

# Modelling across extremal dependence classes

J. L. Wadsworth<sup>1</sup>, J. A. Tawn<sup>1</sup>, A. C. Davison<sup>2</sup> and D. M. Elton<sup>1</sup>

<sup>1</sup>Lancaster University, UK

<sup>2</sup>Ecole Polytechnique Fédérale de Lausanne, Switzerland

November 16, 2015

## Abstract

Different dependence scenarios can arise in multivariate extremes, entailing careful selection of an appropriate class of models. In bivariate extremes, the variables are either asymptotically dependent or are asymptotically independent. Most available statistical models suit one or other of these cases, but not both, resulting in a stage in the inference that is unaccounted for, but can substantially impact subsequent extrapolation. Existing modelling solutions to this problem are either applicable only on sub-domains, or appeal to multiple limit theories. We introduce a unified representation for bivariate extremes that encompasses a wide variety of dependence scenarios, and applies when at least one variable is large. Our representation motivates a parametric model that encompasses both dependence classes. We implement a simple version of this model, and show that it performs well in a range of settings.

**Keywords:** asymptotic independence, censored likelihood, conditional extremes, dependence modelling, extreme value theory, multivariate regular variation.

## 1 Introduction

The first challenge faced when modelling extremes of two or more variables is to decide which type of dependence they exhibit. There are two possibilities in the bivariate case. For a random vector  $(Z_1, Z_2)$ , with marginal distributions  $F_1, F_2$ , define the limiting probability

$$\chi = \lim_{u \rightarrow 1} \mathbb{P}\{F_1(Z_1) > u \mid F_2(Z_2) > u\}, \quad (1.1)$$

if it exists. The pair  $(Z_1, Z_2)$  are termed *asymptotically dependent* if  $\chi > 0$ , and *asymptotically independent* if  $\chi = 0$ . In higher dimensions the situation becomes more complicated; Wadsworth and Tawn (2013) outline the idea of  $k$ -dimensional joint tail dependence, which is summarized by  $\sum_{i=0}^{k-2} \binom{k}{i}$  limits such as (1.1). For this reason, we focus on bivariate data, but discuss higher dimensional cases in Section 7.

It is important to detect the appropriate dependence class because most models for bivariate extremes encompass one or the other, but not both. Classical multivariate extreme value theory (e.g., Resnick, 1987, Chapter 5) yields asymptotic dependence models (Coles and Tawn, 1991; de Haan and de Ronde, 1998). Its first stage is usually to transform variables to a common marginal distribution. Suppose that  $(X_P, Y_P) = [\{1 - F_1(Z_1)\}^{-1}, \{1 - F_2(Z_2)\}^{-1}]$  have marginal standard Pareto distributions (interpreted asymptotically, if  $F_1, F_2$  are discontinuous). In the asymptotic dependence case the basic modelling principle is that for an arbitrary pair of norms  $\|\cdot\|_a$  and  $\|\cdot\|_b$ , the pseudo angular and radial variables

$$\mathbf{W} = (X_P, Y_P) / \|(X_P, Y_P)\|_a, \quad R = \|(X_P, Y_P)\|_b, \quad (1.2)$$

become independent in the limit, in the sense that

$$\lim_{t \rightarrow \infty} \mathbb{P}\{\mathbf{W} \in B, R > t(r+1) \mid R > t\} = H(B)(r+1)^{-1}, \quad r \geq 0, \quad B \subset S^a := \{\mathbf{w} \in \mathbb{R}_+^2 : \|\mathbf{w}\|_a = 1\}, \quad (1.3)$$

for continuity sets of the limit measure  $H$ . The limit holds for both dependence classes, but is only useful under asymptotic dependence: under any form of asymptotic independence,  $H(\cdot)$  is a discrete two-point

distribution that places atoms of probability on the endpoints of the continuous arc  $\mathcal{S}^a$ ,  $(0,1)/\|(0,1)\|_a$ ,  $(1,0)/\|(1,0)\|_a$ . Since  $\|\cdot\|_a$  is arbitrary, we henceforth use the  $L_1$ -norm,  $\|\cdot\|_1$ , and redefine  $H$  to be the limiting distribution of  $W = X_P/(X_P + Y_P)$ , with  $H(w) = H([0,w])$  ( $0 \leq w \leq 1$ ). Under asymptotic dependence,  $H$  has mass on the interior of  $[0,1]$  and likelihood-based statistical modelling typically assumes the existence of a *spectral density*,  $h(w) = dH(w)/dw$  (Coles and Tawn, 1991). One common goal of multivariate extreme value modelling is to estimate probabilities such as  $\mathbb{P}\{(Z_1, Z_2) \in A\}$ , where the set  $A$  is extreme in at least one margin. Under asymptotic dependence, this is aided by inference on  $h$ , and the independent limit distribution of the scaling appearing in (1.3).

The degeneracy of  $H$  under asymptotic independence occurs because (1.1) implies that the very largest values of  $Z_1$  or  $Z_2$ , and hence of  $X_P$  or  $Y_P$ , occur singly, pushing all the mass of  $W$  to the boundaries of the interval  $[0,1]$ . This is due to the heavy tails of Pareto random variables: since the high quantiles on the Pareto scale are very large, one of  $X_P$  and  $Y_P$  will dominate the other when  $R$  is extreme.

This argument suggests that the choice of margins is central to simplifying extremal dependence modelling. Thus, rather than (1.3), we assume that there exist a common marginal distribution  $F : (0, x^F) \rightarrow [0,1]$ , where  $x^F \leq \infty$  is the upper endpoint of the support, a norm  $\|\cdot\|_*$ , and normalization functions  $a(t) > 0$  and  $b(t)$ , such that the positive random variables  $(X, Y) = [F^{-1}\{F_1(Z_1)\}, F^{-1}\{F_2(Z_2)\}]$  satisfy

$$\lim_{t \rightarrow \infty} \mathbb{P} \left\{ \frac{X}{X+Y} \leq w, \|(X, Y)\|_* > a(t)r + b(t) \mid \|(X, Y)\|_* > b(t) \right\} = J(w)\bar{K}(r), \quad r \geq 0, \quad (1.4)$$

at continuity points of  $J$ , where  $J$  is a non-degenerate probability distribution having mass on the interior of  $[0,1]$ , and  $\bar{K}$  is the survivor function of the generalized Pareto, GP( $\sigma, \lambda$ ), distribution. That is,

$$\bar{K}(r) = (1 + \lambda r/\sigma)_+^{-1/\lambda}, \quad r \geq 0, \quad \sigma > 0, \quad \lambda \in \mathbb{R}, \quad a_+ = \max(a, 0); \quad (1.5)$$

the case  $\lambda = 0$  is interpreted as the limit  $\bar{K}(r) = \exp(-r/\sigma)$ . In (1.4),  $a(t)$  and  $b(t)$  are the same as in the theory for univariate extremes for the variable  $\|(X, Y)\|_*$ ; see Chapter 1 of Leadbetter et al. (1983), for example. When  $(Z_1, Z_2)$  are asymptotically dependent and  $F(\cdot) = 1 - (\cdot)^{-1}$ , so that  $(X, Y)$  have standard Pareto margins, then (1.4) is equivalent to (1.3), with  $a(t) = b(t) = t$  and  $\bar{K}(r) = (1+r)^{-1}$ ; thus  $\sigma = \lambda = 1$ , and the distribution  $J$  in (1.4) equals  $H$  as defined following (1.3). When  $(Z_1, Z_2)$  are asymptotically independent, then a marginal  $F$  with a lighter tail is required to obtain a distribution  $J$  placing mass in  $(0,1)$ . The extremal dependence is then described by the combination of  $J$ ,  $\|\cdot\|_*$  and  $\lambda$ . Section 3 contains further discussion of the meaning and interpretation of (1.4), and motivates it with a variety of examples.

Under asymptotic dependence, the norms used in transformation (1.2) to  $\mathbf{W}$  and  $R$  are arbitrary and need not be the same. In (1.4), we have again defined a pseudo angular and radial transformation

$$W = X/(X+Y), \quad R = \|(X, Y)\|_*, \quad (1.6)$$

where for later simplicity we use the  $L_1$ -norm in the definition of  $W$ , but the norm  $\|\cdot\|_*$  defining  $R$  must be chosen so that the limit (1.4) holds. The inverse of (1.6) is

$$(X, Y) = R \left( \frac{W}{\|(W, 1-W)\|_*}, \frac{1-W}{\|(W, 1-W)\|_*} \right). \quad (1.7)$$

When assumption (1.4) holds, we see from (1.7) that for large  $R$  the variables  $(X, Y)$  behave as if the angular component  $(W/\|(W, 1-W)\|_*, (1-W)/\|(W, 1-W)\|_*)$  is randomly scaled by an independent generalized Pareto variable. However, it is not straightforward to exploit this statistically, because the flexibility in (1.4) stems from not having specified the margins  $F$  in which we make the pseudo radial-angular transformation. Nonetheless, the dependence structure defined by (1.7) must describe a rich variety of extremal dependencies, and motivates a copula model, described in Section 4, that we can apply to both asymptotically dependent and asymptotically independent data. This model can indeed capture many extremal dependence structures, reproducing the entire ranges of common summary statistics for extremal dependence in both dependence classes.

In Section 2 we review current statistical methods for bivariate extremes, focussing on those providing a non-trivial treatment of asymptotic independence. In Section 3 we present examples to illustrate assumption (1.4), and discuss further the interpretation of the limit assumption. Section 4 introduces a statistical

model and describes its dependence properties. Inference approaches are developed in Section 5, with some simulations to assess how well a given version of the model can estimate rare event probabilities, and in Section 6 we apply our model to oceanographic data previously analyzed using both dependence structures. We conclude the article by outlining extensions to higher dimensions and discussing related issues.

## 2 Existing methodology incorporating asymptotic independence

Many inferential approaches for extremal dependence assume the applicability of equation (1.3) with asymptotic dependence; see for example Coles and Tawn (1991), Einmahl et al. (1997), de Haan and de Ronde (1998), Mikosch (2005) and Sabourin and Naveau (2014). Ledford and Tawn (1997) noted a gap in the theory for practical treatment of asymptotic independence and introduced the *coefficient of tail dependence*,  $\eta \in (0, 1]$ . For  $(X_P, Y_P)$  as defined in Section 1, this coefficient may be defined through the equation

$$\mathbb{P}(X_P > tx, Y_P > ty) = \mathcal{L}(tx, ty)t^{-1/\eta}(xy)^{-1/2\eta}, \quad tx, ty \geq 1, \quad (2.1)$$

where  $\mathcal{L}$  is bivariate slowly varying at infinity, i.e.,  $\mathcal{L}(tx, ty)/\mathcal{L}(t, t) \rightarrow d\{x/(x+y)\}$ ,  $t \rightarrow \infty$ , with  $d: (0, 1) \rightarrow (0, \infty)$  termed the *ray dependence function*, depending only on the ray  $q := x/(x+y)$ . When  $\eta = 1$  and  $\mathcal{L}(t, t) \not\rightarrow 0$  as  $t \rightarrow \infty$  we obtain asymptotic dependence, but otherwise there is asymptotic independence.

Setting  $x = y = 1$  in (2.1) gives  $\mathbb{P}(X_P > t, Y_P > t) = \mathcal{L}(t, t)t^{-1/\eta}$ . Under asymptotic dependence,  $\eta = 1$  and the dependence is summarized by the parameter  $\chi = \lim_{t \rightarrow \infty} \mathcal{L}(t, t) > 0$ . Under asymptotic independence,  $\chi = 0$  and  $\eta \leq 1$  summarizes the degree of dependence.

The parameters  $\chi$  and  $\eta$  do not explain all the features of the extremal dependence of  $(Z_1, Z_2)$ . Under asymptotic dependence, the function  $d(q)$  prescribes how to scale  $(xy)^{-1/2}$  in order to find joint survivor probabilities across different rays,  $q \in [0, 1]$  in Pareto margins. When  $\chi > 0$ , the link between  $d$  and  $H$ , as defined following equation (1.3), is

$$d(q) = \frac{2}{\chi} \int_0^1 \min \left\{ w \left( \frac{1-q}{q} \right)^{1/2}, (1-w) \left( \frac{q}{1-q} \right)^{1/2} \right\} dH(w). \quad (2.2)$$

By definition,  $d(1/2) = 1$ , so  $\chi = 2 \int_0^1 \min(w, 1-w) dH(w)$ . Ramos and Ledford (2009) offered a characterization of the function  $d(q)$  when  $\eta \neq 1$ , beginning with the limit assumption

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_P > tx, Y_P > ty \mid X_P > t, Y_P > t) = d\{x/(x+y)\}(xy)^{-1/2\eta}, \quad x, y \geq 1. \quad (2.3)$$

In this case we may write

$$d(q) = \eta \int_0^1 \min \left\{ w \left( \frac{1-q}{q} \right)^{1/2}, (1-w) \left( \frac{q}{1-q} \right)^{1/2} \right\}^{1/\eta} dH_\eta(w), \quad (2.4)$$

where  $H_\eta$  is the *hidden angular measure*, characterized in Ramos and Ledford (2009); see also Resnick (2002, 2006) and Das and Resnick (2014) for further details of this framework of *hidden regular variation*. Suitable parametric models for  $H_\eta$  give probability models for simultaneously extreme random variables on regions of the form  $(X_P, Y_P) \in (v, \infty)^2$  for large  $v$ ; see Ramos and Ledford (2009) for examples.

Unfortunately the Ramos–Ledford–Tawn approach is applicable only within regions where both variables are large. However, under asymptotic independence, the variables  $(X_P, Y_P)$  do not grow in their joint extremes at the same rate as their marginal extremes, so these may not be the regions of most practical interest. Wadsworth and Tawn (2013) provided an alternative representation for multivariate tail probabilities, allowing study of regions where one variable may be larger than the other. Their assumption was

$$\mathbb{P}(X_P > t^\beta, Y_P > t^\gamma) = L(t; \beta, \gamma)t^{-\kappa(\beta, \gamma)}, \quad \beta, \gamma \geq 0, \max(\beta, \gamma) > 0, \quad (2.5)$$

where the function  $\kappa$  is homogeneous of order 1, and the function  $L(\cdot; \beta, \gamma)$  is slowly varying at infinity, i.e., for all  $a > 0$ ,  $\lim_{t \rightarrow \infty} L(ta; \beta, \gamma)/L(t; \beta, \gamma) = 1$ . Under asymptotic independence  $\kappa$  was shown to display structure similar to that provided by  $d$  under asymptotic dependence. Representation (2.5) is useful for

estimation of joint survivor probabilities when one variable may be much larger than the other, although the inferential methodology of Wadsworth and Tawn (2013) does not easily extend to regions more general than joint survivor regions. Example 2 in Section 3 covers some special cases of this set-up.

Heffernan and Tawn (2004) developed a very general modelling assumption that we present in the adapted form of Heffernan and Resnick (2007). For  $(X_E, Y_E) = (-\log\{1 - F_1(Z_1)\}, -\log\{1 - F_2(Z_2)\})$  with (asymptotically) standard exponential marginal distributions, they assume the existence of a non-degenerate  $G$  in

$$\lim_{t \rightarrow \infty} \mathbf{P} \left\{ \frac{X_E - b(Y_E)}{a(Y_E)} \leq x, Y_E > t + y \mid Y_E > t \right\} = G(x)e^{-y}, \quad y \geq 0. \quad (2.6)$$

Inference under (2.6) is semiparametric, as the functions  $a(Y_E)$  and  $b(Y_E)$  are typically chosen to be  $Y_E^\alpha, \beta Y_E$ ,  $\alpha \in (-\infty, 1)$ ,  $\beta \in [0, 1]$ , for non-negative dependence, and  $G$  is estimated nonparametrically. Asymptotic dependence arises in the model only when  $\alpha = 0$ ,  $\beta = 1$ , and then any structure is captured through  $G$ . Once more the limiting independence of the normalized  $Y_E$  and  $\{X_E - b(Y_E)\}/a(Y_E)$  is crucial to the inference. This method is a very flexible approaches to multivariate extreme value modelling, though we address some of its drawbacks with the representation (1.4) and the associated model to be developed in Section 4. One problem is that when conditioning on different variables, consistency of the resulting models is an unresolved issue (Liu and Tawn, 2014). The need for nonparametric estimation of  $G$  may be viewed as a strength or weakness, but can lead to difficulties in estimating non-zero probabilities (Peng and Qi, 2004; Wadsworth and Tawn, 2013).

Like the methods described above, the new approach described in Section 4 is suitable for both asymptotically dependent and asymptotically independent data. However, it is motivated by a single limit representation, and may be applied when either variable is large. Moreover, our framework allows a smooth transition across the dependence class boundary, in a sense to be described in Section 4.3.

### 3 Limit Assumption

In Section 3.1 we provide a condition that is equivalent to (1.4) under additional smoothness assumptions. This condition is useful to illustrate applicability of (1.4) when these extra assumptions are met. In Section 3.2 we discuss flexibility in how the limit may be exploited, and then discuss the interpretation of the limit assumption. A variety of examples are presented in Section 3.3.

#### 3.1 Alternative Condition

Suppose that  $(X, Y) = [F^{-1}\{F_1(Z_1)\}, F^{-1}\{F_2(Z_2)\}]$  are continuous random variables with a joint density, so this is also true for  $(R, W)$ , as defined in (1.6). This assumption is more restrictive than necessary, but it facilitates development and is often reasonable. Let  $c(u_1, u_2)$  denote the density of the *copula*, i.e., the density of  $\{F(X), F(Y)\} = \{F_1(Z_1), F_2(Z_2)\}$ . Then, with  $f$  denoting the density of  $F$ , the joint density of  $(X, Y)$  is  $f_{X,Y}(x, y) = c\{F(x), F(y)\}f(x)f(y)$ . The Jacobian of the transformation from  $(X, Y)$  to  $(R, W)$  as defined in (1.6) is  $r\|(w, 1 - w)\|_*^{-2}$ , and the density  $f_{R,W}(r, w)$  of  $(R, W)$  equals

$$c \left[ F \left\{ \frac{rw}{\|(w, 1 - w)\|_*} \right\}, F \left\{ \frac{r(1 - w)}{\|(w, 1 - w)\|_*} \right\} \right] f \left\{ \frac{rw}{\|(w, 1 - w)\|_*} \right\} f \left\{ \frac{r(1 - w)}{\|(w, 1 - w)\|_*} \right\} \frac{r}{\|(w, 1 - w)\|_*}. \quad (3.1)$$

To demonstrate applicability of (1.4), we use the following simpler condition, which is valid when the relevant densities and limits exist. In Appendix A we show that under mild assumptions (1.4) is implied by

$$\lim_{t \rightarrow \infty} \mathbf{P}\{W \leq w \mid R = b(t)\} = J(w), \quad (3.2)$$

with  $b(t) = F_R^{-1}(1 - 1/t)$ , the  $1 - 1/t$  quantile of  $R$ ; or, terms of the joint density function  $f_{R,W}(r, w)$ ,

$$\int_0^w f_{R,W}\{b(t), v\} dv \sim J(w) \int_0^1 f_{R,W}\{b(t), v\} dv, \quad t \rightarrow \infty. \quad (3.3)$$

Thus, when integration over the  $W$  coordinate does not affect the rate at which the joint density decays in  $r$  as  $r \rightarrow r^F := \sup\{r : F_R(r) < 1\}$ , then condition (3.2), and hence (1.4), is satisfied. Expression (3.1) shows how the transformed margins, defined by  $F, f$ , and the copula,  $c$ , interact for (3.3) to apply.

In order to study the domain of attraction of the radial variable  $R$ , we assume differentiability of its density  $f_R(r)$ , and define the reciprocal hazard function  $h_R(r) := \{1 - F_R(r)\}/f_R(r)$ . If  $\lim_{r \rightarrow \infty} h'_R(r) =: \lambda \in (-\infty, \infty)$  then  $R$  lies in the domain of attraction of the GP distribution with shape parameter  $\lambda$  (Pickands, 1986). Moreover if one takes  $b(t) = F_R^{-1}(1 - 1/t)$ , and  $a(t) = h_R\{b(t)\}$ , then  $\sigma = 1$  in (1.5), i.e.

$$\lim_{t \rightarrow \infty} \frac{1 - F_R\{a(t)r + b(t)\}}{1 - F_R\{b(t)\}} = \bar{K}(r) = (1 + \lambda r)_+^{-1/\lambda}, \quad r \geq 0.$$

### 3.2 Uniqueness of limits

In general, for a given copula, no unique choice of marginal distribution  $F$  leads to assumption (1.4) being satisfied. Consider, for example, the independence copula, with  $c(u_1, u_2) = 1, (u_1, u_2) \in [0, 1]^2$ . The following cases are all covered by (1.4):

- (i) gamma margins, with shape parameter  $\alpha > 0$ . Then  $R = \|(X_G, Y_G)\|_* = X_G + Y_G$ , has a GP(1, 0) limit. The limiting distribution for  $W$  is Beta( $\alpha, \alpha$ );
- (ii) Weibull margins, with shape parameter  $\alpha > 1$ . Then  $R = \|(X_W, Y_W)\|_* = (X_W^\alpha + Y_W^\alpha)^{1/\alpha}$ , has a GP(1, 0) limit. The limiting distribution for  $W$  has density  $j(w) \propto w^{\alpha-1}(1-w)^{\alpha-1}\{w^\alpha + (1-w)^\alpha\}^{-2}$ ;
- (iii) uniform(0, 1) margins. Then  $R = \|(X_U, Y_U)\|_* = \max(X_U, Y_U)$ , has a GP(1, -1) limit. The limiting distribution for  $W$  has density  $j(w) \propto \max(w, 1-w)^{-2}$ ;
- (iv) truncated Gaussian margins. Then  $R = \|(X_N, Y_N)\|_* = (X_N^2 + Y_N^2)^{1/2}$ , has a GP(1, 0) limit. The limiting distribution for  $W$  has density  $j(w) \propto \{w^2 + (1-w)^2\}^{-1}$ .

The corresponding marginal densities may all be expressed as  $f(x) = x^\beta e^{-x^\gamma} \gamma / \Gamma\{(\beta + 1)/\gamma\}$ , with (i)  $\beta = \alpha - 1, \gamma = 1$ ; (ii)  $\beta = \alpha - 1, \gamma = \alpha$ ; (iii)  $\beta = 0, \gamma \rightarrow \infty$ ; and (iv)  $\beta = 0, \gamma = 2$ . In each case the norm  $\|\cdot\|_*$  is the  $L_\gamma$  norm, and the resulting density for  $W$  satisfies

$$j(w) \propto w^\beta (1-w)^\beta / \{w^\gamma + (1-w)^\gamma\}^{(2\beta+2)/\gamma} = w^\beta (1-w)^\beta / \|(w, 1-w)\|_*^{2\beta+2}, \quad 0 < w < 1,$$

demonstrating a link between the margins of  $(X, Y)$ , the norm  $\|\cdot\|_*$ , and the distribution  $J(w)$ .

This lack of uniqueness also applies to multivariate regularly varying random vectors with asymptotically dependent copulas: equal heavy-tailed margins with any positive shape parameter will give a convergence as in (1.4), and the resulting distribution of  $W$  will depend on this shape parameter and the norm used to define  $R$ ; see Example 1 of Section 3.3. Hence in considering how the distribution  $J$  describes the extremal dependence, one must simultaneously consider  $\lambda, \|\cdot\|_*$  and  $J$ . In convergence (1.3), by contrast, the effect of the margins is removed by standardization, and the extremal dependence depends only on  $H$  and the norm used to define  $R$ .

The necessity of considering  $\lambda, \|\cdot\|_*$  and  $J$  together can be more clearly seen by observing what convergence (1.4) implies for that of the normalized  $(X, Y)$ . Multiplying  $\{\|(X, Y)\|_* - b(t)\}/a(t)$  by  $(X, Y)/\|(X, Y)\|_*$ , and conditioning on the event  $\mathcal{B} = \{\|(X, Y)\|_* > b(t)\}$ , the continuous mapping theorem gives that on  $\mathcal{B}$ ,

$$\frac{(X, Y)}{a(t)} - \frac{b(t)}{a(t)} \frac{(X, Y)}{\|(X, Y)\|_*} \xrightarrow{d} R^*(W_1, W_2), \quad t \rightarrow \infty, \quad (3.4)$$

with  $W_1 = W/\|(W, 1-W)\|_*$ ,  $W_2 = (1-W)/\|(W, 1-W)\|_*$ . Here  $\xrightarrow{d}$  denotes convergence in distribution, and  $R^* \sim \text{GP}(1, \lambda)$  is a random variable with survivor function  $\bar{K}$ . Equations (1.4), (1.7) and (3.4) suggest that for large  $t$  we have the approximate distributional equality on  $\mathcal{B}$ ,

$$(X, Y) \stackrel{d}{\approx} \{a(t)R^* + b(t)\}(W_1, W_2). \quad (3.5)$$

Therefore the extremes of  $(X, Y)$  are described by the combination of the shape parameter  $\lambda$ , the norm  $\|\cdot\|_*$  defining the sphere  $\mathcal{S}^*$  on which  $(W_1, W_2)$  live, and the distribution  $J$  giving the density of  $W$  on  $[0, 1]$ .

### 3.3 Examples

We present three broad classes of examples, assuming throughout that derivatives of second order terms are also second order.

**Example 1.** Suppose that  $(X, Y)$  have  $\alpha$ -Pareto margins,  $P(X > x) = x^{-\alpha}$ ,  $x > 1$ , and that  $P(X > tx, Y > ty)$  is a differentiable bivariate regularly varying function of index  $-\alpha$  as  $t \rightarrow \infty$ . Then one can write

$$P(X > tx, Y > ty) = \{1 + o(1)\}\delta^{(\alpha)}(tx, ty) = \{\chi + o(1)\}d^{(\alpha)}\{x/(x+y)\}(xy)^{-\alpha/2}t^{-\alpha}, \quad t \rightarrow \infty, \quad tx, ty > 1,$$

with  $\chi > 0$  as in (1.1),  $\delta^{(\alpha)}$  a homogeneous function of order  $-\alpha$ , and  $d^{(\alpha)}$  the associated ray dependence function, discussed in Section 2. Such examples are asymptotically dependent. Then taking  $\|\cdot\|_* = \|\cdot\|$ , an arbitrary norm, yields

$$f_{R,W}(r, w) = \{1 + o(1)\}r^{-1-\alpha}\delta_{12}^{(\alpha)}\left\{\frac{w}{\|(w, 1-w)\|}, \frac{1-w}{\|(w, 1-w)\|}\right\}\frac{1}{\|(w, 1-w)\|^2}, \quad r \rightarrow \infty;$$

here  $\delta_{12}^{(\alpha)}$ , the joint derivative of  $\delta^{(\alpha)}$ , is homogeneous of order  $-\alpha - 2$ . The reciprocal hazard function of  $R$  satisfies  $h_R(r) = r\{1/\alpha + o(1)\}$  ( $r \rightarrow \infty$ ), so the limiting distribution of normalized exceedances of  $R$  is generalized Pareto with  $\lambda = 1/\alpha$ . The limiting density of  $W$  is

$$j(w) \propto \delta_{12}^{(\alpha)}(w, 1-w)\|(w, 1-w)\|^\alpha, \quad w \in (0, 1).$$

**Example 2.** Suppose that  $(X, Y)$  have standard exponential margins, and that for a constant  $C > 0$ ,

$$P(X > tx, Y > ty) = \{C + o(1)\}\exp\{-\kappa(x, y)t\}, \quad t \rightarrow \infty, \quad x, y > 0,$$

where  $\kappa : (0, \infty)^2 \rightarrow (0, \infty)$  is a differentiable positive homogeneous function that defines a norm. This special case of the set-up of Wadsworth and Tawn (2013) is satisfied by the Morgenstern, inverted extreme value, Ali–Mikhail–Haq, and Pareto copulas, amongst others; see Heffernan (2000) for a summary of their extremal dependence properties. All such examples are asymptotically independent, with  $\eta = 1/\kappa(1, 1)$ . Let  $\kappa_i$  denote the partial derivative of  $\kappa$  with respect to its  $i$ th argument, and similarly let  $\kappa_{12}$  denote the joint derivative. Taking  $\|(x, y)\|_* = \kappa(x, y)$  gives

$$f_{R,W}(r, w) = \{C + o(1)\}\exp(-r)\left\{\frac{\kappa_1(w, 1-w)\kappa_2(w, 1-w)}{\kappa(w, 1-w)^2}r - \frac{\kappa_{12}(w, 1-w)}{\kappa(w, 1-w)}\right\}, \quad r \rightarrow \infty,$$

which satisfies condition (3.3). Furthermore, since the reciprocal hazard function  $h_R(r) = 1 + o(1)$  as  $r \rightarrow \infty$ ,  $\lambda = 0$ : normalized exceedances of  $R$  have a limiting exponential distribution. The limiting density of  $W$  as  $r \rightarrow \infty$  is

$$j(w) \propto \frac{\kappa_1(w, 1-w)\kappa_2(w, 1-w)}{\kappa(w, 1-w)^2}, \quad w \in (0, 1).$$

**Example 3.** Let  $(X, Y)$  be elliptically distributed, truncated to the positive quadrant, so one can write

$$(X, Y) = Q\Sigma^{1/2}(U_1, U_2),$$

with  $\Sigma^{1/2}$  the Cholesky factor of a positive-definite matrix,  $(U_1, U_2)$  lying on the part of the unit circle such that  $\Sigma^{1/2}(U_1, U_2)$  lies in the positive quadrant, and  $Q$  a random variable known as the generator. Then the norm  $\|(x, y)\|_* = \{(x, y)\Sigma^{-1}(x, y)^T\}^{1/2}$  returns the variable  $Q$ , i.e.,  $R = Q$ . Thus we have exact independence of  $R$  and  $W$ , and the density of  $W$  is

$$j(w) \propto \|(w, 1-w)\|_*^{-2} = (1 - \rho^2)\{w^2 - 2\rho w(1-w) + (1-w)^2\}^{-1}, \quad w \in (0, 1).$$

The exact form of the limiting distribution for exceedances of  $R$  depends on  $Q$ : Abdous et al. (2005) consider extremes of elliptical distributions and provide details on the domain of attraction of the generator. The variables  $X$  and  $Y$  are asymptotically dependent if and only if  $Q$  has regularly varying tails (Hult and Lindskog, 2002). This links precisely to the asymptotic dependence features described in Section 4.2. As highlighted by Example 1, the norm  $\|\cdot\|_*$  may be chosen arbitrarily if  $Q$  has a heavy tail, though an advantage of the norm  $\|(x, y)\|_*$  is that independence is exact, rather than asymptotic, in the sense of equation (1.4). The Gaussian is the best-known elliptical distribution; its extremes are asymptotically independent, with  $R$  having the Weibull density  $f_R(r) = re^{-r^2/2}$  ( $r > 0$ ); thus  $\lambda = 0$ .

Like elliptical copulas, Archimedean survival copulas have a radial-angular decomposition, with the pseudo-angles being uniformly distributed on  $[0, 1]$  (McNeil and Nešlehová, 2009). Thus (1.4) is satisfied whenever the radial variable falls into the domain of attraction of a generalized Pareto distribution.

### 3.4 Application of (1.4)

In order to apply (1.4) directly, one must know the (class of) margins  $F$ , and the (class of) norm  $\|\cdot\|_*$ , to which it applies. The basis of statistical procedures assuming asymptotic dependence is that any choice of heavy-tailed margins and norm will lead to a limit, and so that choice is arbitrary. If asymptotic dependence cannot be assumed, then the correct class of marginal distributions and the correct norm must be chosen, and this makes direct exploitation of (1.4) challenging. One might choose among marginal classes based on some measure of fit, but this would not account for uncertainty in the dependence class. For this reason we aim to construct a model having the essential features of (3.5).

## 4 Model

### 4.1 Introduction

We use the observations of Section 3, and in particular equation (3.5), to motivate a model that can capture both asymptotic dependence and asymptotic independence. Consider the dependence structure of

$$(A, B) = S(V_1, V_2),$$

$$(V_1, V_2) = (V, 1 - V) / \|(V, 1 - V)\|_m \in \mathcal{S}^m = \{\mathbf{v} \in \mathbb{R}_+^2 : \|\mathbf{v}\|_m = 1\}, \quad V \sim F_V \perp\!\!\!\perp S \sim \text{GP}(1, \lambda), \quad (4.1)$$

where  $F_V$  is a distribution defined on  $[0, 1]$ . The norm  $\|\cdot\|_m$  and distribution  $F_V$  are modelling choices;  $\lambda$  and any parameters of  $F_V$  are to be inferred. Model (4.1) reflects the structure of (3.5), which provides an asymptotic representation of the extremes of a wide variety of dependence structures. As we show in Section 4.2, the dependence structure of (4.1) is broad enough to capture both types of extremal dependence structures. Although (4.1) is motivated by (3.5), we adopt different notation in order to emphasize that the former is a modelling approach rather than an assumption on the underlying random vector.

Model (4.1) has parameters that are common to the margins and dependence structure, but we are interested only in exploiting its copula,

$$C(u_1, u_2) = F_{A,B}\{F_A^{-1}(u_1), F_B^{-1}(u_2)\}, \quad (4.2)$$

where,  $F_{A,B}$ ,  $F_A$ , and  $F_B$  are the joint and marginal distribution functions of (4.1). We refer to  $F_A, F_B$  as *pseudo-marginals* throughout, as they are unrelated to the true marginals of the observable random vector, reflecting only those in which the factorization (4.1) holds best for the extremes.

Representation (3.5) holds when a suitable pseudo-radial variable is large. By analogy, it is reasonable to assume that (4.1) holds only when some norm of the variables is large. This will be implemented in our inference strategy, explained in Section 5. Thus, if the observed vector  $(Z_1, Z_2)$  has joint distribution function  $F_{1,2}$ , then we suppose for all sufficiently extreme observations that  $F_{1,2}(z_1, z_2) \approx C\{F_1(z_1), F_2(z_2)\}$ , with  $C$  as in (4.2). Finally note that the fact that  $A$  and  $B$  may have different margins is not incompatible with the spirit of (3.5), as the margins therein are those of  $(X, Y)$  given that  $\|(X, Y)\|_* > 0$ , which may be unequal if the dependence structure is asymmetric.

### 4.2 Extremal dependence properties

We detail the extremal dependence properties of the model (4.1) under some mild restrictions on the types of norm considered and the support of  $V$ . Proofs of all propositions may be found in Appendix A. The following conditions on  $\|\cdot\|_m$  are imposed throughout this section.

**Condition 1** (Symmetry).  $\|(x, y)\|_m = \|(y, x)\|_m$ .

**Condition 2** (Boundary).  $\|(x, y)\|_m \geq \|(x, y)\|_\infty$ .

**Condition 3** (Equality with  $L_\infty$ ).  $\|(x_0, y_0)\|_m = \|(x_0, y_0)\|_\infty$  for some  $x_0 \neq y_0$ .

These conditions specify ranges for the marginal projections  $V_1 = V/\|(V, 1 - V)\|_m$ , and  $V_2 = (1 - V)/\|(V, 1 - V)\|_m$  to be  $[0, 1]$ . In particular the mapping  $T : [0, 1] \rightarrow [0, 1]$  given by  $T(v) = v/\|(v, 1 - v)\|_m$  is surjective. Condition 3 imposes that if equality with  $\|\cdot\|_\infty$  occurs at  $(1, 1)$ , then since we must also have equality somewhere off the diagonal, the norm must behave locally like  $\|\cdot\|_\infty$  around  $(1, 1)$ , by convexity. This specifically rules out cases such as  $\|(x, y)\|_m = \max\{ax + (1 - a)y, ay + (1 - a)x\}$ ,  $a > 1$ , for which  $\|(1, 1)\|_m = 1$ , but which does not behave locally like the  $L_\infty$  norm; these can induce dependence properties different from those claimed under Condition 3.

We focus on the dependence measures  $\chi$  (equation (1.1)) and  $\eta$  (equation (2.1)) and the function  $\kappa$  (equation (2.5)). These were defined following a transformation of the variables to standard Pareto margins, but for exposition of calculation, here we will exploit the equivalence  $\mathbb{P}(X_P > t^\beta, Y_P > t^\gamma) = \mathbb{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\}$ , where  $q_i(t) := F_i^{-1}(1 - 1/t)$  ( $t \geq 1$ ,  $i \in \{A, B\}$ ) is the  $1 - 1/t$  quantile function. Wadsworth and Tawn (2013) show that under asymptotic dependence, if (2.5) holds, then  $\kappa(\beta, \gamma) \equiv \max(\beta, \gamma)$ , whereas more interesting structures are obtained under asymptotic independence. The dependence structure of asymptotically dependent distributions is described by the ray dependence function  $d$  or distribution  $H$  in equation (2.2). We discuss these below, also giving the corresponding quantities for the Ramos–Ledford framework under hidden regular variation.

The marginal and joint survivor functions are key to the study of dependence. The former can be expressed as  $\mathbb{P}(A > x) = \mathbb{E}\{\mathbb{P}(SV_1 > x \mid V_1)\}$  and  $\mathbb{P}(B > y) = \mathbb{E}\{\mathbb{P}(SV_2 > y \mid V_2)\}$ , where, noting the link between  $(V_1, V_2)$  and  $V$ ,  $\mathbb{E}$  denotes expectation with respect to  $V$ . This provides

$$\mathbb{P}(A > x) = \mathbb{E}\left\{\left(1 + \lambda x/V_1\right)_+^{-1/\lambda}\right\}, \quad \mathbb{P}(B > y) = \mathbb{E}\left\{\left(1 + \lambda y/V_2\right)_+^{-1/\lambda}\right\}. \quad (4.3)$$

The joint survivor function can likewise be expressed as

$$\mathbb{P}(A > x, B > y) = \mathbb{E}\left[\left\{1 + \lambda \max(x/V_1, y/V_2)\right\}_+^{-1/\lambda}\right]. \quad (4.4)$$

Below we present  $\chi$ ,  $\eta$  and  $\kappa(\beta, \gamma)$  for the different ranges of  $\lambda$ , and types of norm under consideration. For all cases we assume:

**Assumption 1.** The distribution function of  $V$ ,  $F_V : [0, 1] \rightarrow [0, 1]$ , is continuous and strictly increasing.

Equivalently the measure associated to  $F_V$  has no point masses and its support is the entire unit interval. With  $T$  as defined above, define  $v' := \inf\{v \in [0, 1] : T(v) = 1\}$  and  $v'' := \sup\{v \in [0, 1] : T(v) = 1\}$ .

**Case 1** ( $\lambda > 0$ ). Define the positive quantity

$$\chi_\lambda = \mathbb{E}\left[\min\left\{V_1^{1/\lambda}/\mathbb{E}(V_1^{1/\lambda}), V_2^{1/\lambda}/\mathbb{E}(V_2^{1/\lambda})\right\}\right]. \quad (4.5)$$

**Proposition 1.** *If  $\beta, \gamma > 0$ , then*

$$\mathbb{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\} = t^{-\max(\beta, \gamma)} \theta_{\beta, \gamma}(t), \quad t \geq 1,$$

where  $\theta_{\beta, \gamma}$  is slowly varying at infinity. Furthermore,  $\theta_{\beta, \gamma}(t) \rightarrow \chi_\lambda$  as  $t \rightarrow \infty$  if  $\beta = \gamma$ , and  $\theta_{\beta, \gamma}(t) \rightarrow 1$  otherwise.

It is an immediate corollary that  $\eta = 1$ , and  $\chi = \chi_\lambda > 0$ . However, for any fixed  $F_V$ , as  $\lambda \rightarrow 0$  the dependence weakens to asymptotic independence, by the following:

**Proposition 2.** *Given a fixed  $F_V$ ,  $\chi_\lambda \rightarrow 0$  as  $\lambda \rightarrow 0^+$ .*

**Remark 1.** The ray dependence function (2.2) for  $\lambda > 0$  is

$$d(q) = \frac{1}{\chi_\lambda} \mathbb{E}\left[\min\left\{\frac{V_1^{1/\lambda}}{\mathbb{E}(V_1^{1/\lambda})} \left(\frac{1-q}{q}\right)^{1/2}, \frac{V_2^{1/\lambda}}{\mathbb{E}(V_2^{1/\lambda})} \left(\frac{q}{1-q}\right)^{1/2}\right\}\right], \quad q \in [0, 1].$$



If  $F_V$  has a Lebesgue density  $f_V$ , the associated spectral density  $h(w) = dH(w)/dw$  is given by

$$h(w; \lambda, f_V) = \frac{1}{2} \frac{\lambda^{1-1/\lambda} w^{\lambda-1} (1-w)^{\lambda-1} \mu_1^\lambda \mu_2^\lambda}{\|(w\mu_1)^\lambda, ((1-w)\mu_2)^\lambda\|_m^{1/\lambda} \{(w\mu_1)^\lambda + ((1-w)\mu_2)^\lambda\}^2} \times f_V \left\{ \frac{(\mu_1 w)^\lambda}{(w\mu_1)^\lambda + ((1-w)\mu_2)^\lambda} \right\},$$

with  $\mu_1 = \mathbf{E}(V_1^{1/\lambda})/\lambda^{1/\lambda}$ ,  $\mu_2 = \mathbf{E}(V_2^{1/\lambda})/\lambda^{1/\lambda}$ . This satisfies  $\int_0^1 wh(w; \lambda, f_V)dw = 1/2$ , a necessary moment constraint on  $H$ , even if  $\int_0^1 v f_V(v) dv \neq 1/2$ . Justification for these forms is given in the Supplementary Material.

**Case 2** ( $\lambda = 0$ ).

**Proposition 3.** *Let  $\beta, \gamma > 0$ , and define  $\omega := \beta/(\beta + \gamma)$ . Then*

$$P\{A > q_A(t^\beta), B > q_B(t^\gamma)\} = t^{-\kappa(\beta, \gamma)} \theta_{\beta, \gamma}(t), \quad t \geq 1,$$

where  $\theta_{\beta, \gamma}$  is slowly varying at infinity, and

$$\kappa(\beta, \gamma) = \begin{cases} \|(\beta, \gamma)\|_m, & \omega \in [1 - v', v'] \\ \|(\beta, \gamma)\|_\infty, & \text{otherwise.} \end{cases}$$

It is an immediate corollary that  $\eta = \|(1, 1)\|_m^{-1}$ . When  $\eta < 1$  then  $\chi = 0$ , i.e., we have asymptotic independence. When  $\eta = 1$ , then  $\chi = \lim_{t \rightarrow \infty} \theta_{\beta, \gamma}(t)$  when  $\beta = \gamma$ . Proposition 8 in Appendix A states that this limit is still zero, i.e., we still have asymptotic independence.

**Case 3** ( $\lambda < 0$  and  $\|(1, 1)\|_m = \|(1, 1)\|_\infty$ ). For this case only, we further assume:

**Assumption 2.**  $F_V$  is continuously differentiable near  $1/2$  with  $F'_V(1/2) > 0$ .

**Proposition 4.** *If  $\beta, \gamma > 0$ , then*

$$P\{A > q_A(t^\beta), B > q_B(t^\gamma)\} = t^{-\kappa(\beta, \gamma)} \theta_{\beta, \gamma}(t), \quad t \geq 1,$$

where  $\kappa(\beta, \gamma) = (1 + \lambda) \max(\beta, \gamma) - \lambda(\beta + \gamma)$  and  $\theta_{\beta, \gamma}$  is slowly varying at infinity with

$$\lim_{t \rightarrow \infty} \theta_{\beta, \gamma}(t) = \frac{F'_V(1/2)}{4} \times \begin{cases} m_+^\lambda m_-^{-1}, & \beta < \gamma, \\ \{\min(m_+, m_-)^\lambda - \frac{1+\lambda}{1-\lambda} \max(m_+, m_-)^\lambda\} \max(m_+, m_-)^{-1}, & \beta = \gamma, \\ m_-^\lambda m_+^{-1}, & \beta > \gamma, \end{cases} \quad (4.6)$$

for  $m_+ = P(V \in [v', v''])$  and  $m_- = P(V \in [1 - v'', 1 - v'])$ .

A corollary when  $\beta = \gamma$  is that  $\eta = (1 - \lambda)^{-1}$ . Since  $\eta < 1$  we must have  $\chi = 0$ , asymptotic independence.

**Remark 2.** The ray dependence function (2.4) in this case is

$$d(q) = \{q(1 - q)\}^{\frac{1-\lambda}{2}} \frac{\min\{qm_+, (1 - q)m_-\}^\lambda \max\{qm_+, (1 - q)m_-\}^{-1} - \frac{1+\lambda}{1-\lambda} \max\{qm_+, (1 - q)m_-\}^{\lambda-1}}{\min(m_+, m_-)^\lambda \max(m_+, m_-)^{-1} - \frac{1+\lambda}{1-\lambda} \max(m_+, m_-)^{\lambda-1}};$$

see the Supplementary Material. The density of the associated measure  $H_\eta$  can be calculated as in Beirlant et al. (2004, Section 9.5.3).

**Case 4** ( $\lambda < 0$  and  $\|(1, 1)\|_m > \|(1, 1)\|_\infty$ ). In this case  $\chi = 0$ , but the regular variation assumptions (2.1) and (2.5) are not satisfied. The marginal densities have upper endpoint  $-1/\lambda$ , i.e.,  $q_A(t^\beta), q_B(t^\gamma) \rightarrow -1/\lambda$  as  $t \rightarrow \infty$ , but the upper endpoint of the joint survivor function is strictly below  $-1/\lambda$ , as can be seen by substituting  $x = q_A(t^\beta)$ ,  $y = q_B(t^\gamma)$  in (4.4); this probability will be exactly zero whenever

$$\max\{q_A(t^\beta)/V_1, q_B(t^\gamma)/V_2\} \geq -1/\lambda, \quad (V_1, V_2) \in \mathcal{S}^m. \quad (4.7)$$

For  $a, b, c, d > 0$ ,  $\max(a/b, c/d) \leq \max(a, c)/\min(b, d)$ , yielding  $\max(a, c) \geq \min(b, d) \max(a/b, c/d)$ , so

$$\max\{q_A(t^\beta)/V_1, q_B(t^\gamma)/V_2\} \geq \min\{q_A(t^\beta), q_B(t^\gamma)\} \max(1/V_1, 1/V_2).$$

Moreover  $\max(1/V_1, 1/V_2) = 1/\min(V_1, V_2) \geq \|(1, 1)\|_m$ , since  $\min(V_1, V_2)$  is largest when  $V = 1/2$ . Combining these two observations we have

$$\max\{q_A(t^\beta)/V_1, q_B(t^\gamma)/V_2\} \geq \min\{q_A(t^\beta), q_B(t^\gamma)\} \|(1, 1)\|_m \rightarrow -\|(1, 1)\|_m/\lambda > -1/\lambda, \quad t \rightarrow \infty,$$

so there is a  $t_0 < \infty$  such that (4.7) is satisfied for all  $t > t_0$ . It follows that  $\chi = 0$ , whereas  $\eta$  and  $\kappa(\beta, \gamma)$  are ill-defined.

Propositions 1, 3, 4 and Remark 1 show how different combinations of  $\lambda$ ,  $F_V$  and  $\|\cdot\|_m$  influence extremal dependence properties, under the assumed conditions on the support of  $V$  and type of norm. To summarize: asymptotic dependence is present when  $\lambda > 0$ , with the dependence then described by  $d(q)$  given in Remark 1, determined by  $\lambda$ ,  $F_V$  and  $\|\cdot\|_m$ . Asymptotic independence is present when  $\lambda \leq 0$ ; for  $\lambda = 0$ ,  $\kappa$  is determined by the shape of  $\|\cdot\|_m$ , while for  $\lambda < 0$ , hidden regular variation only arises if  $\|(1, 1)\|_m = 1$ . Overlap in dependence structures might seem to arise when  $\lambda = 0$  and  $\|(\beta, \gamma)\|_m = \delta(\beta + \gamma) + (1 - \delta)\max(\beta, \gamma)$ ,  $\delta \in (0, 1]$ , since this matches the case  $\lambda \in [-1, 0)$  and  $\|(1, 1)\|_m = \|(1, 1)\|_\infty$ . However, Proposition 4 shows that in general the slowly varying function arising when  $\lambda < 0$  depends on the properties of the norm  $\|\cdot\|_m$  used for a fixed distribution  $F_V$ , whereas the slowly varying function arising when  $\lambda = 0$  cannot change in this way.

### 4.3 Transition between dependence classes

Due to the focus on limits such as (1.1), the classification between asymptotic dependence and asymptotic independence is viewed as dichotomous: either the joint and marginal survivor probabilities decay at the same rate or they do not. Where existing modelling approaches are suitable for both dependence types, the transition between them occurs on the boundary of a parameter space, inducing an undesirable discontinuity in the extremal dependence features. For example, consider  $\chi(u) := \mathbb{P}\{F_1(Z_1) > u \mid F_2(Z_2) > u\}$  ( $u \in [0, 1]$ ). In the Ramos–Ledford–Tawn approach, when  $\eta = 1$  there is an instant “jump” to  $\chi(u) \equiv \chi > 0$  for all  $u$  above the level at which the model is assumed to hold, whereas when  $\eta < 1$ ,  $\chi(u) \rightarrow 0$  as  $u \rightarrow 1$ . Similarly in the Heffernan–Tawn model, when  $\alpha = 0, \beta = 1$ , the value of  $\chi(u) \equiv 1 - \int_0^\infty G(-v)e^{-v} dv$  for all  $u$  above the level at which the model is assumed to hold, where  $G$  is as in limit (2.6), whereas  $\chi(u) \rightarrow 0$  for all other values of  $(\alpha, \beta)$ . Consequently, *any* decrease in an empirically estimated  $\chi(u)$  suggests that asymptotic independence will be inferred under the Ramos–Ledford–Tawn and Heffernan–Tawn models.

An elegant feature of model (4.1) is the smoothness of the transitions across dependence classes in  $\lambda$ , and the fact that asymptotic independence or dependence does not occur at boundary points for  $\lambda$ . In particular when  $\lambda \rightarrow 0^+$ , the function  $\chi_\lambda$  defined in (4.5) tends to zero, and the value of the function  $\chi(u) \equiv \chi_\lambda(u)$  discussed in Section 4.3 may depend on  $u$  in regions where the model holds, thereby smoothing out some of the discontinuity discussed above. Furthermore, if  $\|(1, 1)\|_m = 1$ , achieved if we set  $\|\cdot\|_m = \|\cdot\|_\infty$ , then  $\chi_\lambda \rightarrow 0$  as  $\lambda \rightarrow 0^+$  and  $\eta$  decreases from 1 at  $\lambda = 0$  towards 0 as  $\lambda \rightarrow -\infty$ . In this sense the model smoothly interpolates across the dependence classes. We will adopt these modelling choices in Section 5.

## 5 Inference

### 5.1 Likelihood and parameterization

We now consider fitting (4.1) as a dependence model for extreme bivariate data by likelihood methods. Let  $F_A, F_B$  and  $f_A, f_B > 0$  denote the pseudo-marginal distribution and density functions respectively, and let  $f_{A,B}$  denote the joint density of  $(A, B)$ . The density corresponding to the copula  $C(u_1, u_2)$  is

$$c(u_1, u_2) = \frac{f_{A,B}\{F_A^{-1}(u_1), F_B^{-1}(u_2)\}}{f_A\{F_A^{-1}(u_1)\}f_B\{F_B^{-1}(u_2)\}}, \quad 0 \leq u_1, u_2 \leq 1.$$

Recall that  $(V_1, V_2) = (V, 1-V)/\|(V, 1-V)\|_m$ ; we assume that  $V$  has a Lebesgue density (thus Assumptions 1 and 2 are satisfied), denoted by  $f_V$ . Using the independence of  $S$  and  $V$  we obtain the joint density

$$f_{A,B}(x, y) = \frac{\|(x, y)\|_m}{(x+y)^2} \{1 + \lambda\|(x, y)\|_m\}_+^{-1/\lambda-1} f_V\left(\frac{x}{x+y}\right), \quad x, y > 0.$$

The pseudo-marginal density and distribution functions required to compute  $c(u_1, u_2)$  are not explicit, requiring numerical evaluation of a one-dimensional integral.

We only wish to use model (4.1) for extreme dependence, so we must censor non-extreme data. Since the margins and dependence have a common parameterization, it is only straightforward to censor on regions that remain of the same form under marginal transformation. We therefore choose to censor data for which the maximum value on the uniform marginal scale is less than some  $u$  close to 1. This translates to the uncensored variables having  $\max(A, B)$  large, and by equivalence of norms, any  $\|(A, B)\|_m$  will also be large. Thus the likelihood that we use for independent pairs  $(u_{1,1}, u_{2,1}), \dots, (u_{1,n}, u_{2,n})$  with uniform margins is

$$L(\zeta) = \prod_{i: \max(u_{1,i}, u_{2,i}) > u} c(u_{1,i}, u_{2,i}; \zeta) \prod_{i: \max(u_{1,i}, u_{2,i}) \leq u} C(u, u; \zeta), \quad (5.1)$$

with  $\zeta$  a parameter vector. In practice the data must be transformed to uniform margins using the probability integral transform. One possibility is semiparametric transformation, using the empirical distribution below a high threshold and the asymptotically-motivated generalized Pareto distribution above it (Coles and Tawn, 1991). A simpler alternative is to use the empirical distribution function throughout. The properties of censored two-stage parametric and semiparametric maximum likelihood estimators of copula parameters are explored in Shih and Louis (1995).

In this implementation, we constrain  $\lambda \leq 1$ . In order to fit the model, points must be transformed onto  $A, B$  pseudo-margins using numerical inversion; if  $\lambda$  is large, then numerical instabilities may arise because the pseudo-margins are heavy-tailed. Considering the form of  $h(\cdot; \lambda, f_V)$  given in Remark 1, this still yields a slightly richer class of spectral densities than those defined simply by  $f_V$ . The complete set of parameters is determined by the choice of  $f_V$  and any parameterization of the norm  $\|\cdot\|_m$ . Below we take

$$V \sim \text{Beta}(\alpha, \alpha) \quad \text{and} \quad \|\cdot\|_m = \|\cdot\|_\infty, \quad (5.2)$$

giving  $\zeta = (\lambda, \alpha)$ . The beta distribution is chosen for its simplicity and flexibility of shape, but might be replaced by other distributions. As mentioned in Section 4.2, (5.2) permits all possible  $\chi$  and  $\eta$  values; it also provides a simple model for the dependence structure in both asymptotic independence and dependence frameworks, through the attainable forms of  $\kappa$ , and ray dependence function  $d$ . Although (5.2) represents a misspecification for each of the dependence structures to be used in Section 5.3, our numerical results suggest that it works reasonably well.

Recalling Section 3.2, the choice of a fixed norm in model (4.1) is not as restrictive as might first appear. Since the extremal dependence depends on the combination of  $\lambda$ ,  $\|\cdot\|_m$  and the distribution of  $V$ , the fixing of the norm can be offset by the other model elements to yield a good representation of the data anyway.

An R package for fitting and checking model (4.1), `EVcopula`, is available at [www.lancaster.ac.uk/~wadswojl/](http://www.lancaster.ac.uk/~wadswojl/).

## 5.2 Parameter Identifiability

The parameters  $(\lambda, \alpha)$  of the model defined by (4.1) and (5.2) exhibit negative association, as increasing either parameter whilst fixing the other gives stronger dependence. When the data derive from an asymptotically dependent random vector exhibiting multivariate regular variation, this trade-off may be particularly strong, because each  $\lambda > 0$  leads to a spectral density (in the sense described in Section 1, derived using standard Pareto margins and the  $L_1$  norm)  $h(\cdot; \lambda, f_V)$ , as detailed in Remark 1. With the modelling choices in (5.2) the spectral density  $h(\cdot; \lambda, f_V)$  simplifies to

$$h(w; \lambda, \alpha) = \frac{1}{2\mu} \frac{\lambda^{1-1/\lambda} w^{\lambda\alpha-1} (1-w)^{\lambda\alpha-1}}{\{w^\lambda + (1-w)^\lambda\}^{2\alpha} \max(w, 1-w)} \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}, \quad 0 < w < 1,$$

with  $\mu = \mu_1 = \mu_2$  due to symmetry. A dominant factor in maximum likelihood estimation of  $(\lambda, \alpha)$  is thus the combination of these parameters giving a spectral density most similar to the underlying truth. Although the parameters have different roles, in practice there are many combinations that yield a similar  $h(\cdot; \lambda, \alpha)$ . To determine if the resulting identifiability issues matter in applications, we suggest inspection of the joint log-likelihood surface for  $(\lambda, \alpha)$ ; we implement this in Section 6. More generally for other norms and choices of  $f_V$ , parameter identifiability must be considered.

### 5.3 Simulation

For three different dependence structures, we estimate the probability of lying in rectangular-shaped sets  $(u_1, v_1) \times (u_2, v_2)$  on the copula scale, where  $u_1 < v_1$ ,  $u_2 < v_2$ , and  $u_2$  represents an extreme quantile. We call them Set 1, ..., Set 5, with  $(u_1, v_1) = (0.05, 0.2), (0.2, 0.4), (0.4, 0.6), (0.6, 0.8), (0.8, 0.9999)$ , and  $(u_2, v_2) = (0.995, 0.99995)$ , respectively. We compare our method to that of Heffernan and Tawn (2004), the only other approach easily able to estimate probabilities when the components may not both be extreme.

We simulate 100 replicate samples of size 1000 from (i) the bivariate extreme value distribution with symmetric logistic dependence structure (Coles and Tawn, 1991); (ii) the inverted copula of (i) (Ledford and Tawn, 1997); and (iii) the bivariate normal distribution. The first is asymptotically dependent, and the others are asymptotically independent. We use dependence parameters  $\{0.2, 0.4, 0.6, 0.8\}$ , representing decreasing dependence for (i) and (ii) and increasing dependence for (iii): we label the dependence levels from 1–4 in order of increasing strength. The censoring threshold in likelihood (5.1) was  $u = 0.95$ , and the data were transformed to uniformity using the empirical distribution function. Estimation for the Heffernan and Tawn (2004) method was based on all data whose  $Y$  coordinate exceeded a 90% quantile threshold.

Table 1 displays the root mean squared errors (RMSEs) of the log of all non-zero estimated probabilities. For our model, we define a probability to be zero if its estimate is less than twice machine epsilon in R, since numerical procedures are involved in the calculations; this can occasionally produce negative numbers, which we also set to zero. The number of probabilities estimated as zero is also provided in the table. Overall the new model produces estimates with lower RMSEs than the Heffernan–Tawn model. Any exceptions arise when the Heffernan–Tawn model estimates only a very few non-zero probabilities. In general, estimation for sets closer to one of the axes is better when the dependence is lower. This seems natural as when dependence is high, few if any points in a dataset will be observed near the axes. Both models perform poorly under strong dependence for sets near the axes. Future work could explore whether a more sophisticated implementation of our approach, such as allowing different  $f_V, \|\cdot\|_m$ , or changing the censoring scheme, improves this.

As a diagnostic for the model fit, we also consider the extremal dependence functions  $\chi(u)$ , defined in Section 4.3, and  $\bar{\chi}(u) := 2 \log(1 - u) / \log\{P(F_1(Z_1) > u, F_2(Z_2) > u)\} - 1$  for  $u \in (0.9, 0.999)$  (Coles et al., 1999). As  $u \rightarrow 1$ ,  $\chi(u) \rightarrow \chi$ , as in (1.1), whilst  $\bar{\chi}(u) \rightarrow \bar{\chi} = 2\eta - 1 \in [-1, 1]$ . The value of  $\chi$  thus gives some discrimination between different asymptotically dependent copulas, whilst  $\bar{\chi}$  can discriminate between different asymptotically independent copulas. As functions of  $u$ ,  $\chi(u)$  and  $\bar{\chi}(u)$  are useful for checking model fits under either dependence scenario. In the Supplementary Material, Figure 1 displays pointwise medians and 90% confidence intervals of  $\chi(u), \bar{\chi}(u)$  for each dependence structure and for both methods of inference. Small biases of the new model are typically offset by lower variability and better performance away from the diagonal, i.e., away from the region on which  $\chi(u)$  and  $\bar{\chi}(u)$  focus.

## 6 Environmental application

We consider an oceanographic dataset comprising measurements of wave height, surge and wave period recorded at Newlyn, U.K., filtered to correspond to a 15-hour time window for approximate temporal independence, and previously analyzed by Coles and Tawn (1994), Bortot et al. (2000) and Coles and Pauli (2002). Coles and Tawn (1994) noted the presence of seasonality, which was not taken into account in their, or subsequent, analyses; for ease of comparison we also ignore it. Coles and Tawn (1994) used an asymptotically dependent model for these data, whilst Bortot et al. (2000) used an asymptotically independent Gaussian tail model. Coles and Pauli (2002) employed a mixture-type model, able to encompass both dependence types, with asymptotic dependence arising at a boundary point. The literature appears to have reached a consensus that there is strong, but not overwhelming, evidence for asymptotic dependence between wave height and surge, and fairly strong evidence for asymptotic independence between the other two pairs.

Here we fit the simple symmetric model (5.2), with dependence threshold  $u = 0.95$  in likelihood (5.1). Marginal transformations to uniformity were carried out using the semiparametric procedure of Coles and Tawn (1991) described in Section 5.1, but the dependence parameter estimates were almost the same using the fully empirical marginal transformation.

Table 2 gives maximum likelihood estimates and confidence intervals for the dependence parameters. The estimate  $\hat{\lambda} = 0.54$  suggests asymptotic dependence between wave height and surge, whilst the values  $\hat{\lambda} = -0.21$  and  $\hat{\lambda} = -0.43$  indicate asymptotic independence for the pairs involving period; this is supported

Table 1: RMSEs of non-zero log-probabilities and number of zero estimated probabilities for the new model (New) and Heffernan and Tawn model (HT) for dependence structures (i)–(iii).

Dep. / Method	Level	RMSE					Number of zeroes				
		Set 1	Set 2	Set 3	Set 4	Set 5	Set 1	Set 2	Set 3	Set 4	Set 5
(i) / New	1	0.47	0.39	0.33	0.28	0.095	1	0	0	0	0
	2	1.70	1.00	0.71	0.52	0.023	0	0	0	0	0
	3	5.30	4.30	3.30	1.90	0.0011	41	2	0	0	0
	4	13.00	11.00	6.90	8.90	0.0009	95	97	85	62	0
(i) / HT	1	1.70	1.40	1.40	0.69	0.17	45	19	8	0	0
	2	1.10	1.10	1.00	1.50	0.033	98	87	57	18	0
	3	–	4.40	3.70	2.00	0.02	100	99	99	89	0
	4	–	–	–	–	0.018	100	100	100	100	0
(ii) / New	1	0.25	0.15	0.13	0.10	0.16	0	0	0	0	0
	2	0.53	0.27	0.24	0.19	0.11	0	0	0	0	0
	3	2.50	1.30	0.66	0.41	0.043	5	0	0	0	0
	4	6.90	5.20	3.60	1.90	0.0041	20	11	7	0	0
(ii) / HT	1	1.60	0.93	0.40	0.29	0.31	16	4	0	0	0
	2	0.90	1.30	1.30	0.55	0.20	56	20	1	0	0
	3	2.10	0.92	1.20	1.20	0.067	94	73	30	1	0
	4	–	–	–	1.20	0.02	100	100	100	82	0
(iii) / New	1	0.24	0.17	0.14	0.13	0.20	0	0	0	0	0
	2	0.52	0.38	0.29	0.18	0.17	0	0	0	0	0
	3	1.20	0.75	0.56	0.35	0.095	1	0	0	0	0
	4	3.60	2.10	1.40	0.88	0.016	19	3	0	0	0
(iii) / HT	1	1.70	0.86	0.41	0.29	0.37	13	2	0	0	0
	2	1.20	1.40	0.79	0.33	0.21	51	15	1	0	0
	3	2.70	1.30	1.40	1.00	0.10	93	69	27	1	0
	4	–	–	2.30	1.50	0.021	100	100	96	54	0

Table 2: Maximum likelihood estimates and 95% profile likelihood confidence intervals for  $(\lambda, \alpha)$ , for the pairs of oceanographic variables.

	Height–Surge	Height–Period	Surge–Period
$\hat{\lambda}$	0.54 (0.26, 0.90)	–0.21 (–0.40, –0.03)	–0.43 (–0.66, –0.16)
$\hat{\alpha}$	0.58 (0.40, 0.81)	2.21 (1.58, 3.06)	0.68 (0.50, 0.90)

by the confidence intervals. The likelihood surfaces plotted in Figure 1 show that the parameters are identifiable and give an appreciation of the joint asymptotic confidence regions. Figure 2 shows the empirical and fitted functions  $\chi(u)$  and  $\bar{\chi}(u)$ , which suggest a reasonable fit to the data. Fits from the Heffernan–Tawn model are also displayed, conditioned on each variable in turn, and show potential discrepancies in the inferred strength of the dependence; by having only a single model, we can avoid such discrepancies and the need to decide which variable should be chosen for conditioning upon.

A further diagnostic is presented in Figures 3 (a) and (b), where “fitted” values of  $\hat{S} = \max(\hat{A}, \hat{B})$ , and  $\hat{V} = \hat{A}/(\hat{A} + \hat{B})$  are plotted for the pairs height–surge and period–surge, on a uniform scale. Plots for height–period are similar to those for period–surge, and hence are omitted. Here  $(\hat{A}, \hat{B}) = [\hat{F}_A^{-1}\{\tilde{F}_1(Z_1)\}, \hat{F}_B^{-1}\{\tilde{F}_2(Z_2)\}]$ , where  $\hat{F}_A = \hat{F}_B$  is the fitted common pseudo-marginal distribution, and  $\tilde{F}_1, \tilde{F}_2$  are the estimated true marginals. Points are plotted corresponding to  $(\hat{S}, \hat{V})$  where  $\hat{S}$  exceeds its 90% quantile. A lack of discernible patterns in Figures 3 (a) and (b) suggests that independence of  $S$  and  $V$  is a reasonable approximation. For comparison, Figures 3 (c) and (d) show equivalent plots with  $\max(X_P, Y_P)$  and  $X_P/(X_P + Y_P)$  on a uniform scale,  $(X_P, Y_P) = [\{1 - \tilde{F}_1(Z_1)\}^{-1}, \{1 - \tilde{F}_2(Z_2)\}^{-1}]$ ; this would be the approach to modelling under asymptotic dependence (Coles and Tawn, 1991). The patterns in Figure 3 (c) suggest that asymptotic dependence

is plausible, but that a higher threshold is required for independence of  $\max(X_P, Y_P)$  and  $X_P/(X_P + Y_P)$ . Figure 3 (d) shows that these variables would be dependent at any finite threshold.

## 7 Extensions and discussion

We have provided an alternative limit representation for bivariate extremes, which motivates a statistical model that can capture a wide spectrum of asymptotically dependent and asymptotically independent behaviour. An obvious question concerns extensions to higher dimensions. Assumption (1.4) is indeed simple to extend to the multivariate case: in some common margins,  $F$ , the vector of positive random variables  $\mathbf{X} = (X_1, \dots, X_k) = [F^{-1}\{F_1(Z_1)\}, \dots, F^{-1}\{F_k(Z_k)\}]$  satisfies

$$\lim_{t \rightarrow \infty} \mathbb{P} \left( \frac{\mathbf{X}}{\sum_{i=1}^k X_i} \leq \mathbf{w}, \|\mathbf{X}\|_* > a(t)r + b(t) \mid \|\mathbf{X}\|_* > b(t) \right) = J(\mathbf{w})\bar{K}(r), \quad r \geq 0, \quad (7.1)$$

at continuity points of  $J$ , with  $J$  placing mass on the interior of  $\mathcal{S}_{k-1}^1 = \{\mathbf{w} \in \mathbb{R}_+^k : \|\mathbf{w}\|_1 = 1\}$ , and  $\bar{K}$  as in (1.5). This is a more general assumption than multivariate regular variation, the  $k$ -dimensional extension of (1.3), that underpins much of classical multivariate extreme value theory (de Haan and de Ronde, 1998).

However, the practical applicability of assumption (7.1) in higher dimensions is more limited than in the bivariate case. The assumption that the distribution of  $\mathbf{W} := \mathbf{X} / \sum_{i=1}^k X_i$  has mass on the interior of  $\mathcal{S}_{k-1}^1$  requires a certain regularity in the multivariate dependence structure, which is present in many theoretical examples, such as in the multivariate extensions of Examples 1–3, but often absent in datasets. For example, the data analyzed in Section 6 exhibited asymptotic dependence between one pair of variables, but asymptotic independence between the other two pairs. The only existing model which can handle this is that of Heffernan and Tawn (2004). However there are obvious issues with the curse of dimensionality when using a semiparametric model for higher dimensions. The simulation study in Section 5 demonstrated a tendency for the semiparametric distribution estimator not to cover all parts of the plane, and this drawback would be exacerbated in higher dimensions.

We have assumed throughout that the radial variable  $R = \|(X, Y)\|_*$  is defined by a norm, following the development of much of classical multivariate extreme value theory. In fact the convexity property does not appear necessary, and some recent articles on multivariate extremes have shifted focus on to positive homogeneous functions rather than norms (e.g. Dombry and Ribatet, 2015; Scheffler and Stoev, 2015). For our model the convexity property of  $\|\cdot\|_m$  was used in some of the derivations; further work could explore more deeply the consequences of relaxing this assumption.

A simple extension to the practical modelling introduced in Sections 5 and 6 is to allow an asymmetric dependence structure. Our theoretical results in Section 4 already cover this scenario, but for simplicity of implementation we assumed the distribution of  $V$  to be symmetric, so that the pseudo-marginals of  $A, B$  were equal. As noted in Remark 1, the implied  $H$  incorporates the necessary moment constraint for any  $F_V$ .

In essence our approach is intermediate between assuming multivariate regular variation and the approach of Heffernan and Tawn (2004). With the former, both the marginal distribution and the form of the normalization of each marginal variable, i.e.,  $\mathbf{X}_P / \|\mathbf{X}_P\|$ , are fixed. This is restrictive, but allows for simpler characterization of the consequences of the assumption. With the latter, the margins are fixed to be of exponential type, but the form of the normalization of each marginal variable,  $\{\mathbf{X}_E - \mathbf{b}(Y_E)\} / \mathbf{a}(Y_E)$ , is not fixed. This permits great flexibility in the variety of distributions that satisfy the assumption, but leaves  $k$  possible limits, each with  $2(k-1)$  parameters to estimate, and a  $(k-1)$ -dimensional empirical distribution. Our main assumption does not fix the form of the margins, but does fix the form of the normalization of the variables  $\mathbf{X} / \|\mathbf{X}\|_*$ . This offers greater flexibility than multivariate regular variation, and although less flexible than the model of Heffernan and Tawn (2004) has the benefit of giving only a single limit. In the bivariate case, model (4.1), inspired by (1.4), permits inference across both extremal dependence classes, with a smooth transition between them.

### Acknowledgements

This work was undertaken whilst JLW was based at EPFL and the University of Cambridge. We thank the Swiss National Science Foundation for funding, and the referees and associate editor for comments that have

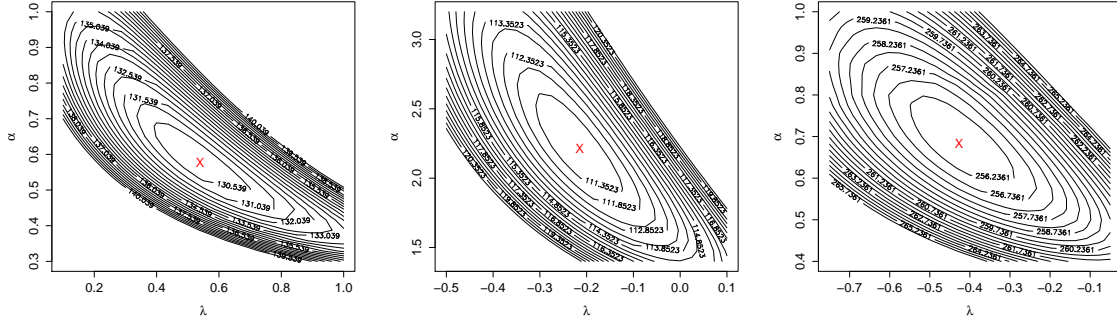


Figure 1: Negative log-likelihood surfaces for height–surge, height–period and surge–period, respectively. Contours are in steps of 0.5. Crosses show maximum likelihood estimates.

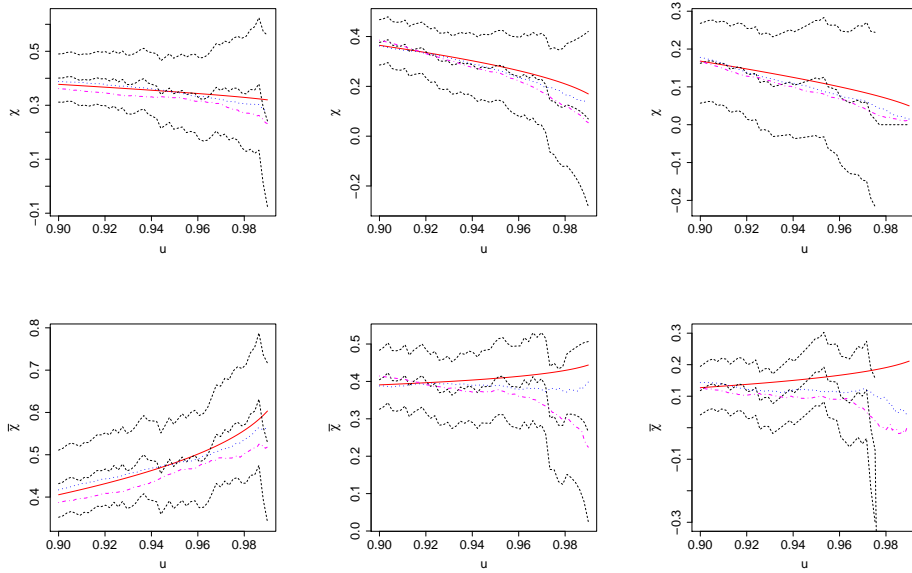


Figure 2: Empirical (dashed lines, with approximate 95% pointwise confidence intervals) and fitted (solid line) estimates of  $\chi(u)$  (top row) and  $\bar{\chi}(u)$  (bottom row), for  $u \in (0.9, 0.99)$ . From left to right the pairs are height–surge, height–period and period–surge. Dot-dash lines: Heffernan–Tawn fit conditioning on the first variable of the pair. Dotted lines: Heffernan–Tawn fit conditioning on the second variable of the pair.

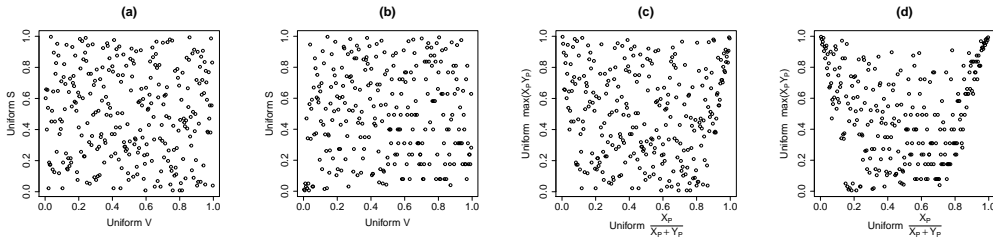


Figure 3: Fitted  $\hat{S}$  and  $\hat{V}$ , on a uniform scale, for (a) height–surge, and (b) period–surge. For comparison,  $\max(X_P, Y_P)$  and  $X_P/(X_P + Y_P)$  defined from the variables transformed to a standard Pareto scale are given for the same pairs in (c) and (d), respectively. Points which are aligned on the  $\hat{S} / \max(X_P, Y_P)$  axis are due to rounding of the data.

greatly improved the work.

## A Auxiliary results and proofs

### A.1 Link between (1.4) and (3.2)

**Proposition 5.** *Let  $W = X/(X + Y)$ ,  $R = \|(X, Y)\|_*$ , and assume that  $W$  and  $R$  have a joint density. Further assume  $R$  to be in the domain of attraction of a generalized Pareto distribution, with normalization functions  $a(t) > 0$ ,  $b(t)$ . Then, provided that the limit on the right exists,*

$$\lim_{t \rightarrow \infty} P\{W \leq w, R > a(t)r + b(t) \mid R > b(t)\} = J(w)\bar{K}(r) \quad \Leftrightarrow \quad \lim_{t \rightarrow \infty} P\{W \leq w \mid R = b(t)\} = J(w).$$

*Proof.* Right to left: The statement on the right is equivalent to

$$\lim_{t \rightarrow \infty} \frac{\frac{\partial}{\partial b(t)} P\{W \leq w, R > b(t)\}}{\frac{\partial}{\partial b(t)} P\{R > b(t)\}} = J(w).$$

Since both  $\lim_{t \rightarrow \infty} P\{W \leq w, R > b(t)\}$  and  $\lim_{t \rightarrow \infty} P\{R > b(t)\}$  equal zero, but the ratio of the derivatives has limit  $J(w)$ , the general form of l'Hôpital's rule states that

$$\lim_{t \rightarrow \infty} \frac{P\{W \leq w, R > b(t)\}}{P\{R > b(t)\}} = J(w).$$

Consequently, as  $t \rightarrow \infty$ ,

$$P\{W \leq w, R > a(t)r + b(t) \mid R > b(t)\} = \frac{P\{W \leq w, R > a(t)r + b(t)\} P\{R > a(t)r + b(t)\}}{P\{R > a(t)r + b(t)\} P\{R > b(t)\}} \rightarrow J(w)\bar{K}(r).$$

Left to right: Set  $r = 0$  in the left-hand statement, yielding

$$\lim_{t \rightarrow \infty} \frac{P\{W \leq w, R > b(t)\}}{P\{R > b(t)\}} = J(w)\bar{K}(0),$$

and note that  $\bar{K}(0) = 1$ . Then applying l'Hôpital's rule again provides

$$\lim_{t \rightarrow \infty} \frac{\frac{\partial}{\partial b(t)} P\{W \leq w, R > b(t)\}}{\frac{\partial}{\partial b(t)} P\{R > b(t)\}} = J(w).$$

□

### A.2 Proofs of Propositions 1–4

We prove Propositions 1–4, giving the values of  $\chi$ ,  $\eta$  and  $\kappa$  claimed in Section 4.2. Omitted proofs of Lemmas 1 and 2, and Propositions 6, 7, 8 and 9 are provided in the Supplementary Material. The following lemma on inversion of regularly varying functions will be useful throughout.

**Lemma 1.** *Suppose  $\beta > 0$  and  $\phi$  is a slowly varying function such that  $s \mapsto s^{-\beta}\phi(s)$  defines a continuous strictly decreasing function from  $[s_0, \infty)$  onto  $(0, 1]$  for some  $s_0$ . Then we can find a slowly varying function  $u$  defined on  $[1, \infty)$  such that  $s^{-\beta}\phi(s) = t^{-\beta}$  whenever  $s = tu^{1/\beta}(t)$ . Furthermore  $u(t) \rightarrow c$  as  $t \rightarrow \infty$  iff  $\phi(s) \rightarrow c$  as  $s \rightarrow \infty$  (here  $c$  can be any value in the extended range  $[0, +\infty]$ ).*

The slowly varying functions  $\phi^{-1/\beta}$  and  $u^{1/\beta}$  are de Bruijn conjugates.

Define  $\tau : [0, 1] \rightarrow [1, \infty]$  as the reciprocal of  $T : [0, 1] \rightarrow [0, 1]$  defined in Section 4.2, i.e.,  $\tau(v) = \|(v, 1 - v)\|_m/v$ , so that  $\tau(V) = 1/V_1$  and  $\tau(1 - V) = 1/V_2$ . Using this notation equation (4.3) becomes

$$P(A > x) = \int_0^1 \{1 + \lambda x \tau(v)\}_+^{-1/\lambda} dF_V(v), \quad P(B > y) = \int_0^1 \{1 + \lambda y \tau(1 - v)\}_+^{-1/\lambda} dF_V(v), \quad (\text{A.1})$$



where the upper endpoint of the support is  $\Lambda = +\infty$  if  $\lambda \geq 0$  and  $\Lambda = -1/\lambda$  if  $\lambda < 0$ ; and (4.4) becomes

$$\mathbb{P}(A > x, B > y) = \int_0^1 [1 + \lambda \max\{x\tau(v), y\tau(1-v)\}]_+^{-1/\lambda} dF_V(v) \quad (\text{A.2a})$$

$$= \int_0^{x/(x+y)} \{1 + \lambda x\tau(v)\}_+^{-1/\lambda} dF_V(v) + \int_{x/(x+y)}^1 \{1 + \lambda y\tau(1-v)\}_+^{-1/\lambda} dF_V(v). \quad (\text{A.2b})$$

The expressions  $x \mapsto \mathbb{P}(A > x)$  and  $y \mapsto \mathbb{P}(B > y)$  define continuous strictly decreasing functions from  $[0, \Lambda)$  onto  $(0, 1]$ ; this observation can be used to help justify the conditions for Lemma 1 when it is used below.

From Condition 2,  $\tau(v) \geq (1-v)/v > 1$  for  $v < 1/2$ , while Conditions 1 and 3 imply  $\tau(v) = 1$  for some  $v \in [1/2, 1]$ . Set  $\Omega_0 = \{v \in [0, 1] : \tau(v) = 1\}$ . Now  $\tau(v) = \tilde{\tau}(1/v)$  where  $\tilde{\tau} : [1, \infty] \rightarrow [1, \infty]$  is the continuous convex function defined by  $\tilde{\tau}(u) = \|(1, u-1)\|_m$ . It follows that  $\Omega_0$  is a closed subinterval of  $[1/2, 1]$ , so  $\Omega_0 = [v', v'']$  with  $v', v''$  as defined in Section 4.2. Also note that  $1/2 \leq v' \leq v'' \leq 1$ ,

$$\tau \text{ is strictly decreasing on } [0, v'] \text{ and strictly increasing on } [v'', 1], \quad (\text{A.3})$$

and

$$v \leq \frac{x}{x+y} \iff yv \leq x(1-v) \iff y\tau(1-v) \leq x\tau(v). \quad (\text{A.4})$$

The quantities  $m_+$  and  $m_-$  as given in Proposition 4 can be expressed

$$m_+ = \int_{\Omega_0} dF_V(v) = F_V(v'') - F_V(v') \quad \text{and} \quad m_- = \int_{1-\Omega_0} dF_V(v) = F_V(1-v') - F_V(1-v'');$$

by Assumption 1,  $m_+, m_- > 0$  iff  $v' \neq v''$ . We proceed with Cases 1–3 in turn, firstly by establishing the form of the quantile functions  $q_A(t^\beta)$  and  $q_B(t^\gamma)$ , followed by proofs of the main Propositions concerning the behaviour of the joint survivor functions.

### A.2.1 Case 1: $\lambda > 0$

Recall the positive quantities  $\mu_1, \mu_2$  defined in Remark 1; these can be expressed  $\mu_1 = \lambda^{-1/\lambda} \int_0^1 \tau^{-1/\lambda}(v) dF_V(v)$ , and  $\mu_2 = \lambda^{-1/\lambda} \int_0^1 \tau^{-1/\lambda}(1-v) dF_V(v)$ .

**Proposition 6.** *Let  $\beta, \gamma > 0$ . Then there exist slowly varying functions  $l_A, l_B$  such that  $q_A(t^\beta) = t^{\lambda\beta} l_A(t)$  and  $q_B(t^\gamma) = t^{\lambda\gamma} l_B(t)$  for all  $t \geq 1$ . Furthermore  $l_A(t) \rightarrow \mu_1^\lambda$  and  $l_B(t) \rightarrow \mu_2^\lambda$  as  $t \rightarrow \infty$ .*

*Proof of Proposition 1.* Firstly suppose  $\beta = \gamma$ . From (A.2a),  $\theta_{\beta, \gamma}$  as defined in Proposition 1, is

$$\theta_{\beta, \gamma}(t) = \int_0^1 [t^{-\lambda\beta} + \lambda \max\{l_A(t)\tau(v), l_B(t)\tau(1-v)\}]_+^{-1/\lambda} dF_V(v).$$

Since  $\tau \geq 1$ , and  $l_A(t)$  (or  $l_B(t)$ ) has a non-zero limit as  $t \rightarrow \infty$ , we can bound  $\max\{l_A(t)\tau(v), l_B(t)\tau(1-v)\}$  uniformly away from 0 for all sufficiently large  $t$ . Furthermore Proposition 6 implies  $\max\{l_A(t)\tau(v), l_B(t)\tau(1-v)\} \rightarrow \max\{\mu_1^\lambda \tau(v), \mu_2^\lambda \tau(1-v)\}$  as  $t \rightarrow \infty$ . Applying dominated convergence and using the definitions of  $\mu_1$  and  $\mu_2$  then gives

$$\lim_{t \rightarrow \infty} \theta_{\beta, \gamma}(t) = \int_0^1 [\lambda \max\{\mu_1^\lambda \tau(v), \mu_2^\lambda \tau(1-v)\}]^{-1/\lambda} dF_V(v) = \chi_\lambda.$$

The fact that this limit is non-zero implies  $\theta_{\beta, \gamma}$  is slowly varying. Now assume  $\beta < \gamma$  (the case  $\beta > \gamma$  can be handled similarly). Then

$$r(t) := \frac{q_A(t^\beta)}{q_A(t^\beta) + q_B(t^\gamma)} = \left\{ 1 + t^{\lambda(\gamma-\beta)} \frac{l_B(t)}{l_A(t)} \right\}^{-1} \rightarrow 0, \quad t \rightarrow \infty.$$

If  $v \leq r(t)$  then (A.4) gives

$$\begin{aligned} q_A(t^\beta)\tau(v) \geq q_B(t^\gamma)\tau(1-v) &\implies 1 + \lambda q_A(t^\beta)\tau(v) \geq 1 + \lambda q_B(t^\gamma)\tau(1-v) > \lambda q_B(t^\gamma) > 0 \\ \implies 0 < \{1 + \lambda q_B(t^\gamma)\tau(1-v)\}_+^{-1/\lambda} - \{1 + \lambda q_A(t^\beta)\tau(v)\}_+^{-1/\lambda} &\leq \{\lambda q_B(t^\gamma)\}_+^{-1/\lambda}. \end{aligned}$$

Combined with (A.1) and (A.2b) we thus have

$$\begin{aligned}
0 &\leq \mathbf{P}\{B > q_B(t^\gamma)\} - \mathbf{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\} \\
&= \int_0^{r(t)} \left[ \{1 + \lambda q_B(t^\gamma)\tau(1-v)\}_+^{-1/\lambda} - \{1 + \lambda q_A(t^\beta)\tau(v)\}_+^{-1/\lambda} \right] dF_V(v) \\
&\leq \int_0^{r(t)} \{\lambda q_B(t^\gamma)\}^{-1/\lambda} dF_V(v) = t^{-\gamma} \{\lambda l_B(t)\}^{-1/\lambda} F_V\{r(t)\}.
\end{aligned}$$

The continuity of  $F_V$  at 0 gives  $F_V\{r(t)\} \rightarrow F_V(0) = 0$  as  $t \rightarrow \infty$ . Since  $\mathbf{P}\{B > q_B(t^\gamma)\} = t^{-\gamma}$  we then get  $\mathbf{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\} = t^{-\gamma}\{1 + o(1)\}$  as  $t \rightarrow \infty$ . The result follows.  $\square$

*Proof of Proposition 2.* Note that by Condition 3,  $v'' > 1/2$ . From (4.5) we get  $\chi_\lambda \leq \mathcal{R}_- + \mathcal{R}_+$  where

$$\mathcal{R}_- = \frac{\int_0^{1/2} \tau^{-1/\lambda}(v) dF_V(v)}{\int_0^1 \tau^{-1/\lambda}(v) dF_V(v)} \quad \text{and} \quad \mathcal{R}_+ = \frac{\int_{1/2}^1 \tau^{-1/\lambda}(1-v) dF_V(v)}{\int_0^1 \tau^{-1/\lambda}(1-v) dF_V(v)}.$$

Now  $\tau(v) \geq 1$  with equality iff  $v \in \Omega_0$ . Since  $\Omega_0 \subseteq [1/2, 1]$  dominated convergence then gives

$$\lim_{\lambda \rightarrow 0^+} \int_0^{1/2} \tau^{-1/\lambda}(v) dF_V(v) = 0 \quad \text{and} \quad \lim_{\lambda \rightarrow 0^+} \int_{1/2}^1 \tau^{-1/\lambda}(v) dF_V(v) = \int_{\Omega_0} dF_V(v) = m_+.$$

If  $v' = 1/2 < v''$  then  $m_+ > 0$  so  $\mathcal{R}_- \rightarrow 0$  as  $\lambda \rightarrow 0^+$ . Otherwise  $v' > 1/2$ , in which case we can find  $\delta > 0$  so that  $\tau(v) \geq 1 + \delta$  when  $v \in [0, 1/2]$ . Setting  $I_\delta = \{v \in [0, 1] : \tau(v) \leq 1 + \delta/2\}$  we then get

$$\mathcal{R}_- \leq \frac{\int_0^{1/2} (1 + \delta)^{-1/\lambda} dF_V(v)}{\int_{I_\delta} (1 + \delta/2)^{-1/\lambda} dF_V(v)} \leq \frac{(1 + \delta)^{-1/\lambda}}{(1 + \delta/2)^{-1/\lambda} C_\delta} = C_\delta^{-1} \rho^{1/\lambda}$$

where  $\rho = 1 - \delta/(2 + 2\delta) \in (0, 1)$  and  $C_\delta := \int_{I_\delta} dF_V(v) > 0$  (positivity follows from Assumption 1 and the fact that the interval length  $|I_\delta| > 0$ ). As  $\lambda \rightarrow 0^+$ ,  $\rho^{1/\lambda} \rightarrow 0$  and hence  $\mathcal{R}_- \rightarrow 0$ . A similar argument shows  $\mathcal{R}_+ \rightarrow 0$ .  $\square$

### A.2.2 Case 2: $\lambda = 0$

Let  $\beta, \gamma > 0$  and set  $\omega = \beta/(\beta + \gamma) \in (0, 1)$ . Then  $\beta\tau(\omega) = \|(\beta, \gamma)\|_m = \gamma\tau(1 - \omega)$  while (A.4) gives

$$\nu(v) := \max\{\beta\tau(v), \gamma\tau(1 - v)\} = \begin{cases} \beta\tau(v) & \text{if } 0 \leq v \leq \omega, \\ \gamma\tau(1 - v) & \text{if } \omega \leq v \leq 1. \end{cases} \quad (\text{A.5})$$

The function  $\nu$  is a positive, continuous and convex function on  $[0, 1]$ , with  $\nu(0) = +\infty = \nu(1)$ . Set  $\widehat{\nu} := \min\{\nu(v) : v \in [0, 1]\}$  and  $\Omega := \{v \in [0, 1] : \nu(v) = \widehat{\nu}\}$ ; in particular,  $\Omega$  is a non-empty closed subinterval of  $[0, 1]$ . The general shape of  $\nu$  and key properties of  $\widehat{\nu}$  and  $\Omega$  can be deduced from (A.3):

*C1:*  $\omega \in [1 - v', v']$ . Then  $\beta\tau(v)$  is strictly decreasing on  $[0, \omega]$ ,  $\gamma\tau(1 - v)$  is strictly increasing on  $[\omega, 1]$  and these quantities are equal when  $v = \omega$ . It follows that  $\Omega = \{\omega\}$  and  $\widehat{\nu} = \beta\tau(\omega) = \gamma\tau(1 - \omega) = \|(\beta, \gamma)\|_m$ .

*C2:*  $\omega \in (v', 1)$ . Then  $\nu(v) = \beta\tau(v)$  is strictly decreasing on  $[0, v']$  and  $\nu(v) = \beta\tau(v) = \beta$  (a constant) on  $[v', \min\{\omega, v''\}]$ . Also  $\beta\tau(v)$  is strictly increasing on  $[v'', 1]$  and  $\omega > v' \geq 1 - v'$  so  $\gamma\tau(1 - v)$  is strictly increasing and not less than  $\beta\tau(v)$  on  $[\omega, 1]$ ; hence  $\nu(v) = \max\{\beta\tau(v), \gamma\tau(1 - v)\}$  is strictly increasing on  $[\min\{\omega, v''\}, 1]$ . It follows that  $\Omega = [v', \min\{\omega, v''\}]$  and  $\widehat{\nu} = \beta = \|(\beta, \gamma)\|_\infty$  (note that,  $\omega > v' \geq 1/2$  which implies  $\beta > \gamma$ ).

*C3:*  $\omega \in (0, 1 - v')$ . By a similar argument to C2,  $\Omega = [\max\{\omega, 1 - v''\}, 1 - v']$  and  $\widehat{\nu} = \gamma = \|(\beta, \gamma)\|_\infty$ .

The main results in this case are built from the following lemma.

**Lemma 2.** Suppose  $a : [0, 1] \rightarrow [0, \infty]$  is continuous,  $u$  is regularly varying at infinity with index  $\rho > 0$ , and  $I, I_s \subseteq [0, 1]$  for  $s \geq 0$  is a collection of closed intervals with the interval length  $|I| > 0$  and  $I_s \rightarrow I$  as  $s \rightarrow \infty$ . Define  $\phi$  by

$$\phi(s) = \int_{I_s} u^{-a(v)}(s) dF_V(v)$$

for each  $s \geq 0$ , and set  $\alpha = \min\{a(v) : v \in I\}$ . Then  $\phi$  is regularly varying with index  $-\alpha\rho$ .

Note that by  $I_s \rightarrow I$  we mean that the Hausdorff distance between  $I_s$  and  $I$  tends to 0; equivalently, the end points of  $I_s$  converge to the end points of  $I$ .

**Proposition 7.** Let  $\beta, \gamma > 0$ . Then there exist slowly varying functions  $l_A, l_B$  such that  $q_A(t^\beta) = \log\{t^\beta l_A(t)\}$  and  $q_B(t^\gamma) = \log\{t^\gamma l_B(t)\}$  for all  $t \geq 1$ . Furthermore  $l_A, l_B$  are continuous, take values in  $[m_+, 1]$  and  $[m_-, 1]$  respectively, and satisfy  $l_A(t) \rightarrow m_+$  and  $l_B(t) \rightarrow m_-$  as  $t \rightarrow \infty$ .

*Proof of Proposition 3.* Setting

$$r(t) = \frac{q_A(t^\beta)}{q_A(t^\beta) + q_B(t^\gamma)} = \frac{\beta \log t + \log l_A(t)}{(\beta + \gamma) \log t + \log l_A(t)l_B(t)} \quad (\text{A.6})$$

we have  $r(t) \rightarrow \omega$  as  $t \rightarrow \infty$  (note that  $l_A$  and  $l_B$  are slowly varying). Furthermore (A.2b) gives

$$\begin{aligned} \mathbb{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\} &= \int_0^{r(t)} e^{-\tau(v) \log\{t^\beta l_A(t)\}} dF_V(v) + \int_{r(t)}^1 e^{-\tau(1-v) \log\{t^\gamma l_B(t)\}} dF_V(v) \\ &= \int_0^{r(t)} \{t^\beta l_A(t)\}^{-\tau(v)} dF_V(v) + \int_{r(t)}^1 \{t^\gamma l_B(t)\}^{-\tau(1-v)} dF_V(v). \end{aligned} \quad (\text{A.7})$$

Now assume  $\beta \leq \gamma$  (the case  $\beta \geq \gamma$  can be handled similarly). Then  $\omega \leq 1/2 \leq v'$  so (A.3) gives  $\min\{\tau(v) : v \in [0, \omega]\} = \tau(\omega) = \|(\beta, \gamma)\|_m / \beta$ . Furthermore,  $\Omega \subseteq [\omega, 1]$  (recall the description of  $\nu$  at the beginning of this section) so  $\min\{\tau(1-v) : v \in [\omega, 1]\} = \gamma^{-1} \min\{\nu(v) : v \in [\omega, 1]\} = \widehat{\nu} / \gamma$ . Lemma 2 can now be applied to show that the integrals on the right hand side of (A.7) are regularly varying functions, the first with index  $-\|(\beta, \gamma)\|_m \leq -\widehat{\nu}$  and the second with index  $-\widehat{\nu}$ . By the forms of  $\widehat{\nu}$  described in C1-C3 immediately preceding Lemma 2, the result follows.  $\square$

The fact that  $\chi = 0$  when  $\eta = 1$  in this case is given by the following.

**Proposition 8.** If  $v' \neq v''$  (equivalently  $m_+, m_- > 0$ ) and  $1 - v' \leq \omega \leq v'$  then  $\lim_{t \rightarrow \infty} \theta_{\beta, \gamma}(t) = 0$ .

### A.2.3 Case 3: $\lambda < 0$ , $\|(1, 1)\|_m = \|(1, 1)\|_\infty$ , with Assumption 2

**Proposition 9.** Let  $\beta, \gamma > 0$ . Then there exist slowly varying functions  $l_A, l_B$  such that  $q_A(t^\beta) = \Lambda - t^{\lambda\beta} l_A(t)$  and  $q_B(t^\gamma) = \Lambda - t^{\lambda\gamma} l_B(t)$  for all  $t \geq 1$ . Furthermore  $l_A(t) \rightarrow \Lambda m_+^\lambda$  and  $l_B(t) \rightarrow \Lambda m_-^\lambda$  as  $t \rightarrow \infty$ .

Let  $\Delta$  be a neighbourhood of  $1/2$  on which  $F'_V$  is continuous; in particular,  $dF_V(v) = F'_V(v) dv$  for  $v \in \Delta$ .

*Proof of Proposition 4.* Set  $r(t) = q_A(t^\beta) / \{q_A(t^\beta) + q_B(t^\gamma)\}$  so (A.2b) gives  $\mathbb{P}\{A > q_A(t^\beta), B > q_B(t^\gamma)\} = \mathcal{I}_- + \mathcal{I}_+$  where

$$\mathcal{I}_- = \int_0^{r(t)} \{1 + \lambda q_A(t^\beta) \tau(v)\}_+^{-1/\lambda} dF_V(v) \quad \text{and} \quad \mathcal{I}_+ = \int_{r(t)}^1 \{1 + \lambda q_B(t^\gamma) \tau(1-v)\}_+^{-1/\lambda} dF_V(v).$$

To consider  $\mathcal{I}_-$  firstly set  $v_-(t) = q_A(t^\beta) / \{q_A(t^\beta) + \Lambda\}$ . Since  $q_B(t^\gamma) < \Lambda$  and  $q_A(t^\beta), q_B(t^\gamma) \rightarrow \Lambda$  as  $t \rightarrow \infty$  we get  $v_-(t) < r(t)$  while  $v_-(t), r(t) \rightarrow 1/2$  as  $t \rightarrow \infty$ . As  $v'' > 1/2$  we can then choose  $T_0$  so that  $[v_-(t), r(t)] \subseteq [1 - v'', v''] \cap \Delta$  whenever  $t \geq T_0$ . For  $t \geq T_0$  it follows that  $\tau(v) = \max\{(1-v)/v, 1\}$  when  $v \in [v_-(t), r(t)]$ ; in particular  $\tau\{v_-(t)\} = \Lambda / q_A(t^\beta)$ . Furthermore (A.3) implies  $\tau(v)$  is decreasing on  $[0, r(t)]$ . For  $v \in [v_-(t), r(t)]$  we thus have

$$1 + \lambda q_A(t^\beta) \tau(v) > 0 \implies \tau(v) < \frac{1}{-\lambda q_A(t^\beta)} = \tau\{v_-(t)\} \implies v > v_-(t).$$

Therefore

$$\begin{aligned}\mathcal{I}_- &= \int_{v_-(t)}^{r(t)} \left\{ 1 + \lambda q_A(t^\beta) \max\left(\frac{1-v}{v}, 1\right) \right\}_+^{-1/\lambda} dF_V(v) \\ &= \int_{v_-(t)}^{r(t)} \min \left\{ 1 + \lambda q_A(t^\beta) \frac{1-v}{v}, 1 + \lambda q_A(t^\beta) \right\}_+^{-1/\lambda} dF_V(v).\end{aligned}$$

Consider the new variable  $u = \{1 + \lambda q_B(t^\gamma)\}^{-1} \{1 + \lambda q_A(t^\beta)(1-v)/v\}$ , and its inverse  $v = -\lambda q_A(t^\beta)[1 - \lambda q_A(t^\beta) - \{1 + \lambda q_B(t^\gamma)\}u]^{-1}$ . We have  $u = 0$  (respectively  $u = 1$ ) when  $v = v_-(t)$  (respectively  $v = r(t)$ ). Thus

$$\mathcal{I}_- = \{1 + \lambda q_B(t^\gamma)\}^{1-1/\lambda} \int_0^1 \min \left\{ u, \frac{1 + \lambda q_A(t^\beta)}{1 + \lambda q_B(t^\gamma)} \right\}^{-1/\lambda} G_t^-(u) du \quad (\text{A.8a})$$

$$= \{1 + \lambda q_B(t^\gamma)\} \{1 + \lambda q_A(t^\beta)\}^{-1/\lambda} \int_0^1 \min \left\{ \frac{1 + \lambda q_B(t^\gamma)}{1 + \lambda q_A(t^\beta)} u, 1 \right\}^{-1/\lambda} G_t^-(u) du, \quad (\text{A.8b})$$

where

$$G_t^-(u) = F' \left[ \frac{-\lambda q_A(t^\beta)}{1 - \lambda q_A(t^\beta) - \{1 + \lambda q_B(t^\gamma)\}u} \right] \frac{-\lambda q_A(t^\beta)}{[1 - \lambda q_A(t^\beta) - \{1 + \lambda q_B(t^\gamma)\}u]^2}.$$

As  $t \rightarrow \infty$  we have  $-\lambda q_A(t^\beta), -\lambda q_B(t^\gamma) \rightarrow 1$  so  $1 - \lambda q_A(t^\beta) - \{1 + \lambda q_B(t^\gamma)\}u \rightarrow 2$ , uniformly for  $u \in [0, 1]$ . Using Assumption 2 it follows that  $G_t^-(u) \rightarrow F'_V(1/2)/4 =: \Gamma$ , uniformly for  $u \in [0, 1]$ . Hence the integrands in (A.8) are uniformly bounded for all sufficiently large  $t$ . Proposition 9 gives

$$1 + \lambda q_A(t^\beta) = m_+^\lambda t^{\lambda\beta} \{1 + o(1)\} \quad \text{and} \quad 1 + \lambda q_B(t^\gamma) = m_-^\lambda t^{\lambda\gamma} \{1 + o(1)\} \quad \text{as } t \rightarrow \infty. \quad (\text{A.9})$$

If  $\beta < \gamma$ : As  $t \rightarrow \infty$  we have  $\{1 + \lambda q_A(t^\beta)\}/\{1 + \lambda q_B(t^\gamma)\} \rightarrow +\infty$  by (A.9), so applying dominated convergence to (A.8b) gives

$$\mathcal{I}_- = m_-^{\lambda-1} t^{\lambda\gamma-\gamma} \int_0^1 u^{-1/\lambda} \Gamma du \{1 + o(1)\} = O(t^{\lambda\gamma-\gamma}) = o(t^{\lambda\beta-\gamma}).$$

If  $\beta > \gamma$ : As  $t \rightarrow \infty$  we have  $\{1 + \lambda q_B(t^\gamma)\}/\{1 + \lambda q_A(t^\beta)\} \rightarrow +\infty$  by (A.9), so applying dominated convergence to (A.8b) gives

$$\mathcal{I}_- = m_-^\lambda t^{\lambda\gamma} m_+^{-1} t^{-\beta} \int_0^1 1^{-1/\lambda} \Gamma du \{1 + o(1)\} = \Gamma m_-^\lambda m_+^{-1} t^{\lambda\gamma-\beta} \{1 + o(1)\}.$$

If  $\beta = \gamma$ : As  $t \rightarrow \infty$  we have  $\{1 + \lambda q_A(t^\beta)\}/\{1 + \lambda q_B(t^\gamma)\} \rightarrow m_+^\lambda/m_-^\lambda$  by (A.9), so applying dominated convergence to (A.8a) gives

$$\mathcal{I}_- = m_-^{\lambda-1} t^{\lambda\beta-\beta} \int_0^1 \min \left( u, \frac{m_+^\lambda}{m_-^\lambda} \right)^{-1/\lambda} \Gamma du \{1 + o(1)\}.$$

When  $m_+ \leq m_-$  this becomes  $\mathcal{I}_- = -\Gamma\lambda(1-\lambda)^{-1} m_-^{\lambda-1} t^{\lambda\beta-\beta} \{1 + o(1)\}$ . When  $m_+ \geq m_-$  we get

$$\mathcal{I}_- = \Gamma m_-^{\lambda-1} \left( \int_0^{\frac{m_+^\lambda}{m_-^\lambda}} u^{-1/\lambda} du + \int_{\frac{m_+^\lambda}{m_-^\lambda}}^1 \frac{m_-}{m_+} du \right) t^{\lambda\beta-\beta} \{1 + o(1)\} = \Gamma \left( m_-^\lambda - \frac{m_+^\lambda}{1-\lambda} \right) m_+^{-1} t^{\lambda\beta-\beta} \{1 + o(1)\}.$$

A similar calculation for  $\mathcal{I}_+$  leads to

$$\mathcal{I}_+ = \begin{cases} \Gamma m_+^\lambda m_-^{-1} t^{\lambda\beta-\gamma} \{1 + o(1)\} & \text{if } \beta < \gamma, \\ o(t^{\lambda\gamma-\beta}) & \text{if } \beta > \gamma, \\ \Gamma \left( m_+^\lambda - \frac{1}{1-\lambda} m_-^\lambda \right) m_-^{-1} t^{\lambda\beta-\beta} \{1 + o(1)\} & \text{if } \beta = \gamma \text{ and } m_+ \leq m_-, \\ \Gamma \frac{-\lambda}{1-\lambda} m_+^{\lambda-1} t^{\lambda\beta-\beta} \{1 + o(1)\} & \text{if } \beta = \gamma \text{ and } m_+ \geq m_-, \end{cases}$$

as  $t \rightarrow \infty$ . This is combined with  $\mathcal{I}_-$  to give (4.6). As the limit is non-zero in all cases,  $\theta_{\beta,\gamma}$  is slowly varying.  $\square$

## References

- Abdous, B., Fougères, A.-L., and Ghoudi, K. (2005). Extreme behaviour for bivariate elliptical distributions. *The Canadian Journal of Statistics*, 33(3):317–334.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes*. Wiley.
- Bortot, P., Coles, S. G., and Tawn, J. A. (2000). The multivariate Gaussian tail model: an application to oceanographic data. *Journal of the Royal Statistical Society, Series C*, 49(1):31–49.
- Coles, S. G., Heffernan, J. A., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.
- Coles, S. G. and Pauli, F. (2002). Models and inference for uncertainty in extremal dependence. *Biometrika*, 89(1):183–196.
- Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B*, 53(2):377–392.
- Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes – an application to structural design (with discussion). *Journal of the Royal Statistical Society, Series C*, 43(1):1–48.
- Das, B. and Resnick, S. (2014). Generation and detection of multivariate regular variation and hidden regular variation. <http://arxiv.org/abs/1403.5774>.
- de Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45.
- Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and its interface*, to appear.
- Einmahl, J., de Haan, L., and Sinha, A. (1997). Estimation of the spectral measure of an extreme-value distribution. *Stoch. Proc. Appl.*, 70:143–171.
- Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes*, 3(3):279–290.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571.
- 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*, 66(3):497–546.
- Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34(3):587–608.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York.
- Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B*, 59(2):475–499.
- Liu, Y. and Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, 127:19–35.
- McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas,  $d$ -monotone functions and  $\ell_1$ -norm symmetric distributions. *Annals of Statistics*, 37(5B):3059–3097.
- Mikosch, T. (2005). How to model multivariate extremes if one must? *Statistica Neerlandica*, 59(3):324–338.
- Peng, L. and Qi, Y. (2004). Discussion of *A conditional approach for multivariate extreme values*, by J. E. Heffernan and J. A. Tawn. *Journal of the Royal Statistical Society, Series B*, 66(3):541–542.

- Pickands, J. (1986). The continuous and differentiable domains of attraction of the extreme value distributions. *Ann. Probab.*, 14(3):996–1004.
- Ramos, A. and Ledford, A. W. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society, Series B*, 71(1):219–241.
- Resnick, S. I. (1987). *Extremes Values, Regular Variation and Point Processes*. Springer Verlag, New York.
- Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.
- Resnick, S. I. (2006). *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- Sabourin, A. and Naveau, P. (2014). Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization. *Computational Statistics and Data Analysis*, 71:542–567.
- Scheffler, H.-P. and Stoev, S. (2015). Implicit extremes and implicit max-stable laws. <http://arxiv.org/abs/1411.4688>.
- Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399.
- Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.