coles, S.: An introduction to statistical modeling of extreme values. Springer, London (2001)
- Cook, J.D.: Basic properties of the soft maximum. *bepress*, Working Paper Series 70 (2011). <https://biostats.bepress.com/cgi/viewcontent.cgi?article=1073&context=mdandersonbiostat>
- Cooley, D., Davis, R.A., Naveau, P.: Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. *Ann. Appl. Stat.* **6**(4), 1406–1429 (2012)
- Davison, A.C., Huser, R.: Statistics of extremes. *Annual Rev. Stat. Appl.* **2**, 203–235 (2015)
- de Haan, L., Ferreira, A.: Extreme value theory: an introduction. Springer, New York (2006)
- de Haan, L., Resnick, S.I.: Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**(4), 317–337 (1977)
- Dombry, C., Engelke, S., Oesting, M.: Bayesian inference for multivariate extreme value distributions. *Electronic J. Stat.* **11**(2), 4813–4844 (2017)
- Dunn, P.K., Smyth, G.K.: Randomized quantile residuals. *J. Comp. Graph. Stat.* **5**(3), 236–244 (1996)
- Eastoe, E.F., Tawn, J.A.: Modelling non-stationary extremes with application to surface level ozone. *J. Royal Stat. Soc. Ser. C* **58**(1), 25–45 (2009). <https://doi.org/10.1111/j.1467-9876.2008.00638.x>
- Einmahl, J.H., Yang, F., Zhou, C.: Testing the multivariate regular variation model. *J. Bus. Econ. Stat.* **39**(4), 907–919 (2021)
- Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events for insurance and finance. Springer, New York (1997)

- Galasso, B., Zemel, Y., de Carvalho, M.: Bayesian semiparametric modelling of phase-varying point processes. *Electronic J. Stat.* **16**(1), 2518–2549 (2022)
- Guan, Z.: Efficient and robust density estimation using Bernstein type polynomials. *J. Nonparam. Stat.* **28**(2), 250–271 (2016)
- Gudendorf, G., Segers, J.: Extreme-value copulas. In: *Copula theory and its applications*, pp. 127–145. Springer, Berlin, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-12465-5\\_6](https://doi.org/10.1007/978-3-642-12465-5_6)
- Guillem, A.I.G.: Structure de dépendance des lois de valeurs extrêmes bivariées. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics* **330**(7), 593–596 (2000)
- Haario, H., Saksman, E., Tamminen, J.: Componentwise adaptation for high dimensional MCMC. *Comput. Stat.* **20**(2), 265–273 (2005)
- Hanson, T.E., de Carvalho, M., Chen, Y.: Bernstein polynomial angular densities of multivariate extreme value distributions. *Stat. Prob. Lett.* **128**, 60–66 (2017)
- Hüsler, J., Li, D.: Testing asymptotic independence in bivariate extremes. *J. Stat. Plann. Infer.* **139**(3), 990–998 (2009)
- Katz, R.W.: Statistical methods for nonstationary extremes. In: *Extremes in a changing climate*, pp. 15–37. Springer (2013)
- Kiriliouk, A., Rootzén, H., Segers, J., Wadsworth, J.L.: Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics* **61**(1), 123–135 (2019)
- Koenker, R., Bassett, G.: Regression quantiles. *Econometrica* **46**, 33–50 (1978)
- Mhalla, L., Chavez-Demoulin, V., Dupuis, D.J.: Causal mechanism of extreme river discharges in the upper danube basin network. *J. Royal Stat. Soc. Ser. C* **69**(4), 741–764 (2020)
- Nelsen, R.B.: *An introduction to copulas*, 2nd ed. Springer, New York (2006)
- Padoan, S.A., Rizzelli, S.: Consistency of Bayesian inference for multivariate max-stable distributions. *Ann. Stat.* (2022)
- Pickands, J.: Multivariate extreme value distributions. In: *Proc. 43th Sess. Int. Statist. Inst.*, pp. 859–878 (1981)
- Ramos, A., Ledford, A.: A new class of models for bivariate joint tails. *J. Royal Stat. Soc. Ser. B* **71**(1), 219–241 (2009). <https://doi.org/10.2307/20203885>
- Resnick, S.I.: *Heavy-tail phenomena: Probabilistic and statistical modeling*. Springer, New York (2007)
- Sabourin, A., Naveau, P.: Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Comput. Stat. Data Anal.* **71**, 542–567 (2014)
- Stephenson, A., Tawn, J.: Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92**(1), 213–227 (2005)
- Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B* **58**(1), 267–288 (1996)
- van der Vaart, A.W.: *Asymptotic statistics*. Cambridge University Press, Cambridge, UK (1998)
- Wang, H., Tsai, C.L.: Tail index regression. *J. Am. Stat. Assoc.* **104**(487), 1233–1240 (2009)
- Wang, Y., Stoev, S.A.: Conditional sampling for spectrally discrete max-stable random fields. *Adv. Appl. Prob.* **43**(2), 461–483 (2011)
- Yee, T.W., Stephenson, A.G.: Vector generalized linear and additive extreme value models. *Extremes* **10**(1–2), 1–19 (2007)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.