

# Parsimonious Factor Models for Asymmetric Dependence in Multivariate Extremes

Pavel Krupskii\* and Boris Béranger†

December 5, 2025

## Abstract

Modelling multivariate extreme events is essential when extrapolating beyond the range of observed data. Parametric models that are suitable for real-world extremes must be flexible—particularly in their ability to capture asymmetric dependence structures—while also remaining parsimonious for interpretability and computationally scalable in high dimensions. Although many models have been proposed, it is rare for any single construction to satisfy all of these requirements. For instance, the popular Hüsler-Reiss model ([Hüsler and Reiss, 1989](#)) is limited to symmetric dependence structures. In this manuscript, we introduce a class of additive factor models and derive their extreme-value limits. This leads to a broad and tractable family of models characterised by a manageable number of parameters. These models naturally accommodate asymmetric tail dependence and allow for non-stationary behaviour. We present the limiting models from both the componentwise-maxima and Peaks-over-Thresholds perspectives, via the multivariate extreme value and multivariate generalized Pareto distributions, respectively. Simulation studies illustrate identifiability properties based on existing inference methodologies. Finally, applications to summer temperature maxima in Melbourne, Australia, and to weekly negative returns from four major UK banks demonstrate improved fit compared with the Hüsler-Reiss model.

## 1 Introduction

The modelling of multivariate extreme events is a critical task in many disciplines — including finance, insurance, hydrology, climate science, and engineering — where a sound understanding

---

\*School of Mathematics and Statistics, University of Melbourne. [pavel.krupskiy@unimelb.edu.au](mailto:pavel.krupskiy@unimelb.edu.au)

†School of Mathematics and Statistics, UNSW Sydney. [b.beranger@unsw.edu.au](mailto:b.beranger@unsw.edu.au)

of the probability of such events is highly consequential. Extreme Value Theory ([de Haan and Ferreira, 2006](#); [Resnick, 1987](#)) provides a robust theoretical framework for addressing this class of problems. The most common approach is to specify a parametric model and use it to extrapolate beyond the range of the observed data in order to quantify the risk of extreme events.

While the theory for univariate random variables is well established, numerous challenges remain when extending these methods to two or more variables. The extremeness of a random vector can be defined via the componentwise maxima — commonly referred to as the block maxima approach — or, alternatively, by considering any joint observation that exceeds a high threshold, known as the multivariate Peaks-over-Threshold approach. Both approaches are introduced in detail in [Section 2](#), where their connections are also highlighted.

Since the marginal distributions can be readily modelled using univariate extreme value theory, we assume throughout this manuscript that the margins have been standardised, allowing us to concentrate solely on the extremal dependence structure. Although several mechanisms exist for constructing parametric models of extremal dependence, much of the literature focuses on the logistic ([Gumbel, 1960](#)) and Hüsler–Reiss ([Hüsler and Reiss, 1989](#)) models. The logistic model employs a single dependence parameter, making it too rigid for many real-world applications, while the Hüsler–Reiss model cannot capture asymmetric dependence structures—in particular, situations where some components are more strongly dependent than others, as commonly observed in rainfall, temperature, and financial data. The extremal- $t$  ([Opitz, 2013](#)) model uses elliptical distributions to extend the Hüsler–Reiss model but possesses the same drawbacks. [Padoan \(2011\)](#) and [Beranger et al. \(2019\)](#) proposed asymmetric extensions of the Hüsler–Reiss and extremal- $t$  models but do require some sum-to-zero conditions on the skewness parameters. Non-exhaustive overviews

of existing models can be found in [Beirlant et al. \(2004\)](#); [Dey and Yan \(2016\)](#); [Naveau and Segers \(2024\)](#). The goal of this manuscript is to develop a new class of models capable of representing asymmetric extremal dependence structures.

Factor copula models ([Krupskii and Joe, 2013](#)) provide a flexible framework for modelling data that exhibit tail dependence — situations in which two or more variables can simultaneously take extreme values — as well as tail asymmetry. To ensure computational scalability in high dimensions, structural assumptions on how the latent factors relate to the observed variables are often imposed, leading to more parsimonious model formulations ([Krupskii and Joe, 2015](#)). When applied to finite-dimensional distributions, these constructions can also be used to model spatial data.

[Krupskii et al. \(2018a\)](#) introduced a common factor copula model tailored for spatial applications that is interpretable and scales well to high dimensions; however, it does not accommodate permutation asymmetry, a feature frequently observed in practice. [Lee and Joe \(2018\)](#) proposed factor copulas for multivariate extremes under a tree structure, and [Krupskii and Genton \(2021\)](#) further extended this framework to conditional normal models. Yet, it remains unclear whether these approaches can adequately represent substantial permutation asymmetry. Moreover, they rely on numerical integration for likelihood evaluation, which becomes computationally demanding in high dimensions. More recently, [Kiriliouk et al. \(2024\)](#) adapted vine constructions to create highly flexible models for multivariate extremes. Despite their versatility, these models lack parsimony, and inference becomes prohibitive in very high-dimensional settings.

In this manuscript, we develop a broad class of additive factor copula models that extends the construction of [Krupskii et al. \(2018a\)](#). We derive the corresponding limiting distributions and show that they permit permutation asymmetry between any pair of variables. We also identify par-

simonious sub-models that are particularly well suited for capturing asymmetric tail dependence. The corresponding multivariate extreme value and multivariate generalized Pareto distributions are derived such that the modelling of block maxima and threshold exceedances is possible. The proposed class offers different types of extremal dependence structure for the variable of interests through models that can assume zero or non-zero mass at one or more of the boundaries of the support, meaning that joint extremes of a vector of interest may not always occur simultaneously (see e.g. [Goix et al., 2017](#); [Simpson et al., 2020](#)). Furthermore, given that common factor ( $V_0$ , see [Section 3](#)) affects all the variables and thus introduces extremal dependence in a small area under the influence of this factor, strengthens the interpretability of the proposed models.

The remainder of the paper is organised as follows. [Section 2](#) reviews essential background on multivariate extremes from the perspectives of Block Maxima and Peaks-over-Thresholds, and introduces the Hüsler–Reiss model, which serves as our primary benchmark. [Section 3](#) presents the main contributions of the paper: we define a class of additive factor models, derive their limiting distributions, and investigate their tail properties. [Section 3](#) summarises inference methods for both componentwise maxima and threshold exceedances, and includes simulation studies demonstrating that accurate parameter estimation is achievable even in higher dimensions. [Section 5](#) provides two empirical analyses that highlight the increased flexibility of the proposed models compared with the Hüsler–Reiss model. Finally, [Section 6](#) offers concluding remarks. All technical proofs are deferred to the [Appendix A](#).

## 2 Background

Let  $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d})^\top, i \in \{1, \dots, n\}$  be  $n$  independent and identically distributed (i.i.d.) replicates of a random vector  $\mathbf{X} \in \mathbb{R}^d$  with continuous cumulative distribution function (cdf)  $F$  (in short  $\mathbf{X} \sim F$ ) with margins  $F_1, \dots, F_d$ . By [Sklar \(1959\)](#)'s representation theorem,  $F$  can be represented as  $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ , where  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$  and  $C : [0, 1]^d \rightarrow [0, 1]$  is a copula. By continuity of the marginal distribution, the above can be re-written as

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \quad \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d,$$

highlighting that a copula is a restriction of a multivariate distribution with standard uniform margins to the unit hypercube.

### 2.1 Extremes as Block Maxima

Let  $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d})^\top$  denote the vector component-wise maxima, where  $M_{n,j} = \max_{i=1, \dots, n} X_{i,j}$ . Assume there exist sequences of normalizing constants  $\mathbf{a}_n = (a_{n,1}, \dots, a_{n,d}) \in \mathbb{R}_+^d$  and  $\mathbf{b}_n = (b_{n,1}, \dots, b_{n,d}) \in \mathbb{R}^d$  such that, as  $n \rightarrow \infty$ , the location-scale standardized vector of maxima,  $\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$  converges in distribution to a random vector  $\mathbf{Y}$  with joint distribution  $G$  and non-degenerate margins,  $G_j, j = 1, \dots, d$ . Let  $\boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\sigma} \in \mathbb{R}_+^d \setminus \{\mathbf{0}\}$  and  $\boldsymbol{\xi} \in \mathbb{R}^d$  denote the vectors of marginal location, scale and shape parameters. We then say that the random vector  $\mathbf{X}$  lies in the maximum domain of attraction (MDA) of  $\mathbf{Y}$ . The limiting distribution  $G$  is a multivariate extreme value distribution whose margins belong to the family of generalized extreme value (GEV) distributions (see, e.g., [de Haan and Ferreira, 2006](#), Chap. 6). Consequently, the marginal distributions of  $\mathbf{X}$  are in the MDA of the GEV family, which encompasses the Weibull, Gumbel, and Fréchet

classes. The limiting distribution of the rescaled component-wise maxima can be expressed as

$$G(\mathbf{x}) = C_{\text{EV}}(G_1(x_1), \dots, G_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d \quad (1)$$

where  $C_{\text{EV}}$  is the limiting copula defined as

$$C_{\text{EV}}(\mathbf{u}) = \lim_{n \rightarrow \infty} C(u_1^{1/n}, \dots, u_d^{1/n})^n, \quad \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \quad (2)$$

and called an extreme value (EV) copula (Beirlant et al., 2004, Chap. 8). The copula  $C$  is said to be in the domain of attraction of  $C_{\text{EV}}$ . Extreme value copulas possess the max-stability property meaning that for any integer  $k \geq 1$ ,  $C_{\text{EV}}(u_1^{1/k}, \dots, u_d^{1/k})^k = C_{\text{EV}}(u_1, \dots, u_d)$  (see e.g., Gudendorf and Segers, 2010). When studying multivariate extremes from the perspective of copulas, it is convenient to characterise the EV copula through the stable tail dependence function  $\ell : \mathbb{R}_+^d \rightarrow \mathbb{R}_+$ , defined as

$$\lim_{n \rightarrow \infty} n \left\{ 1 - C \left( 1 - \frac{x_1}{n}, \dots, 1 - \frac{x_d}{n} \right) \right\} = -\log C_{\text{EV}}(e^{-x_1}, \dots, e^{-x_d}) \equiv \ell(\mathbf{x}), \quad (3)$$

where  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}_+^d$  (Drees and Huang, 1998; Huang, 1992). It can be seen that (3) is equivalent to (2) after taking logarithms and applying a linear expansion. This yields the representation

$$G(\mathbf{x}) = \exp \{ -\ell(-\log G_1(x_1), \dots, -\log G_d(x_d)) \}. \quad (4)$$

The tail dependence function is homogeneous and convex, and satisfies

$$\max(\mathbf{x}) \leq \ell(\mathbf{x}) \leq \sum_{i=1}^d x_i,$$

with the lower and upper bounds corresponding respectively to complete dependence (i.e.,  $G(\mathbf{x}) = \min\{G_1(x_1), \dots, G_d(x_d)\}$ ) and asymptotic independence (i.e.,  $G(\mathbf{x}) = \prod_{i=1}^d G_i(x_i)$ ). At the copula

level, this is equivalent to

$$\prod_{i=1}^d u_i \leq C_{\text{EV}}(\mathbf{u}) \leq \max(u_1, \dots, u_d).$$

By homogeneity, it is sufficient to restrict attention to the unit simplex

$$\mathcal{S} = \{\mathbf{w} \in [0, 1]^d : \|\mathbf{w}\|_1 = 1\},$$

on which the Pickands dependence function  $A : \mathcal{S} \rightarrow [1/d, 1]$  is defined. The tail dependence function then admits the representation

$$\ell(\mathbf{x}) = \sum_{i=1}^d x_i A(\mathbf{w}), \quad w_i = \frac{x_i}{\|\mathbf{x}\|_1}, \quad \mathbf{x} \in \mathbb{R}_+^d \setminus \{\mathbf{0}\}.$$

By properties of the tail dependence function  $\ell$ , the Pickands function  $A$  is also homogeneous and convex, and satisfies

$$\max(w_1, \dots, w_d) \leq A(\mathbf{w}) \leq 1, \quad \mathbf{w} \in \mathcal{S}.$$

Extremal dependence can also be quantified through the coefficient of upper tail dependence (Joe, 1997; Li, 2009), defined by

$$\chi_U = \lim_{u \rightarrow 0} \Pr \{X_i > F_i^{-1}(1 - u), \forall i \neq j | X_j > F_j^{-1}(1 - u)\} = \lim_{u \rightarrow 0} \frac{\bar{C}(1 - u, \dots, 1 - u)}{u},$$

where  $\bar{C}$  denotes the survival copula. This quantity links directly to the tail dependence and Pickands functions via

$$\chi_U = d - \ell(1, \dots, 1) = d \{1 - A(1/d, \dots, 1/d)\}.$$

Independence in the tails corresponds to  $\chi_U = 0$ , whereas complete dependence occurs when  $\chi_U = d - 1$ . The quantity  $\vartheta = \ell(1, \dots, 1) = dA(1/d, \dots, 1/d)$  is known as the extremal coefficient, and ranges between 1 (complete dependence) and  $d$  (asymptotic independence).

## 2.2 Extremes as Peaks-over-Thresholds

Another approach to modelling multivariate extremes characterises extreme events through threshold exceedances rather than through componentwise maxima. More precisely, assume the convergence in distribution of  $\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$  to  $\mathbf{Y}$  as introduced above. Using the notation  $\mathbf{X} \not\leq \mathbf{b}_n$  to indicate that  $\mathbf{X} \leq \mathbf{b}_n$  does not hold — i.e., there exists at least one  $j \in \mathcal{I} = \{1, \dots, d\}$  such that  $X_j > b_{n,j}$  — we have

$$\frac{\mathbf{X} - \mathbf{b}_n}{\mathbf{a}_n} \vee \boldsymbol{\eta} \Big| \mathbf{X} \not\leq \mathbf{b}_n \xrightarrow{d} \tilde{\mathbf{Y}}, \quad \text{as } n \rightarrow \infty,$$

where  $\tilde{\mathbf{Y}}$  follows a multivariate Generalized Pareto distribution,  $\vee$  denotes the componentwise maximum operator, and  $\boldsymbol{\eta}$  is the vector of lower endpoints of the marginal distributions  $G_1, \dots, G_d$  (Rootzén et al., 2018a). Let  $H$  denote the distribution function of the multivariate Generalized Pareto distribution associated with  $G$ , defined through

$$H(\mathbf{x}) = \frac{1}{\log G(\mathbf{0})} \log \left\{ \frac{G(\mathbf{x} \wedge \mathbf{0})}{G(\mathbf{x})} \right\}, \quad 0 < G(\mathbf{0}) < 1,$$

for  $\mathbf{x} > \boldsymbol{\eta}$ ; otherwise  $H(\mathbf{x}) = 0$  whenever  $x_i < \eta_i$  for at least one  $i \in \mathcal{I}$  (Rootzén and Tajvidi, 2006).

Here,  $\wedge$  denotes the componentwise minimum operator. Using the representation (4), we obtain

$$H(\mathbf{x}) = \ell \left[ \boldsymbol{\pi} \left\{ 1 + \frac{\boldsymbol{\xi}}{\tilde{\boldsymbol{\sigma}}} (\mathbf{x} \wedge \mathbf{0}) \right\}^{-1/\boldsymbol{\xi}} \right] - \ell \left\{ \boldsymbol{\pi} \left( 1 + \frac{\boldsymbol{\xi}}{\tilde{\boldsymbol{\sigma}}} \mathbf{x} \right)^{-1/\boldsymbol{\xi}} \right\},$$

where  $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} - \boldsymbol{\xi}\boldsymbol{\mu}$ ,  $\boldsymbol{\pi} = \boldsymbol{\tau}/\ell(\boldsymbol{\tau})$ , and  $\tau_j = (1 + \xi_j \mu_j / \sigma_j)^{-1/\xi_j}$ ,  $j \in \mathcal{I}$  (Prop. 2 Rootzén et al., 2018b).

When  $\boldsymbol{\xi} = \mathbf{0}$ , this representation simplifies to

$$H(\mathbf{x}) = \ell \left\{ \boldsymbol{\pi} \exp \left( -\frac{\mathbf{x} \wedge \mathbf{0}}{\boldsymbol{\sigma}} \right) \right\} - \ell \left\{ \boldsymbol{\pi} \exp \left( -\frac{\mathbf{x}}{\boldsymbol{\sigma}} \right) \right\},$$

with  $\tau_j = \exp\{\mu_j/\sigma_j\}$ . By construction, each  $j$ -th conditional marginal  $X_j - u_j | X_j > u_j$  follows a univariate Generalized Pareto distribution with parameters  $\tilde{\sigma}_j, \xi_j$ . A transformation to unit Pareto

margins then yields  $H(\mathbf{x}) = \{\ell(\mathbf{x}^{-1} \wedge \mathbf{1}) - \ell(\mathbf{x}^{-1})\} \ell(\mathbf{1})^{-1}$ . These results imply that, for a sufficiently high threshold  $\mathbf{u}$ , the distribution of threshold exceedances  $\mathbf{X} - \mathbf{u} | \mathbf{X} \not\leq \mathbf{u}$  can be approximated by a multivariate Generalized Pareto distribution, which is fully characterised by the stable tail dependence function  $\ell$ . Since the support of  $H$  is contained in the set of all  $\mathbf{x}$  such that  $x_j > \eta_j$  for all  $j$  and  $x_j > 0$  for at least one  $j$ , the density of  $\tilde{\mathbf{Y}}$  is given by

$$h(\mathbf{x}) = -\frac{\partial^d}{\partial x_1 \cdots \partial x_d} \ell \left\{ \pi \left( 1 + \frac{\gamma}{\bar{\sigma}} \mathbf{x} \right)^{-1/\gamma} \right\}, \quad \mathbf{x} \in (\boldsymbol{\eta}, \infty).$$

The distribution  $H$  may also place mass on lower-dimensional faces of  $(\boldsymbol{\eta}, \infty)$ . For instance, the probability mass assigned to the hyperplane  $\{\mathbf{x} : x_j = \eta_j\}, j \in \mathcal{I}$  is

$$H_j(\{\eta_j\}) = \lim_{x_j \rightarrow \infty} \ell(\pi_1, \dots, \pi_{j-1}, \pi_j x_j, \pi_{j+1}, \dots, \pi_d) - \pi_j x_j = \sum_{k \in \mathcal{I}_{-j}} \pi_k \frac{\partial \ell}{\partial x_k}(\pi_j \mathbf{e}_j), \quad (5)$$

where  $\mathcal{I}_{-j} = \mathcal{I} \setminus \{j\}$  and  $\partial \ell / \partial x_k$  denotes the first-order derivative of  $\ell$  with respect to the  $k$ -th component and  $\mathbf{e}_j$  is the  $j$ -th  $d$ -dimensional canonical vector (see [Rootzén et al., 2018b](#), Proposition 17).

In the case of unit Pareto margins,  $\pi_j = 1/\ell(\mathbf{1})$  and (5) simplifies to

$$H_j(\{\eta_j\}) = \lim_{x_j \rightarrow \infty} \{\ell(1, \dots, 1, x_j, 1, \dots, 1) - x_j\} \ell(\mathbf{1})^{-1} = \lim_{v \rightarrow 0} \frac{\partial^{d-1}}{\partial \mathbf{x}_{-j}} \ell(v, \dots, v, 1, v, \dots, v) \ell(\mathbf{1})^{-1},$$

where the second equality follows from the homogeneity property of  $\ell$ . For a comprehensive treatment of the multivariate Peaks-over-Thresholds approach, the reader is referred to [Naveau and Segers \(2024\)](#).

## 2.3 The Hüsler-Reiss distribution, [Hüsler and Reiss \(1989\)](#)

Let  $\mathbf{X}$  be a zero-mean, unit-variance Gaussian random vector with correlations satisfying  $(1 - \rho_{ij}(n)) \log n \rightarrow \lambda_{ij}^2$  as  $n \rightarrow \infty$ . The limiting marginal distributions are

$$G_j(x) = \exp \{-\exp\{-x\}\}, \quad x \in \mathbb{R}, \quad j \in \mathcal{I}$$

and the corresponding EV copula is given by

$$C_{\text{EV}}(\mathbf{u}) = \exp \left( \sum_{j=1}^d \log u_j \Phi_{d-1} \left[ \left\{ \lambda_{ij} + \log \left( \frac{\log u_j}{\log u_i} \right) / (2\lambda_{ij}) \right\}_{i \in \mathcal{I}_{-j}} ; \bar{\Sigma}^{(j)} \right] \right), \quad \mathbf{u} \in [0, 1]^d,$$

where  $\Phi_d(\cdot; \Sigma)$  denotes the cumulative distribution function (cdf) of a  $d$ -dimensional normal distribution with covariance (or correlation) matrix  $\Sigma$  and  $\bar{\Sigma}^{(j)}$  is the partial correlation matrix with entries  $(\bar{\Sigma}^{(j)})_{i,k} = (\lambda_{k,j}^2 + \lambda_{i,j}^2 - \lambda_{k,i}^2) / (2\lambda_{k,j}\lambda_{i,j})$ ,  $i \neq k$ . The stable tail dependence function is then

$$\ell(\mathbf{x}) = \sum_{j=1}^d x_j \Phi_{d-1} \left[ \left\{ \lambda_{ij} + \log(x_j/x_i) / (2\lambda_{ij}) \right\}_{i \in \mathcal{I}_{-j}} ; \bar{\Sigma}^{(j)} \right], \quad \mathbf{x} \in \mathbb{R}_+^d,$$

from which the multivariate extreme value distribution  $G(\mathbf{x})$  and the extremal coefficient  $\vartheta = \sum_{j=1}^d \Phi_{d-1} \left( (\lambda_{ij})_{i \in \mathcal{I}_{-j}} ; \bar{\Sigma}^{(j)} \right)$  are obtained. In particular,  $\lambda_{i,j} \rightarrow 0$  for all  $i, j \in \mathcal{I}, i \neq j$  implies  $\vartheta \rightarrow 1$  (complete dependence), whereas  $\lambda_{ij} \rightarrow \infty$  implies  $\vartheta \rightarrow d$  (independence).

Assuming unit Pareto marginals, the density of the multivariate Generalized Pareto distribution is given by (Engelke et al., 2015)

$$h(\mathbf{x}) = \phi_{d-1} \left[ \left\{ \lambda_{ik} + \log(x_i/x_k) / (2\lambda_{ik}) \right\}_{i \in \mathcal{I}_{-k}} ; \bar{\Sigma}^{(k)} \right] \left\{ \ell(\mathbf{1}) x_k^2 \prod_{i \in \mathcal{I}_{-k}} x_i 2\lambda_{ik} \right\}^{-1},$$

where  $\phi_d(\cdot; \Sigma)$  denotes the probability density function (pdf) of a  $d$ -dimensional normal distribution with covariance (or correlation) matrix  $\Sigma$ . Interestingly, this expression does not depend on  $k$ .

In the bivariate case ( $d = 2$ ), it can be shown that

$$\dot{\ell}_k(x_1, x_2) = \Phi \left\{ \lambda + \log \left( \frac{x_k}{x_j} \right) / (2\lambda) \right\}, \quad j = \mathcal{I}_k.$$

Since the unit Pareto margins imply  $\eta_1 = \eta_2 = 0$ , from (5) we have  $H_1(\{0\}) = H_2(\{0\}) = 0$ , meaning that, for the Hüsler-Reiss distribution,  $H$  does not place mass on the boundaries of its support.

Figure 1 illustrates various aspects of the bivariate Hüsler–Reiss distribution. In particular, the top and bottom left panels show the density of the EV copula  $c_{EV}(\mathbf{u}) = \partial^d / \partial \mathbf{u} C_{EV}(\mathbf{u})$ , and the stable tail dependence function  $\ell(\mathbf{x})$  for  $\lambda = 1$ . For the same parameter setting, the top middle panel displays the density of the multivariate EV distribution  $g(\mathbf{x})$ , with Gumbel margins, with yellow dots representing 200 random draws from the component-wise maxima.

Considering 200,000 observations drawn from a bivariate normal distribution whose margins have been standardised to unit Pareto scale, and setting a threshold  $\mathbf{u}$  at the 0.999 marginal quantiles, the top right panel shows approximately 300 observations of  $\mathbf{X}/\mathbf{u}$  with at least one component above 1 (yellow dots), together with the log-density of the multivariate Generalized Pareto distribution,  $\log H(x_1, x_2)$ .

Graphical summaries of the dependence are provided in the bottom middle and right panels of Figure 1, showing the extremal coefficient  $\vartheta$  as a function of  $\lambda$  and the Pickands dependence function for  $\lambda = 0.5, 1$  and  $2$ . These plots clearly illustrate that the level of dependence decreases as  $\lambda$  increases. Note that the extremal coefficient, Pickands dependence function, and extreme value density  $g$  were computed using the R package (Beranger et al., 2025; Beranger and Padoan, 2025).

### 3 A broad class of additive factor models for modelling extremes

Consider the  $d$ -dimensional random vectors  $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$  and  $\mathbf{Z}^* = (Z_1^*, \dots, Z_d^*)^\top$  defined such that  $\{\mathbf{Z}^\top, (\mathbf{Z}^*)^\top\}^\top$  follows a multivariate normal distribution with zero mean, unit variance, and correlation matrix  $\mathbf{P}$ . We construct the random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  by defining each

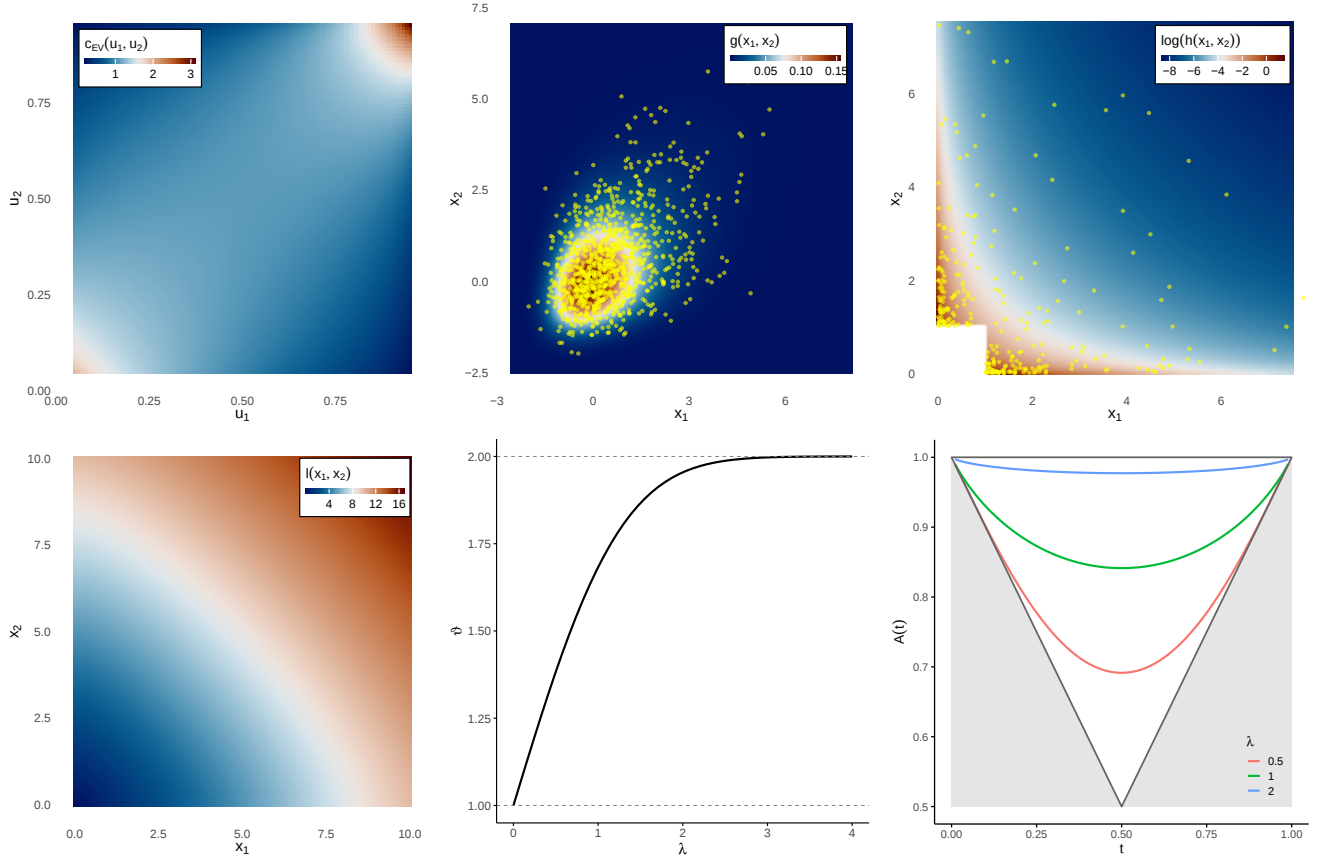


Figure 1: Left to right, top to bottom: density of the extreme value copula  $c_{EV}$ , density of the bivariate extreme value distribution  $g(\mathbf{x})$ , density of the bivariate Generalized Pareto distribution with unit Pareto margins (on log scale,  $\log h(\mathbf{x})$ ) and stable tail dependence function  $\ell(\mathbf{x})$  of the Hüsler-Reiss model with  $\lambda = 1$ . The remaining panels display the extremal coefficient  $\vartheta$  as function of  $\lambda$  and the Pickands dependence function for  $\lambda = 0.5, 1$  and  $2$ .

marginal random variable  $X_i$  through the following additive factor models

$$X_i = Z_i + (1/\alpha_i) \cdot V_0 \cdot \mathbf{1}(Z_i^* > c_i), \quad i \in \mathcal{I}, \quad (6)$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^\top \in \mathbb{R}_+^d$ , the common factor  $V_0$  is introduced to model upper tail dependence and  $\mathbf{c} = (c_1, \dots, c_d)^\top$  is a vector of real-valued truncation parameters. If  $\mathbf{c} = -\infty$  (no truncation), then the model (6) simplifies to the factor copula model for spatial data (Krupskii et al., 2018a). For such model, considering an exponential factor  $V_0$  yields the Hüsler-Reiss distribution for which the

non-trivial EV limiting distribution can be obtained. However, the associated bivariate marginal distributions are permutation symmetric which can be a restrictive assumption in many applications. We now study the EV limiting distributions for the model (6), and show that this model can generate more flexible dependence structures including permutation asymmetry.

### 3.1 Extreme-value limiting distributions

In this section, we consider the additive factor model defined in (6) and examine its limiting distributions under various assumptions for the common factor  $V_0$  and under different dependence or independence regimes between  $\mathbf{Z}$  and  $\mathbf{Z}^*$ .

**Proposition 1** *Consider the random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  defined in (6). If the common factor  $V_0$  follows a continuous distribution with cdf  $F_0$  such that for any constants  $M \in \mathbb{R}$ ,  $k > 0$  and some  $\alpha_0 \in (0, 0.5)$ ,  $\gamma < 0$  we have*

$$F_0 \{F_0^{-1}(1 - k/n) + M(\ln n)^{\alpha_0}\} = 1 - k/n + o(1/n), \quad n \rightarrow \infty,$$

and

$$F_0 \{F_0^{-1}(1 - k/n) + y\} \geq 1 - (k/n) \cdot \exp(\gamma y), \quad \text{if} \quad -|M| \ln n < y < -|M|(\ln n)^{\alpha_0}, \quad n \rightarrow \infty,$$

then the stable tail dependence function associated with  $\mathbf{X}$  is given by

$$\ell_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathcal{A} \subset \mathcal{I}, \mathcal{B} = \mathcal{I} - \mathcal{A}} \left( \max_{i \in \mathcal{A}} \frac{x_i}{\zeta_i} \right) \cdot \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}})$$

where we recall that  $\mathcal{I} - \mathcal{A} = \mathcal{I} \setminus \{\mathcal{A}\}$ , the indices  $\mathcal{A}$  and  $\mathcal{B}$  indicate the corresponding subvectors, and

$$\zeta_i = \Phi(-c_i), i \in \mathcal{I}.$$

**Proof:** See Appendix A.1. □

**Corollary 1** *The assumptions on the cdf of the common factor  $V_0$  stated above in Proposition 1 are fulfilled by the Pareto distribution with  $F_0(x) = 1 - x^{-\beta}$ ,  $\beta > 0$  and  $x > 1$ , and the Weibull distribution with  $F_0(x) = 1 - \exp(-x^\beta)$ ,  $\beta \in (0, 1)$  and  $x > 0$ .*

**Proof:** Let  $|M|(\ln n)^{\alpha_0} < |y| < |M| \ln n$ , where  $\alpha_0 \in (0, 0.5)$ . For the Pareto distribution,  $F_0^{-1}(1 - k/n) = (n/k)^{1/\beta}$  and

$$\begin{aligned} F_0 \{F_0^{-1}(1 - k/n) + y\} &= 1 - \frac{k}{n} \cdot \left\{ 1 + y \left( \frac{k}{n} \right)^{1/\beta} \right\}^{-\beta} = 1 - \frac{k}{n} + \beta y \cdot \left( \frac{k}{n} \right)^{1+1/\beta} + o(n^{-1-1/\beta}) \\ &= 1 - \frac{k}{n} + o(1/n). \end{aligned}$$

The second condition holds for any  $\gamma < 0$  provided that  $n$  is large enough:

$$1 - \frac{k}{n} \exp(\gamma y) \leq 1 - \frac{k}{n} - \frac{\gamma k}{n} y \leq 1 - \frac{k}{n} + \frac{k\gamma}{n}.$$

For the Weibull distribution,  $F_0(1 - k/n) = \{\ln(n/k)\}^{1/\beta}$  and

$$\begin{aligned} F_0 \{F_0^{-1}(1 - k/n) + y\} &= 1 - \exp \left[ \left( \ln \frac{k}{n} \right) \cdot \left\{ 1 + y \left( \ln \frac{n}{k} \right)^{-1/\beta} \right\}^\beta \right] \\ &= 1 - \frac{k}{n} \exp \left[ -\beta y \left( \ln \frac{n}{k} \right)^{1-1/\beta} + o \left\{ y \left( \ln \frac{n}{k} \right)^{1-1/\beta} \right\} \right]. \end{aligned}$$

If  $y = M(\ln n)^{\alpha_0}$ , then

$$y \left( \ln \frac{n}{k} \right)^{1-1/\beta} = o(1) \quad \text{and} \quad F_0 \{F_0^{-1}(1 - k/n) + y\} = 1 - \frac{k}{n} + o(1/n) \quad \text{as } n \rightarrow \infty,$$

provided that  $\alpha_0 < 1/\beta - 1$ . On the other hand, if  $-|M|(\ln n)^{\alpha_0} < y < -|M| \ln n$ , then

$$F_0 \{F_0^{-1}(1 - k/n) + y\} \geq 1 - \frac{k}{n} \exp(-\beta y) \geq 1 - \frac{k}{n} \exp(-\gamma y),$$

provided that  $\beta < \gamma$  and  $n$  is large enough. □

**Corollary 2** *Let  $F_i$  be the cdf of  $X_i$  as defined in (6), then*

- i) If  $V_0 \sim \text{Pareto}(\beta)$ ,  $\beta > 0$  then  $F_i$  belongs to the MDA of the Fréchet distribution with shape  $\beta$ .*
- ii) If  $V_0 \sim \text{Weibull}(\beta)$ ,  $\beta \in (0, 1)$ , then  $F_i$  belongs to the MDA of the Gumbel distribution.*

**Proof:** Consider  $X_{i1}, \dots, X_{in} \sim_{\text{i.i.d.}} F_i$  and from the proof of Proposition 1, recall that

$$\Pr \left\{ X_i < \alpha_i^{-1} F_0^{-1} \left( 1 - \frac{x_i}{n \zeta_i} \right) \right\} = 1 - \frac{x_i}{n} + o(1/n). \quad (7)$$

- i) With the Pareto cdf  $F_0(x) = 1 - x^{-\beta}$ , (7) can be rewritten as

$$\Pr(X_i < x) = 1 - \frac{\zeta_i}{(\alpha_i x)^\beta} + o(x^{-\beta}),$$

and setting  $a_n = n^{1/\beta} \zeta_i^{1/\beta} / \alpha_i$  yields

$$\Pr \left( \max_{j=1, \dots, n} X_{ij} / a_n < x \right) = \left( 1 - \frac{1}{x^\beta n} \right)^n + o(1).$$

Consequently  $\lim_{n \rightarrow \infty} \Pr(\max_{j=1, \dots, n} X_{ij} / a_n < x) = \exp(-x^{-\beta})$  and the distribution of  $X_i$  belongs to the MDA of the Fréchet distribution.

- ii) With the Weibull cdf  $F_0(x) = 1 - \exp(-x^\beta)$ , (7) can be rewritten as

$$\Pr(X_i < x) = 1 - \zeta_i \exp \{ -(\alpha_i x)^\beta \} + o \left[ \exp \{ -(\alpha_i x)^\beta \} \right],$$

and setting  $a_n = (\alpha_i \beta)^{-1} \{ \ln(\zeta_i n) \}^{1/\beta - 1}$  and  $b_n = \alpha_i^{-1} \{ \ln(\zeta_i n) \}^{1/\beta}$  yields

$$\begin{aligned} \Pr \left\{ \max_{j=1, \dots, n} (X_{ij} - b_n) / a_n < x \right\} &= \left\{ 1 - \left( \frac{1}{n} \right)^{(1 + x a_n / b_n)^\beta} \right\}^n + o(1) \\ &= \left( 1 - \frac{1}{n} \exp \left[ -x \cdot \frac{\ln n}{\ln(\zeta_i n)} + o \left\{ \frac{\ln n}{\ln(\zeta_i n)} \right\} \right] \right)^n + o(1). \end{aligned}$$

Consequently,  $\lim_{n \rightarrow \infty} \Pr \{ \max_{j=1, \dots, n} (X_{ij} - b_n) / a_n < x \} = \exp \{ -\exp(-x) \}$  and the distribution of  $X_i$  belongs to the MDA of the Gumbel distribution.  $\square$

The following result extends Proposition 1 to the case where  $V_0$  is exponentially distributed, under the assumption that  $\mathbf{Z}$  and  $\mathbf{Z}^*$  are independent.

**Proposition 2** *Consider the random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  defined in (6). If  $V_0$  follows a unit exponential distribution and the vectors  $\mathbf{Z}$  and  $\mathbf{Z}^*$  are independent, then the stable tail dependence function associated with  $\mathbf{X}$  is given by*

$$\ell_{\mathbf{X}}(\mathbf{x}) = \sum_{A \subset \mathcal{I}, B = \mathcal{I} - A} \ell_A(\tilde{\mathbf{x}}_A) \cdot \Pr(\mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B)$$

where  $\tilde{x}_i = x_i/\zeta_i$  and  $\zeta_i = \Phi(-c_i)$ ,  $i \in \mathcal{I}$ , and  $\ell_A$  is the stable tail dependence function of the lower-dimensional Hüsler-Reiss distribution which can be obtained as an EV limiting distribution of the vector  $\mathbf{X}_A = \mathbf{Z}_A + (1/\alpha_A) \cdot V_0$ ; see Krupskii et al. (2018b).

**Proof:** See Appendix A.2. □

A generalization of the above result can be obtained by assuming  $\mathbf{Z}$  and  $\mathbf{Z}^*$  are dependent, but at the cost of more involved formula for  $\ell_{\mathbf{X}}$ . Such scenario is considered in the following proposition where, for simplicity, the focus is restricted to the stable tail dependence function for the  $(i, j)$ -margin, denoted  $\ell_{X_i, X_j}$ , as it admits a simpler form.

**Proposition 3** *Consider the random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  defined in (6). If  $V_0$  follows a unit exponential distribution and the vectors  $\mathbf{Z}$  and  $\mathbf{Z}^*$  are dependent, then the stable tail dependence*

function associated with the  $(X_i, X_j), i, j \in \mathcal{I}, i \neq j$ , is given by

$$\ell_{X_i, X_j}(x_i, x_j) = x_i \left\{ 1 - \frac{\Phi_3 \left( \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \ln \frac{\tilde{x}_i}{\tilde{x}_j}, \alpha_i \rho_{j^*, i} - c_j, \alpha_i \rho_{i^*, i} - c_i \right)^\top; \Sigma_{i|j} \right)}{\Phi(\alpha_i \rho_{i^*, i} - c_i)} \right\} \\ + x_j \left\{ 1 - \frac{\Phi_3 \left( \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \ln \frac{\tilde{x}_j}{\tilde{x}_i}, \alpha_j \rho_{i^*, j} - c_i, \alpha_j \rho_{j^*, j} - c_j \right)^\top; \Sigma_{j|i} \right)}{\Phi(\alpha_j \rho_{j^*, j} - c_j)} \right\},$$

where  $\tilde{x}_l = x_l / \Phi(\alpha_l \rho_{l^*, l} - c_l), l \in \{i, j\}$ ,  $\lambda_{ij} = (\alpha_i^2 - 2\rho_{i,j}\alpha_i\alpha_j + \alpha_j^2)^{1/2}$ , and

$$\Sigma_{i|j} = \begin{pmatrix} 1 & \frac{1}{\lambda_{i,j}}(\alpha_j \rho_{j^*, j} - \alpha_i \rho_{j^*, i}) & \frac{1}{\lambda_{i,j}}(\alpha_j \rho_{i^*, j} - \alpha_i \rho_{i^*, i}) \\ \frac{1}{\lambda_{i,j}}(\alpha_j \rho_{j^*, j} - \alpha_i \rho_{j^*, i}) & 1 & \rho_{i^*, j^*} \\ \frac{1}{\lambda_{i,j}}(\alpha_j \rho_{i^*, j} - \alpha_i \rho_{i^*, i}) & \rho_{i^*, j^*} & 1 \end{pmatrix}.$$

Note that the above uses the notation  $\rho_{i,j} = \mathbf{P}_{i,j} = \text{Cor}(Z_i, Z_j)$ ,  $\rho_{j^*, i} = \mathbf{P}_{i, j+d} = \text{Cor}(Z_i, Z_j^*)$  and  $\rho_{i^*, j^*} = \mathbf{P}_{i+d, j+d} = \text{Cor}(Z_i^*, Z_j^*)$ .

**Proof:** See Appendix A.3. □

The limiting distribution characterized by the above tail dependence function is an extension of the Hüsler-Reiss distribution (see Section 2.3) that allows for asymmetric dependence structures and we therefore refer to it as the skew-Hüsler-Reiss distribution. It is seen that the standard Hüsler-Reiss distribution is recovered when either  $c_i = c_j = -\infty$  or  $c_i = c_j$  combined with  $\rho_{i^*, j^*} = 1$  and independence of  $(Z_i, Z_j)^\top$  and  $(Z_i^*, Z_j^*)^\top$ . The asymmetric extension of the Hüsler-Reiss distribution is not unique (see e.g., Beranger et al., 2019; Padoan, 2011; Zhong et al., 2025) however the proposed construction doesn't not require additional assumptions on the (skewness) parameters.

**Remark 1:** Distributions  $F_0$  of the common factor  $V_0$  with lighter tails than exponential, such as the Weibull distribution with the shape parameter  $\beta > 1$  result in asymptotic independence of the components of  $\mathbf{X}$  (Krupskii et al., 2018a), and hence  $\ell(\mathbf{x}) = \sum_{i=1}^d x_i$  in this case.

**Remark 2:** Following from Proposition 3, for a pair  $(X_i, X_j)^\top$ ,

$$\xi_{i:j} := \lim_{x_i \rightarrow 0} \frac{\partial \ell_{\mathbf{X}_{i:j}}((x_i, x_j)^\top)}{\partial x_i} = 1 - \frac{\Phi_2((\alpha_i \rho_{j^*,i} - c_j, \alpha_i \rho_{i^*,i} - c_i)^\top; \rho_{i^*,j^*})}{\Phi(\alpha_i \rho_{i^*,i} - c_i)},$$

which is a strictly positive quantity unless  $\rho_{i^*,j^*} = 1$  and  $\alpha_i \rho_{i^*,i} - c_i \leq \alpha_i \rho_{j^*,i} - c_j$ . From (5), the bivariate generalized Pareto distribution has a point mass at one or both boundaries. On the other hand,  $\xi_{i:j} = 0$  for any  $i \neq j$  if  $c_1 = c_2 = \dots = c_d = c_0$  and  $Z_1^* = Z_2^* = \dots = Z_d^*$ . In this case, the stable tail dependence function simplifies to

$$\begin{aligned} \ell_{X_i, X_j}(x_i, x_j) = & x_i \frac{\Phi_2\left(\left(\frac{\lambda_{ij}}{2} + \frac{1}{\lambda_{ij}} \ln \frac{\tilde{x}_i}{\tilde{x}_j}, \alpha_i \rho_{i,0} - c_0\right)^\top; \frac{1}{\lambda_{ij}}(\alpha_i \rho_{i,0} - \alpha_j \rho_{j,0})\right)}{\Phi(\alpha_i \rho_{i,0} - c_0)} \\ & + x_j \frac{\Phi_2\left(\left(\frac{\lambda_{ij}}{2} + \frac{1}{\lambda_{ij}} \ln \frac{\tilde{x}_j}{\tilde{x}_i}, \alpha_j \rho_{j,0} - c_0\right)^\top; \frac{1}{\lambda_{ij}}(\alpha_j \rho_{j,0} - \alpha_i \rho_{i,0})\right)}{\Phi(\alpha_j \rho_{j,0} - c_0)}, \end{aligned}$$

where  $\rho_{k,0} = \mathbf{P}_{k,d+1}$ . This is a different parameterization of the asymmetric extension of the Hüsler-Reiss distribution by Beranger et al. (2019) with the advantage of not requiring any constraints on the skewness parameters. The next proposition gives the full formula for  $\ell_{\mathbf{X}}$  in this special case.

**Proposition 4** Consider the random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  defined in (6). Assume  $V_0$  follows a unit exponential distribution,  $c_1 = c_2 = \dots = c_d = c_0$ ,  $Z_1^* = Z_2^* = \dots = Z_d^* = Z_0^*$ , and  $\rho_{k,0} = \text{Cor}(Z_k, Z_0^*)$ ,  $k \in \mathcal{I}$ . Then the stable tail dependence function associated with  $\mathbf{X}$  is given by

$$\ell_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^d x_j \Phi_{d-1}^{ESN} \left\{ \left( \frac{\lambda_{ij}}{2} + \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right)_{i \in \mathcal{I}_{-j}}; \bar{\Sigma}^{(j)}, \frac{\bar{\Sigma}^{(j)-1} \boldsymbol{\delta}_j}{\sqrt{1 - \boldsymbol{\delta}_j^\top \bar{\Sigma}^{(j)-1} \boldsymbol{\delta}_j}}, \frac{\alpha_j \rho_{j,0} - c_0}{\sqrt{1 - \boldsymbol{\delta}_j^\top \bar{\Sigma}^{(j)-1} \boldsymbol{\delta}_j}} \right\}$$

where  $\Phi_{d-1}^{ESN}(\cdot; \Sigma, \boldsymbol{\alpha}, \tau)$  is the cdf of the centered extended Skew-Normal (ESN) distribution with covariance matrix  $\Sigma$ , skewness vector  $\boldsymbol{\alpha} \in \mathbb{R}^{d-1}$  and extension parameter  $\tau \in \mathbb{R}$ . Note that here  $\bar{\Sigma}^{(j)}$  is identical to the Hüsler-Reiss model and in addition  $\tilde{x}_j = x_j / \Phi(\alpha_j \rho_{j,0} - c_0)$ ,  $\lambda_{i,j} = (\alpha_i^2 - 2\rho_{i,j}\alpha_i\alpha_j + \alpha_j^2)^{1/2}$  and  $\boldsymbol{\delta}_j = \{(\alpha_i \rho_{i,0} - \alpha_j \rho_{j,0}) / \lambda_{ij}\}_{i \in \mathcal{I}_{-j}}$ .

**Proof:** See Appendix A.4. □

As noted in Section 2, the stable tail dependence function fully characterizes extremal dependence, since the EV copula and the multivariate EV distribution with unit Fréchet margins are given respectively by  $C_{\text{EV}}(\mathbf{u}) = \exp\{-\ell(-\log \mathbf{u})\}$  and  $G(\mathbf{x}) = \exp\{-\ell(1/\mathbf{x})\}$ . Moreover, as discussed above, the corresponding multivariate Pareto model is completely determined by its density on the interior of the support, which is given by the formula:

$$h(\mathbf{x}) = \phi_{d-1}^{\text{ESN}} \left[ \left\{ \frac{\lambda_{ik}}{2} + \frac{1}{\lambda_{ik}} \log \frac{x_i \Phi(\alpha_i \rho_{i,0} - c_0)}{x_k \Phi(\alpha_k \rho_{k,0} - c_0)} \right\}_{i \in \mathcal{I}_{-k}}; \bar{\Sigma}^{(k)}, \frac{\bar{\Sigma}^{(k)-1} \boldsymbol{\delta}_k}{\mathcal{S}_k}, \frac{\alpha_k \rho_{k,0} - c_0}{\mathcal{S}_k} \right] \\ \times \left\{ \ell(\mathbf{1}) x_k^2 \prod_{i \in \mathcal{I}_{-k}} x_i \lambda_{ik} \right\}^{-1}, \quad (8)$$

where  $\mathcal{S}_k = \sqrt{1 - \boldsymbol{\delta}_k^\top \bar{\Sigma}^{(k)-1} \boldsymbol{\delta}_k}$ . The extremal coefficient and the Pickands dependence function are also determined by the stable tail dependence function, via  $\vartheta = \ell(\mathbf{1})$  and  $A(\mathbf{w}) = \ell(\mathbf{x}/\|\mathbf{x}\|_1)$ .

**Remark 3:** A parsimonious parameterization of this model can be obtained by assuming that the covariance matrix of  $(Z_0^*, Z_1, \dots, Z_d)^\top$  is determined by a covariance function, such that  $\rho_{i,0}$  depends on the distance between the corresponding locations  $\mathbf{s}_i$  and  $\mathbf{s}_0$ . In practical spatial applications, the locations  $\mathbf{s}_1, \dots, \mathbf{s}_d$  are typically known, while the coordinate vector  $\mathbf{s}_0$  can be treated as an additional model parameter to be estimated jointly with the other parameters. Even when the underlying covariance function is stationary, the extremal dependence between any two components of the vector  $\mathbf{X}$  is influenced by their respective distances to  $\mathbf{s}_0$ . This induces skewness and renders the model non-stationary. The proximity of the locations  $\mathbf{s}_1, \dots, \mathbf{s}_d$  to  $\mathbf{s}_0$  governs the degree of skewness: larger distances lead to smaller values of  $\boldsymbol{\delta}_j$  and hence weaker skewness, as implied by Proposition 4.

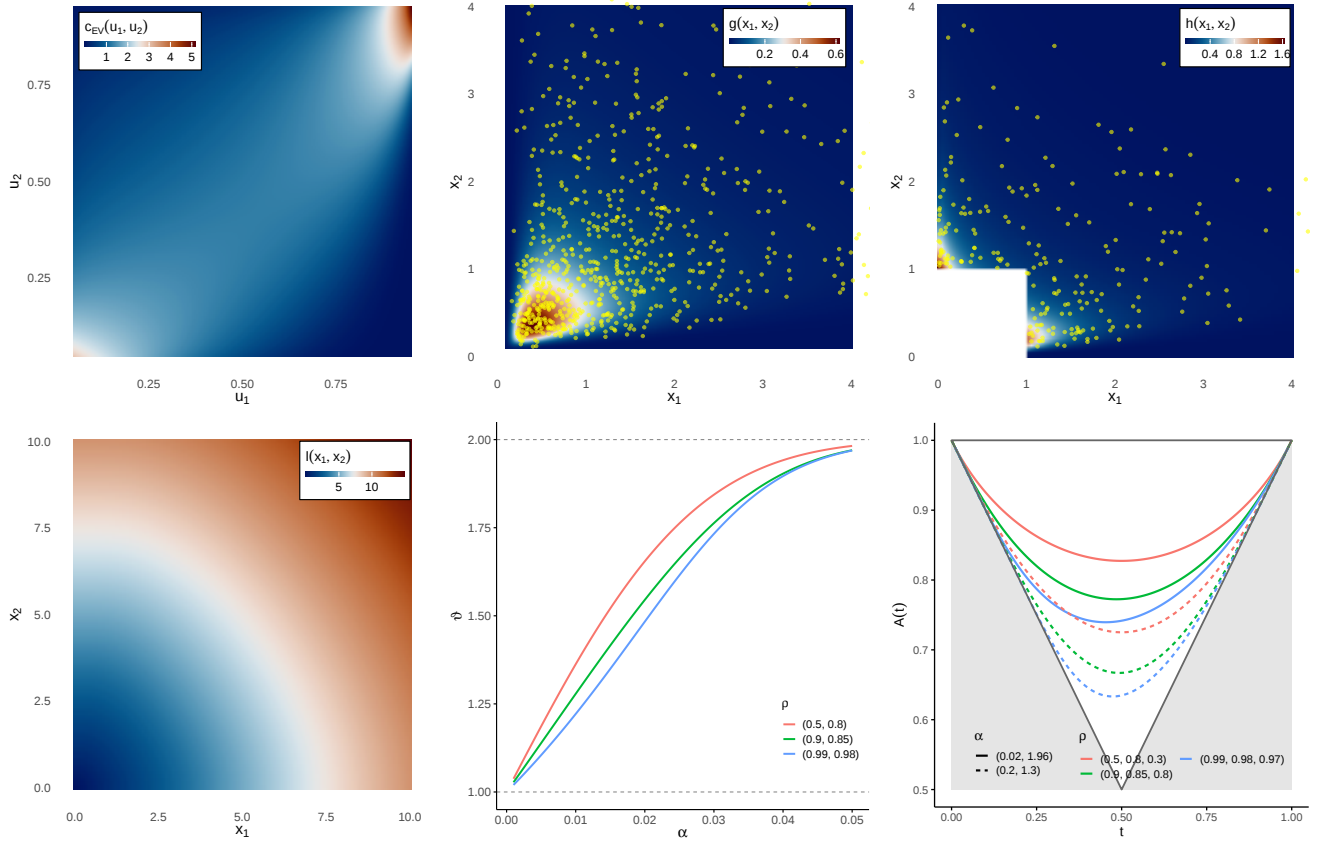


Figure 2: Left to right, top to bottom: density of the extreme value copula  $c_{EV}$ , density of the bivariate extreme value distribution  $g(\mathbf{x})$ , density of the bivariate Generalized Pareto distribution with unit Pareto margins  $\log h(\mathbf{x})$  and stable tail dependence function  $\ell(\mathbf{x})$  of the Proposition 4 model with  $\mathbf{c}_0 = 0.8$ ,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$  and correlation  $\boldsymbol{\rho} = (\rho_{1,2}, \rho_{1,0}, \rho_{2,0}) = (0.99, 0.98, 0.97)$ . The remaining panels display for several choices of  $\boldsymbol{\rho}$  and  $\boldsymbol{\alpha}$ , the extremal coefficient  $\vartheta$  as function of  $\alpha$  with  $\boldsymbol{\alpha} = (\alpha, 98\alpha)$  and the Pickands dependence function with  $c_0 = 0.8$ .

### 3.2 Tail properties

We now delve into the tail properties of the models defined in Section 3.1 and derive insights about their characteristics and potential flexibility.

In the context of Proposition 1, the limiting EV copula corresponds to the Marshall-Olkin distribution (Marshall and Olkin, 1967), and it depends on the correlation matrix of the vector  $\mathbf{Z}^*$  and the vector of truncation parameters  $\mathbf{c}$ . Indeed, the correlation matrix of  $\mathbf{Z}$  has no impact on

the copula since the heavy-tailed factor  $V_0$  dominates the Gaussian component  $\mathbf{Z}$  in the upper tail. With  $\mathbf{c} = -\infty$ , the EV copula simplifies to the comonotonicity copula, i.e.  $C_{\text{EV}}(\mathbf{u}) = \min(\mathbf{u})$ . For example, when  $d = 2$ , the stable tail dependence function of  $\mathbf{X} = (X_1, X_2)$  is

$$\ell_{\mathbf{X}}(\mathbf{x}) = (1 - \alpha)x_1 + (1 - \beta)x_2 + \max(\alpha x_1, \beta x_2),$$

where  $\alpha = \Pr(Z_2^* > c_2 | Z_1^* > c_1) \in [0, 1]$  and  $\beta = \Pr(Z_1^* > c_1 | Z_2^* > c_2) \in [0, 1]$ . The corresponding EV copula is therefore  $C_{\text{EV}}(u_1, u_2) = \min(u_1^{1-\alpha}u_2, u_1u_2^{1-\beta})$  and its density  $c_{\text{EV}}(u_1, u_2) = (1 - \alpha)u_1^\alpha$  if  $u_1 > u_2^{\beta/\alpha}$  and  $(1 - \beta)u_2^\beta$  otherwise. The multivariate EV density is given by  $g(\mathbf{x}) = \exp\{-\ell(1/\mathbf{x})\}(1 - \alpha)(1 - \beta)/(x_1x_2)^2$  while the bivariate generalized Pareto distribution does not admit a density (since  $\partial^2\ell_{\mathbf{X}}/\partial x_1\partial x_2 = 0$ ) but admits mass  $(\alpha \wedge \beta)/(2 - (\alpha \vee \beta))$  along the line  $x_2 = \frac{\alpha}{\beta}x_1$  and at the boundaries of the support  $H_1(\{0\}) = (1 - \beta)/(2 - (\alpha \vee \beta))$ ,  $H_2(\{0\}) = (1 - \alpha)/(2 - (\alpha \vee \beta))$ . Illustrations are provided in Appendix A.5, Figure 7.

Models with an exponential factor preserve substantial flexibility in the tail behaviour, unlike those based on a heavy-tailed distribution for  $V_0$  considered in Proposition 1. However, Proposition 3 shows that the resulting bivariate stable tail dependence function requires evaluating a trivariate normal cdf, making extensions to higher dimensions computationally demanding. By contrast, Proposition 2 demonstrates that assuming independence between  $\mathbf{Z}$  and  $\mathbf{Z}^*$  yields models that are far more tractable, as they rely only on lower-dimensional Hüsler–Reiss distributions while still offering substantial tail flexibility, including the ability to capture asymmetry. We now turn to the following submodel.

**Submodel 1:** Assume that  $Z_1^* = Z_2^* = \dots = Z_d^* = Z^*$  and  $(Z_1, \dots, Z_d)^\top$  is independent of  $Z^*$ .

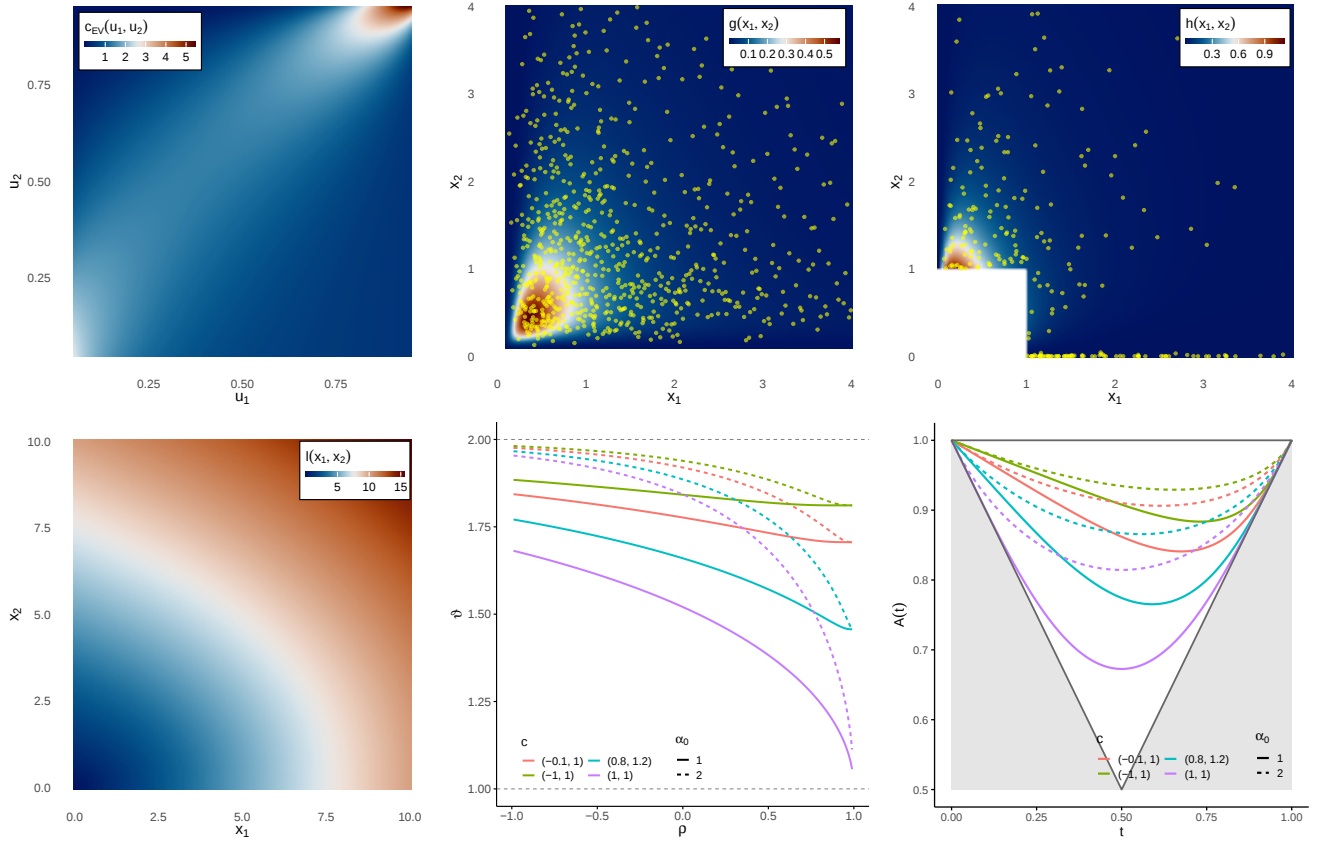


Figure 3: Left to right, top to bottom: density of the extreme value copula  $c_{EV}$ , density of the bivariate extreme value distribution  $g(\mathbf{x})$ , density of the bivariate Generalized Pareto distribution with unit Pareto margins  $\log h(\mathbf{x})$  and stable tail dependence function  $\ell(\mathbf{x})$  of Submodel 1 with  $\mathbf{c} = (0.8, 1.2)$ ,  $\alpha_0 = 1$  and correlation  $\rho_{1,2} = 0.6$ . The remaining panels display for several choices of  $\mathbf{c}$  and  $\alpha$ , the extremal coefficient  $\vartheta$  as function of  $\rho_{1,2}$  and the Pickands dependence function with  $\rho = 0.6$ .

From Proposition 2, for  $i, j \in \mathcal{I}, i \neq j$ ,

$$\begin{aligned} \ell_{X_i, X_j}(x_i, x_j) = & x_i \left\{ 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_i}{\tilde{x}_j} \right) \cdot \min(\zeta_j / \zeta_i, 1) \right\} \\ & + x_j \left\{ 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right) \cdot \min(\zeta_i / \zeta_j, 1) \right\}, \end{aligned}$$

where  $\tilde{x}_l = x_l / \zeta_l$  and  $\zeta_l = \Phi(-c_l)$  for  $l \in \{i, j\}$ . Note that  $\ell_{X_i, X_j}$  only involves univariate normal cdfs thus making this model computationally tractable (see Section 4). As previously mentioned, the EV copula and the multivariate EV distribution with unit Fréchet margins can be easily recovered

from the stable tail dependence function while its partial derivatives can be used to compute the corresponding densities as well as the density of the multivariate Generalized Pareto distribution. Regarding the latter, we have

$$\frac{\partial}{\partial x_k} \ell_{X_i, X_j}(x_i, x_j) = 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_k}{\tilde{x}_l} \right) \cdot \min(\zeta_l / \zeta_k, 1), \quad k \in \{i, j\}, \quad l \neq k,$$

which, from (5), yields  $H_i(\{0\}) = \left[ 1 - \min \left( \frac{\zeta_i}{\zeta_j}, 1 \right) \right] / \ell_{X_i, X_j}(1, 1)$ . This implies that  $H$  places mass on a single boundary determined by the ratio  $\zeta_i / \zeta_j$ , and therefore by the values of  $c_i$  and  $c_j$ ; in particular, when  $c_i = c_j$ , no boundary mass arises. The density in the interior of the support  $(\boldsymbol{\eta}, \infty)$  is then

$$h(x_i, x_j) = \frac{\phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right) \cdot \min(\zeta_j / \zeta_i, 1)}{\lambda_{ij} x_i^2 x_j \ell_{X_i, X_j}(1, 1)}.$$

For simplicity and to avoid over parametrisation, from now on we will consider  $\alpha_1 = \dots = \alpha_d = \alpha_0$ . Reading from left to right and top to bottom, Figure 3 illustrates the EV copula density  $c_{\text{EV}}(u_1, u_2)$ , the multivariate EV density  $g(x_1, x_2)$ , the multivariate generalized Pareto density  $h(x_1, x_2)$ , the stable tail dependence function  $\ell_{\mathbf{X}}(x_1, x_2)$ , the extremal coefficient  $\vartheta$ , and the Pickands dependence function  $A(t)$  for Submodel 1 with  $\mathbf{c} = (0.8, 1.2)$ ,  $\alpha_0 = 1$ , and correlation  $\rho_{1,2} = 0.6$ . Since  $c_1 < c_2$ , we have  $\zeta_1 > \zeta_2$ , implying that  $H_1(0) = 0$  (no mass along  $x_1 = 0$ ) and  $H_1(0) \neq 0$  (mass along  $x_2 = 0$ ). The resulting asymmetry is clearly reflected in the Pickands dependence function, which is directly influenced by the parameters  $\mathbf{c}$ .

Instead of imposing  $Z_1^* = \dots = Z_d^*$  as in Submodel 1, an extension is to allow the components of  $\mathbf{Z}^*$  to be equicorrelated, leading to the following submodel.

**Submodel 2:** Assume that  $\mathbf{Z}^* = (Z_1^*, Z_2^*, \dots, Z_d^*)^\top$  has equicorrelation structure with  $\rho_{i^*, j^*} = \rho^*$  for  $i^* \neq j^*$  and  $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$  is independent of  $\mathbf{Z}^*$ . This model extends Submodel 1, which arises

as the special case  $\rho^* = 1$ . From Proposition 2, for  $i, j \in \mathcal{I}, i \neq j$ ,

$$\begin{aligned} \ell_{X_i, X_j}(x_i, x_j) = & x_i \left[ 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_i}{\tilde{x}_j} \right) \cdot \Phi_2 \left\{ (-c_i, -c_j)^\top; \rho^* \right\} / \zeta_i, \right] \\ & + x_j \left[ 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right) \cdot \Phi_2 \left\{ (-c_i, -c_j)^\top; \rho^* \right\} / \zeta_j \right], \end{aligned}$$

where  $\tilde{x}_l = x_l / \zeta_l$ ,  $\zeta_l = \Phi(-c_l)$  for  $l \in \{i, j\}$  as before. Similarly to Submodel 1, all functions associated with this EV distribution can be derived. In particular, for the generalized Pareto distribution, we obtain

$$\frac{\partial}{\partial x_k} \ell_{X_i, X_j}(x_i, x_j) = 1 - \Phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_k}{\tilde{x}_l} \right) \cdot \Phi_2 \left\{ (-c_i, -c_j)^\top; \rho^* \right\} / \zeta_k, \quad k \in \mathcal{I}, l = \mathcal{I}_{-k},$$

which, from (5), yields  $H_i(\{0\}) = \left[ 1 - \Phi_2 \left\{ (-c_i, -c_j)^\top; \rho^* \right\} / \zeta_i \right] / \ell_{X_i, X_j}(1, 1)$ . This implies that  $H$  places mass on both boundaries. These masses are equal when  $c_i = c_j$  and vanish when  $c_i = c_j = -\infty$ . Finally, the density in the interior of the support  $(\mathbf{0}, \infty)$  is given by

$$h(x_i, x_j) = \frac{\phi \left( -\frac{\lambda_{ij}}{2} - \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right) \cdot \Phi_2 \left\{ (-c_i, -c_j)^\top; \rho^* \right\}}{\lambda_{ij} \zeta_i x_i^2 x_j \ell_{X_i, X_j}(1, 1)}.$$

Illustrations for Submodel 2 are shown in Figure 8 of Appendix A.6, where a higher degree of independence among the extremes is observed when  $\rho^* = 0$  compared to  $\rho^* = 1$  (Submodel 1, Figure 3). Additionally, the multivariate generalized Pareto distribution now assigns mass to both boundaries ( $x_1 = 0$  and  $x_2 = 0$ ).

## 4 Inference

The explicit form of the stable tail dependence function can be used to estimate the parameters of the limiting EV distribution of the vector  $\mathbf{X}$  in (6) for the general case. Depending on whether extremes are defined via component-wise maxima or threshold exceedances, the limiting distribution may be

$G$  or  $H$ , respectively. In this section, we briefly review some inference approaches and demonstrate how they can be applied to estimate the parameters of Submodels 1 and 2.

## 4.1 Inference for Block Maxima data

Consider a sample  $\mathbf{X}_n \equiv \{(\mathcal{X}_{k1}, \dots, \mathcal{X}_{kd})^\top\}_{k=1}^n$  from the distribution  $G$  with univariate marginal cdfs  $G_1, \dots, G_d$  and pdfs  $g_1, \dots, g_d$ . Following Section 2.1, we define the EV copula cdf of the  $(i, j)$ -th marginal as  $C_{\text{EV}}^{(i,j)}(u_i, u_j) = \exp\{-\ell_{X_i, X_j}(x_i, x_j)\}$ , where  $x_i = -\log u_i$  and  $x_j = -\log u_j$ , and its density by

$$c_{\text{EV}}^{(i,j)}(u_i, u_j) = \frac{\exp\{-\ell_{X_i, X_j}(x_i, x_j)\}}{u_i u_j} \cdot \left\{ \frac{\partial \ell_{X_i, X_j}(x_i, x_j)}{\partial x_i} \frac{\partial \ell_{X_i, X_j}(x_i, x_j)}{\partial x_j} - \frac{\partial^2 \ell_{X_i, X_j}(x_i, x_j)}{\partial x_i \partial x_j} \right\}.$$

From (1), the distribution of the  $(i, j)$ -th marginal is  $G^{(i,j)}(x_i, x_j) = C_{\text{EV}}^{(i,j)}\{G_i(x_i), G_j(x_j)\}$  and its density

$$g^{(i,j)}(x_i, x_j) = \frac{\partial^2}{\partial x_i \partial x_j} G^{(i,j)}(x_i, x_j) = c_{\text{EV}}^{(i,j)}\{G_i(x_i), G_j(x_j)\} \cdot g_i(x_i) g_j(x_j),$$

so that the pairwise log-likelihood for the sample  $\mathbf{X}_n$  is

$$\mathcal{L}(\mathbf{X}_n; \boldsymbol{\theta}) = \sum_{k=1}^n \sum_{i < j} \log c_{\text{EV}}^{(i,j)}\{G_i(\mathcal{X}_{ki}; \boldsymbol{\theta}_M), G_j(\mathcal{X}_{kj}; \boldsymbol{\theta}_M); \boldsymbol{\theta}_D\} + (d-1) \sum_{k=1}^n \sum_{i=1}^d \log g_i(\mathcal{X}_{ki}; \boldsymbol{\theta}_M),$$

where  $\boldsymbol{\theta} = (\boldsymbol{\theta}_M^\top, \boldsymbol{\theta}_D^\top)^\top$  with  $\boldsymbol{\theta}_M$  and  $\boldsymbol{\theta}_D$  respectively being the vectors of marginal and dependence parameters. The composite maximum likelihood estimator of  $\boldsymbol{\theta}$ ,  $\hat{\boldsymbol{\theta}}_{\text{CMLE}}$ , is obtained by maximizing  $\ell(\mathbf{X}_n; \boldsymbol{\theta})$ . This estimator is consistent and asymptotically normal under standard regularity conditions (Lindsay, 1998). Alternatively,  $\boldsymbol{\theta}_M$  can be estimated using the marginal likelihood:

$$\mathcal{L}_M(\mathbf{X}_n; \boldsymbol{\theta}_M) = \sum_{k=1}^n \sum_{i=1}^d \log g_i(\mathcal{X}_{ki}; \boldsymbol{\theta}_M),$$

and the dependence parameters can be obtained by maximizing the joint likelihood:

$$\mathcal{L}_D(\mathbf{X}_n; \boldsymbol{\theta}_D) = \sum_{k=1}^n \sum_{i < j} \log c_{\text{EV}}^{(i,j)}\{G_i(\mathcal{X}_{ki}; \hat{\boldsymbol{\theta}}_M), G_j(\mathcal{X}_{kj}; \hat{\boldsymbol{\theta}}_M); \boldsymbol{\theta}_D\},$$

where  $\widehat{\boldsymbol{\theta}}_M$  is the estimator of  $\boldsymbol{\theta}_M$  obtained in the first step. This two-step estimator of  $\boldsymbol{\theta}$  is more computationally efficient, especially for large  $d$ , and is consistent and asymptotically normal (Joe and Xu, 1996).

In the general case, higher-dimensional marginals can be incorporated to construct a composite maximum likelihood estimator. While this approach is more computationally demanding, it can provide greater efficiency than the pairwise maximum likelihood estimator introduced above.

We now conduct a simulation study to assess the finite sample performance of the pairwise likelihood estimator. We generate  $N = 500$  datasets made of  $n = 250$  and  $n = 1000$  component-wise maxima drawn from the EV limiting distribution of Submodel 1 considering a parsimonious parameterization with spatial covariance structure for the vector  $(Z_1, \dots, Z_d)^\top$ . More precisely, we use  $\alpha_1 = \dots = \alpha_d = \alpha = 1/3$ ,  $c_i = \beta_0 + \beta_1 \mathbf{s}_{i1} + \beta_2 \mathbf{s}_{i2} = 0.5 + 1.5 \mathbf{s}_{i1} - 0.5 \mathbf{s}_{i2}$  where  $\{(\mathbf{s}_{i1}, \mathbf{s}_{i2})\}_{i=1}^d$  is a set of  $d = 20, 30, 50$  coordinates randomly selected in  $(0, 1)^2$  for each generated dataset, and  $\rho_{i,j} = \exp\{-(\text{dst}_{i,j}/p_1)^{p_2}\} = \exp\{-(\text{dst}_{i,j}/1.5)\}$  where  $\text{dst}_{i,j}$  is the Euclidean distance between two points with the coordinates  $(\mathbf{s}_{i1}, \mathbf{s}_{i2})$  and  $(\mathbf{s}_{j1}, \mathbf{s}_{j2})$ . Assuming the marginals are known, we estimate  $\boldsymbol{\theta}_C = (p_0, p_1, \alpha, \beta_0, \beta_1, \beta_2)^\top$  and report the root mean squared errors (RMSEs) for the proposed pairwise likelihood estimator in Table 1.

|          | $p_0$     | $p_1$     | $\alpha$  | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $d = 20$ | 1.62/0.95 | 0.12/0.03 | 0.07/0.06 | 0.11/0.05 | 0.06/0.03 | 0.02/0.01 |
| $d = 30$ | 1.52/0.74 | 0.11/0.09 | 0.07/0.05 | 0.09/0.05 | 0.06/0.03 | 0.02/0.01 |
| $d = 50$ | 1.38/0.63 | 0.06/0.03 | 0.07/0.05 | 0.09/0.05 | 0.05/0.03 | 0.02/0.01 |

Table 1: RMSEs of the copula parameter estimates of the EV distribution of Submodel 1 obtained using the pairwise likelihood approach. The true values are  $p_0 = 1.5, p_1 = 1, \alpha = 1/3, \beta_0 = 0.5, \beta_1 = 1.5, \beta_2 = -0.5$ . The results are based on 500 simulated datasets with  $N = 250/N = 1000$  component-wise maxima replicates and  $d = 20, 30, 50$  randomly selected locations in  $(0, 1)^2$ .

It is seen that the pairwise likelihood approach yields pretty accurate parameter estimates,

except for  $p_0$ , and more accurate estimates can be obtained with larger sample size, as expected.

We now conduct a simulation study to assess the finite sample performance of the pairwise likelihood estimator for Submodel 2. We generate  $N = 500$  datasets made of  $n = 250$  and  $n = 1000$  component-wise maxima drawn from the EV limiting distribution of submodel 2, and we use the same parameter values as for Submodel 1 above with the addition of  $\rho = 0.8$ . For comparison, we also assess the finite-sample performance of the triplewise likelihood estimator for Submodel 2 under the same simulation settings. The stable tail dependence function for this model has a simple analytical form in the trivariate case, rendering the triplewise likelihood estimator computationally tractable. Assuming again the marginals to be known, we estimate  $\boldsymbol{\theta}_C = (p_0, p_1, \alpha, \beta_0, \beta_1, \beta_2, \rho)^\top$  using the proposed pairwise and triplewise likelihood estimators. The corresponding RMSEs for both estimators, along with their relative efficiency, are reported in Table 2.

|          | $p_0$     | $p_1$     | $\alpha$  | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\rho$    |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $d = 20$ | 1.78/1.25 | 0.27/0.19 | 0.35/0.35 | 0.13/0.10 | 0.09/0.05 | 0.05/0.04 | 0.03/0.02 |
|          | 1.83/1.20 | 0.15/0.12 | 0.10/0.06 | 0.11/0.05 | 0.07/0.03 | 0.02/0.01 | 0.03/0.01 |
|          | 1.03/0.97 | 0.54/0.63 | 0.28/0.17 | 0.84/0.49 | 0.74/0.57 | 0.53/0.30 | 0.85/0.40 |
| $d = 30$ | 1.63/0.92 | 0.17/0.11 | 0.25/0.23 | 0.09/0.04 | 0.06/0.04 | 0.03/0.01 | 0.02/0.01 |
|          | 1.48/0.96 | 0.13/0.10 | 0.08/0.05 | 0.08/0.04 | 0.05/0.02 | 0.02/0.01 | 0.02/0.01 |
|          | 0.91/1.05 | 0.78/0.97 | 0.31/0.23 | 0.90/0.79 | 0.81/0.63 | 0.77/0.81 | 0.71/0.63 |
| $d = 50$ | 1.41/0.73 | 0.21/0.17 | 0.08/0.06 | 0.09/0.04 | 0.05/0.02 | 0.02/0.01 | 0.02/0.01 |
|          | 1.30/0.64 | 0.13/0.02 | 0.07/0.04 | 0.09/0.03 | 0.05/0.02 | 0.02/0.01 | 0.01/0.01 |
|          | 0.92/0.88 | 0.61/0.12 | 0.95/0.74 | 0.99/0.90 | 0.97/0.95 | 0.97/0.94 | 0.83/0.98 |

Table 2: RMSEs of the copula parameter estimates of the EV distribution of Submodel 2 obtained using the pairwise likelihood approach (first line), triplewise likelihood approach (second line), and relative efficiency of the two estimators (third line). The true values are  $p_0 = 1.5, p_1 = 1, \alpha = 1/3, \beta_0 = 0.5, \beta_1 = 1.5, \beta_2 = -0.5, \rho = 0.8$ . The results are based on 500 simulated datasets with  $N = 250/N = 1000$  component-wise maxima replicates and  $d = 20, 30, 50$  randomly selected locations in  $(0, 1)^2$ .

The pairwise likelihood approach produces relatively poor estimates of  $p_0, p_1$ , and  $\alpha$  in this setting, suggesting that these parameters may be weakly identifiable, particularly for small  $d$ . The

estimation accuracy improves as  $d$  increases, and the parameters  $\beta_0, \beta_1, \beta_2$ , which govern skewness, as well as  $\rho$ , are estimated much more precisely. In contrast, the triplewise likelihood method provides substantially more accurate estimates for all parameters except  $p_0$ , with notable improvements for  $p_1$  and  $\alpha$ . However, the relative improvement diminishes as  $d$  becomes larger.

## 4.2 Inference for Peaks-over-Thresholds data

The aim of this section is to introduce an inference methodology for the proposed multivariate Generalized Pareto distributions of Proposition 4, Submodels 1 and 2. Since the focus is on the estimation accuracy of the dependence structure, the marginal distributions are assumed to be unit Pareto.

Let us consider a sample  $\mathbf{X}_{1:n} \equiv \{\mathbf{X}_k = (\mathcal{X}_{k1}, \dots, \mathcal{X}_{kd})^\top\}_{k=1}^n$  drawn from the original distribution  $F$ . After transforming the margins to the unit Pareto scale and selecting a high threshold  $\mathbf{u} = (u_1, \dots, u_d)$ , we retain the subset of samples  $\mathbf{X}_l$  for which there exists at least one  $j \in \mathcal{I}$  such that  $\mathcal{X}_{lj} > u_j$ , with  $l \in \{1, \dots, m\}$  and  $m \ll n$ . The resulting sample,  $\mathbf{Y}_{1:m} \equiv \{\mathbf{Y}_l = (\mathcal{Y}_{l1}, \dots, \mathcal{Y}_{ld})^\top\}_{l=1}^m$  is therefore assumed to be drawn from the distribution  $H$ .

A first approach for performing inference on multivariate Generalized Pareto models is to directly use the known density  $h$  derived in Section 2.2, yielding the (full) likelihood:

$$\mathcal{L}(\mathbf{Y}_{1:m}; \boldsymbol{\theta}_D) = \sum_{l=1}^m \log h(\mathbf{Y}_l; \boldsymbol{\theta}_D), \quad (9)$$

which enables simpler and more efficient inference than for Block Maxima models, which, as discussed in the previous section, lack a tractable analytic expression for their densities, particularly in high dimensions. This approach is especially useful for models that allow all variables to be simultaneously extreme and place no mass on the boundaries of the support, as in the model derived

in Proposition 4.

A censored likelihood approach is sometimes employed when the multivariate Pareto distribution may not be reliable over the full set of threshold exceedances (Rootzén et al., 2018a). Since the goal of this section is primarily to demonstrate that model parameters can be accurately recovered, the censored likelihood is not considered here, although in some cases it can help reduce bias (Huser et al., 2016). This is partly motivated by the fact that certain likelihood contributions would otherwise require numerical integration.

Focusing on the model of Proposition 4, we first generate  $d = 5$  random locations within  $(0, 1)^2$  and set  $\mathbf{s}_0$  at the center of the study region, i.e.,  $(0.5, 0.5)$ . We then generate  $n = 200,000$  observations from the model using the parsimonious parametrization suggested in Remark 3, where the correlations are defined as  $\rho_{i,j} = \exp \{-(\text{dst}_{i,j}/p_1)^{p_2}\}$ , with  $p_1 = 0.5$ ,  $p_2 = 1.5$  (including for  $j = 0$ ),  $\alpha_1 = \dots = \alpha_5 = \alpha_0 = 1.2$ , and  $c_0 = 0.8$ .

For simplicity, we assume  $\mathbf{s}_0$  is known and maximize the likelihood (9), with  $h$  given by (8), with respect to  $\boldsymbol{\theta}_D = (p_1, p_2, \alpha_0, c_0)$ . The estimation procedure is performed for threshold levels corresponding to the 99%, 99.5%, and 99.9% empirical quantiles and is repeated 500 times. Figure 4 presents boxplots summarizing the distribution of the estimates for each parameter and threshold level. As the threshold increases, the bias in the estimates decreases, which is consistent with the fact that the multivariate generalized Pareto is the limiting distribution of threshold exceedances. Increasing the threshold from the 99.5% to the 99.9% quantile also increases the variability of the estimates, although both thresholds yield approximately unbiased estimates. This suggests that the 99.9% quantile is too high, providing insufficient data to accurately estimate the model parameters. Overall, the 99.5% threshold appears to offer a good balance between bias and variance.

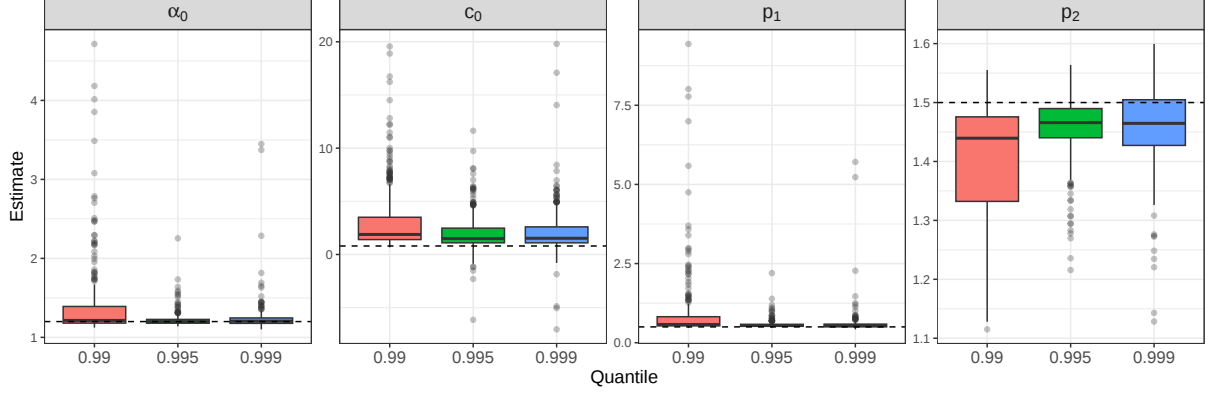


Figure 4: Boxplots of the parameter estimates  $\hat{\alpha}_0$ ,  $\hat{c}_0$ ,  $\hat{p}_1$  and  $\hat{p}_2$  (left to right) from the Proposition 4 model with  $d = 5$  and where  $\mathbf{s}_0 = (0.5, 0.5)$  is assumed fixed and the exceedance thresholds are set at the 99%, 99.5% and 99.9% empirical quantiles. The true parameter values are indicated by dashed horizontal lines.

For models such as Submodels 1 and 2, the likelihood (9) is, however, unsuitable since there is positive mass on the boundaries and omitting it would lead to overestimation of the probability of joint extremes (Thibaud and Opitz, 2015). We therefore adapt the strategy of Beranger et al. (2017), which was originally developed for the unit simplex. Since this approach does not scale well with the dimension  $d$ , we restrict our attention to the bivariate case ( $d = 2$ ). In this context, the density  $h(\mathbf{x})$  is defined on the interior  $(0, \infty)^2$ , while point masses  $H_1(0)$  and  $H_2(0)$  are present on the boundaries  $\{\mathbf{x} \in [0, \infty)^2 : x_1 = 0\}$  and  $\{\mathbf{x} \in [0, \infty)^2 : x_2 = 0\}$ , respectively. Given that data are never exactly located on the boundary, define

$$h^\epsilon(x_1, x_2) = \begin{cases} h(x_1, x_2) & \text{if } (x_1, x_2) \in (\epsilon, \infty)^2 \\ H_1(\{0\}) & \text{if } x_1 < \epsilon \\ H_2(\{0\}) & \text{if } x_2 < \epsilon \end{cases},$$

for some small  $\epsilon$  and the likelihood is then

$$\begin{aligned} \mathcal{L}^\epsilon(\mathbf{y}_{1:m}; \boldsymbol{\theta}_D) &= \sum_{l=1}^m \log h^\epsilon(\mathbf{y}_l; \boldsymbol{\theta}_D) \\ &= \sum_{l=1}^{m_{12}} \log(h(\mathbf{y}_l^\epsilon; \boldsymbol{\theta}_D)) + m_1 H_1(\{0\}; \boldsymbol{\theta}_D) + m_2 H_2(\{0\}; \boldsymbol{\theta}_D), \end{aligned} \quad (10)$$

where  $\{\mathcal{Y}_l^\epsilon\}_{l=1}^{m_{12}}$  is the restriction of  $\mathcal{Y}_{1:m}$  to  $(\epsilon, \infty)^2$  and  $m_{12} + m_1 + m_2 = m$ . For a review and comparison of estimation methods for multivariate extremes, we refer the reader to [Huser et al. \(2016\)](#). We note also that [Goix et al. \(2017\)](#) and [Simpson et al. \(2020\)](#) have proposed methodologies to estimate extremal mass at the boundaries.

We now investigate the performance of the inference approach described above, focusing on Submodel 1, where the true parameter values are the same as in [Figure 3](#), i.e.,  $c = (0.8, 1.2)$ ,  $\alpha_0 = 1$ , and  $\rho_{1,2} = 0.6$ . In its current form, the parameters of this bivariate model are not identifiable; we therefore reduce it to a two-parameter model with  $\lambda = \sqrt{\alpha_0^2(1 - \rho_{1,2})} \approx 0.8050$  and  $\zeta^* = \zeta_1/\zeta_2 \approx 1.8411$ . We generate  $n = 10,000, 100,000$ , and  $200,000$  observations from  $\mathbf{X}$  and consider exceedances over the 98%, 99%, and 99.5% empirical quantiles for  $n = 10,000$ , and 99%, 99.9%, and 99.95% for the larger sample sizes. The model is fitted using [\(10\)](#) with  $\epsilon = 0.025, 0.05$ , and  $0.1$ , with each experiment repeated 500 times. Boxplots of the estimates of  $\lambda$  and  $\zeta^*$  are reported in [Figure 5](#).

It is observed that, as the quantile threshold increases, the multivariate Generalized Pareto model provides a better approximation, resulting in more accurate estimates. Additionally, since the model places mass on one boundary, a reasonably high value of  $\epsilon$  (around 0.05–0.1) is necessary to ensure that observations are appropriately allocated to each subspace. For  $n = 10,000$ , the threshold cannot be set too high, as this would result in too few exceedances (on average  $m \approx 77$  for the 99.5% quantile), leading to slight bias. For larger sample sizes, a threshold at the 99.9% quantile appears sufficient, as the estimates are stable.

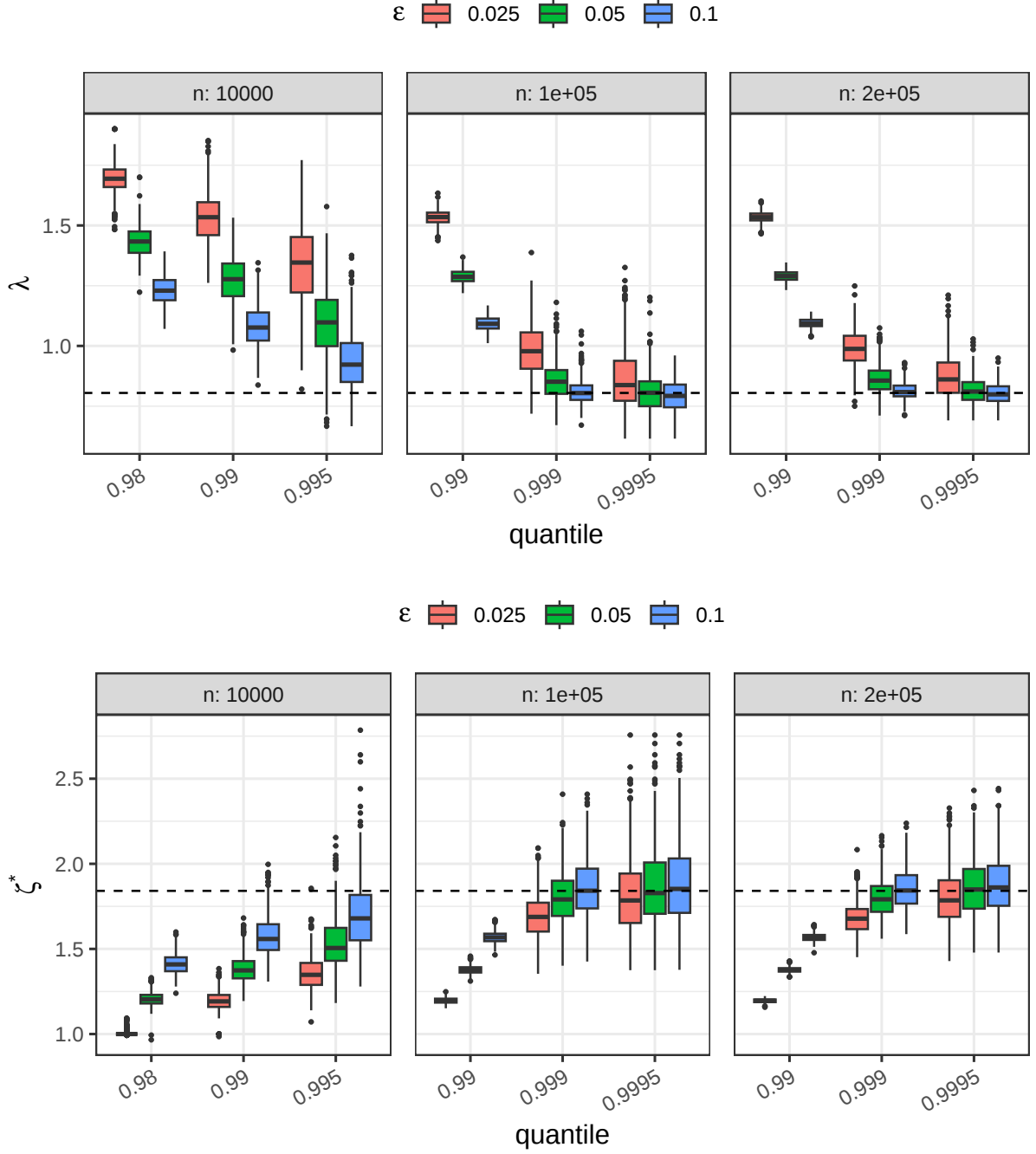


Figure 5: Boxplots of the parameter estimates  $\hat{\lambda}$  (top) and  $\hat{\zeta}^*$  (bottom) for Submodel 1 in dimension  $d = 2$ , obtained using the likelihood (10) with  $\epsilon = 0.025, 0.05$  and  $0.1$ . Exceedance thresholds set at the 98%, 99% and 99.5% empirical quantiles for  $n = 10000$  (left panels) and at the 99%, 99.9% and 99.95% quantiles for  $n = 100000$  and  $200000$  (middle and right panels). The true parameter values are indicated by dashed horizontal lines.

## 5 Illustrations on real datasets

In this section, we demonstrate that asymmetric models offer substantial flexibility gains over symmetric models, such as Hüsler-Reiss, when modeling extremal dependence. In particular, we apply the models introduced in Section 3.1 to the analysis of two datasets: the first requires a block maxima approach in relatively high dimensions, while the second uses the Peaks-over-Thresholds approach in more moderate dimensions.

### 5.1 Summer temperature maxima in Melbourne, Australia

This dataset comprises 50 summer temperature maxima in Melbourne, Australia, recorded over the extended summer period from August to April (August 1961 to April 2011) across a study region of 90 locations arranged on a  $0.15^\circ$  grid in a 9-by-10 layout. The data are available in the R package `ExtremalDep` (Beranger et al., 2025) as the object `heat`, and an analysis using max-stable processes is provided by Beranger et al. (2021).

In this high-dimensional setting ( $d = 90$  spatial locations), we consider four models corresponding to different extreme-value limiting distributions of the random vector  $\mathbf{X}$ , as defined in (6). In all cases, the covariance matrix of the vector  $\mathbf{Z}$  is specified using the powered-exponential covariance function. The models are defined as follows:

- 1 Submodel 1 from Section 3.2, with  $c_i = 0$  for  $i \in \mathcal{I}$ , corresponding to the Hüsler-Reiss distribution.
- 2 Submodel 1 from Section 3.2, with  $c_i = \beta_0 + \beta_1 \mathbf{s}_{i1} + \beta_2 \mathbf{s}_{i2}$ ,  $i \in \mathcal{I}$ , where  $\{(\mathbf{s}_{i1}, \mathbf{s}_{i2})\}_{i=1}^d$  denote the spatial coordinates of the  $d = 90$  locations.

- 3 Submodel 2 from Section 3.2, with  $c_i = \beta_0 + \beta_1 \mathbf{s}_{i1} + \beta_2 \mathbf{s}_{i2}$ ,  $i \in \mathcal{I}$ , where  $\{(\mathbf{s}_{i1}, \mathbf{s}_{i2})\}_{i=1}^d$  are the coordinates of the  $d = 90$  locations.
- 4 Model from Proposition 4, in which the covariance matrix of  $(Z_0^*, Z_1, \dots, Z_d)^\top$  is also determined by the powered-exponential covariance function.

Following insights drawn from the simulation studies in Section 4.1, these models are fitted using the pairwise likelihood estimator since no substantial loss of efficiency is anticipated when  $d = 90$ . The goodness of fit of model  $m \in \{1, 2, 3, 4\}$  is evaluated by computing the RMSE defined as

$$\text{RMSE}_m(t) = \sqrt{\frac{1}{|S|} \sum_{(i,j) \in S} \{\hat{A}_{X_i, X_j}(t) - \hat{A}_{X_i, X_j}^m(t)\}^2},$$

where  $\hat{A}_{X_i, X_j}^m(t)$  represents model-based estimates of the Pickands dependence function  $A_{X_i, X_j}(t)$  for each pair of locations  $\mathbf{s}_i$  and  $\mathbf{s}_j$  and  $\hat{A}_{X_i, X_j}^{\text{emp}}(t)$  is its corresponding nonparametric estimate.

Figure 6 shows the RMSE values for different sets of pairs,  $S$ . While all models perform similarly

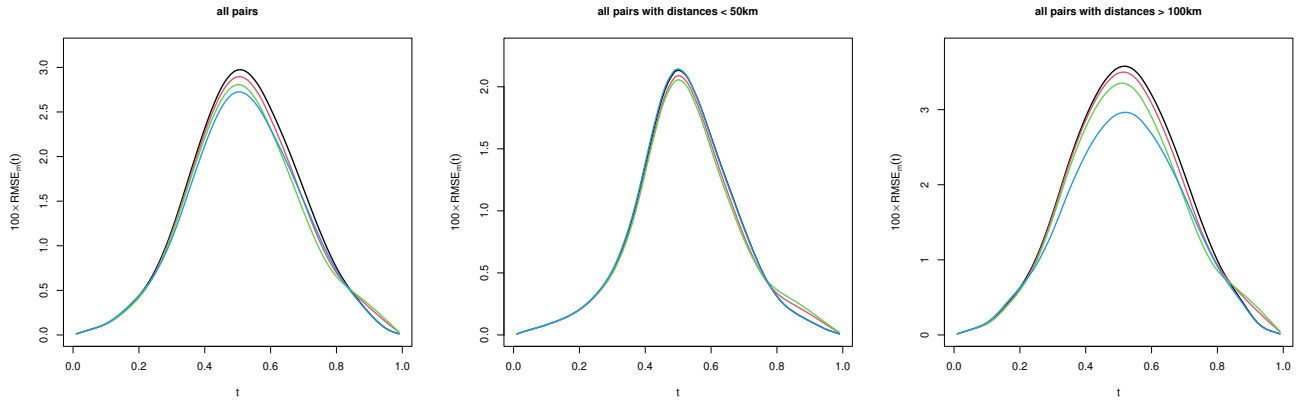


Figure 6:  $\text{RMSE}_m(t)$  values for  $t \in (0, 1)$  for models 1-4 (black, red, green, and blue lines, respectively), computed for all pairs of variables (left), for pairs separated by less than 50 km (middle), and for pairs separated by more than 100 km (right).

at short distances, Models 2-4 — whose tail-dependence functions are permutation-asymmetric —

achieve smaller RMSEs than Model 1 at larger distances, indicating nonstationarity in the data.

We also compute aggregated RMSEs defined by

$$\text{RMSE}_m = \int_0^1 \text{RMSE}_m(t) dt$$

as well as  $\text{RMSE}_m(0.5)$ , which reflects the accuracy of the estimated upper tail-dependence coefficients. Table 3 reports the RMSE values for the different sets of pairs,  $S$ .

| Model   | $100 \times \text{RMSE}_m / 100 \times \text{RMSE}_m(0.5)$ |              |               |
|---------|------------------------------------------------------------|--------------|---------------|
|         | (all pairs)                                                | (dst < 50km) | (dst > 100km) |
| $m = 1$ | 1.24/2.97                                                  | 0.73/2.14    | 1.57/3.56     |
| $m = 2$ | 1.20/2.90                                                  | 0.71/2.09    | 1.53/3.49     |
| $m = 3$ | 1.15/2.81                                                  | 0.71/2.06    | 1.46/3.35     |
| $m = 4$ | 1.14/2.73                                                  | 0.73/2.15    | 1.34/2.95     |

Table 3:  $100 \cdot \text{RMSE}_m / 100 \cdot \text{RMSE}_m(0.5)$  values for Models 1-4, computed for all pairs of variables (left), for pairs separated by less than 50 km (middle), and for pairs separated by more than 100 km (right).

Again, all models perform similarly at short distances, while Models 2–4 achieve smaller RMSEs than Model 1 at larger distances. Among these, Model 4 attains the lowest RMSEs, showing a 17–21% reduction relative to Model 1 at larger distances, indicating a better overall fit to the data.

## 5.2 UK Bank Returns

This second illustration focuses on the analysis of weekly negative raw returns for stocks of four major UK banks (HSBC, Lloyds, RBS, and Barclays) recorded between 29 October 2007 and 17 October 2016, yielding  $n = 470$  observations. A multivariate Peaks-over-Thresholds approach was applied to this dataset by Kirilouk et al. (2019), and we adopt their proposed marginal threshold selection at the 83% quantile, leaving 149 observations for which at least one component exceeds the threshold.

Since our primary interest lies in modeling extremal dependence, the margins are standardized to unit Pareto, and we consider both the Hüsler-Reiss model and its asymmetric counterpart as defined in Proposition 4. As demonstrated earlier, these models are fully defined in the interior of the support, which allows inference to be conducted directly using (9), a simpler and more elegant approach than using (10) when  $d = 4$ .

For the model in Proposition 4, we assume  $\alpha_1 = \dots = \alpha_d = \alpha_0$ , which leads to the following parametrization:  $\lambda_{ij} = \alpha_0 (2(1 - \rho_{ij}))^{1/2}$  and  $\tau_j = \alpha_0 \rho_{j,0} - c_0$ . Thus, we estimate the following 10-dimensional parameter vector:

$$\boldsymbol{\theta}_D = (\lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}, \lambda_{34}, \tau_1, \tau_2, \tau_3, \tau_4)^\top,$$

while the Hüsler-Reiss model is recovered by setting  $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ . Table 4 presents the

| Model        | $\lambda_{12}$ | $\lambda_{13}$ | $\lambda_{14}$ | $\lambda_{23}$ | $\lambda_{24}$ | $\lambda_{34}$ | $\tau_1$       | $\tau_2$        | $\tau_3$       | $\tau_4$        |
|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|----------------|-----------------|
| Hüsler-Reiss | 1.10<br>(0.05) | 1.04<br>(0.04) | 1.03<br>(0.04) | 0.94<br>(0.05) | 0.96<br>(0.04) | 0.96<br>(0.05) | —<br>—         | —<br>—          | —<br>—         | —<br>—          |
| Prop. 4      | 1.34<br>(0.14) | 1.07<br>(0.05) | 1.10<br>(0.12) | 1.39<br>(0.17) | 1.06<br>(0.08) | 1.10<br>(0.16) | 0.30<br>(0.47) | −0.59<br>(0.60) | 0.66<br>(0.47) | −0.12<br>(0.61) |

Table 4: Parameter estimates, along with parametric bootstrap estimates of their standard deviations, for the Hüsler–Reiss and Proposition 4 models fitted to the UK Bank dataset. The data were transformed to unit Pareto margins, with thresholds set at the 83% marginal quantiles.

estimated parameters for both models, along with bootstrap estimates of their standard deviations. The  $\lambda_{ij}$  estimates are broadly similar across models, whereas Model 4 shows larger bootstrap standard deviations, as expected due to the additional parameters being estimated.

Although the individual confidence intervals for each  $\tau_j$  appear to include zero, both the AIC and BIC are substantially lower when allowing for asymmetry (1722 vs. 1825 for AIC and 1752 vs. 1843 for BIC), indicating a better fit for the Proposition 4 model. Furthermore, since the

models are nested, this conclusion is supported by a likelihood ratio test of the null hypothesis  $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ , which yields a  $p$ -value less than 0.001.

## 6 Conclusion

In this paper, we introduced a new class of additive factor models, derived their corresponding extreme-value limiting distributions, and studied their tail behavior. By extending the common factor model of [Krupskii et al. \(2018a\)](#), we were able to introduce permutation asymmetry in the dependence structure of extremes across multiple random variables, thereby increasing modeling flexibility compared to the popular (symmetric) Hüsler–Reiss model, which can be recovered as a special case. Extremes are modeled from both the Block Maxima and Peaks-over-Thresholds perspectives, making the approach applicable to a wide range of multivariate extremes problems. Several sub-models are derived from the proposed family — some capturing joint extremes of the random vector of interest, and others not — illustrating their modeling capabilities. All considered models share the appealing property of being able to accommodate non-stationarity.

Simulation experiments are included to demonstrate that the model parameters can be reliably recovered, while two analyses of real datasets —consisting of Block Maxima and Peaks-over-Thresholds data — showcase an improved quality of fit compared to the Hüsler–Reiss model. Although the primary focus of this manuscript is on multivariate extremes, some illustrations are also provided in the context of spatial data. Many of the proposed models assign positive mass to some or all of the boundaries of their support, which makes inference challenging in moderate to high dimensions. A direction for future research would therefore be to develop a class of parsimonious factor models that, by construction, are fully characterized within the interior of their support.

In addition, multi-factor models can better capture dependence over large spatial domains and therefore extensions of the proposed factor models could also be a topic of future research.

## Acknowledgements

Boris Béranger was supported by the Australian Research Council (ARC) through the Discovery Project Scheme (DP220103269).

## A Appendix

### A.1 Proof of Proposition 1

Consider a vector  $\mathbf{w} = (w_1, \dots, w_d)^\top$  with  $w_i = \alpha_i^{-1} F_0^{-1} \{1 - x_i/(n\zeta_i)\}$  for  $i = 1, \dots, d$ . We find:

$$\begin{aligned} \Pr(\mathbf{X} < \mathbf{w}) &= \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} \Pr\{\mathbf{Z}_{\mathcal{A}} + (1/\alpha_{\mathcal{A}}) \cdot V_0 < \mathbf{w}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}} < \mathbf{w}_{\mathcal{B}}, \mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}\} \\ &= \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} \Pr\{\mathbf{Z}_{\mathcal{A}} + (1/\alpha_{\mathcal{A}}) \cdot V_0 < \mathbf{w}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}\} + o(1/n). \end{aligned}$$

Denote  $i^* = \operatorname{argmin}_{i \in \mathcal{A}}(w_i \alpha_i)$ . Note that

$$\begin{aligned} P_{\mathcal{A}} &:= \Pr\{\mathbf{Z}_{\mathcal{A}} + (1/\alpha_{\mathcal{A}}) \cdot V_0 < \mathbf{w}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}\} \\ &= \int_x \Pr(\mathbf{Z}_{\mathcal{A}} < \mathbf{w}_{\mathcal{A}} - y/\alpha_{\mathcal{A}}, \mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}) dF_0(y) = P_{\mathcal{A},1} + P_{\mathcal{A},2} + P_{\mathcal{A},3}, \end{aligned}$$

where, with  $w_{L,n} = w_{i^*}\alpha_{i^*} - (\ln n)^{\alpha_0}$  and  $w_{U,n} = w_{i^*}\alpha_{i^*} + (\ln n)^{\alpha_0}$ ,

$$\begin{aligned}
P_{A,1} &= \int_{y < w_{L,n}} \Pr(\mathbf{Z}_A < \mathbf{w}_A - y/\alpha_A, \mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) dF_0(y) \\
&= \int_{y < w_{L,n}} \Pr(\mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) dF_0(y) - v_n \\
&= F_0(w_{i^*}\alpha_{i^*}) \cdot \Pr(\mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) - v_n + o(1/n), \\
P_{A,2} &= \int_{w_{L,n}}^{w_{U,n}} \Pr(\mathbf{Z}_A < \mathbf{w}_A - y/\alpha_A, \mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) dF_0(y) \\
&< F_0(w_{U,n}) - F_0(w_{L,n}) = o(1/n), \\
P_{A,3} &:= \int_{y > w_{U,n}} \Pr(\mathbf{Z}_A < \mathbf{w}_A - y/\alpha_A, \mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) dF_0(y) \\
&< \Phi\{-(\ln n)^{\alpha_0}/\alpha_{i^*}\} \bar{F}_0(w_{U,n}) = o(1/n).
\end{aligned}$$

Here,

$$\begin{aligned}
v_n &= \int_{y < w_{L,n}} \Pr(\mathbf{Z}_A > \mathbf{w}_A - y/\alpha_A, \mathbf{Z}_A^* > \mathbf{c}_A, \mathbf{Z}_B^* < \mathbf{c}_B) dF_0(y) \\
&< \int_{y < w_{L,n}} \Phi(y/\alpha_{i^*} - w_{i^*}) dF_0(y) = v_{n,1} + v_{n,2},
\end{aligned}$$

where, using the integration by parts formula, we get:

$$\begin{aligned}
v_{n,1} &= -\{1 - F_0(w_{L,n})\} \Phi\{-(\ln n)^{\alpha_0}\} = o(1/n), \\
v_{n,2} &= \alpha_{i^*}^{-1} \int_{y < w_{L,n}} \{1 - F_0(y)\} \phi(y/\alpha_{i^*} - w_{i^*}) dy = \int_{y < -(\ln n)^{\alpha_0}} \{1 - F_0(z\alpha_{i^*} + w_{i^*}\alpha_{i^*})\} \phi(z) dz \\
&= \int_{-(\ln n)^{\alpha^*}}^{-(\ln n)^{\alpha_0}} \{1 - F_0(z\alpha_{i^*} + w_{i^*}\alpha_{i^*})\} \phi(z) dz + o(1/n) \\
&\leq \frac{x_{i^*}}{n\zeta_{i^*}} \int_{-(\ln n)^{\alpha^*}}^{-(\ln n)^{\alpha_0}} \exp(\gamma z\alpha_{i^*}) \phi(z) dz + o(1/n) \\
&= \frac{x_{i^*}}{n\zeta_{i^*}} \exp(0.5\gamma^2\alpha_{i^*}^2) [\Phi\{-(\ln n)^{\alpha_0} - \gamma\alpha_{i^*}\} - \Phi\{-(\ln n)^{\alpha^*} - \gamma\alpha_{i^*}\}] + o(1/n) = o(1/n),
\end{aligned}$$

where  $\alpha^* \in (0.5, 1)$  is a constant. It follows that

$$\begin{aligned}\Pr(\mathbf{X} < \mathbf{w}) &= \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} F_0\{\min_{i \in \mathcal{A}}(w_i \alpha_i)\} \cdot \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}) + o(1/n) \\ &= \left\{1 - \max_{i \in \mathcal{A}} \frac{x_i}{n \zeta_i}\right\} \cdot \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}) + o(1/n).\end{aligned}$$

In particular,

$$\Pr(X_i < w_i) = \{1 - x_i/(n \zeta_i)\} \Pr(Z_i^* > c_i) + \Pr(Z_i^* < c_i) + o(1/n) = 1 - x_i/n + o(1/n),$$

and

$$\ell_{\mathbf{X}}(\mathbf{x}) = \lim_{n \rightarrow \infty} n\{1 - \Pr(\mathbf{X} < \mathbf{w})\} = \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} \left(\max_{i \in \mathcal{A}} \frac{x_i}{\zeta_i}\right) \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}).$$

The proof is now complete. □

## A.2 Proof of Proposition 2

Let  $F_i(z)$  be the cdf of  $X_i$  as before. We can write

$$\begin{aligned}F_i(z) &= \Pr(X_i < z) = \Pr(X_i < z, Z_i^* > c_i) + \Pr(X_i < z, Z_i^* < c_i) \\ &= \int_{\mathbb{R}^2} \Pr\{V_0 < \alpha_i(z - Z_i), Z_i^* > c_i | Z_i = z_i, Z_i^* = z_i^*\} \phi_{\rho_{i,i^*}}(z_i, z_i^*) dz_i dz_i^* + \Pr(Z_i < z, Z_i^* < c_i) \\ &= \int_{c_i}^{\infty} \int_{-\infty}^z [1 - \exp\{-\alpha_i(z - z_i)\}] \phi_{\rho_{i,i^*}}(z_i, z_i^*) dz_i dz_i^* + \Phi_{\rho_{i,i^*}}(z, c_i) \\ &= \Phi(z) - \exp(0.5\alpha_i^2 - \alpha_i z) \Phi_{-\rho_{i,i^*}}(\alpha_i \rho_{i,i^*} - c_i, z - \alpha_i).\end{aligned}$$

As  $z \rightarrow \infty$ , we have  $\Phi(z) \rightarrow 1$  and  $\Phi_{-\rho_{i,i^*}}(\alpha_i \rho_{i,i^*} - c_i, z - \alpha_i) \rightarrow \Phi(\alpha_i \rho_{i,i^*} - c_i)$ . Solving

$$1 - \exp(0.5\alpha_i^2 - \alpha_i z) \Phi(\alpha_i \rho_{i,i^*} - c_i) = 1 - x/n,$$

we obtain  $z = z_i(x/n) = 0.5\alpha_i + \alpha_i^{-1} \ln(n/x) + \alpha_i^{-1} \ln \Phi(\alpha_i \rho_{i,i^*} - c_i) = 0.5\alpha_i + \alpha_i^{-1} \ln(n \zeta_i/x)$  since  $\rho_{i,i^*} = 0$ . It is easy to see that  $F_i(z_i(x/n)) = 1 - x/n + o(1/n)$  as  $n \rightarrow \infty$ .

Now let  $\tilde{F}_i$  be the cdf of  $\tilde{X}_i = Z_i + (1/\alpha_i) \cdot V_0$ , and let  $w_i = z_i(x_i/n)$ , then  $\tilde{F}_i(w_i) = 1 - x_i/(\zeta_i n) + o(1/n)$ . It follows that

$$\begin{aligned} \Pr(\mathbf{X} < \mathbf{w}) &= \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} \Pr\{\mathbf{Z}_{\mathcal{A}} + (1/\alpha_{\mathcal{A}}) \cdot V_0 < \mathbf{w}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}} < \mathbf{w}_{\mathcal{B}}, \mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}\} \\ &= \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} \Pr\{\tilde{\mathbf{X}}_{\mathcal{A}} = \mathbf{Z}_{\mathcal{A}} + (1/\alpha_{\mathcal{A}}) \cdot V_0 < \mathbf{w}_{\mathcal{A}}\} \cdot \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}) + o(1/n), \end{aligned}$$

and

$$n\{1 - \Pr(\mathbf{X} < \mathbf{w})\} = \sum_{\mathcal{A} \in \mathcal{I}, \mathcal{B} = \mathcal{A}^c} n\{1 - \Pr(\tilde{\mathbf{X}}_{\mathcal{A}} < \mathbf{w}_{\mathcal{A}})\} \cdot \Pr(\mathbf{Z}_{\mathcal{A}}^* > \mathbf{c}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}^* < \mathbf{c}_{\mathcal{B}}) + o(1).$$

To conclude the proof, we notice that  $n\{1 - \Pr(\tilde{\mathbf{X}}_{\mathcal{A}} < \mathbf{w}_{\mathcal{A}})\} \rightarrow \ell_{\mathcal{A}}(\tilde{\mathbf{x}}_{\mathcal{A}})$  (Krupskii et al., 2018b).  $\square$

### A.3 Proof of Proposition 3

Without loss of generality, consider  $i = 1$  and  $j = 2$ . Let  $F_{12}$  be the joint cdf of  $(X_1, X_2)^\top$  as defined in (6), and  $F_1, F_2$  be the respective marginal cdfs. From the proof of Proposition 2, if we let  $z_l(x/n) = 0.5\alpha_l + \alpha_l^{-1} \ln(n/x) + \alpha_l^{-1} \ln \Phi(\alpha_l \rho_{l,l^*} - c_l)$ , then  $F_l(z_l(x/n)) = 1 - x/n + o(1/n)$  as  $n \rightarrow \infty$ .

Denote  $w_l = z_l(x_l/n)$  and  $\mathbf{z} = (z_1, z_2, z_1^*, z_2^*)^\top$ . Let  $\Phi_{i,j,k^*}$  be the cdf of the vector  $(Z_i, Z_j, Z_k^*)^\top$  and  $\Phi_{i,j,k^*,l^*}, \phi_{i,j,k^*,l^*}$  be the cdf (pdf, respectively) of the vector  $(Z_i, Z_j, Z_k^*, Z_l^*)^\top$ . We find:

$$F_{12}(w_1, w_2) = \Pr(X_1 < w_1, X_2 < w_2) = A_{12} + A_1 + A_2 + A_0,$$

where

$$\begin{aligned}
A_{12} &:= \Pr(X_1 < w_1, X_2 < w_2, Z_1^* > c_1, Z_2^* > c_2) \\
&= \int_{S_{12}} (1 - \exp[\max\{\alpha_1(z_1 - w_1), \alpha_2(z_2 - w_2)\}]) \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} \\
&= \Phi_{\rho_{1,2}}(w_1, w_2) - \Phi_{1,2,1^*}(w_1, w_2, c_1) - \Phi_{1,2,2^*}(w_1, w_2, c_2) + \Phi_{1,2,1^*,2^*}(w_1, w_2, c_1, c_2) \\
&\quad - \int_{S_{1|2}} \exp\{\alpha_1(z_1 - w_1)\} \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} - \int_{S_{2|1}} \exp\{\alpha_2(z_2 - w_2)\} \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z}, \\
A_1 &:= \Pr(X_1 < w_1, X_2 < w_2, Z_1^* > c_1, Z_2^* < c_2) = \int_{S_1} [1 - \exp\{\alpha_1(z_1 - w_1)\}] \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} \\
&= \Phi_{1,2,2^*}(w_1, w_2, c_2) - \Phi_{1,2,1^*,2^*}(w_1, w_2, c_1, c_2) - \int_{S_1} \exp\{\alpha_1(z_1 - w_1)\} \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z}, \\
A_2 &:= \Pr(X_1 < w_1, X_2 < w_2, Z_1^* < c_1, Z_2^* > c_2) = \int_{S_2} [1 - \exp\{\alpha_2(z_2 - w_2)\}] \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} \\
&= \Phi_{1,2,1^*}(w_1, w_2, c_1) - \Phi_{1,2,1^*,2^*}(w_1, w_2, c_1, c_2) - \int_{S_1} \exp\{\alpha_1(z_1 - w_1)\} \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z}, \\
A_0 &:= \Pr(X_1 < w_1, X_2 < w_2, Z_1^* < c_1, Z_2^* < c_2) = \Phi_{1,2,1^*,2^*}(w_1, w_2, c_1, c_2),
\end{aligned}$$

and  $S_{12} = \{\mathbf{z} : z_1 < w_1, z_2 < w_2, z_1^* > c_1, z_2^* > c_2\}$ ,  $S_l = \{\mathbf{z} : z_1 < w_1, z_2 < w_2, z_l^* > c_l, z_{-l}^* < c_{-l}\}$ .

Here,  $S_{l|-l} = \{\mathbf{z} \in S_{1,2} : \alpha_l(z_l - w_l) > \alpha_{-l}(z_{-l} - w_{-l})\} = \{\mathbf{z} \in S_{1,2} : z_l > \alpha_{-l}z_{-l}/\alpha_l + \xi_{l,-l}\}$ , where

$$\xi_{l,-l} = \alpha_l^{-1} \ln(x_{-l}/x_l) + 0.5\alpha_l - 0.5\alpha_{-l}^2/\alpha_l + \alpha_l^{-1} \ln\{\zeta_l/\zeta_{-l}\},$$

with  $\zeta_l = \Phi(\alpha_l \rho_{l,l^*} - c_l)$ ,  $l \in \{1, 2\}$ . It implies that

$$\begin{aligned}
1 - C_{X_1, X_2}(1 - x_1/n, 1 - x_2/n) &= \exp(-\alpha_1 w_1) \int_{S_{1^*}^* \cup S_{1|2}^*} \exp(\alpha_1 z_1) \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} \\
&\quad + \exp(-\alpha_2 w_2) \int_{S_{2^*}^* \cup S_{2|1}^*} \exp(\alpha_2 z_2) \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} + o(1/n),
\end{aligned}$$

where  $S_l^* = \{\mathbf{z} : z_l^* > c_l, z_{-l}^* < c_{-l}\}$ ,  $S_{l|-l}^* = \{\mathbf{z} : z_l > \alpha_{-l}z_{-l}/\alpha_l + \xi_{l,-l}, z_1^* > c_1, z_2^* > c_2\}$ .

Note that  $\exp(-\alpha_l w_l) = \frac{x_l}{n\zeta_l} \cdot \exp(-0.5\alpha_l^2)$ , and

$$\begin{aligned} I_l &:= \int_{S_l^* \cup S_{l-l}^*} \exp(\alpha_l z_l) \phi_{1,2,1^*,2^*}(\mathbf{z}) d\mathbf{z} \\ &= \exp(0.5\alpha_l^2) \left\{ \zeta_l - \Phi_{\Sigma_{l-l}} \left( -\frac{\lambda_{12}}{2} - \frac{1}{\lambda_{12}} \ln \frac{\tilde{x}_l}{\tilde{x}_{-l}}, \alpha_l \rho_{-l^*,l} - c_{-l}, \alpha_l \rho_{l^*,l} - c_l \right) \right\}. \end{aligned}$$

The result of proposition now follows by taking the limit:

$$\ell_{X_1, X_2}(x_1, x_2) = \lim_{n \rightarrow \infty} n \left\{ 1 - C_{X_1, X_2} \left( 1 - \frac{x_1}{n}, 1 - \frac{x_2}{n} \right) \right\}. \quad \square$$

#### A.4 Proof of Proposition 4

Consider a vector  $\mathbf{w} = (w_1, \dots, w_d)^\top$  with  $w_l = w_l(n) = 0.5\alpha_l + \alpha_l^{-1} \ln(n/x_l) + \alpha_l^{-1} \ln \Phi(\alpha_l \rho_{l,0} - c_0)$  for  $l = 1, \dots, d$ , then from the proof of Proposition 2,  $F_l(w_l) = 1 - x_l/n + o(1/n)$ , where  $F_l$  is the cdf of  $X_l$  as before. We can write:

$$\Pr(\mathbf{X} < \mathbf{w}) = I_1(\mathbf{w}) + I_2(\mathbf{w}),$$

where

$$I_1(\mathbf{w}) = \Pr(\mathbf{X} < \mathbf{w}, Z_0^* < c_0) = \Pr(\mathbf{Z} < \mathbf{w}, Z_0^* < c_0) = \Phi(c_0) + o(1/n),$$

and, using the integration by parts formula, we find:

$$\begin{aligned} I_2(\mathbf{w}) &= \Pr(\mathbf{X} < \mathbf{w}, Z_0^* > c_0) = \Pr(\mathbf{Z} + (1/\boldsymbol{\alpha}) \cdot V_0 < \mathbf{w}, -Z_0^* < -c_0) \\ &= \int_0^\infty \Phi_{d+1} \left( (\mathbf{w} - x/\boldsymbol{\alpha}, -c_0)^\top; \Sigma_0 \right) \exp(-x) dx \\ &= \Phi_{d+1} \left( (\mathbf{w}, -c_0)^\top; \Sigma_0 \right) - \sum_{j=1}^d I_{3,j}(\mathbf{w}) = \Phi(-c_0) - \sum_{j=1}^d I_{3,j}(\mathbf{w}) + o(1/n), \end{aligned}$$

where  $\Sigma_0$  is the correlation matrix of  $(\mathbf{Z}^\top, -Z_0^*)^\top$ . By changing the variables, we get:

$$\begin{aligned}
I_{3,j}(\mathbf{w}) &= (1/\alpha_j) \int_0^\infty \Phi_d \left( \left[ \left\{ \frac{w_i - x/\alpha_i - \rho_{ij}(w_j - x/\alpha_j)}{\tilde{\rho}_{ij}} \right\}_{i \in \mathcal{I}_{-j}}, \frac{-c_0 - \rho_{j,0}(w_j - x/\alpha_j)}{\tilde{\rho}_{j,0}} \right]^\top; \Sigma_{0|j} \right) \\
&\quad \times \phi(w_j - x/\alpha_j) \exp(-x) dx \\
&= \frac{\tilde{x}_j}{n} \int_{-\infty}^{w_j - \alpha_j} \Phi_d \left( \left[ \left\{ \frac{\varrho_{ij} + (\alpha_j/\alpha_i - \rho_{ij})z}{\tilde{\rho}_{ij}} \right\}_{i \in \mathcal{I}_{-j}}, \frac{-c_0 + \rho_{j,0}\alpha_j + \rho_{j,0}z}{\tilde{\rho}_{j,0}} \right]^\top; \Sigma_{0|j} \right) \phi(z) dz \\
&= \frac{\tilde{x}_j}{n} \Phi_d \left[ \left\{ \left( \frac{\lambda_{ij}}{2} + \frac{1}{\lambda_{ij}} \ln \frac{\tilde{x}_j}{\tilde{x}_i} \right)_{i \in \mathcal{I}_{-j}}; \alpha_j \rho_{j,0} - c_0 \right\}^\top; \tilde{\Sigma}^{(j)} \right] + o(1/n).
\end{aligned}$$

Here,  $\Sigma_{0|j}$  is the partial correlation matrix of  $(\mathbf{Z}^\top, -Z_0^*)^\top$  conditional on  $Z_j$ ,  $\tilde{\rho}_{ij} = (1 - \rho_{ij}^2)^{1/2}$ ,  $\tilde{\rho}_{j,0} = (1 - \rho_{j,0}^2)^{1/2}$ , and  $\varrho_{ij} = (0.5/\alpha_i) \cdot \lambda_{ij}^2 + (1/\alpha_i) \cdot \ln(\tilde{k}_j/\tilde{k}_i)$ . The stable tail dependence function is obtained by taking the limit:

$$\begin{aligned}
\ell_{\mathbf{X}}(\mathbf{x}) &= \lim_{n \rightarrow \infty} n \{1 - \Pr(\mathbf{X} < \mathbf{w})\} \\
&= \sum_{j=1}^d \tilde{x}_j \Phi_d \left[ \left\{ \left( \frac{\lambda_{ij}}{2} + \frac{1}{\lambda_{ij}} \log \frac{\tilde{x}_j}{\tilde{x}_i} \right)_{i \in \mathcal{I}_{-j}}; \alpha_j \rho_{j,0} - c_0 \right\}^\top; \tilde{\Sigma}^{(j)} \right],
\end{aligned}$$

and the result of proposition now follows by recognizing the cdf of the extended Skew-Normal distribution [Arellano-Valle and Genton \(see 2010\)](#). □

## A.5 Illustrations of the Marshall-Olkin distribution

Consider the Marshall–Olkin model with  $c = (1.2, 2)$  and correlation  $\rho = 0.6$  between  $Z_1^*$  and  $Z_2^*$ .

The top-left panel of Figure [7](#) illustrates the EV copula through its density

$$c_{\text{EV}}(u_1, u_2) = \begin{cases} (1 - \alpha)u_1^\alpha, & u_1 > u_2^{\beta/\alpha} \\ (1 - \beta)u_2^\beta, & \text{otherwise,} \end{cases}$$

with the black solid line representing the edge case  $u_2 = u_1^{\alpha/\beta}$ . The top-middle panel shows the multivariate EV density  $g(\mathbf{x}) = \exp\{-\ell(1/\mathbf{x})\}(1 - \alpha)(1 - \beta)/(x_1 x_2)^2$ , where the yellow dots correspond to 3,000 randomly generated component-wise maxima. In the top-right panel, the density

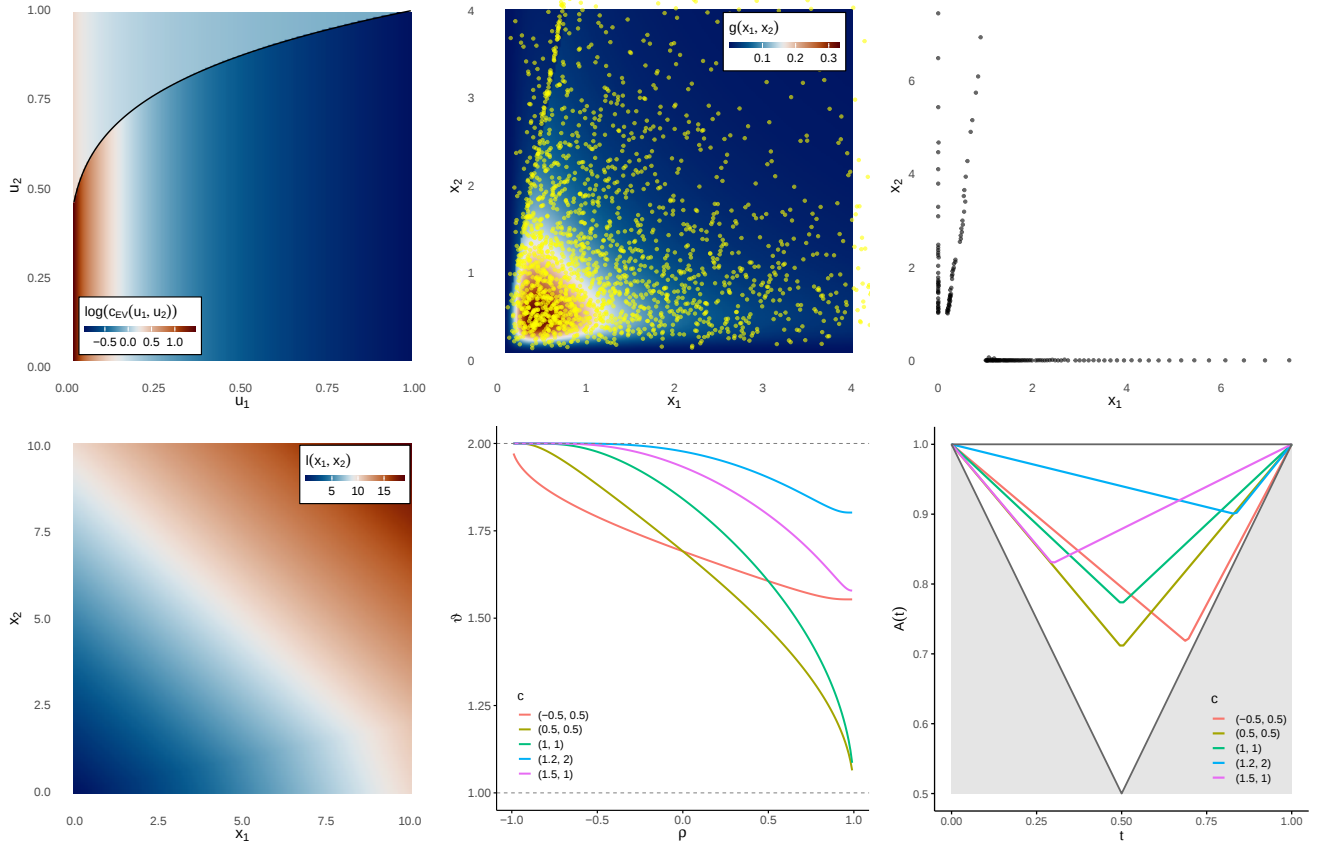


Figure 7: Left to right, top to bottom: density of the extreme value copula (on log scale,  $\log c_{EV}$ ), density of the bivariate extreme value distribution  $g(\mathbf{x})$ , exceedances above the  $(1, 1)$  threshold on unit Pareto margins, and the stable tail dependence function  $\ell(\mathbf{x})$  of the Marshall-Olkin model with  $c = (1.2, 2)$  and correlation  $\rho = 0.6$  between  $Z_1^*$  and  $Z_2^*$ . The remaining panels display for several choices of  $c$ , the extremal coefficient  $\vartheta$  as function of  $\rho$  and the Pickands dependence function with  $\rho = 0.6$ .

of the bivariate generalized Pareto is intentionally left blank (except for some simulated points) because the model does not admit a closed-form density. The bottom-left panel depicts the stable tail dependence function, while the bottom-middle and bottom-right panels provide intuition about the type of dependence structure captured by this model through the extremal coefficient and the Pickands dependence function.

## A.6 Illustrations of Submodel 2

Consider Submodel 2 with  $\mathbf{c} = (0.8, 1.2)$ ,  $\alpha_0 = 1$  and correlation  $\rho_{1,2} = 0.6$ . The key difference with the illustrations of Submodel 1 from Figure 3 is reduced from  $\rho^* = 1$  to  $\rho^* = 0$ , which leads to a higher degree of independence among the extremes. The six panels of Figure 8 can be directly compared with those of Figure 3. We note that the multivariate generalized Pareto distribution now assigns mass to both boundaries ( $x_1 = 0$  and  $x_2 = 0$ ) as displayed in the top-right panel.

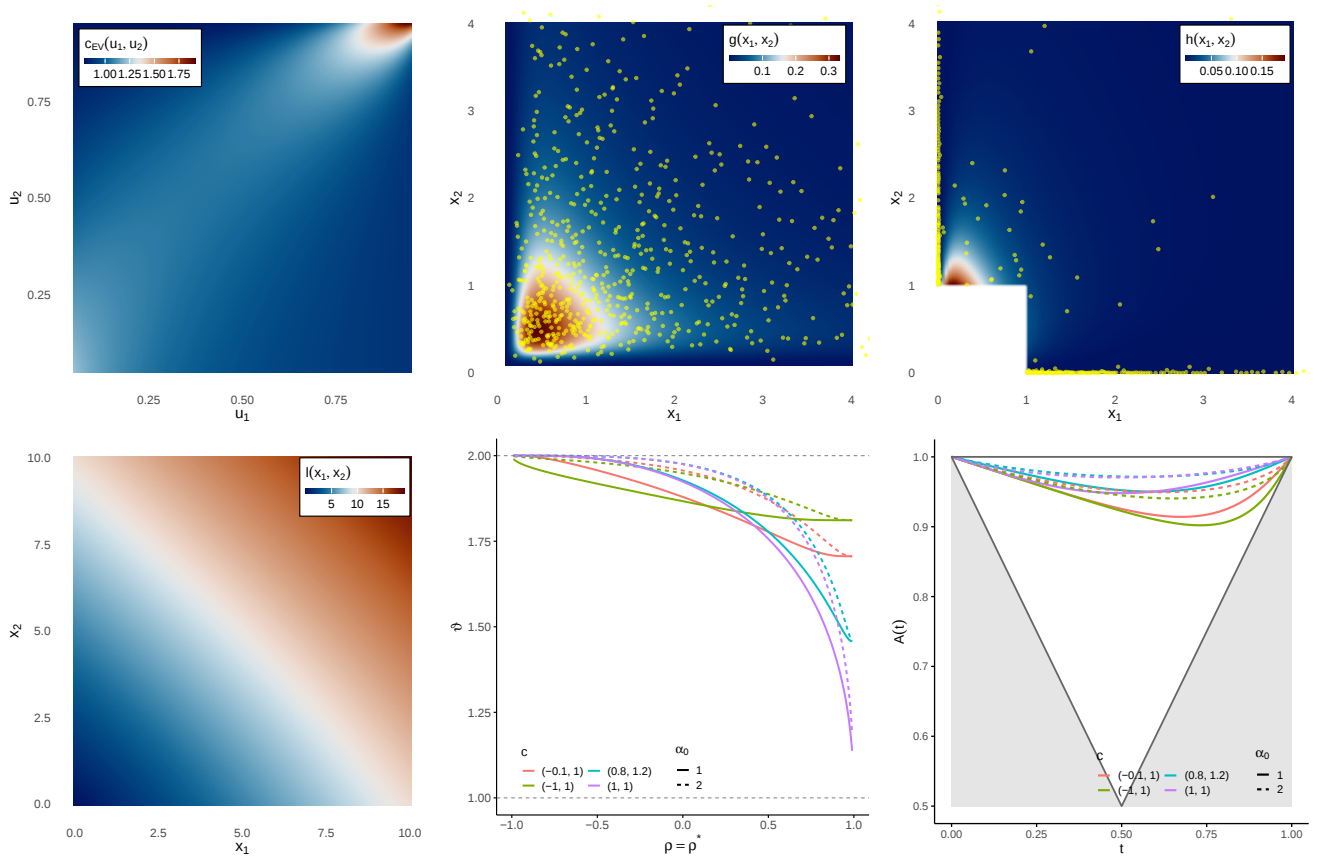


Figure 8: Left to right, top to bottom: density of the extreme value copula (on log scale,  $\log c_{EV}$ ), density of the bivariate extreme value distribution  $g(\mathbf{x})$ , exceedances above the  $(1, 1)$  threshold on unit Pareto margins, and the stable tail dependence function  $\ell(\mathbf{x})$  of the Submodel 2 with  $\mathbf{c} = (0.8, 1.2)$ ,  $\alpha_0 = 1$ , correlation  $\rho = \rho_{1,2} = 0.6$  and  $\rho^* = 0$ . The remaining panels display for several choices of  $\mathbf{c}$ , the extremal coefficient  $\vartheta$  as function of  $\rho (= \rho^*)$  and the Pickands dependence function with  $\rho = 0.6$ .

## References

- Arellano-Valle, R.B., Genton, M.G., 2010. Multivariate extended skew- $t$  distributions and related families. *Metron* 68, 201–234.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L., 2004. *Statistics of extremes: theory and applications*. John Wiley & Sons.
- Beranger, B., Padoan, S., Marcon, G., 2025. *ExtremalDep: Extremal Dependence Models*. R package version 1.0.0.
- Beranger, B., Padoan, S., Xu, Y., Sisson, S., 2019. Extremal properties of the multivariate extended skew-normal distribution, part b. *Statistics & Probability Letters* 147, 105 – 114.
- Beranger, B., Padoan, S.A., 2025. Modelling extremal dependence in high-dimensions: A practical perspective. [2412.13453](#).
- Beranger, B., Padoan, S.A., Sisson, S.A., 2017. Models for extremal dependence derived from skew-symmetric families. *Scandinavian Journal of Statistics* 44, 21–45. [10.1111/sjos.12240](#).
- Beranger, B., Stephenson, A.G., Sisson, S.A., 2021. High-dimensional inference using the extremal skew- $t$  process. *Extremes* 24, 653–685.
- Dey, D.K., Yan, J. (Eds.), 2016. *Extreme Value Modeling and Risk Analysis: Methods and Applications*. CRC Press, Boca Raton, FL.
- Drees, H., Huang, X., 1998. Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* 64, 25–47.

- Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M., 2015. Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77, 239–265.
- Goix, N., Sabourin, A., Cléménçon, S., 2017. Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis* 161, 12–31.
- Gudendorf, G., Segers, J., 2010. Extreme-value copulas, in: Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (Eds.), *Copula Theory and Its Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 127–145.
- Gumbel, E.J., 1960. Bivariate exponential distributions. *Journal of the American Statistical Association* 55, 698–707.
- de Haan, L., Ferreira, A., 2006. *Extreme Value Theory*. Springer-Verlag, New York.
- Huang, X., 1992. Statistics of bivariate extreme values. Dissertation, thesis. Zugl.: Rotterdam, Univ., Diss., 1991.
- Huser, R., Davison, A.C., Genton, M.G., 2016. Likelihood estimators for multivariate extremes. *Extremes* 19, 79–103.
- Hüsler, J., Reiss, R. D., 1989. Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters* 7, 283–286.
- Joe, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

- Joe, H., Xu, James J., 1996. The estimation method of inference functions for margins for multivariate models. Technical Report # 166 .
- Kiriliouk, A., Lee, J., Segers, J., 2024. X-vine models for multivariate extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 87, 579–602.
- Kirilouk, A., Rootzén, H., Segers, J., Wadsworth, J.L., 2019. Peaks over thresholds modeling with multivariate generalized pareto distributions. *Technometrics* 61, 123–135.
- Krupskii, P., Genton, M.G., 2021. Conditional normal extreme-value copulas. *Extremes* 24(3), 403–431.
- Krupskii, P., Huser, R., Genton, M.G., 2018a. Factor copula models for replicated spatial data. *Journal of the American Statistical Association* 521, 467–479.
- Krupskii, P., Joe, H., 2013. Factor copula models for multivariate data. *Journal of Multivariate Analysis* 120, 85–101.
- Krupskii, P., Joe, H., 2015. Structured factor copula models: theory, inference and computation. *Journal of Multivariate Analysis* 138, 53–73.
- Krupskii, P., Joe, H., Lee, D., Genton, M.G., 2018b. Extreme value limit of the convolution of exponential and multivariate normal distributions: Link to the Hüsler-Reiss distribution. *Journal of Multivariate Analysis* 163, 80–95.
- Lee, D., Joe, H., 2018. Multivariate extreme value copulas with factor and tree dependence structures. *Extremes* 21(1), 147–176.

- Li, H., 2009. Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis* 100, 243–256.
- Lindsay, B., 1998. Composite likelihood methods. *Contemporary Mathematics* 80, 220–239.
- Marshall, W. A., Olkin, I., 1967. A multivariate exponential distribution. *Journal of the American Statistical Association* 62, 30–44.
- Naveau, P., Segers, J., 2024. Multivariate extreme value theory. [2412.18477](#).
- Opitz, T., 2013. Extremal  $t$  processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis* 122, 409–413.
- Padoan, S.A., 2011. Multivariate extreme models based on underlying skew- and skew-normal distributions. *Journal of Multivariate Analysis* 102, 977 – 991.
- Resnick, S.I., 1987. *Extreme Values, Regular Variation, and Point Processes*. Springer Series in Operations Research and Financial Engineering, Springer-Verlag, New York.
- Rootzén, H., Segers, J., L. Wadsworth, J., 2018a. Multivariate peaks over thresholds models. *Extremes* 21, 115–145.
- Rootzén, H., Segers, J., Wadsworth, J.L., 2018b. Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis* 165, 117–131.
- Rootzén, H., Tajvidi, N., 2006. Multivariate generalized Pareto distributions. *Bernoulli* 12, 917–930.

- Simpson, E.S., Wadsworth, J.L., Tawn, J.A., 2020. Determining the dependence structure of multivariate extremes. *Biometrika* 107, 513–532.
- Sklar, A., 1959. Fonctions de répartition à  $n$  dimensions et leurs marges. Institute of Statistics of the University of Paris 8, 229–231.
- Thibaud, E., Opitz, T., 2015. Efficient inference and simulation for elliptical Pareto processes. *Biometrika* 102, 855–870.
- Zhong, P., Sisson, S.A., Beranger, B., 2025. Fast and flexible inference for spatial extremes [2407.13958](#).