Multivariate peaks-over-threshold with latent variable representations of generalized Pareto vectors

Jean-Noel Bacro, Carlo Gaetan, Thomas Opitz, Gwladys Toulemonde

Multivariate peaks-over-threshold with latent variable representations of generalized Pareto vectors

Jean-Noel Bacro[1], Carlo Gaetan[2], Thomas Opitz[3], Gwladys Toulemonde[1,4]
[1] IMAG, Montpellier University, France
[2] Dipartimento di Scienze Ambientali, Informatica e Statistica,
Ca' Foscari University of Venice, Venice, Italy
[3] Biostatistics and Spatial Processes, INRAE, Avignon, France
[4] LEMON, Inria, Montpellier, France

January 26, 2023

Abstract

Generalized Pareto distributions with positive tail index arise from embedding a Gamma random variable for the rate of an exponential distribution. In this paper, we exploit this property to define a flexible and statistically tractable modeling framework for multivariate extremes based on componentwise ratios between any two random vectors with exponential and Gamma marginal distributions. To model multivariate threshold exceedances, we propose hierarchical constructions using a latent random vector with Gamma margins, whose Laplace transform is key to obtaining the multivariate distribution function. The extremal dependence properties of such constructions, covering asymptotic independence and asymptotic dependence, are studied. We detail two useful parametric model classes: the latent Gamma vectors are sums of independent Gamma components in the first construction (called the convolution model), whereas they correspond to chi-squared random vectors in the second construction. Both of these constructions exhibit asymptotic independence, and we further propose a parametric extension (called beta-scaling) to obtain asymptotic dependence. We demonstrate good performance of likelihood-based estimation of extremal dependence summaries for several scenarios through a simulation study for bivariate and trivariate Gamma convolution models, including a hybrid model mixing bivariate subvectors with asymptotic dependence and independence.

Keywords : Latent variable; Multivariate censoring; Multivariate extreme-value theory; Multivariate generalized Pareto distribution; Pairwise likelihood; Threshold exceedance

1 Introduction

In this paper, we call Multivariate Generalized Pareto Distribution (MGPD) any probability distribution of a random vector with its margins following the univariate Generalized Pareto Distribution (GPD) given as

$$\mathrm{GPD}(y;\xi,\sigma) = \begin{cases} 1-(1+\xi y/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ 1-\exp(-y/\sigma), & \xi = 0, \end{cases}$$

for $y > 0$ with scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$ (also called the tail index), where $a_+ = \max(a, 0)$. Strictly speaking, the term MGPD is often used to refer to the class of distributions arising as limits for appropriately defined threshold exceedances in random vectors (Rootzén and Tajvidi, 2006; Falk et al., 2011; Kiriliouk et al., 2019). These distributions have univariate GPD tails, but the marginal distributions can differ from the GPD for lower values. They are used as natural statistical models for high threshold exceedances in extreme-value analysis through approaches known as multivariate peaks-over-threshold. Extensions of the limit theory to stochastic processes and relevant statistical applications were developed by Ferreira and De Haan (2014); Thibaud and Opitz (2015); Palacios-Rodriguez et al. (2020); Opitz et al. (2021); de Fondeville and Davison (2020). The MGPDs in our more general sense represent a variant of the wide class of multivariate Pareto distributions (modulo marginal transformations), which have been proven useful for numerous statistical applications; see the review of Arnold (2014).

The multivariate generalized Pareto distributions arising as limits in multivariate extreme-value theory are characterized by the property of peaks-over-threshold (POT) stability. As such, they provide useful models for the joint tail of several variables when data are asymptotically dependent, such that the strength of dependence does not decrease at increasingly high quantiles. In practice, however, the limit behavior is often not yet manifest in observed data. Then, it is useful to develop sub-asymptotic models that allow us to better capture the rate of convergence to the limit. This is important since many relevant environmental data show empirical characteristics corresponding to asymptotic independence, such that dependence ultimately vanishes at the most extreme quantiles but is still present at the observed extreme quantiles (Coles et al., 1999; Heffernan and Tawn, 2004; Davison et al., 2013; Tawn et al., 2018, e.g.,); see also the recent review of Huser and Wadsworth (2022). If this is the case, appropriate modeling and extrapolation of co-occurrence probabilities of extreme events requires capturing residual dependence in observed extreme events. Several flexible models allowing for asymptotic independence in threshold exceedances of multivariate data have been proposed (Ledford and Tawn, 1997; Wadsworth and Tawn, 2013; Wadsworth et al., 2017; Engelke et al., 2019, e.g.,), although the proposed approaches may suffer from weaknesses; for example, being constructions that are not natural and lack easy interpretation because they rely on a copula framework with arbitrary marginal transformation, or showing tractability issues owing to computer-intensive integrals arising from latent variable structures or censoring schemes. Moreover, a variety of approaches have been developed in the setting of temporal data (e.g., Bortot and Gaetan, 2014; Noven et al., 2018), spatial data (e.g., Wadsworth and Tawn, 2012; Opitz, 2016; Huser et al., 2017; Huser and Wadsworth, 2019) and spatio-temporal data (e.g. Bacro et al., 2020; Castro-Camilo et al., 2021), some of them allowing to smoothly bridge asymptotic dependence and independence (e.g., Huser et al., 2017; Huser and Wadsworth, 2019). Models based on certain kernel convolutions of a latent gamma process have been used by Noven et al. (2018) for time series and by Bacro et al. (2020) for space-time processes; their multivariate (i.e., finite-dimensional) distributions correspond to one of the constructions proposed in the following.

We here study a general construction principle for multivariate distributions with generalized Pareto margins and positive tail index. The assumption of a positive tail index is without loss of generality in practice since any generalized Pareto marginal distribution can be achieved through an appropriate probability integral transform available in analytical form. The approach is motivated from a classical mixture representation of the heavy-tailed GPD with $\xi > 0$ (Reiss and Thomas, 2007, p.157): by embedding a gamma-distributed random variable $G \sim \Gamma(\alpha, \beta)$ with shape $\alpha > 0$ and rate $\beta > 0$ for the rate of an exponential variable, independent of G, we obtain a random variable following a generalized Pareto distribution (GPD) with shape $1/\alpha > 0$ and scale parameter $\beta/\alpha > 0$:

$$X \mid G \sim \mathrm{Exp}(G), \quad G \sim \Gamma(\alpha, \beta) \Rightarrow X \sim \mathrm{GPD}(1/\alpha, \beta/\alpha). \tag{1}$$

Since the mixture takes place over the rate parameter, we get the equivalent ratio representation $X \stackrel{d}{=} E/G$ with $E \sim \mathrm{Exp}(1)$, $E \perp G$. The $\mathrm{GPD}(1/\alpha, \beta/\alpha)$ is equivalent to a classical Pareto distribution with shape α and scale β/α but with a location shift such that the lower endpoint of the support corresponds to 0.

We capitalize on the ratio representation and propose to construct nonnegative random vectors $\boldsymbol{X} = (X_1, \ldots, X_D)$ with such GPD margins as componentwise ratios of two independent D-dimensional random vectors: $\boldsymbol{E} = (E_1, \ldots, E_D)$ with exponentially distributed margins, $\boldsymbol{G} = (G_1, \ldots, G_D)$ with gamma-distributed margins, i.e.,

$$\boldsymbol{X} = \boldsymbol{E}/\boldsymbol{G}, \quad E_j \sim \mathrm{Exp}(1), \quad G_j \sim \Gamma(\alpha_j, \beta_j),\ \alpha_j > 0,\ \beta_j > 0, \quad j = 1, \ldots, D. \tag{2}$$

We refer to the distribution of $\boldsymbol{X}$ in (2) as a Gamma-driven Pareto distribution. A special case arises when $G_j \equiv G_1,\ j = 1, \ldots, D$, and is known as multivariate Pareto distribution of type IV (modulo some marginal transformations), and this class further includes the multivariate variants of Pareto distributions known as Type I, II, III as special cases; see Arnold (2014, 2015). Multivariate type I constructions go back to Mardia (1962). In the construction (2), asymptotic dependence in $1/\boldsymbol{G}$ is equivalent to asymptotic dependence in $\boldsymbol{X}$, a result that follows from adapting the result on multivariate regular variation provided by Theorem 3 of Fougeres and Mercadier (2012).

For constructing tractable statistical models, we will focus on the case where the exponential random variables E_j, $j = 1, \ldots, D$ are mutually independent, such that multivariate dependence is due to the dependence of the components of the random vector $\boldsymbol{G}$ with Gamma margins. Equivalently, we consider the hierarchical model where the dependent Gamma-distributed random variables in $\boldsymbol{G}$ are embedded for the rate parameter of the exponential distribution. The multivariate distribution of $\boldsymbol{X}$ can then be expressed through Laplace transforms of components of G.

We will study distribution properties of two specific constructions leading to mutual asymptotic independence between all components of $\boldsymbol{X}$. In the first construction, each component G_j is the sum of independent Gamma-distributed variables with the same rate, and dependence arises when the same gamma variables appear in the construction of different components G_j. In the second construction, we define $\boldsymbol{G}$ as a random vector with multivariate chi-squared distribution, such that univariate distributions are a special case of the Gamma-distribution. Finally, we will present a general extension based on a random scaling of $\boldsymbol{G}$ (using a Beta-distributed random variable) to establish asymptotically dependent random vectors within the structure of (2).

For parametric statistical inference, we discuss censoring schemes using threshold exceedances and develop censored pairwise likelihood estimators. Satisfactory estimation performance will be illustrated through a simulation study.

In the remainder of the paper, distributional properties of Gamma-driven Pareto distributions are derived in Section 2. Parametric model classes with asymptotic independence are presented in Section 3, with an extension to asymptotic dependence in Section 3.3.1. We develop pairwise censored likelihood estimation in Section 4, where we also discuss different censoring schemes. Results of a simulation study are presented in Section 5. Proofs for some of the derived formulas are deferred to the Appendix.

2 Extremal dependence in exponential rate embeddings

We first consider the general hierarchical structure where a nonnegative random vector $\boldsymbol{G}$ with arbitrary margins is embedded for the rate parameter of the exponential distribution. We therefore assume mutually independent E_j, $j = 1, \ldots, D$ in (2). In this setting, the Laplace transform of $\boldsymbol{G}$, and of subvectors of $\boldsymbol{G}$, is key to deriving the distribution function and likelihoods.

2.1 Multivariate distribution functions of exponential rate embeddings

We provide the formula of the multivariate distribution function for general constructions where a random vector $\boldsymbol{G}$ is embedded for the rate parameter of the exponential distribution. The Laplace transform of a nonnegative random vector $\boldsymbol{Y}$ is always well-defined and is given by

$$\mathcal{L}_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}[\exp(-\boldsymbol{t}\boldsymbol{Y})], \quad \boldsymbol{t}, \boldsymbol{Y} \in \mathbb{R}_+^D,$$

where $\mathbf{0} = (0, \ldots, 0)^T$. We note that $\mathbb{E}[\exp(-tG)]$ is the survival function of an exponential distribution with a random variable G embedded for its rate, such that the univariate distribution functions of components of $\boldsymbol{X}$ are given by

$$F_{X_j}(x) = 1 - \int \exp(-gx)\, F_{G_j}(\mathrm{d}g) = 1 - \mathcal{L}_{G_j}(x), \quad x \geq 0, \quad j = 1, \ldots, D.$$

Specifically with Gamma embeddings, we get

$$G_j \sim \Gamma(\alpha_j, \beta_j) \quad \Rightarrow \quad F_{X_j}(x) = 1 - \mathcal{L}_{G_j}(x) = 1 - (1 + x/\beta_j)^{-\alpha_j}, \quad x \geq 0. \tag{3}$$

More generally, the joint survivor function $\overline{F}_{\boldsymbol{X}_\mathcal{I}}(\boldsymbol{x}) = \Pr(\boldsymbol{X}_\mathcal{I} > \boldsymbol{x}_\mathcal{I})$ for any subset of components $\emptyset \neq \mathcal{I} \subset \{1, \ldots, D\}$ is

$$\overline{F}_{\boldsymbol{X}_\mathcal{I}}(\boldsymbol{x}_\mathcal{I}) = \int \Pr(\boldsymbol{X}_\mathcal{I} > \boldsymbol{x}_\mathcal{I} \mid \boldsymbol{G}_\mathcal{I} = \boldsymbol{g}_\mathcal{I}) P_{\boldsymbol{G}_\mathcal{I}}(\mathrm{d}\boldsymbol{g}_\mathcal{I}) = \mathcal{L}_{\boldsymbol{G}_\mathcal{I}}(\boldsymbol{x}_\mathcal{I}), \quad \boldsymbol{x}_\mathcal{I} \geq \mathbf{0}. \tag{4}$$

The multivariate distribution function of $\boldsymbol{X}$ is then available through an inclusion-exclusion approach:

$$\Pr(\boldsymbol{X} \leq \boldsymbol{x}) = 1 - \sum_{|\mathcal{I}|=1} \mathcal{L}_{G_i}(x_i) + \sum_{|\mathcal{I}|=2} \mathcal{L}_{\boldsymbol{G}_\mathcal{I}}(\boldsymbol{x}_\mathcal{I}) - \sum_{|\mathcal{I}|=3} \mathcal{L}_{\boldsymbol{G}_\mathcal{I}}(\boldsymbol{x}_\mathcal{I}) + \ldots + (-1)^D \mathcal{L}_{\boldsymbol{G}}(\boldsymbol{x}), \quad \boldsymbol{x} \geq \boldsymbol{0}. \quad (5)$$

Several constructions of bivariate or multivariate Gamma random vectors with closed-form expressions for Laplace transforms are discussed in Kotz et al. (2005). In Section 3, we will focus on two particularly flexible constructions.

2.2 General results for extremal dependence properties

We first derive general expressions for standard extremal dependence summaries. Given two functions g_1, g_2 with $g_2(x) \neq 0$ for $x > x_0$, we write $g_1(x) \sim g_2(x)$ if $g_1(x)/g_2(x) \to 1$ for $x \to \infty$. To study extremal dependence properties, it is helpful to abstract away from marginal distributions F_{X_j} of the components X_j of $\boldsymbol{X}$, and to normalize them to a standard Pareto scale $X_j^\star$, such that $\Pr(X_j^\star > x) = 1/x$, $x > 1$ (Klüppelberg and Resnick, 2008). For Gamma-driven Pareto random vectors, we use the following probability integral transform:

$$X_j^\star = \frac{1}{1 - F_{X_j}(X_j)} = (1 + X_j/\beta_j)^{\alpha_j}, \quad j = 1, \ldots, D. \quad (6)$$

For a general multivariate random vector $\boldsymbol{X} = (X_1, \ldots, X_D)$ with joint distribution F and marginal distributions $F_1, \ldots, F_D$, we consider the coefficient

$$\chi_D(u) = \frac{\Pr(F_1(X_1) > u, \ldots, F_D(X_D) > u)}{\Pr(F_D(X_D) > u)} = \frac{\Pr(X_1^\star > 1/(1-u), \ldots, X_D^\star > 1/(1-u))}{1-u}, \quad u \in (0,1). \quad (7)$$

In the bivariate case, it corresponds to the level-dependent tail correlation coefficient

$$\chi(u) = \frac{\Pr(F_1(X_1) > u, F_2(X_2) > u)}{\Pr(F_2(X_2) > u)} = \frac{\Pr(X_1^\star > 1/(1-u), X_2^\star > 1/(1-u))}{1-u}, \quad u \in (0,1). \quad (8)$$

The limit $\chi = \lim_{u \to 1^-} \chi(u)$ (Sibuya, 1960; Coles et al., 1999), if it exists, is called tail correlation coefficient. If $\chi > 0$, X_1 and X_2 are said to be asymptotically dependent, while asymptotic independence corresponds to $\chi = 0$. For exponential rate embeddings $\boldsymbol{X}$, $\chi(u)$ is given as

$$\chi(u) = \frac{\mathcal{L}_{(G_1,G_2)}(F_{X_1}^{-1}(u), F_{X_2}^{-1}(u))}{\mathcal{L}_{G_2}(F_{X_2}^{-1}(u))}. \quad (9)$$

We illustrate the utility of (9) for a specific construction principle of (G_1, G_2). Suppose that G_1, G_2 are given by a bivariate symmetric one-factor construction with $G_1 = V_0 + V_1$ and $G_2 = V_0 + V_2$, where V_0, V_1, V_2 are any mutually independent nonnegative random variables satisfying $V_1 \overset{d}{=} V_2$. Then, we obtain

$$\chi = \lim_{t\to\infty} \frac{\mathcal{L}_{V_0}(2t)}{\mathcal{L}_{V_0}(t)} \mathcal{L}_{V_1}(t). \tag{10}$$

If $\mathrm{pr}(V_1 > 0) > 0$, then $\mathcal{L}_{V_1}(t) \to 0$ as t tends to ∞, which shows asymptotic independence ($\chi = 0$) in this case. If $V_1 = V_2 \equiv 0$ and $g(t) = \mathcal{L}_{V_0}(t)$ is a regularly varying function with index $\rho \le 0$, i.e., $g(tx)/g(t) \to x^\rho$ for any $x > 0$ as $t \to \infty$, we obtain $\chi = 2^\rho$. Finally, if $V_1 = 0$ and $\mathcal{L}_{V_0}(t)$ is rapidly varying, i.e., $g(tx)/g(t) \to \infty$ for any $x > 0$ as $t \to \infty$, then $\chi = 0$.

Under asymptotic independence, the coefficient of tail dependence $\eta_2 \in (0, 1]$ of Ledford and Tawn (1997), arising in the following bivariate joint tail representation, allows for a finer characterization of the joint tail decay rate:

$$\Pr(X_1^\star \ge x, X_2^\star \ge x) \sim \ell_2(x) x^{-1/\eta_2}, \quad x \to \infty, \tag{11}$$

where $\ell_2 > 0$ is a slowly varying function such that $\ell_2(tx)/\ell_2(t) \to 1$ as $t \to \infty$. Under asymptotic dependence, we have $\eta_2 = 1$ and $\ell_2(x) \sim \chi$, while $\eta_2 < 1$ entails asymptotic independence. The coefficient η_2 is directly related to the alternative coefficient $\overline{\chi} = 2\eta_2 - 1 \in (-1, 1]$ introduced by Coles et al. (1999). The coefficient η_2 is readily generalized to D dimensions by assuming

$$\Pr(X_1^\star \ge x, \dots, X_D^\star \ge x) \sim \ell_D(x) x^{-1/\eta_D}, \tag{12}$$

with $\eta_D \in (0, 1]$ and slowly varying function $\ell_D > 0$. For mutually independent components X_j, $j = 1, \dots, D$, we get $\eta_D = 1/D$. When different dependence structures lead to the same value of η_D, the comparison of the slowly varying functions may be instructive to detect differences in the joint tail weights. We define the level-dependent versions $\eta_D(u)$ and $\overline{\chi}(u) = 2\eta_2(u) - 1$ of η_D and $\overline{\chi}$, respectively, as follows:

$$\eta_D(u) = \frac{\log(1-u)}{\log\left(\Pr(X_1^\star \ge 1/(1-u), \dots, X_D^\star \ge 1/(1-u))\right)}. \tag{13}$$

For exponential rate embeddings $\boldsymbol{X}$, $\eta_D(u)$ is given as

$$\eta_D(u) = \frac{\log(1-u)}{\log\left(\mathcal{L}_{\boldsymbol{G}}\left(F_{X_1}^{-1}(u), \dots, F_{X_D}^{-1}(u)\right)\right)}. \tag{14}$$

Another way to study dependence among extremes is via the distribution resulting from conditioning on a fixed component exceeding a high threshold. Denote by $\boldsymbol{X}_{-j}$ the vector $\boldsymbol{X}$ with component X_j removed. A simple expression for the probabilities of conditional extremes $\Pr(\boldsymbol{X}_{-j} > \boldsymbol{x}_{-j} \mid X_j > x_j)$ (i.e., in the sense of Heffernan and Tawn, 2004) is available thanks to the expression of the joint survivor function of $\boldsymbol{X} = \boldsymbol{E}/\boldsymbol{G}$ through the Laplace transform of $\boldsymbol{G}$:

$$\Pr(\boldsymbol{X}_{-j} > \boldsymbol{x}_{-j} \mid X_j > x_j) = \frac{\mathcal{L}_{\boldsymbol{G}}(\boldsymbol{x})}{\mathcal{L}_{G_j}(x_j)}. \tag{15}$$

If computation of the multivariate Laplace transform is easy, calculating the above probability is straightforward. The conditional probabilities given a large value of x_j can be used as building blocks for the conditional extreme value models of Heffernan and Tawn (2004).

2.3 Extremal dependence of Gamma-driven Pareto vectors

To study the foregoing dependence summaries for the specific case of Gamma-driven Pareto vectors $\boldsymbol{X}$, without loss of generality we fix the rate of the Gamma distributions to 1, such that $G_j \sim \Gamma(\alpha_j, 1)$, $j = 1, \ldots, D$. For any bivariate Gamma-driven Pareto vector $\boldsymbol{X}$, the quantile-dependent coefficients (8) and (13) can then be expressed through the Laplace transforms of the Gamma vector $\boldsymbol{G}$ according to (9) and (14). In the case of complete dependence, $G_1 = \ldots = G_D \sim \Gamma(\alpha, 1)$ in $\boldsymbol{G}$, we insert the Laplace transform in (3) and obtain asymptotic dependence with $\eta_D = 1$ and $\chi = 2^{-\alpha}$.

We further evoke the general result that asymptotic dependence among components of $1/\boldsymbol{G}$ (with Gamma-margins G_j) is equivalent to asymptotic dependence of the same components in $\boldsymbol{X}$, which follows from Theorem 1 of Fougeres and Mercadier (2012), established in the framework of multivariate regular variation. Specifically, asymptotic dependence arises when the Gamma variables G_j are perfectly dependent, i.e., if $G_j = G_1$ for $j = 2, \ldots, D$, almost surely.

The following section sheds light on extreme-value properties for several general construction principles of parametric models.

3 Parametric families of Gamma-driven Pareto vectors

For parametric modeling, we focus on the dependence structures in the vector $\boldsymbol{G}$ with Gamma margins. We require flexibility with respect to tail properties of the Gamma-driven Pareto distribution of $\boldsymbol{X}$, as well as simple and easily interpretable expressions for joint survival functions of $\boldsymbol{X}$. Distribution details for various parametric families of multivariate Gamma distributions are presented in Kotz et al. (2005). Here we study two multivariate constructions, based on convolutions of Gamma variables in the first case and using chi-squared random vectors in the second case. These constructions yield asymptotic independence for the Gamma-driven Pareto distribution, and we further propose multiplying a common Beta-distributed scaling factor to the Gamma vector for achieving flexible asymptotically dependent models while preserving the Gamma-driven Pareto structure.

3.1 Gamma convolution models

Gamma distributions are convolution-stable, i.e., for independent $V_j \sim \Gamma(\alpha_j, \beta)$, $j = 1, \ldots, D$, we obtain $\sum_{j=1}^{D} V_j \sim \Gamma(\sum_{j=1}^{D} \alpha_j, \beta)$. With this property, we can define a vector $\mathbf{G}$ with Gamma margins where each component is a sum of Gamma-distributed terms, some of these terms being shared with other components. This construction defines a class of asymptotically independent Gamma-driven Pareto distributions, except for the special case when two components of $\boldsymbol{G}$ are fully dependent.

3.1.1 Construction and distribution

Consider independent random variables $E_j \sim \text{Exp}(1)$, $j = 1, \ldots, D$, and $V_i \sim \Gamma(\alpha_i, 1)$, $i = 0, \ldots, m$, $m \geq 1$. We define

$$X_j = \frac{E_j}{\sum_{i=0}^{m} \delta_{ji} V_i} \tag{16}$$

with indicator variables $\delta_{ij} \in \{0, 1\}$ satisfying $\sum_i \delta_{ji} > 0$ for all $j = 1, \ldots, D$. For the sake of identifiability of the components V_i, we assume that vectors $(\delta_{1,i}, \ldots, \delta_{D,i})$ are different for different indices i. Univariate cumulative distribution functions are

$$F_j(x) = 1 - \int \exp(-vx)\, \mathrm{P}_{\sum_{i=0}^{m} \delta_{ji} V_i}(\mathrm{d}v) = 1 - \mathcal{L}_{\sum_{i=0}^{m} \delta_{ji} V_i}(x) = 1 - (1+x)^{-\sum_i \delta_{ji} \alpha_i}, \quad j = 1, \ldots, D,$$

where we write P_X for the probability distribution of a random variable X.

The joint survivor function has the following representation:

$$\begin{aligned}\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) &= \int \Pr(X_1 > x_1, \ldots, X_D > x_D \mid V_0 = v_0, \ldots, V_m = v_m) \prod_{i=0}^{m} P_{V_i}(\mathrm{d}v_i) \\ &= \int \prod_{j=1}^{D} \exp\left(-x_j \sum_{i=0}^{m} \delta_{ji} v_i\right) \prod_{i=0}^{m} P_{V_i}(\mathrm{d}v_i) \\ &= \prod_{i=0}^{m} \int \exp\left(-v_i \sum_{j=1}^{D} \delta_{ji} x_j\right) P_{V_i}(\mathrm{d}v_i) \\ &= \prod_{i=0}^{m} \mathcal{L}_{V_i}\left(\sum_{j=1}^{D} \delta_{ji} x_j\right). \end{aligned} \tag{17}$$

The cumulative distribution function can then be obtained through the inclusion-exclusion formula (5). A parsimonious construction of Pareto models driven by a Gamma convolution is achieved by using a single factor V_0 common to all components, while the other independent Gamma-distributed variables V_i, $i = 1, \ldots, D$ arise only in one of components: $\delta_{j0} = 1$ and $\delta_{ji} = \mathbb{I}(i = j)$ for $j = 1, \ldots, D$ (Mathai and Moschopoulos, 1991), with $\mathbb{I}(\cdot)$ the indicator function.

3.1.2 Asymptotic behavior of joint exceedances

We study bivariate properties by setting $D = 2$ and considering $\boldsymbol{X} = (X_1, X_2) = (E_1/(V_0 + V_1), E_2/(V_0 + V_2))$ with $E_1, E_2 \sim \mathrm{Exp}(1)$ and $V_0 \sim \Gamma(\alpha_0, 1)$, $V_1 \sim \Gamma(\alpha_1, 1)$, $V_2 \sim \Gamma(\alpha_2, 1)$. Here, we allow for shape zero in the Gamma distribution by setting $V \equiv 0$ if $V \sim \Gamma(0, 1)$. We assume the following minimal constraints on the parameter space: $\alpha_j \geq 0$, $j = 0, 1, 2$ with $\alpha_0 + \alpha_1 > 0$ and $\alpha_0 + \alpha_2 > 0$.

Given a probability level $u \in (0, 1)$, we consider the corresponding quantile $t_j := t_j(u) = F_{X_j}^{-1}(u) = (1-u)^{-1/(\alpha_0+\alpha_j)} - 1$, $j = 1, 2$. From (17), we obtain the level-dependent tail correlation (8) for the symmetric case $\alpha_1 = \alpha_2$ as

$$\chi(u) = \mathcal{L}_{V_0}\left(\frac{t_1}{1+t_2}\right) \mathcal{L}_{V_1}(t_1) = \Pr(W_0 > t_1 + t_2 | W_0 > t_2) \Pr(W_1 > t_1), \tag{18}$$

where W_0 and W_1 are two independent generalized-Pareto-distributed random variables with scale 1 and shapes $1/\alpha_0$ and $1/\alpha_1$, respectively. For $\alpha_0 > 0$ and $\alpha_1 > 0$, $\boldsymbol{X}$ is asymptotically independent. To gain further insight into the joint tail decay under asymptotic independence with general $\alpha_1, \alpha_2 \geq 0$, we consider the normalized variables $X_j^\star = (1+X_j)^{\alpha_0+\alpha_j}$, $j = 1, 2$. The joint normalized survivor function is then equal to

$$\Pr(X_1^\star \geq x, X_2^\star \geq x) = x^{-(\alpha_1/(\alpha_0+\alpha_1)+\alpha_2/(\alpha_0+\alpha_2))} \left(x^{1/(\alpha_0+\alpha_1)} + x^{1/(\alpha_0+\alpha_2)} - 1\right)^{-\alpha_0}. \tag{19}$$

In the symmetric case with $\alpha_1 = \alpha_2$, we get the asymptotic behavior

$$\Pr(X_1^\star \geq x, X_2^\star \geq x) \sim 2^{-\alpha_0} x^{-\frac{\alpha_0+2\alpha_1}{\alpha_0+\alpha_1}} = 2^{-\alpha_0} x^{-\left(1+\frac{1}{1+\alpha_0/\alpha_1}\right)}. \tag{20}$$

Therefore, $\eta_2 = (\alpha_0 + \alpha_1)/(\alpha_0 + 2\alpha_1)$, and any value $\eta_2 \in [0.5, 1]$ can be attained. The slowly varying function ℓ_2 in (11) behaves as $2^{-\alpha_0}$. For $\alpha_0 = 0$, we get full independence of X_1 and X_2 such that $\eta_2 = 0.5$. For $\alpha_1 = \alpha_2 = 0$, we have asymptotic dependence where $\eta_2 = 1$. Moreover, for any fixed value $\alpha_0 > 0$, we can attain any value $\eta_2 \in (0.5, 1)$. From both (19) and (20), we see that α_0 has a strong influence on the slowly varying function $\ell(x)$. In the case of asymmetry $\alpha_1 \neq \alpha_2$, we write the joint survivor function as

$$\begin{aligned}\Pr(X_1^* \geq x, X_2^* \geq x) \quad &= \quad x^{-\left(\frac{\alpha_1}{\alpha_0+\alpha_1}+\frac{\alpha_2}{\alpha_0+\alpha_2}\right)} \times \\ &\left(x^{1/(\alpha_0+\min(\alpha_1,\alpha_2))}\left[1 + \underbrace{x^{1/(\alpha_0+\max(\alpha_1,\alpha_2))-1/(\alpha_0+\min(\alpha_1,\alpha_2))}}_{\to 0\ (x\to\infty)}\right] - 1\right)^{-\alpha_0}.\end{aligned}$$

Therefore, if $\alpha_1 < \alpha_2$ without loss of generality,

$$\Pr(X_1^* \geq x, X_2^* \geq x) \sim x^{-\left(1+\frac{\alpha_2}{\alpha_0+\alpha_2}\right)}, \tag{21}$$

and $\eta_2 = (\alpha_0 + \alpha_2)/(\alpha_0 + 2\alpha_2)$; for fixed α_0, only the larger value α_2 determines η_2, whereas the smaller value α_1 contributes to the slowly varying part $\ell(x)$.

Finally, for any dimension $D \geq 2$, following (7), we obtain the following general result:

$$\chi_D(u) = \frac{\Pr(X_1 > t_1, \ldots, X_D > t_D)}{\Pr(X_D > t_D)} = \mathcal{L}_{V_0}\left(\frac{\sum_{j=1}^{D-1} t_j}{1+t_D}\right) \prod_{j=1}^{D-1} \mathcal{L}_{V_j}(t_j), \tag{22}$$

where $t_j := t_j(u) = F_{X_j}^{-1}(u) = (1-u)^{-1/(\alpha_0+\alpha_j)} - 1$, $j = 1, \ldots, D$. This can be rewritten as

$$\chi(u) = \Pr\left(W_0 > \frac{\sum_{j=1}^{D-1} t_j}{1+t_D}\right) \prod_{j=1}^{D-1} \Pr(W_j > t_j) = \Pr\left(W_0 > \sum_{j=1}^{D} t_j | W_0 > t_D\right) \prod_{j=1}^{D-1} \Pr(W_j > t_j), \tag{23}$$

where the variables W_j are independent Pareto-distributed random variables with parameters $1/\alpha_j$, $j = 0, \ldots, D$. Moreover, using the Laplace transforms, we get

$$\Pr(X_1^* \geq x, \ldots, X_D^* \geq x) = \left(\sum_{j=1}^{D} x^{\frac{1}{\alpha_0+\alpha_j}} - (D-1) \right)^{-\alpha_0} x^{-\sum_{j=1}^{D} \frac{\alpha_j}{\alpha_0+\alpha_j}}. \tag{24}$$

When $\alpha_1 = \alpha_2 = \ldots = \alpha_D \equiv \alpha$, then $\eta_D = (\alpha_0 + \alpha)/(\alpha_0 + D\alpha)$. In the general case, let $\alpha_{j_0} = \min_{1 \leq j \leq D} \alpha_j$; then we get

$$\Pr(X_1^* \geq x, \ldots, X_D^* \geq x) \sim x^{-\left(1+\sum_{j=1, j \neq j_0}^{D} \frac{\alpha_j}{\alpha_0+\alpha_j}\right)}.$$

We highlight that the Gamma convolution approach allows constructing multivariate models mixing pairs with asymptotic dependence and asymptotic independence. Let us consider a simple trivariate example: if $V_1 = V_2$ are identical and independent of V_3, we get asymptotic dependence in (X_1, X_2), while (X_1, X_3) and (X_2, X_3) show asymptotic independence. In Section 3.3, we will first put forward a general method to establish asymptotically dependent random vectors, and then we will propose an approach to construct random vectors that contain some pairs of random variables with asymptotic dependence while others are asymptotically independent.

3.1.3 Conditional extremes

For the conditional extremes representation (15), the latent Gamma convolution model yields

$$\Pr(\boldsymbol{X}_{-j} > \boldsymbol{x}_{-j} \mid X_j > u) = \mathcal{L}_{V_0}\left(\frac{\sum_{i \neq j} x_i}{1+u}\right) \prod_{i \neq j} \mathcal{L}_{V_i}(x_i). \tag{25}$$

The first factor on the right-hand side is smaller than 1 for $\min_{i \neq j} x_i > 0$, such that the joint tail of the conditional distribution is lighter than that of the independent variables $\tilde{X}_i = E_i/G_i$, $i \neq j$. When the threshold u of X_j tends to infinity in (25), this factor on the right-hand side tends to one, such that the limiting conditional distribution is given by independent components possessing the distribution of $\tilde{X}_i$, $i \neq j$. This limit distribution corresponds to generalized Pareto margins but without any dependence, where the components possess relatively heavier tails than without conditioning.

More generally, when considering unconditional probabilities (i.e., $j = D + 1$ and $u = 0$ above), the joint tail probabilities of $\boldsymbol{X}$ are always smaller than those of $\tilde{X}_i$, $i = 1, \ldots, D$. If $\alpha_0 > 0$, this shows that the joint tail of $\boldsymbol{X}$ (with dependent components) is lighter than the joint tail of $\tilde{\boldsymbol{X}}$ (with independent components $\tilde{X}_j$ each having larger tail index than X_i).

3.2 Multivariate chi-squared models

Based on $2\alpha \in \mathbb{N}$ i.i.d. standard Gaussian vectors $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iD})$, $i = 1, \ldots, 2\alpha$, with correlation matrix Σ (supposed to be invertible) we construct a random vector $\boldsymbol{G}$ with chi-squared distribution of 2α degrees of freedom through the componentwise sum

$$\boldsymbol{G} = \frac{1}{2} \sum_{i=1}^{2\alpha} \boldsymbol{Z}_i^2. \tag{26}$$

Then $\boldsymbol{G}$ has $\Gamma(\alpha, 1)$-marginal distributions, and as before we can define a Gamma-driven Pareto vector

$$\boldsymbol{X} = \boldsymbol{E}/\boldsymbol{G}, \quad \boldsymbol{E} = (E_1, \ldots, E_D), \quad E_j \overset{iid}{\sim} \text{Exp}(1). \tag{27}$$

Borrowing results from McCullagh and Møller (2006, Section 2.1.2) on permanental processes, we can write the joint survival function of $\boldsymbol{X}$ as

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{E} \exp(-x_1 G_1 - \ldots - x_D G_D) = \left| I_D + \text{diag}(x_1, \ldots, x_D)\, \Sigma \right|^{-\alpha}, \tag{28}$$

where I_D denotes the $D \times D$ identity matrix and $\text{diag}(x_1, \ldots, x_D)$ is a diagonal matrix with elements x_j, $j = 1, \ldots, D$, on the diagonal. From (28), we get $\overline{F}_{\boldsymbol{X}}(x, \ldots, x) \sim c x^{-D\alpha}$ with a constant $c = c(D, \alpha, \Sigma) > 0$. Therefore, $\eta_D = 1/D$, the coefficient of tail dependence for classical independence. Consider $D = 2$, and denote by ρ the correlation coefficient. Then $c(D, \alpha, \rho) = (1 - \rho^2)^{-\alpha}$, and the factor $\overline{F}_{\boldsymbol{X}}(x, x)/\overline{F}_{X_1}^2(x, x) \sim c \geq 1$ indicates the heavier joint tail of $\boldsymbol{X}$ compared to classical independence.

3.3 Asymptotic dependent pairs of variables

Below, we first propose an extension called Beta-scaling to ensure asymptotic dependence between all components of the random vector $\boldsymbol{X}$. Then, to obtain both asymptotic dependence and asymptotic independence between different pairs of components of $\boldsymbol{X}$, we apply Beta-scaling only to a subvector of $\boldsymbol{X}$.

3.3.1 Asymptotic dependence through Beta-scaling

Lewis et al. (1989) detail a technique of random rescaling within the class of Gamma distributions. If $B \sim \text{Beta}(\alpha_1, \alpha_2)$ is independent of $G \sim \Gamma(\alpha_1 + \alpha_2, \beta)$ where $\alpha_1, \alpha_2 > 0$, then $BG \sim \Gamma(\alpha_1, \beta)$. We here suppose that the Gamma random vector $\boldsymbol{G}$ has identical marginal distributions $G_j \sim \Gamma(\alpha, \beta)$,

$j = 1, \ldots, D$. If we take $B \sim \mathrm{Beta}(\tilde{\alpha}, \alpha - \tilde{\alpha})$ independent of $\boldsymbol{G}$ with $0 < \tilde{\alpha} < \alpha$, the randomly rescaled vector $B(G_1, \ldots, G_D)$ has $\Gamma(\tilde{\alpha}, \beta)$-margins. The inverse Beta variable $1/B$ satisfies $P(1/B > x) \sim cx^{-\tilde{\alpha}}$, i.e., it is regularly varying with index $\tilde{\alpha}$, with constant $c = 1/\mathrm{Be}(\tilde{\alpha}, \alpha - \tilde{\alpha})$ where Be denotes the Beta function. Since the distribution of a Gamma-driven Pareto vector $\boldsymbol{E}/\boldsymbol{G}$ is regularly varying with index $\alpha > \tilde{\alpha}$, Breiman (1965)'s lemma can be used to show asymptotic dependence in the resulting construction (to which we refer as Beta-scaling):

$$\boldsymbol{X} = \frac{\boldsymbol{E}}{B\boldsymbol{G}}, \quad E_j \sim \mathrm{Exp}(1),\ G_j \sim \Gamma(\alpha, \beta),\ B \sim \mathrm{Beta}(\tilde{\alpha}, \alpha - \tilde{\alpha}),\ j = 1, \ldots, D, \tag{29}$$

where $\boldsymbol{E}$, $\boldsymbol{G}$ and B are mutually independent. Breiman's lemma allows calculating the so-called tail copula, which for $\boldsymbol{x} > \boldsymbol{0}$ is given as follows (see Engelke et al., 2019, Proposition 1):

$$\lambda_{\mathbf{X}}(\boldsymbol{x}) = \lim_{t \to \infty} t \Pr\left(\min_{j=1,\ldots,D} X_j^\star > t x_j \right) = C^{-1}\, \mathbb{E}\left[\min_{j=1,\ldots,D} \frac{E_j^{\tilde{\alpha}}}{x_j G_j^{\tilde{\alpha}}} \right], \tag{30}$$

where the expression of the constant C

$$C = \mathbb{E}\left[\frac{E_1^{\tilde{\alpha}}}{G_1^{\tilde{\alpha}}} \right] = \frac{\Gamma(\alpha - \tilde{\alpha})\Gamma(1 + \tilde{\alpha})}{\Gamma(\alpha)},$$

follows from the formula for partial moments of generalized Pareto distributions (Arnold, 2015, Formula 3.3.8). We observe $C \to \infty$ when $\tilde{\alpha} \uparrow \alpha$. For mutually independent exponential variables E_j, Beta-scaling inserts a new level in the hierarchical structure by defining a random scale of the Gamma variables, and we use the following conditional expectation argument to simplify expression (30):

$$\Pr\left(\min_{j=1,\ldots,D} \frac{E_j^{\tilde{\alpha}}}{G_j^{\tilde{\alpha}}} > x \mid (G_1, \ldots, G_D) \right) = \exp\left(-x^{1/\tilde{\alpha}} \sum_{j=1}^{D} G_j \right), \tag{31}$$

such that $M = \min_{j=1,\ldots,D} \frac{E_j^{\tilde{\alpha}}}{G_j^{\tilde{\alpha}}}$, conditionally on $G_1, \ldots, G_D$, follows a Weibull distribution with scale $1/(G_1 + \ldots + G_D)^{\tilde{\alpha}}$ and Weibull index $1/\tilde{\alpha}$. From Engelke et al. (2019, Proposition 1), we get

$$\begin{aligned}
\chi &= C^{-1}\, \mathbb{E}\left[\min\left(\frac{E_1^{\tilde{\alpha}}}{G_1^{\tilde{\alpha}}}, \frac{E_2^{\tilde{\alpha}}}{G_2^{\tilde{\alpha}}} \right) \right] \\
&= C^{-1}\, \mathbb{E}\left[\mathbb{E}\left[\min\left(\frac{E_1^{\tilde{\alpha}}}{G_1^{\tilde{\alpha}}}, \frac{E_2^{\tilde{\alpha}}}{G_2^{\tilde{\alpha}}} \right) \mid (G_1, G_2) \right] \right] \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha - \tilde{\alpha})} \mathbb{E}\left[(G_1 + G_2)^{-\tilde{\alpha}} \right].
\end{aligned}$$

The survival function of vectors with structure $\boldsymbol{X} = B^{-1}\tilde{\boldsymbol{X}}$, where $\tilde{\boldsymbol{X}} = \boldsymbol{E}/\boldsymbol{G}$, is given as

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{E}_B\left[\overline{F}_{\tilde{\boldsymbol{X}}}(B\boldsymbol{x}) \right] = \int_0^1 \overline{F}_{\tilde{\boldsymbol{X}}}(b\boldsymbol{x}) f_B(b)\, \mathrm{d}b, \quad f_B(b) = \frac{\Gamma(\alpha)}{\Gamma(\tilde{\alpha})\Gamma(\alpha - \tilde{\alpha})} b^{\tilde{\alpha}-1}(1-b)^{\alpha - \tilde{\alpha} - 1}, \tag{32}$$

and the probability density function of $\boldsymbol{X}$ therefore is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{E}_B B^D f_{\tilde{\boldsymbol{X}}}(B\boldsymbol{x}) = \int_0^1 f_{\tilde{\boldsymbol{X}}}(b\boldsymbol{x}) b^D f_B(b)\,\mathrm{d}b. \tag{33}$$

We calculate values of (32) and (33) through univariate integration if no closed-form formulas are available.

3.3.2 Hybrid construction mixing pairs with asymptotic dependence and asymptotic independence

Let $\mathbf{X}$ be a Gamma-driven Pareto vector and assume that $\mathbf{X} = (\mathbf{X}_{\mathcal{I}}, \mathbf{X}_{\tilde{\mathcal{I}}})$ where $\mathcal{I}$ designates any nonempty subset of $\{1, \ldots, D\}$ and $(\mathcal{I}, \tilde{\mathcal{I}})$ is a partition of $\{1, \ldots, D\}$. Let the components of $\mathbf{X}$ be defined as

$$X_j = \begin{cases} \frac{E_j}{G_j}, & j \in \mathcal{I}, \\ \frac{E_j}{BG_j}, & j \in \tilde{\mathcal{I}}, \end{cases}$$

where the two D-dimensional vectors $\mathbf{E} = (E_1, \ldots, E_D)$ and $\mathbf{G} = (G_1, \ldots, G_D)$ have margins distributed as Exp(1) and $\Gamma(\alpha, 1)$, $\alpha > 0$, respectively; $B \sim \text{Beta}(\tilde{\alpha}, \alpha - \tilde{\alpha})$ with $0 < \tilde{\alpha} < \alpha$; finally, $\boldsymbol{E}$, $\boldsymbol{G}$ and B are mutually independent.

From the previous pairwise dependence properties, we see that any pair of components (X_i, X_j) is asymptotically independent if $(i, j) \notin \mathcal{I} \times \mathcal{I}$ and $(1/G_i, 1/G_j)$ is asymptotically independent; in all other pairs, we have asymptotic dependence.

4 Statistical inference with the Peaks-Over-Threshold approach

We distinguish two conceptually different approaches to modeling tail behavior in data. For the vector $\boldsymbol{X}$, an event is considered as extreme with respect to a threshold vector $\boldsymbol{u}$

(A) when $\boldsymbol{X} \not\leq \boldsymbol{u}$, i.e., when at least one component of $\boldsymbol{X}$ exceeds its marginal threshold,

or

(B) when $\boldsymbol{X} > \boldsymbol{u}$, i.e., when all components of $\boldsymbol{X}$ exceed their corresponding marginal thresholds.

Accordingly, the probability of an extreme event is given by $p_{\max}(\boldsymbol{u}) = 1 - \mathrm{pr}(\boldsymbol{X} \le \boldsymbol{u})$, or $p_{\min}(\boldsymbol{u}) = \mathrm{pr}(\boldsymbol{X} > \boldsymbol{u})$. For a Gamma-driven Pareto vector $\boldsymbol{X}$, we have

$$p_{\min}(\boldsymbol{u}) = \mathcal{L}_{\boldsymbol{G}}(\boldsymbol{u}), \quad p_{\max}(\boldsymbol{u}) = 1 - F_{\boldsymbol{X}}(\boldsymbol{u}), \qquad \boldsymbol{u} > \boldsymbol{0}.$$

Suppose that we have observed n i.i.d. copies $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iD})$, $i = 1, \ldots, n$, of the vector $\boldsymbol{X}$. Deriving an analytical expression of the full likelihood may be difficult for large D since it involves D-fold differentiation of the Laplace transform $\mathcal{L}_{\boldsymbol{G}}$ in the joint distribution function (5). Pairwise-likelihood inference (Lindsay, 1988; Varin et al., 2011) uses a simpler product of pairwise likelihoods, ignoring dependence across pairs. Given a pair (X_1, X_2) of any two components of $\mathbf{X}$, the vector $\boldsymbol{\theta}$ collects the unknown parameters in the bivariate distribution function $F(x_1, x_2) = F_{X_1,X_2}(x_1, x_2; \boldsymbol{\theta})$, and different censored likelihoods arise for options (A) and (B) from above:

Jointly censored pairwise likelihood. In case **(A)**, we consider the pairwise likelihood contribution of data $\tilde{x}_j = x_j \mathbf{1}(\boldsymbol{x} \not\le \boldsymbol{u}) + u_j \mathbf{1}(\boldsymbol{x} \le \boldsymbol{u})$, $j = 1, \ldots, D$, given for $(\tilde{x}_1, \tilde{x}_2)$ (without loss of generality) as

$$L_A(\boldsymbol{\theta}; \tilde{x}_1, \tilde{x}_2) = \begin{cases} F(u_1, u_2; \boldsymbol{\theta}) & \text{if } x_1 \le u_1, x_2 \le u_2, \\ \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2; \boldsymbol{\theta}) & \text{otherwise.} \end{cases}$$

Partially censored pairwise likelihood. In case **(B)**, we consider the pairwise likelihood contribution of data $\tilde{x}_j = \max\{x_j, u_j\}$, $j = 1, \ldots, D$, leading to the following contribution of $(\tilde{x}_1, \tilde{x}_2)$ (without loss of generality):

$$L_B(\boldsymbol{\theta}; \tilde{x}_1, \tilde{x}_2) = \begin{cases} F(u_1, u_2; \boldsymbol{\theta}) & \text{if } x_1 \le u_1, x_2 \le u_2 \\ \frac{\partial}{\partial x_1} F(x_1, u_2; \boldsymbol{\theta}) & \text{if } x_1 > u_1, x_2 \le u_2, \\ \frac{\partial}{\partial x_2} F(u_1, x_2; \boldsymbol{\theta}) & \text{if } x_1 \le u_1, x_2 > u_2, \\ \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2; \boldsymbol{\theta}) & \text{if } x_1 > u_1, x_2 > u_2. \end{cases}$$

The previous formulas can be written using the Laplace transform, namely

$$L_A(\boldsymbol{\theta}; \tilde{x}_1, \tilde{x}_2) = \begin{cases} 1 - \mathcal{L}_{G_1}(u_1) - \mathcal{L}_{G_2}(u_2) + \mathcal{L}_{G_{12}}(u_1, u_2) & \text{if } x_1 \le u_1, x_2 \le u_2, \\ \frac{\partial^2}{\partial x_1 \partial x_2} \mathcal{L}_{G_{12}}(x_1, x_2) & \text{otherwise,} \end{cases}$$

and

$$L_B(\boldsymbol{\theta};\tilde{x}_1,\tilde{x}_2)=\begin{cases}1-\mathcal{L}_{G_1}(u_1)-\mathcal{L}_{G_2}(u_2)+\mathcal{L}_{G_{12}}(u_1,u_2) & \text{if } x_1\le u_1, x_2\le u_2,\\ -\frac{\partial}{\partial x_1}\mathcal{L}_{G_1}(x_1)+\frac{\partial}{\partial x_1}\mathcal{L}_{G_{12}}(x_1,u_2) & \text{if } x_1> u_1, x_2\le u_2,\\ -\frac{\partial}{\partial x_2}\mathcal{L}_{G_2}(x_2)+\frac{\partial}{\partial x_2}\mathcal{L}_{G_{12}}(u_1,x_2) & \text{if } x_1\le u_1, x_2> u_2\\ \frac{\partial^2}{\partial x_1\partial x_2}\mathcal{L}_{G_{12}}(x_1,x_2) & \text{if } x_1> u_1, x_2> u_2,\end{cases}$$

where $\mathcal{L}_{G_j}(t)=\mathrm{E}\left(e^{-tG_j}\right)$ and $\mathcal{L}_{G_{12}}(t_1,t_2)=\mathrm{E}\left(e^{-t_1G_1-t_2G_2}\right)$ denote the univariate and bivariate Laplace transform, respectively, of (G_1,G_2).

In summary, the pairwise likelihood is given by

$$PL_a(\boldsymbol{\theta})=\prod_{i=1}^{n}\prod_{j=1}^{D}\prod_{k>j}^{D}L_a(\boldsymbol{\theta};x_{ij},x_{ik}),\qquad a\in\{A,B\}. \tag{34}$$

Its maximization yields a consistent estimator $\widehat{\boldsymbol{\theta}}_a$ under mild conditions, and standard errors for $\widehat{\boldsymbol{\theta}}_a$ can be obtained using the "sandwich" matrix $\widehat{H_a}^{-1}\widehat{J_a}\widehat{H_a}^{-1}$, where

$$\widehat{H_a}=-\sum_{i=1}^{n}\sum_{j=1}^{D}\sum_{k>j}^{D}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log L_a(\boldsymbol{\theta};x_{ij},x_{ik})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}_a}}$$

and

$$\widehat{J_a}=\sum_{i=1}^{n}\sum_{j=1}^{D}\sum_{k>j}^{D}\frac{\partial}{\partial\boldsymbol{\theta}}\log L_a(\boldsymbol{\theta};x_{ij},x_{ik})\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log L_a(\boldsymbol{\theta};x_{ij},x_{ik})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}_a}}\right)^\top.$$

5 Simulation study

We illustrate the performance of pairwise-likelihood inference in Section 4 using simulations of the Gamma convolution construction in Section 3.1, in combination with Beta-scaling proposed in Section 3.3.1 to generate asymptotic dependence. We focus on both bivariate and trivariate vectors. Three settings are investigated: trivariate asymptotically independent vectors; bivariate asymptotically dependent vectors; hybrid trivariate vectors exhibiting both pairwise asymptotic dependence and asymptotic independence. With Beta-scaling, the marginal distributions of the latent Gamma vector must have the same shape parameter. To preserve flexibility, a copula approach using marginal transformations allows bypassing this issue, as illustrated in the bivariate asymptotically dependent and hybrid trivariate cases below.

For each set of parameters, 1500 vectors are generated. Parameters are estimated using the jointly and partially censored pairwise likelihoods PL_A and PL_B in (34), respectively, with marginal thresholds chosen as the empirical 80% quantiles. Our results are quite similar for the two PL-types, and we here report them only for PL_B. PL maximisation is coded in C and runs in parallel with the R library *parallel*. Optimisation requires approximately 10 seconds for the trivariate asymptotically independent case, and 1 minute for the hybrid trivariate case, using a 2.8GHz machine with 4 cores and 16 Gb of memory. The simulation-estimation steps are repeated 500 times for each parameter configuration. We show box-plots and level-dependent dependence functions $\chi(\cdot)$ and $\overline{\chi}(\cdot)$ resulting from the estimated model parameters; we compare them to exact theoretical values computed with the true parameter values and to empirical estimates. We here report results only for specific cases of parameter values and for level-dependent functions restricted to $\overline{\chi}(\cdot)$ for asymptotic independence and to $\chi(\cdot)$ for asymptotic dependence.

5.1 Asymptotically independent construction

We simulate the trivariate random vector (X_1, X_2, X_3) constructed as

$$X_1 = \beta_1 E_1/(G_0 + G_1), \quad X_2 = \beta_2 E_2/(G_0 + G_2), \quad X_3 = \beta_3 E_3/(G_0 + G_3))$$

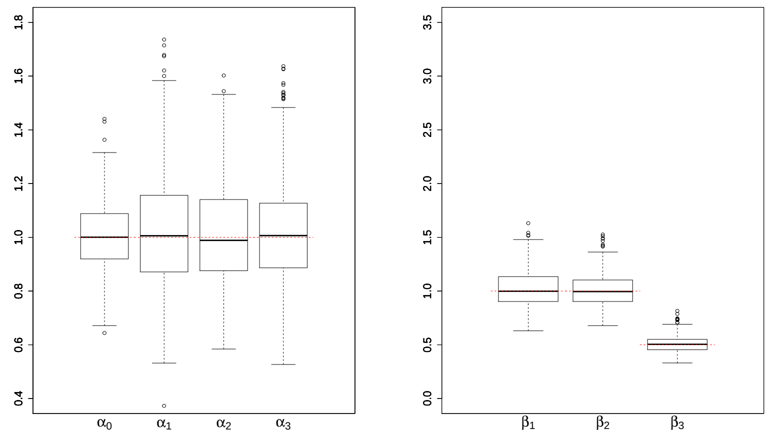
where $\beta_j > 0$, and $E_j \sim \text{Exp}(1)$ and $G_k \sim \Gamma(\alpha_k, 1)$ are independent for $j \in \{1,2,3\}$ and $k \in \{0,1,2,3\}$. We fix the following parameters: $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1$ and $\beta_1 = \beta_2 = 1$, $\beta_3 = 0.5$. Results show that all model parameters are estimated without bias and with moderate uncertainty (see boxplots in Figure 1), and consequently the level-dependent functions $\overline{\chi}(u)$ (Figure 2) are also very well estimated.

5.2 Asymptotically dependent construction

We consider the asymptotically dependent bivariate construction

$$X_1 = E_1/(B(G_0 + G_1)), \quad X_2 = E_2/(B(G_0 + G_2))$$

where $E_j \sim \text{Exp}(1)$, $j \in \{1,2\}$, and $G_k \sim \Gamma(\alpha_k, 1)$, $k \in \{0,1,2\}$, are independent with $\alpha_1 = \alpha_2$ and $B \sim \text{Beta}(\widetilde{\alpha}, \alpha_0 + \alpha_1 - \widetilde{\alpha})$, $\widetilde{\alpha} < \alpha_0 + \alpha_1$. The bivariate random vector (X_1, X_2) has asymptotic dependence. We simulate according to the following parameter configuration: $\widetilde{\alpha} = 0.3$, $\alpha_0 = \alpha_1 = \alpha_2 = 1$. We here present results for the copula approach, i.e., we consider $(T_1(X_1), T_2(X_2))$



(a) (b)

Figure 1: Estimations for simulation setting 1: asymptotically independent trivariate case. Boxplots of the estimation results for (a) α_k, $0 \leq k \leq 3$ and (b) β_j, $1 \leq j \leq 3$. Dotted red lines indicate true values.

corresponding to transformation T_j of the margins of (X_1, X_2) to the GPD, here with parameters chosen as $\xi_1 = \xi_2 = 0.07$ and $\sigma_1 = \sigma_2 = 20$. All marginal and dependence parameters are simultaneously estimated with the constraint $\alpha_1 = \alpha_2$.

As illustrated by Figure 3, the marginal parameters are all well estimated. Concerning the dependence parameters, $\widetilde{\alpha}$ is quite well estimated whereas the α parameters are under-estimated. Despite this estimation bias in some of the parameters, the corresponding function $\chi(\cdot)$ reproduces its true counterpart very satisfactorily (see Figure 4). The under-estimation of the dependance parameters α_i could be due to a lack of identifiability when using the beta-scaling approach but has no particular impact on the tail dependence structure. The underlying true function $\chi(\cdot)$ is correctly represented through the model-based function.

5.3 Hybrid construction with asymptotically independent and dependent pairs

We investigate a hybrid trivariate construction:

$$X_1 = E_1/(B(G_0 + G_1)), \quad X_2 = E_2/(B(G_0 + G_2)), \quad X_3 = E_3/(G_0 + G_3)$$

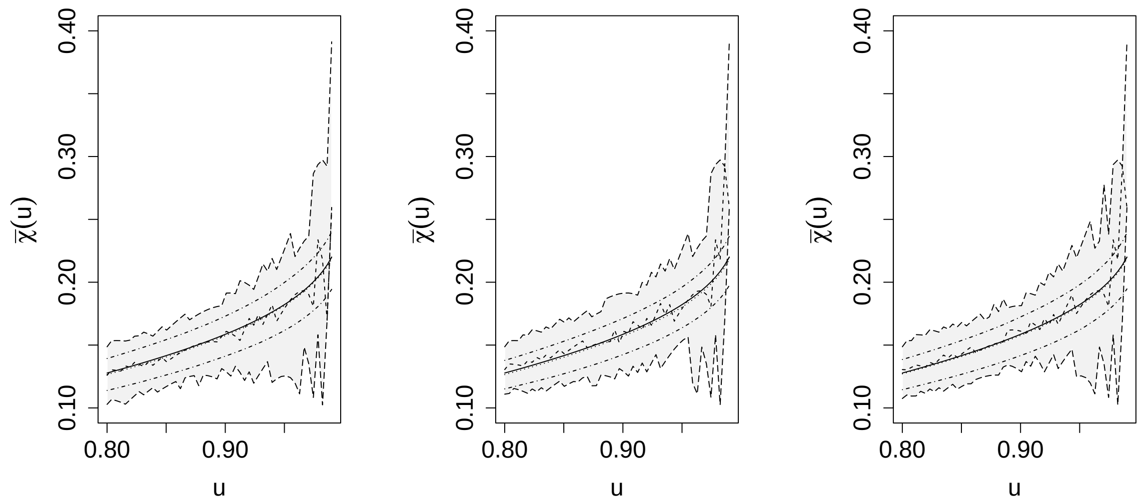
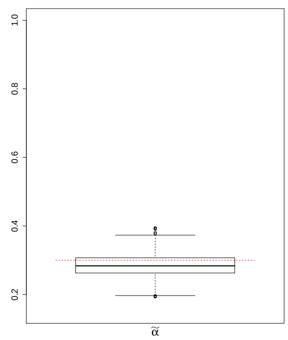
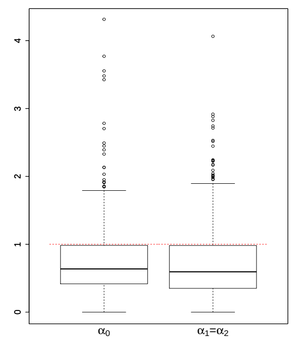
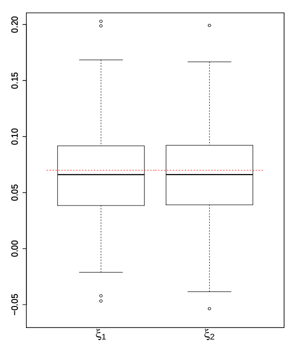


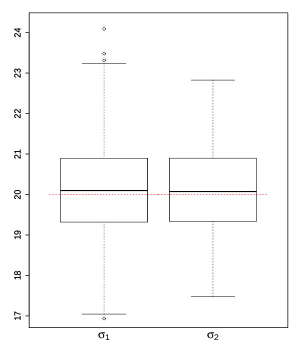
Figure 2: Same setting as in Figure 1 but here showing true (black), empirical (dashed) and fitted (dotted) $\overline{\chi}(u)$ for pairs (X_1, X_2) (left), (X_1, X_3) (middle) and (X_2, X_3) (right). Empirical quantiles at 25% and 75% are displayed in long-dashed and dot-dashed lines for the empirical and model-based estimations, respectively.





(a) (b)





(c) (d)

Figure 3: Parameter estimations for the asymptotically dependent construction. Box-plots of the estimation results for (a) $\widetilde{\alpha}$; (b) α_i $0 \leq i \leq 2$; (c) ξ_i, $i = 1, 2$; (d) σ_i, $i = 1, 2$. The true values $\widetilde{\alpha} = 0.3$, $\alpha_0 = \alpha_1 = \alpha_2 = 1$, $\xi_1 = \xi_2 = 0.07$ and $\sigma_1 = \sigma_2 = 20$ are represented by a horizontal line.

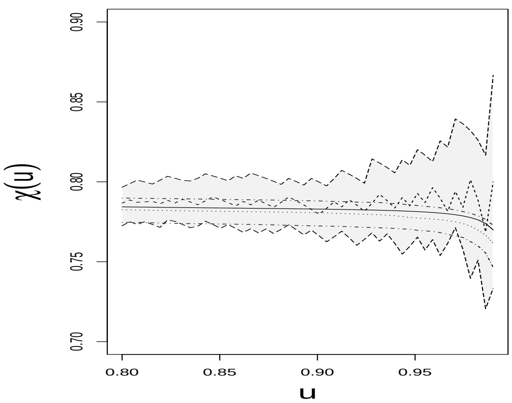


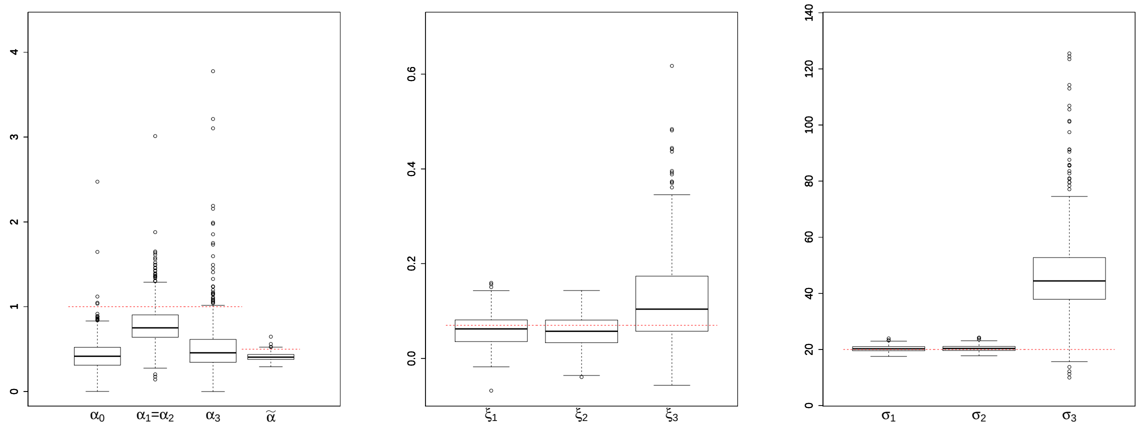
Figure 4: True (black), empirical (dashed) and fitted (dotted) $\chi(\cdot)$ function with $\alpha_0 = \alpha_1 = \alpha_2 = 1$ and $\widetilde{\alpha} = 0.3$). Empirical quantiles at 25% and 75% are displayed in long-dashed and dot-dashed lines for the empirical and model-based estimations, respectively.

where $E_j \sim \text{Exp}(1)$, $j \in \{1, 2, 3\}$, and $G_k \sim \Gamma(\alpha_k, 1)$, $k \in \{0, 1, 2, 3\}$, are independent with $\alpha_1 = \alpha_2 = \alpha_3$ and $B \sim \text{Beta}(\widetilde{\alpha}, \alpha_0 + \alpha_1 - \widetilde{\alpha})$, $\widetilde{\alpha} < \alpha_0 + \alpha_1$.

A direct consequence of this construction is that the pair of variables (X_1, X_2) has asymptotic dependence whereas the two other pairs have asymptotic independence. For the simulation the following parameter configuration is considered: $\widetilde{\alpha} = 0.5$, $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1$. As for the asymptotically dependent bivariate construction we present results for the copula approach with parameters chosen as $\xi_1 = \xi_2 = \xi_3 = 0.07$ and $\sigma_1 = \sigma_2 = \sigma_3 = 20$. All marginal and dependence parameters are simultaneously estimated with the constraint $\alpha_1 = \alpha_2$.

Boxplots of PL_B-based estimations (Figure 5 (b,c)) show that parameters of the marginal distributions are well estimated for components $1, 2$ but overestimated for component 3, where the different role of this component in the dependence structure could be the reason. Dependence parameters exhibit a more or less marked systematic underestimation (see Figure 5 (a)). Therefore, it seems as if the relatively small underestimation of $\tilde{\alpha}$ (leading to too heavy tails in the scaling $1/B$) compensates the biases in the estimation of the other parameters.

As for the case of the asymptotic dependence seen previously, we attribute these biases to a lack of identifiability of some of the model parameters. Nevertheless, the estimated functions $\chi(\cdot)$



(a) (b) (c)

Figure 5: Estimations for simulation setting 2: hybrid trivariate case. Boxplots for the parameters (a) $\widetilde{\alpha}$ and α_k, $0 \leq k \leq 3$, (b) ξ_j, (c) σ_j, $1 \leq j \leq 3$. Dotted red lines indicate true values.

(asymptotic dependence case) and $\overline{\chi}(\cdot)$ (asymptotic independence case) for the bivariate vectors (X_1, X_2), (X_1, X_3) and (X_2, X_3) show very good results (see Figure 6). Again, it appears that all the tail dependence functions are particularly well represented through the model-based functions with only moderate uncertainties.

6 Discussion

We have introduced new models for multivariate threshold exceedances with generalized Pareto margins and flexible, easily tractable expressions of their multivariate distributions. Pairs with asymptotic dependence and independence can be generated in hybrid models through Beta scaling. Estimation through pairwise likelihood is accurate in the asymptotic independent setting of the Gamma convolution approach, but the case of asymptotic dependence with β-scaling requires further investigation due to identifiability issues for the parameter vector. In dimension $D > 2$, models may involve a relatively large number of parameters when it is imposing parameter constraints a priori is not feasible. To select from several competing parameter configurations, we could use the composite likelihood information criterion (CLIC, Varin et al., 2011).

Simulation-based approaches using Markov chain Monte Carlo could come to the rescue for mod-

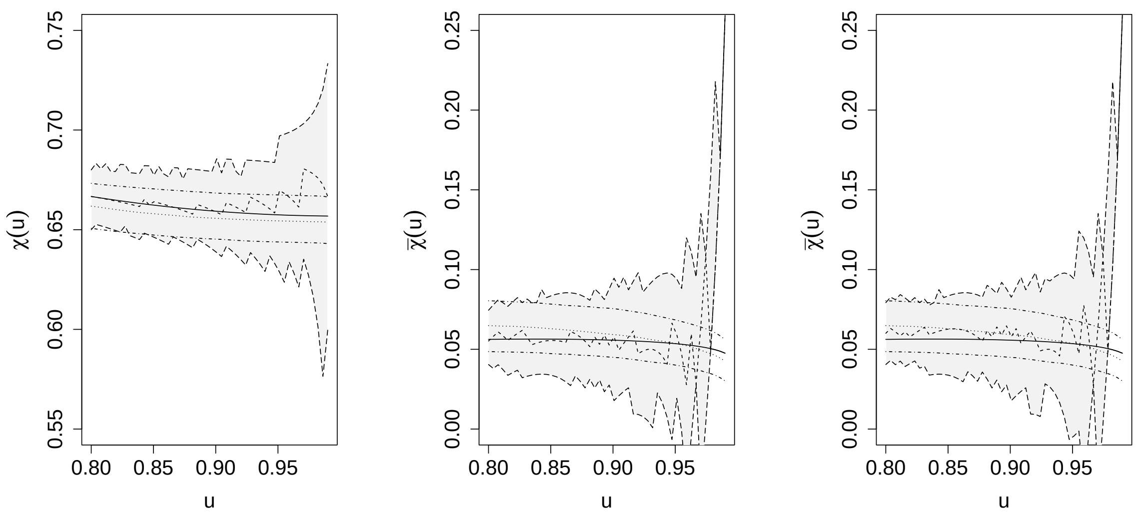


Figure 6: Same as Figure 5, but here showing true (black), empirical (dashed) and fitted (dotted) level-dependent dependence coefficients $\chi(u)$ for the pair (X_1, X_2) (left), and $\overline{\chi}(u)$ for the pairs (X_1, X_3) (middle) and (X_2, X_3) (right). Empirical quantiles at 25% and 75% are displayed in long-dashed and dot-dashed lines for the empirical and model-based estimations, respectively.

els with many parameters. Such inference would further allow bypassing the numerical integration owing to β-scaling by explicitly generating the latent Gamma and Beta variables conditional to data. Conditional simulation of latent variables would further make numerical prediction of unobserved data components straightforward.

We focused on latent Gamma convolution constructions with a single common factor G_0. More sophisticated structures could be studied, including tree-like structure where G_0 is the root and each branch corresponds to adding an independent Gamma variable G_k in some of the components of the latent Gamma vector.

We have detailed constructions leading to generalized Pareto margins. More general models could be obtained based on latent convolutions of nonnegative variables that do no have Gamma distribution but have tractable Laplace transform. Spatial and spatio-temporal extensions could be constructed to generalize the asymptotically independent model of Bacro et al. (2020) towards more flexible dependence regimes. To differentiate extremal dependence of contemporaneous and time-shifted observations, we could apply beta-scaling to achieve asymptotic independence in time while having asymptotic dependence in space.

7 Acknowledgements

The authors were supported by the French national programme LEFE/INSU.

References

Arnold, B. C. (2014), "Univariate and multivariate Pareto models," Journal of Statistical Distributions and Applications, 1(1), 11.

Arnold, B. C. (2015), Pareto Distribution, 2nd edn, New York: Wiley.

Bacro, J.-N., Gaetan, C., Opitz, T., and Toulemonde, G. (2020), "Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation Data," Journal of the American Statistical Association, 115(530), 555–569.

Bortot, P., and Gaetan, C. (2014), "A latent process model for temporal extremes," Scandinavian Journal of Statistics, 41(3), 606–621.

Breiman, L. (1965), "On some limit theorems similar to the arc-sin law," Theory of Probability and its Applications, 10(2), 323–331.

Castro-Camilo, D., Mhalla, L., and Opitz, T. (2021), "Bayesian space-time gap filling for inference on hot spots: an application to Red Sea surface temperatures," Extremes, 24(1), 105–128.

Coles, S., Heffernan, J., and Tawn, J. (1999), "Dependence measures for extreme value analyses," Extremes, 2, 339–365.

Davison, A. C., Huser, R., and Thibaud, E. (2013), "Geostatistics of dependent and asymptotically independent extremes," Mathematical Geosciences, 45(5), 511–529.

de Fondeville, R., and Davison, A. C. (2020), "Functional peaks-over-threshold analysis," arXiv preprint arXiv:2002.02711, .

Engelke, S., Opitz, T., and Wadsworth, J. (2019), "Extremal dependence of random scale constructions," Extremes, 22, 623–666.

Falk, M., Hüsler, J., and Reiss, R. D. (2011), Laws of Small Numbers: Extremes and Rare Events, 3rd edn, Basel, CH: Birkhäuser.

Ferreira, A., and De Haan, L. (2014), "The generalized Pareto process; with a view towards application and simulation," Bernoulli, 20(4), 1717–1737.

Fougeres, A.-L., and Mercadier, C. (2012), "Risk measures and multivariate extensions of Breiman's theorem," Journal of Applied Probability, 49(2), 364–384.

Heffernan, J. E., and Tawn, J. A. (2004), "A conditional approach for multivariate extreme values (with discussion)," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(3), 497–546.

Huser, R., Opitz, T., and Thibaud, E. (2017), "Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures," Spatial Statistics, 21(Part A), 166–186.

Huser, R., and Wadsworth, J. L. (2019), "Modeling spatial processes with unknown extremal dependence class," Journal of the American Statistical Association, 114(525), 434–444.

Huser, R., and Wadsworth, J. L. (2022), "Advances in statistical modeling of spatial extremes," Wiley Interdisciplinary Reviews: Computational Statistics, 14(1), e1537.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2019), "Peaks over thresholds modeling with multivariate generalized Pareto distributions," Technometrics, 61(1), 123–135.

Klüppelberg, C., and Resnick, S. I. (2008), "The Pareto copula, aggregation of risks, and the emperor's socks," Journal of Applied Probability, 45(1), 67–84.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2005), Continuous Multivariate Distributions, Models and Applications: Second Edition, Vol. 1, New York: John Wiley & Sons.

Ledford, A. W., and Tawn, J. A. (1997), "Modelling dependence within joint tail regions," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59(2), 475–499.

Lewis, P., McKenzie, E., and Hugus, D. (1989), "Gamma processes," Stochastic Models, 5(1), 1–30.

Lindsay, B. G. (1988), "Composite likelihood methods," Contemporary Mathematics, 80(1), 221–39.

Mardia, K. V. (1962), "Multivariate Pareto distributions," The Annals of Mathematical Statistics, 33(3), 1008–1015.

Mathai, A. M., and Moschopoulos, P. G. (1991), "On a multivariate gamma," Journal of Multivariate Analysis, 39(1), 135–153.

McCullagh, P., and Møller, J. (2006), "The permanental process," Advances in Applied Probability, 38(4), 873–888.

Noven, R. C., Veraart, A. E. D., and Gandy, A. (2018), "A latent trawl process model for extreme values," The Journal of Energy Markets, 11(3), 1–24.

Opitz, T. (2016), "Modeling asymptotically independent spatial extremes based on Laplace random fields," Spatial Statistics, 16, 1–18.

Opitz, T., Allard, D., and Mariethoz, G. (2021), "Semi-parametric resampling with extremes," Spatial Statistics, 42, 100445.

Palacios-Rodriguez, F., Toulemonde, G., Carreau, J., and Opitz, T. (2020), “Semi-parametric generalized Pareto processes for simulating space-time extreme events,” Stochastic Environmental Research and Risk Assessment, 34, 2033–2052.

Reiss, R., and Thomas, M. (2007), Statistical Analysis of Extreme Values, third edn, Basel: Birkhäuser.

Rootzén, H., and Tajvidi, N. (2006), “Multivariate generalized Pareto distributions,” Bernoulli, 12(5), 917–930.

Sibuya, M. (1960), “Bivariate extreme statistics, I,” Annals of the Institute Statistical Mathematics, 11(3), 195–210.

Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018), “Modelling spatial extreme events with environmental applications,” Spatial Statistics, 28, 39–58.

Thibaud, E., and Opitz, T. (2015), “Efficient inference and simulation for elliptical Pareto processes,” Biometrika, 102(4), 855–870.

Varin, C., Reid, N., and Firth, D. (2011), “An overview of composite likelihood methods,” Statistica Sinica, 21(1), 5–42.

Wadsworth, J. L., and Tawn, J. A. (2012), “Dependence modelling for spatial extremes,” Biometrika, 99(2), 253–272.

Wadsworth, J., and Tawn, J. (2013), “A new representation for multivariate tail probabilities,” Bernoulli, 19(5B), 2689–2714.

Wadsworth, J., Tawn, J. A., Davison, A., and Elton, D. M. (2017), “Modelling across extremal dependence classes,” Journal of the Royal Statistical Society (Series B), 79(1), 149–175.

Appendix

Proof of Equation (10):

$$
\begin{aligned}
\Pr(X_1 > t_1, X_2 > t_2) &= \mathcal{L}_{V_0+V_1,V_0+V_2}(t_1,t_2)\\
&= \mathbb{E}\{\exp(t_1(V_0+V_1)+t_2(V_0+V_2))\}\\
&= \mathcal{L}_{V_0}(t_1+t_2)\mathcal{L}_{V_1}(t_1)\mathcal{L}_{V_2}(t_2)
\end{aligned}
$$

Then,

$$
\chi(t) = \frac{\Pr(X_1 > t, X_2 > t)}{\Pr(X_2 > t_2)}\frac{\mathcal{L}_{V_0+V_1,V_0+V_2}(t,t)}{\mathcal{L}_{V_0+V_2}(t)} = \frac{\mathcal{L}_{V_0}(2t)}{\mathcal{L}_{V_0}(t)}\mathcal{L}_{V_1}(t).
$$

Proof of Equation (18):

$$
\begin{aligned}
\chi(u) &= \frac{\mathcal{L}_{V_0}(t_1+t_2)\mathcal{L}_{V_1}(t_1)}{\mathcal{L}_{V_0}(t_2)}\\
&= \left(\frac{1+t_1+t_2}{1+t_2}\right)^{-\alpha_0}\mathcal{L}_{V_1}(t_1)\\
&= \left(1+\frac{t_1}{1+t_2}\right)^{-\alpha_0}\mathcal{L}_{V_1}(t_1)\\
&= \mathcal{L}_{V_0}\left(\frac{t_1}{1+t_2}\right)\mathcal{L}_{V_1}(t_1)
\end{aligned}
$$

Since

$$
\Pr(W_0 > \frac{t_1}{1+t_2}) = \left(1+\frac{t_1}{1+t_2}\right)^{-\alpha_0} = \left(\frac{1+t_1+t_2}{1+t_2}\right)^{-\alpha_0} = \frac{\Pr(W_0 > t_1+t_2)}{\Pr(W_0 > t_2)},
$$

the proof is fulfilled.

Proof of Equation (19):

$$
\begin{aligned}
\Pr(X_1^\star \geq x, X_2^\star \geq x) &= \Pr((1+X_1)^{\alpha_0+\alpha_1} \geq x, (1+X_2)^{\alpha_0+\alpha_2} \geq x) = \Pr(X_1 \geq x^{\frac{1}{\alpha_0+\alpha_1}}-1, X_2 \geq x^{\frac{1}{\alpha_0+\alpha_2}}-1)\\
&= \mathcal{L}_{V_0+V_1,V_0+V_2}(x^{\frac{1}{\alpha_0+\alpha_1}}-1, x^{\frac{1}{\alpha_0+\alpha_2}}-1)\\
&= \mathcal{L}_{V_0}\left(x^{\frac{1}{\alpha_0+\alpha_1}}+x^{\frac{1}{\alpha_0+\alpha_2}}-2\right)\mathcal{L}_{V_1}(x^{\frac{1}{\alpha_0+\alpha_1}}-1)\mathcal{L}_{V_2}(x^{\frac{1}{\alpha_0+\alpha_2}}-1)\\
&= \left(x^{\frac{1}{\alpha_0+\alpha_1}}+x^{\frac{1}{\alpha_0+\alpha_2}}-1\right)^{-\alpha_0} x^{-\frac{\alpha_1}{\alpha_0+\alpha_1}} x^{-\frac{\alpha_2}{\alpha_0+\alpha_2}}
\end{aligned}
$$

Proof of Equation (22)**:**

$$\chi_D(u) = \frac{\mathcal{L}_{(V_0+V_1,\ldots,V_0+V_D)}(t_1,\ldots,t_D)}{\mathcal{L}_{V_0+V_D}(t_D)} = \frac{\mathcal{L}_{V_0}(\sum_{j=1}^D t_j)\prod_{j=1}^D \mathcal{L}_{V_j}(t_j)}{\mathcal{L}_{V_0}(t_D)\mathcal{L}_{V_D}(t_D)}$$

and since

$$\frac{\mathcal{L}_{V_0}(\sum_{j=1}^D t_j)}{\mathcal{L}_{V_0}(t_D)} = \left(\frac{1+\sum_{j=1}^D t_j}{1+t_D}\right)^{-\alpha_0} = \left(1+\frac{\sum_{j=1}^{D-1} t_j}{1+t_D}\right)^{-\alpha_0} = \mathcal{L}_{V_0}\left(\frac{\sum_{j=1}^{D-1} t_j}{1+t_D}\right),$$

the result follows.

Proof of Equation (23)**:**

same as proof of (18) rewritten in dimension D.

Proof of Equation (24)**:**

$$\Pr(X_1^* > x,\ldots,X_D^* > x) = \Pr(X_1 > x^{\frac{1}{\alpha_0+\alpha_1}} - 1,\ldots,X_D > x^{\frac{1}{\alpha_0+\alpha_D}} - 1)$$

Let $z_j = x^{\frac{1}{\alpha_0+\alpha_j}} - 1$. Then, from (17)

$$\begin{aligned}
\Pr(X_1 > z_1,\ldots,X_D > z_D) &= \mathcal{L}_{V_0}(z_1+\ldots+z_D)\mathcal{L}_{V_1}(z_1)\ldots\mathcal{L}_{V_D}(z_D) \\
&= (1+z_1+\ldots+z_D)^{-\alpha_0}(1+z_1)^{-\alpha_1}\ldots(1+z_D)^{-\alpha_D} \\
&= \left(\sum_{j=1}^D x^{\frac{1}{\alpha_0+\alpha_j}} - (D-1)\right)^{-\alpha_0}\prod_{j=1}^D x^{-\frac{\alpha_j}{\alpha_0+\alpha_j}} \\
&= \left(\sum_{j=1}^D x^{\frac{1}{\alpha_0+\alpha_j}} - (D-1)\right)^{-\alpha_0} x^{-\sum_{j=1}^D \frac{\alpha_j}{\alpha_0+\alpha_j}}
\end{aligned}$$