Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties

Emma S. Simpson[1][*] and Jonathan A. Tawn[2]

[1]Department of Statistical Science, University College London, WC1E 6BT, U.K.

[2]School of Mathematical Sciences, Lancaster University, LA1 4YF, U.K.

September 20, 2024

Abstract

The key to successful statistical analysis of bivariate extreme events lies in flexible modelling of the tail dependence relationship between the two variables. In the extreme value theory literature, various techniques are available to model separate aspects of tail dependence, based on different asymptotic limits. Results from Balkema and Nolde [1] and Nolde [34] highlight the importance of studying the limiting shape of an appropriately-scaled sample cloud when characterising the whole joint tail. We now develop the first statistical inference for this limit set, which has considerable practical importance for a unified inference framework across different aspects of the joint tail. Moreover, Nolde and Wadsworth [35] link this limit set to various existing extremal dependence frameworks. Hence, a by-product of our new limit set inference is the first set of self-consistent estimators for several extremal dependence measures, avoiding the current possibility of contradictory conclusions. In simulations, our limit set estimator is successful across a range of distributions, and the corresponding extremal dependence estimators provide a major joint improvement and small marginal improvements over existing techniques. We consider an application to sea wave heights, where our estimates successfully capture the expected weakening extremal dependence as the distance between locations increases.

Keywords: bivariate extremes; coefficient of asymptotic independence; conditional extremes; extremal dependence structure; gauge function; sample cloud

[*]Corresponding author. Email address: emma.simpson@ucl.ac.uk

1 Introduction

Multivariate extreme value problems are important across a range of subject domains, such as sea level [11], air pollution [23], rainfall [13] and river flow [19]. The typical formulation is to have n independent and identically distributed observations, $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, from a d-dimensional random vector $\boldsymbol{X} \in \mathbb{R}^d$ with unknown joint distribution $F_{\boldsymbol{X}}$. Generally, the aim is to estimate $F_{\boldsymbol{X}}(\boldsymbol{x})$ and $\Pr(\boldsymbol{X} \in A)$ for some $A \subset \mathbb{R}^d$, where $\boldsymbol{x}$ and all the elements in A are large in at least one component of $\boldsymbol{X}$. Our focus here is slightly different; we instead aim to estimate the boundary of a set that in the limit as $n \to \infty$ contains all sample values from $\boldsymbol{X}$ after some appropriate scaling. The purpose of this is two-fold: (i) to identify potentially 'risky' combinations of variables in finite samples and (ii) to provide improved inference for existing extremal dependence measures.

The usual approach to inference in multivariate extremes is to estimate the marginal distributions and dependence structure (copula) with a focus on their behaviour in the tail region. Univariate extreme value methods are well established [8, 14], with dependence modelling being the key challenge. A standard framework is to transform the margins of $\boldsymbol{X}$ to a common distribution based on univariate extreme value models fitted to the individual components. In extreme value analysis, common marginal selections include Fréchet, Gumbel, Exponential and Laplace distributions, with choices made based on theoretical results and/or to highlight particular tail dependence features. Here, our focus is on Exponential marginals, although we discuss possible extensions to other choices in Section 7, and we take $d = 2$ for ease of exposition. That is, we consider variables (X_1, X_2) with $\Pr(X_j < x) = 1 - \exp(-x)$, for $x \geq 0$ and $j = 1, 2$.

Even with common margins, multivariate extremes are often difficult to model due to the lack of a natural ordering [3]. As such, there have been several different asymptotically-motivated modelling approaches and assumptions proposed in the literature, e.g., multivariate regular variation [39], hidden regular variation [27, 40], conditional extremes [23, 22], powered joint tails [50, 15], and mixture structures [20, 44]. The asymptotic arguments differ in each of these cases, covering situations where the growth rates of the margins are the same or different across all variables, and where they can vary with the direction of the extreme region of interest.

Results from [1] on the limiting shape, and associated boundary G, of an appropriately-scaled sample cloud offer a new, more unified, asymptotic approach. Here, the focus is on the scaled sample cloud, corresponding to n independent samples from the joint distribution of (X_1, X_2), denoted by

$$C_n = \{(X_{1,i}/\log n, X_{2,i}/\log n); i = 1, \ldots, n\}. \tag{1}$$

We note that the log-scaling is appropriate here due to $\boldsymbol{X}$ having Exponential margins; we comment on this further in Section 2.1. As $n \to \infty$, the sample C_n converges onto a compact limit set with upper boundary

$$G = \{(x_1, x_2) \in [0,1]^2 : g(x_1, x_2) = 1\}, \tag{2}$$

for a function g termed the gauge function, which will also be defined in Section 2.1. Given that $\boldsymbol{X}$ has Exponential margins, the shape of G is fully determined by the extremal dependence of the variables.

Here, G provides a characterisation of the full joint tail, irrespective of the nature of extremal dependence, whereas the previously mentioned methods offer a separate treatment for different forms of extremal dependence. Balkema and Nolde [2] discuss, at a purely probabilistic level, how this limit set can be used for risk evaluation and tail inference for linear combinations of $\boldsymbol{X}$.

Although G is a limit set boundary, we propose that it can also be used to provide valuable information about the likely boundary set for finite samples. Specifically, the boundary set

$$G_m := \{(x_1, x_2) \in \mathbb{R}_+^2 : g(x_1/\log m, x_2/\log m) = 1\} \tag{3}$$

for $m \in \mathbb{N}$, is not exceeded by any pair in the sample $\{(X_{1,i}, X_{2,i}) : i = 1, \ldots, m\}$ with probability tending to 1 as $m \to \infty$, and with no other curve satisfying this property if it has on it any (x_1, x_2) that is below G_m. Hence, for large m, this boundary set can be used to identify the most risky combinations of variables for a sample size of m, and when $m > n$, extrapolation for the purposes of risk assessment is possible.

The statistical exploitation of this limit theory through inference for G, which is the aim of this paper, has not been explored but has two key benefits. First, there is considerable practical advantage in the estimator $\hat{G}_m$ of G_m, obtained via estimation of (2), using the sample of size n for extrapolation purposes. Secondly, as [35] have shown, different parts of the boundary set G provide valuable geometric interpretations for a number of existing extremal dependence measures, discussed in Section 3.1, which play key roles in inference for the multivariate extremes methods listed above. An additional benefit of estimating G is that it provides new estimators of these dependence measures through exploitation of these theoretical results. It is therefore of interest to compare how each of the resulting estimators, based on a unifying estimate $\hat{G}$ of G, performs individually, relative to existing techniques. Critically, though, we see most value in using $\hat{G}$ to ensure self-consistency over the estimated dependence measures. The previous estimators do not guarantee such self-consistent information, since they are estimated separately using different asymptotic theories.

To illustrate the issue of having a lack of self-consistency in the estimation of extremal dependence features, we consider the most common starting point for a bivariate extreme value analysis: identifying whether (X_1, X_2) are *asymptotically dependent* or *asymptotically independent* [9]. For our Exponentially distributed variables, this can be formalised through the coefficient of asymptotic dependence χ, with

$$\chi = \lim_{x \to \infty} \Pr(X_2 > x \mid X_1 > x) = \lim_{x \to \infty} \frac{\Pr(X_1 > x, X_2 > x)}{\exp(-x)}. \tag{4}$$

A value of $\chi > 0$ ($\chi = 0$) corresponds to asymptotic dependence (asymptotic independence) between X_1 and X_2, respectively. From limit (4), it can be seen that asymptotic dependence occurs when the most extreme values can happen simultaneously in the two variables, and asymptotic independence is when such joint extremes are not possible. Since many asymptotic results for bivariate extremes, and their associated statistical models, are only suitable in one of these situations, distinguishing between them can play a crucial role in selecting a modelling approach. With the existing estimators it is possible that, for example, the coefficient of asymptotic independence of [26] indicates asymptotic dependence ($\chi > 0$)

while the normalising functions in the conditional extremes approach of [23] point towards asymptotic independence ($\chi = 0$). With our proposed estimators, unhelpful contradictions such as this are avoided; we return to this point in Section 5.5.

Finally, of course, there is much more to understanding bivariate extremes than simply distinguishing between $\chi > 0$ and $\chi = 0$, with inference for G helping here as well. For example, irrespective of whether the variables are asymptotically dependent or asymptotically independent, it is possible that one variable could take its largest values while the other is smaller order; a scenario of importance when studying the nature of multivariate extreme events [20, 44]. In such cases, models that are able to capture a mixture structure in the extremal dependence features are required, with [47] giving a first approach. Inference for G provides key diagnostics for identifying this possibility in a way that is self-consistent with respect to χ and other extremal dependence features.

We continue the paper by providing further detail and examples on the asymptotic scaled sample cloud in Section 2, while definitions of various extremal dependence measures and their links to G are covered in Section 3. We introduce our procedure for estimating G in Section 4 and explain how this can be exploited to obtain self-consistent estimators of the considered extremal dependence properties in Section 5. The performance of our estimation procedure is demonstrated through a simulation study and application to sea wave heights in Section 6. We conclude with a discussion in Section 7.

2 The limiting shape of a scaled sample cloud

2.1 General theory

A sample cloud corresponds to n independent replicates from the joint distribution of (X_1, X_2); we denote this by $C_n^* = \{(X_{1,i}, X_{2,i}); i = 1, \ldots, n\}$. The limiting convex hull of C_n^*, as $n \to \infty$, has been previously studied; see for example [17, 6, 12]. Following [35], our focus is on a scaled version of C_n^*, which we denote C_n, as defined in expression (1). We assume that we have standard Exponential marginal distributions, and that the joint density $f(x_1, x_2)$ exists and is non-zero everywhere in $\mathbb{R}_+^2$. The $\log n$ scaling is chosen due to the marginal distributions of X_1 and X_2, as it ensures that $\max_{i=1,\ldots,n} (X_{j,i}/\log n) \xrightarrow{p} 1$, as $n \to \infty$, for $j = 1, 2$; [34] shows that other light-tailed margins could also be used here and would result in different scaling functions being required, as well as affecting the shape of G.

Interest lies in studying the asymptotic shape of C_n, as $n \to \infty$. A useful tool in this task is the gauge function $g(x_1, x_2)$, defined via the relationship

$$-\log f(tx_1, tx_2) \sim t g(x_1, x_2), \qquad t \to \infty, \qquad x_1, x_2 \geq 0; \tag{5}$$

see [34, 35]. As $n \to \infty$, C_n converges onto the compact limit set

$$G_S = \{(x_1, x_2) \subseteq [0, 1]^2 : g(x_1, x_2) \leq 1\}.$$

The set G_S is star-shaped, meaning that for any $(x_1, x_2) \in G_S$, $(vx_1, vx_2) \in G_S$ for all $v \in (0, 1)$. In the following, we assume that the limit set G_S exists.
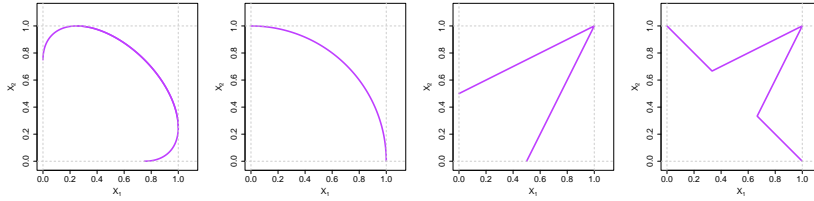
4

Figure 1: The boundary set G for Gaussian, inverted logistic, logistic and asymmetric logistic (left to right) dependence models with Exponential margins. The Gaussian correlation and the (inverted/asymmetric) logistic dependence parameters are set to $\rho = \gamma = 0.5$.

The boundary of G_S, defined by the set G in (2), is of particular interest. Some important properties of G are as follows: (i) due to the log-scaling, G must touch the lines $x_1 = 1$ and $x_2 = 1$ at least once; (ii) if f is non-zero everywhere, G must be a continuous function; and (iii) by limit (5), the gauge function must be homogeneous of order 1, so that, for example, taking $r = x_1 + x_2$ and $w = x_1/r$, with $(x_1, x_2) \in G$, we have $g(rw, r(1-w)) = rg(w, 1-w)$. These properties will be crucial in the development of our inferential approach, presented in Section 4.

2.2 Theoretical examples

The shapes of the limit set G_S and its boundary G are determined by the dependence between the variables. We illustrate these sets through four well-known copula families (see Figure 1) which show that the general properties of G_S and G stated in Section 2.1 hold, as well as providing an indication of the variety of different shapes that G can take. We use these distributions as examples throughout the paper. Given their known extremal dependence properties, these four examples provide insight into how G captures these features, with further details on this link presented in Section 3.2.

The first example is a bivariate Gaussian copula with correlation parameter $\rho \in [0, 1]$. Nolde [34] shows that the gauge function in this case has the form

$$g(x_1, x_2) = (1 - \rho^2)^{-1} \left(x_1 + x_2 - 2\rho x_1^{1/2} x_2^{1/2} \right), \qquad x_1, x_2 \geq 0.$$

In the case of independence between the variables, i.e., when $\rho = 0$, the gauge function is simply $g(x_1, x_2) = x_1 + x_2$. For details on the calculation of the remaining three gauge functions discussed below, see [45]. First, consider an inverted bivariate extreme value copula with logistic model and dependence parameter $\gamma \in (0, 1]$ [see 27], with $\gamma = 1$ corresponding to complete independence, which has a gauge function of the form

$$g(x_1, x_2) = \left(x_1^{1/\gamma} + x_2^{1/\gamma} \right)^\gamma, \qquad x_1, x_2 \geq 0,$$

and a bivariate extreme value copula with logistic model [21], again with dependence parameter $\gamma \in (0, 1]$, having the gauge function

$$g(x_1, x_2) = \begin{cases} \frac{1}{\gamma} \max(x_1, x_2) - \left(\frac{1}{\gamma} - 1 \right) \min(x_1, x_2), & x_1, x_2 \geq 0, \gamma < 1, \\ x_1 + x_2, & x_1, x_2 \geq 0, \gamma = 1. \end{cases}$$

5

There is a discontinuity here between $\gamma = 1$ and $\gamma \to 1$, which is linked to findings about the logistic model in [26]; since $\gamma = 1$ corresponds to independence, which is already covered by the Gaussian copula with $\rho = 0$, we restrict to the $\gamma < 1$ case in the remainder of the paper.

The first two examples in Figure 1 are widely known to correspond to asymptotically independent models, while the third exhibits asymptotic dependence if $\gamma < 1$ [9]. An important link between the bivariate gauge function and extremal dependence comes from the value of $g(1,1)$ [34]. If $g(1,1) > 1$, this corresponds to the case of asymptotic independence with $\chi = 0$, while under asymptotic dependence, with $\chi > 0$, we have $g(1,1) = 1$. This is clearly satisfied by the three examples above. However, it is also possible to have $g(1,1) = 1$ and $\chi = 0$, so care with interpretation is needed.

Finally, we consider a bivariate extreme value copula with asymmetric logistic model [46], where a mixture of extremal dependence features is possible, arising from logistic and independence components. That is, both variables can be simultaneously extreme, putting this model in the asymptotic dependence class, but each of the variables can also be individually large while the other is of smaller order. We provide details on the form of this dependence model in the supplementary material [43, Section I]. Following the discussion of discontinuity in the logistic model above, we again restrict the dependence parameter to $\gamma \in (0,1)$. Moreover, setting the additional model parameters to $\theta_1 = \theta_2 = 0$ yields a logistic model, while setting $\theta_1 = \theta_2 = 1$ recovers the setting of independence; we therefore focus on $\theta_1, \theta_2 \in (0,1)$ in the following. Figure 1 (right panel) also shows the gauge function of this copula, which is of the form

$$g(x_1, x_2) = \min \left\{ (x_1 + x_2); \frac{1}{\gamma} \max(x_1, x_2) - \left(\frac{1}{\gamma} - 1 \right) \min(x_1, x_2) \right\}, \quad x_1, x_2 \geq 0. \tag{6}$$

Here, we once again have $\chi > 0$ and it is clear that $g(1,1) = 1$. Interestingly, the gauge function does not depend on the parameters θ_1 and θ_2 when $\theta_1, \theta_2 \in (0,1)$.

3 Links between the sample cloud boundary and extremal dependence measures

3.1 Bivariate extremal dependence features

We now introduce the tail dependence features that we aim to estimate in a self-consistent way by ultimately exploiting properties of G. In Section 3.2, we explain how each of these dependence features is linked to the set G, using results from [35]. We begin with the coefficient of asymptotic independence, η, of [26], or equivalently $\bar{\chi} = 2\eta - 1$ [9]. Recall that we assume the variables X_1 and X_2 have standard Exponential marginal distributions. Then, $\eta \in (0,1]$ is defined by considering the behaviour of the joint survivor function

$$\Pr(X_1 > x, X_2 > x) \sim \mathcal{L}(e^x) e^{-x/\eta},$$

as $x \to \infty$, where the function $\mathcal{L}$ is slowly varying at infinity. Considering the definition of χ in (4), when $\eta = 1$, this yields $\chi = \lim_{x \to \infty} \mathcal{L}(e^x)$, whereas if $\eta < 1$, we obtain $\chi = \lim_{x \to \infty} \mathcal{L}(e^x) e^{-(1/\eta - 1)x} = 0$. Hence, for $\eta = 1$ and $\mathcal{L}(x) \not\to 0$ as $x \to \infty$, we have $\chi > 0$, and therefore asymptotic dependence. If $\eta < 1$,

or $\eta = 1$ and $\mathcal{L}(x) \to 0$ as $x \to \infty$, we have $\chi = 0$ and the variables are asymptotically independent. Estimation of the coefficient η can contribute towards the classification of tail dependence behaviour in practice, and in turn enable the selection of an appropriate model for the joint extremes.

Wadsworth and Tawn [50] extend the approach of Ledford and Tawn [26] by allowing for different scalings. They consider limiting probabilities of the form

$$\Pr\{X_1 > \omega x, X_2 > (1 - \omega)x\} \sim \mathcal{L}_\omega(e^x)e^{-x\lambda(\omega)},$$

as $x \to \infty$, for $\omega \in [0, 1]$, $\lambda(\omega) \in (0, 1]$ and some $\mathcal{L}_\omega$ that is slowly varying at infinity. The case where $\omega = 1/2$ is linked to the coefficient η by the relation $\eta^{-1} = 2\lambda(1/2)$. Therefore, under asymptotic dependence $\lambda(1/2) = 1/2$, and more generally, for $\omega \in [0, 1]$, we have $\lambda(\omega) = \max(\omega, 1 - \omega)$ in this case. Under complete independence, $\lambda(\omega) = 1$ for all $\omega \in [0, 1]$.

Motivated by the possibility of mixture structures in the extremal dependence features, Simpson et al. [44] introduced a further set of indices related to η. They consider a separate measure for each subset of variables, which describes whether they can be simultaneously large while the other variables are of smaller order. In the bivariate case, there are two measures of interest, denoted by $\tau_1(\delta), \tau_2(\delta) \in (0, 1]$, for $\delta \in [0, 1]$. The measure $\tau_1(\delta)$ is based on a hidden regular variation assumption, see [44] and [35], and in Exponential margins is defined by the relation

$$\Pr(X_1 > x, X_2 \leq \delta x) \sim \mathcal{L}_\delta(e^x)e^{-x/\tau_1(\delta)}, \tag{7}$$

as $x \to \infty$, for $\delta \in [0, 1]$ and some slowly varying function $\mathcal{L}_\delta$. The function $\tau_1(\delta)$ is monotonically increasing with δ, and has $\tau_1(1) = 1$. If there exists any $\delta^* < 1$ such that $\tau_1(\delta^*) = 1$, the variable X_1 can take its largest values while X_2 is of smaller order, otherwise X_1 can only take its largest values when X_2 is also large. The measure $\tau_2(\delta)$ can be defined analogously through the limiting behaviour of $\Pr(X_1 \leq \delta x, X_2 > x)$, as $x \to \infty$.

Finally, we consider the conditional extremes modelling approach of Heffernan and Tawn [23]. Conditional extremes models capture both asymptotic dependence and asymptotic independence, so are widely applicable. The conditional extremes framework requires that the marginal distributions have Exponential upper tails, which is clearly satisfied when (X_1, X_2) have standard Exponential margins. Selecting X_1 as the conditioning variable, we assume there exist functions $a_1(\cdot)$ and $b_1(\cdot) > 0$ such that

$$\left\{ \frac{X_2 - a_1(X_1)}{b_1(X_1)}, X_1 - u \right\} \bigg| X_1 > u \to (Z, E), \qquad \text{as } u \to \infty, \tag{8}$$

where $E \sim Exp(1)$ is independent of Z, and Z represents some non-degenerate residual distribution that places no mass on $\{+\infty\}$ [23, 25]. Under this conditioning, a suitably normalised version of X_2 and exceedances of X_1 above the threshold u become independent as $u \to \infty$. Heffernan and Tawn [23] propose setting $a_1(x) = \alpha_1 x$ and $b_1(x) = x^{\beta_1}$, for $\alpha_1 \in [0, 1]$ and $\beta_1 \in [0, 1)$, and demonstrate that this is a reasonable choice for a range of non-negatively dependent distributions. The case where $\alpha_1 = 1$ and $\beta_1 = 0$ corresponds to asymptotic dependence, while $\alpha_1 < 1$ corresponds to asymptotic independence, with complete independence achieved when $\alpha_1 = 0$, $\beta_1 = 0$ and $Z \sim Exp(1)$. An analogous result holds

Dependence model	η	$\lambda(\omega)$	$\tau_1(\delta) = \tau_2(\delta)$	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$
Gaussian	$(1+\rho)/2$	$\begin{cases} \frac{1-2\rho\{\omega(1-\omega)\}^{1/2}}{1-\rho^2}, & \text{if } t_\omega \geq \rho^2 \\ \max(\omega, 1-\omega), & \text{if } t_\omega < \rho^2 \end{cases}$	$\begin{cases} 1, & \text{if } \delta \geq \rho^2 \\ \frac{1-\rho^2}{1+\delta-2\rho\delta^{1/2}}, & \text{if } \delta < \rho^2 \end{cases}$	ρ^2	$1/2$
Inverted logistic	$2^{-\gamma}$	$\{\omega^{1/\gamma} + (1-\omega)^{1/\gamma}\}^\gamma$	1	0	$1-\gamma$
Logistic	1	$\max(\omega, 1-\omega)$	$\gamma/(1+\gamma\delta-\delta)$	1	0
Asymmetric logistic	1	$\max(\omega, 1-\omega)$	1	1	0

Table 1: Values of $\{\eta, \lambda(\omega), \tau_1(\delta), \tau_2(\delta), \alpha_1, \alpha_2, \beta_1, \beta_2\}$ for a Gaussian model with correlation parameter $\rho \in [0,1]$ and (inverted) logistic models with dependence parameter $\gamma \in (0,1)$. Here, $t_\omega = \min(\omega, 1-\omega)/\max(\omega, 1-\omega)$.

if X_2 is selected as the conditioning variable, with normalising functions $a_2(x) = \alpha_2 x$ and $\beta_2(x) = x^{\beta_2}$ for $\alpha_2 \in [0,1]$ and $\beta_2 \in [0,1)$.

The values of η, $\lambda(\omega)$, $\tau_1(\delta)$, $\tau_2(\delta)$, α_1, α_2, β_1 and β_2 for the copulas considered in Figure 1 are given in Table 1. For the Gaussian distribution, the $\tau_i(\delta)$ ($i = 1, 2$) result for $\delta < \rho^2$ was derived by [35]. For each of the first three copula models, the exchangeability of the variables means that $\tau_1(\delta) = \tau_2(\delta)$, $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. Table 1 also shows this property holds in the asymmetric logistic case when the variables are not exchangeable, i.e., when $\theta_1 \neq \theta_2$. More generally, in Section 3.2, we will see that these equivalences also arise when G is symmetric about the line $x_1 = x_2$, even if the distribution itself is not exchangeable.

3.2   Boundary set interpretation of bivariate extremal dependence features

This section provides a summary of key information from the important theoretical work of Nolde and Wadsworth [35], which established links between G and the extremal dependence features from Section 3.1. Balkema and Nolde [1] were the first to provide results linking the asymptotic shape of a scaled sample cloud to asymptotic independence. Nolde [34] extended this work to show that the coefficient of asymptotic independence, η, is linked to the set G defined in (2) via the relationship

$$\eta = \min\left\{s \in (0,1] : [s, \infty]^2 \cap G = \emptyset\right\}. \tag{9}$$

We can think of this as moving the set $[1, \infty]^2$ along the line $x_2 = x_1$ towards the origin, until it intersects the set G (although this intersection does not have to occur on the diagonal). A pictorial demonstration of this result, and those that follow in this section, is provided in the supplementary material [43, Section A]. A higher dimensional version of (9) was previously used in [45] to study the extremal dependence structure of vine copulas.

To calculate the value of $\lambda(\omega)$, [35] consider sets of the form

$$S_\omega = \{(x_1, x_2) : x_1 \geq \omega/\max(\omega, 1-\omega), x_2 \geq (1-\omega)/\max(\omega, 1-\omega)\},$$

for any fixed $\omega \in [0,1]$. They prove that

$$\lambda(\omega) = \frac{\max(\omega, 1-\omega)}{s_\omega}, \qquad \text{with } s_\omega = \min\{s \in [0,1] : sS_\omega \cap G = \emptyset\}. \tag{10}$$

The set $S_{1/2}$ is equivalent to $[1, \infty]^2$, so that the value of $s_{1/2}$ is η, as we expect since $\lambda(1/2) = 1/(2\eta)$, as discussed in Section 3.1. Nolde and Wadsworth [35] link the set G to $\tau_1(\delta)$ and $\tau_2(\delta)$ by considering sets

$$S_{1,\delta} = \{(x_1, x_2) : x_1 \in (1, \infty], x_2 \in [0, \delta]\},$$

$$S_{2,\delta} = \{(x_1, x_2) : x_1 \in [0, \delta], x_2 \in (1, \infty]\},$$

for $\delta \in [0, 1]$, and showing that, for $i = 1, 2$,

$$\tau_i(\delta) = \min\{s \in (0, 1] : sS_{i,\delta} \cap G = \emptyset\}. \tag{11}$$

Finally, in the conditional extremes framework, [35] show that

$$\alpha_1 = \max\{\tilde{\alpha}_1 \in [0, 1] : g(1, \tilde{\alpha}_1) = 1\}. \tag{12}$$

It is possible to have more than one such $\tilde{\alpha}_1$, but the largest value is needed to avoid Z in (8) placing mass on $\{+\infty\}$. This means we should find the largest value of x_2 where G intersects the line $x_1 = 1$. Considering an analogous definition of α_2, alongside result (9), it is clear that $\eta \geq \max(\alpha_1, \alpha_2)$. If there are ν separate values of $\tilde{\alpha}_1$ in (12), this indicates a ν-component mixture structure in the extremal dependence conditioning on X_1 large, following the representation of [47]. Moreover, the smallest value of δ such that $\tau_1(\delta) = 1$ is given by $\delta^* = \min\{\tilde{\alpha}_1 \in [0, 1] : g(1, \tilde{\alpha}_1) = 1\}$. If $\nu = 1$ with only one value of $\tilde{\alpha}_1$ satisfying $g(1, \tilde{\alpha}_1) = 1$, we have $\delta^* = \alpha_1$; this is the case for the first three examples in Figure 1. The asymmetric logistic example in Figure 1 has $\nu = 2$ and $\tilde{\alpha}_1 \in \{0, 1\}$, yielding $\alpha_1 = 1$ and $\delta^* = 0$. Overall, the set of measures $\{\eta, \max_{\delta<1} \tau_1(\delta), \max_{\delta<1} \tau_2(\delta)\}$ can describe whether the variables (X_1, X_2) can be simultaneously large, whether one variable is large while the other is of smaller order, or whether we have a combination of these cases, but considering the $\tilde{\alpha}_1$ values from (12) can provide additional insight.

Nolde and Wadsworth [35] also present results on the parameter β_1 in the conditional extremes model. They show that

$$g(1, \alpha_1 + u) = 1 + O\left\{u^{1/(1-\beta)}\right\}, \qquad \text{as } u \to 0. \tag{13}$$

The results for α_2 and β_2 are analogous to those presented here.

As well as considering how the extremal dependence features described above can be obtained from the set G, we examine how much information these features can give us about G. The indices $\lambda(\omega), \tau_1(\delta)$ and $\tau_2(\delta)$ ($\omega, \delta \in [0, 1]$) are the most interesting from this perspective, as collectively they can sometimes fully describe G. In the logistic case, only $\tau_1(\delta)$ and $\tau_2(\delta)$ are needed; the opposite is true for the inverted logistic case; whereas for the Gaussian case, they are all required. However, for the asymmetric logistic model, the extremal dependence mixture structure means that collectively, $\lambda(\omega)$ and $\tau_i(\delta)$ ($i = 1, 2$) do not provide sufficient information to deduce the shape of G. Hence, there is value in developing an approach to estimate G directly, to enhance our understanding of the nature of the extremal dependence and to deduce self-consistent estimates of currently studied extremal dependence features. The remainder of the paper is dedicated to this goal.

4 Estimation of the limit set G

4.1 Framework for modelling the boundary set G

For inference, we require the variables (X_1, X_2) to have Exponential margins, which can be achieved in practice by applying rank and probability integral transforms to each variable. Suppose we have n obser-

vations of the random variables Y_i, $i = 1, 2$, denoted by $y_{i,1}, \ldots, y_{i,n}$. The approximate transformation to observations of X_i with Exponential margins is given by

$$x_{i,j} = -\log\left\{1 - \frac{\text{rank}(y_{i,j})}{n+1}\right\}, \tag{14}$$

for $i = 1, 2$ and $j = 1, \ldots, n$. An alternative version of this transformation with a parametric form for the tail is detailed in [10]; this ultimately allows for marginal extrapolation on the original scale of the data, but where such extrapolation is not required (such as in our setting), the rank transform can lead to more robust estimates. Applying transformation (14) as a preliminary step allows us to focus on estimating the extremal dependence features.

We consider (X_1, X_2) in terms of the pseudo-polar coordinates (R, W), with

$$R = X_1 + X_2 > 0, \qquad W = X_1/R \in [0, 1], \tag{15}$$

which we refer to as radial and angular components, respectively. Due to our choice of marginal distribution, these are not the same as the radial and angular components on Fréchet or Pareto scale used by [39] or [10], for instance, but are akin to those in [50]. This polar coordinate formulation lends itself to estimating G since the star-shaped nature of the set G_S means that any ray emanating from the origin, i.e., corresponding to a fixed angle $w \in [0, 1]$, must intersect the set G exactly once. We are therefore interested in the largest possible radial value associated with each angle w, after the $\log n$ scaling of (X_1, X_2), which also induces $\log n$ scaling of R. In practice, having a finite amount of data means that observations of $(X_1, X_2)/\log n$ are not restricted to the set $[0, 1]^2$, unlike the set G that we aim to estimate, and the distribution of the observed radial variable need not have a finite upper endpoint. We propose to estimate G by estimating high quantiles of the distribution of $R \mid W = w$ for all $w \in [0, 1]$, then transforming these estimates to the original coordinates (X_1, X_2) via the relations $X_1 = RW$, $X_2 = R(1 - W)$ and finally scaling them onto $[0, 1]^2$.

A natural candidate for modelling the tail of $R \mid W = w$ is the generalised Pareto distribution (GPD) [14]. That is, for $w \in [0, 1]$ and a suitable threshold u_w, we have

$$\Pr\left(R < r \mid R > u_w, W = w\right) = 1 - \left[1 + \xi(w)\left\{\frac{r - u_w}{\sigma(w)}\right\}\right]_+^{-1/\xi(w)}, \tag{16}$$

for $r > u_w$, $x_+ = \max\{x, 0\}$, scale parameter $\sigma(w) > 0$ and shape parameter $\xi(w) \in \mathbb{R}$. We consider $r_q(w)$, a high quantile of the distribution of $R \mid W = w$, for all $w \in [0, 1]$, where

$$\Pr\{R < r_q(w) \mid W = w\} = q,$$

with $r_q(w) > u_w$. For $\xi(w) \neq 0$ and $\zeta_u(w) = \Pr(R > u_w \mid W = w)$, this is

$$r_q(w) = u_w + \frac{\sigma(w)}{\xi(w)}\left[\left\{\frac{\zeta_u(w)}{1 - q}\right\}^{\xi(w)} - 1\right]. \tag{17}$$

We aim to estimate the set $\{(r_q(w), w) : w \in [0, 1]\}$ for q close to 1.

4.2 Overview of our inference procedure

In Sections 4.3 and 4.4, we propose a method to estimate the set G, by exploiting the framework detailed in Section 4.1. In Section 4.3, a local estimation procedure is proposed, where estimates of the radial quantiles in (17) are obtained for a fixed set of angles by considering radial observations within some appropriate neighbourhood. We can induce an estimate of G from these radial quantiles by applying an appropriate scaling onto the region $[0, 1]^2$; this scaling procedure is discussed in Section 4.3. To achieve a smooth estimate of the set G, we then propose to exploit the generalised additive models (GAMs) framework for the GPD parameters in (16), with the degree of the corresponding splines informed by the results from the local quantile estimates. This is discussed in Section 4.4 and provides our final estimate of G. An algorithmic representation of the full method is provided in the supplementary material [43, Section L].

4.3 Estimation of G: a local approach

Suppose we have n observations of the radial-angular components (R, W), denoted by $(r_1, w_1), \ldots,$ (r_n, w_n). If (X_1, X_2) have a joint density, we will not observe repeated observations of the angular component in a finite sample, so for any given $w \in [0, 1]$, we will not have sufficient observations[1] to fit distribution (16) directly. We propose a two-step approach to address this issue. First, we obtain local estimates of the model parameters $\sigma(w)$ and $\xi(w)$ and corresponding radial quantiles $r_q(w)$ for fixed values of w. From these, we construct a local estimate of the set G, which we denote $\hat{G}^L$. The theoretical examples in Figure 1 showed that the set G may generally be smooth, but can have points of non-differentiability. Motivated by this observation, our second step, detailed in Section 4.4, exploits the GAMs framework to provide smooth estimates of the GPD threshold and scale parameter, and hence of $r_q(w)$, for $w \in [0, 1]$. In Section 4.4, we explain how we use $\hat{G}^L$ to inform the choice of spline basis functions in the GAMs framework and yield an estimate $\hat{G}$ of G.

In the local estimation procedure, we use observed radial values corresponding to the angular components in some neighbourhood of w, defined as

$$\mathcal{R}_w = \{r_i : |w_i - w| \leq \epsilon_w, i \in \{1, \ldots, n\}\}, \tag{18}$$

for some $\epsilon_w > 0$, to be equivalent to the event $W = w$, i.e., this requires the GPD parameters to be approximately constant over $\mathcal{R}_w$. Using standard maximum likelihood estimation techniques with observations from $\mathcal{R}_w$, it is straightforward to estimate the parameters $\sigma(w)$, $\xi(w)$ in (16) and (17), and we suggest setting $\zeta_u(w) = q_u \in [0, q]$ identically for all w with the corresponding threshold u_w estimated empirically. We propose to carry out this estimation at a range of angular values, denoted by $w_1^*, \ldots, w_k^*$, and use these to estimate the radial quantiles in (17), denoting these quantile estimates by $\hat{r}_q(w_j^*)$, $j \in J_k := \{1, \ldots, k\}$. The w_j^* values are not required to belong to the set of observed W values and may be selected so that the associated $\mathcal{R}_{w_j^*}$ sets overlap. As well as k and the values w_j^* ($j \in J_k$), other tuning parameters to be selected here are the quantile levels (q_u, q) and the neighbourhood sizes

---

[1]The rank transformation given in (14) means that a small number of repetitions are possible.

$\epsilon_{w_j^*}$. We discuss our approach to choosing these quantities in the supplementary material [43, Section B], where we propose choosing $\epsilon_{w_j^*}$ separately for each j to ensure $|\mathcal{R}_{w_j^*}| = m$. Our default tuning parameter suggestions are $k = 199$, $m = 100$, $q_u = 0.5$ and $q = 0.999$.

To obtain an estimate of G, denoted $\hat{G}^L$ for the local approach, we first transform the local radial quantile estimates back to the original (X_1, X_2) coordinates by setting

$$\tilde{x}_{1,j} = \hat{r}_q(w_j^*) \cdot w_j^*, \qquad \tilde{x}_{2,j} = \hat{r}_q(w_j^*) \cdot (1 - w_j^*), \qquad j \in J_k.$$

To ensure $G \subset [0,1]^2$, a naive approach would be to simply divide each component by its largest value, i.e., by updating to $\tilde{x}'_{i,j} = \tilde{x}_{i,j} / \max_{j \in J_k} \tilde{x}_{i,j}$, $i = 1, 2$. However, this is strongly influenced by the largest estimates, and does not work well in practice. Instead, we suggest a two-step scaling/truncation approach, initially informed by an existing estimator of η and then corrected to take into account known features of G. We begin by scaling all $(\tilde{x}_{1,j}, \tilde{x}_{2,j})$ values to obtain a preliminary local limit set estimate $\hat{G}_\eta^L = \left\{ (\hat{x}_{1,j}^{G_\eta}, \hat{x}_{2,j}^{G_\eta}) : j \in J_k \right\}$, with the scaling factor chosen so that an estimate of η from the resulting set would match the Hill estimator $\hat{\eta}_H$, see (24), i.e., so that $\hat{G}_\eta^L$ and $[\hat{\eta}_H, \infty]^2$ just intersect. The scaling factor that achieves this is

$$x^* = \hat{\eta}_H \left[\max_{j \in J_k} \left\{ \min(\tilde{x}_{1,j}, \tilde{x}_{2,j}) \right\} \right]^{-1},$$

i.e., we set $\hat{x}_{i,j}^{G_\eta} = x^* \tilde{x}_{i,j}$ for $i = 1, 2$, $j \in J_k$. The resulting estimate $\hat{G}_\eta^L$ is not guaranteed to lie within $[0,1]^2$, nor to intersect the lines $x_1 = 1$ and $x_2 = 1$; the second step corrects for these possibilities, taking each of the arguments separately. In step 2a, for $i = 1, 2$, if $\max_{j \in J_k} \hat{x}_{i,j}^{G_\eta} \geq 1$, we truncate by defining $\hat{x}_{i,j}^G = \min \left\{ 1, \hat{x}_{i,j}^{G_\eta} \right\}$, $j \in J_k$. Alternatively, in step 2b, if $\max_{j \in J_k} \hat{x}_{i,j}^{G_\eta} < 1$, we ensure that our estimate of G intersects the line $x_i = 1$ at least once by dividing the component by its maximum value, setting $\hat{x}_{i,j}^G = \hat{x}_{i,j}^{G_\eta} / \max_{\ell \in J_k} \hat{x}_{i,\ell}^{G_\eta}$. The scaling in step 2b means that an estimate of η obtained from $\hat{G}^L$ (as will be discussed in Section 5.2) may not be exactly equal to $\hat{\eta}_H$, but our approach actually improves the estimates of this coefficient when $\hat{\eta}_H$ has a tendency to underestimate, such as under asymptotic dependence in the logistic model; see Section 6.1. The full scaling/truncation approach is demonstrated pictorially in the supplementary material [43, Section C]. This scaling procedure relies on us choosing an appropriate threshold u_H for $\hat{\eta}_H$; we set this to the 0.95 quantile of the structure variable M in (24), which is also the level used throughout the simulation studies in Section 6.1, comparing different estimators of η.

4.4 Estimation of G: a smoothed approach

There is no guarantee of smoothness in the estimated set $\hat{G}^L$, whatever the choice of tuning parameters. As we have already identified, it may be desirable to impose that our estimator of G is smooth, at least for some subsets of $w \in [0, 1]$. To achieve this, we use $\hat{G}^L$ to inform a 'smoothing' method for the GPD parameters in the second step of our procedure. In particular, we follow the approach of [51] by first carrying out quantile regression to estimate the thresholds u_w at the same quantile level $q_u \in [0, q]$ as in Section 4.3 by exploiting the asymmetric Laplace distribution [53]. The parameters of the asymmetric Laplace distribution are modelled using GAMs. We find that applying this quantile regression approach
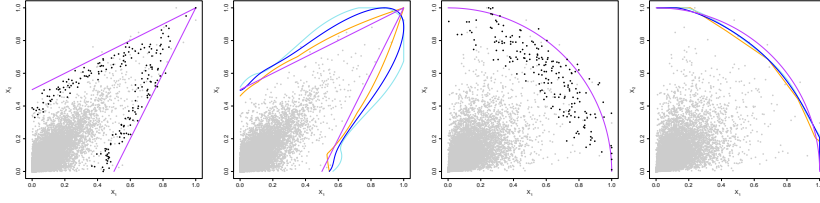
Figure 2: Left: logistic data scaled by $\log n$, with the true set G (purple) and $\hat{G}^L$ (black points). Left-centre: the estimated sets $\hat{G}^{S_1}$ (orange), $\hat{G}^{S_2}$ (light blue) and $\hat{G}^{S_3}$ (dark blue), taking $k = 199$, $m = 100$, $q_u = 0.5$, $q = 0.9$ and $\kappa = 7$. Right-centre and right: equivalent plots for inverted logistic data.

to $R \mid (W = w)$ often gives impossible negative estimates near the largest and smallest observed angles. We overcome this issue by working with $\log R \mid (W = w)$ in this step, and then back-transform to obtain the required radial threshold estimates. For exceedances above u_w, we also assume that the parameter values in the GPD vary smoothly with the value of $w \in [0, 1]$. Again, following the approach of [51], we adopt a GAMs form for the log-scale parameter. Both parts of this approach are implemented in the R package `evgam` [52]. The shape parameter could analogously be allowed to vary smoothly, but this is notoriously difficult to estimate and we found that in tests, anything other than assuming a constant value resulted in increased variance without reducing the bias of our estimates.

We allow the GAMs form for the model parameters to be linear, quadratic or cubic B-splines, with κ knots; for simplicity, the same spline degree and knot positions are chosen for u_w and $\log \sigma(w)$, and corresponding estimates of the quantiles in (17). Once we have estimates of the high radial quantiles, the procedure to obtain an estimate of G is via transformation and scaling/truncation, equivalent to that for $\hat{G}^L$. Although, theoretically, this approach allows for estimates of G across the full range $w \in [0, 1]$, outside the observed angles these can be unreliable, and we restrict our estimates to the observed range, $w \in (w^m, w^M)$. The resulting estimates of G are denoted by $\hat{G}^{S_1}$, $\hat{G}^{S_2}$ and $\hat{G}^{S_3}$ for the linear, quadratic and cubic splines, respectively.

Considering different spline degrees here is useful, as it enables us to capture different shapes of G. To choose between the estimators $\hat{G}^{S_1}$, $\hat{G}^{S_2}$ and $\hat{G}^{S_3}$, we select the spline degree where the estimated radial quantiles at the angles w_j^* ($j \in J_k$) are closest to those from the local procedure, when compared using the mean absolute error. This selection criterion is a reasonable choice, since by forcing an equal number of points in each set $\mathcal{R}_{w_j^*}$, we have ensured there will be approximately the same uncertainty in each radial estimate. The selected estimator $\hat{G}$ (from $\hat{G}^{S_1}$, $\hat{G}^{S_2}$ or $\hat{G}^{S_3}$) becomes our preferred option for modelling G. That is, letting $\hat{r}_q^L(w_j^*)$, $j \in J_k$, denote the local radial quantile estimates and $\hat{r}_q^{S_d}(w_j^*)$, $d = 1, 2, 3$, denote the corresponding estimates of $r_q(w_j^*)$ for $\hat{G}^{S_d}$, we have

$$\hat{G} = \left\{ \hat{G}^{S_d} : d = \underset{d^* \in \{1,2,3\}}{\arg\min} \sum_{j \in J_k} \left| \hat{r}_q^L(w_j^*) - \hat{r}_q^{S_{d^*}}(w_j^*) \right| \right\}. \tag{19}$$

Figure 2 gives a demonstration of the full approach for 10,000 simulated values from a bivariate extreme value model with a logistic model and parameter $\gamma = 1/2$, and its inverted counterpart, with

comparison to the true sets G. The tuning parameters here are chosen for illustrative purposes. It is clear that $\hat{G}^L$ is not smooth and tends to underestimate G, whereas $\hat{G}^{S_1}$, $\hat{G}^{S_2}$ and $\hat{G}^{S_3}$ offer varying levels of improvement. The linear spline performs the best in the logistic case where G has a pointed shape, with the mean absolute errors evaluated in (19) being 55.23, 67.37 and 57.64 for $\hat{G}^{S_1}$, $\hat{G}^{S_2}$ and $\hat{G}^{S_3}$, respectively, so $\hat{G} = \hat{G}^{S_1}$. The errors in the inverted logistic case are 61.95, 61.50 and 61.29, i.e., the cubic spline is slightly preferred as G is smooth, so $\hat{G} = \hat{G}^{S_3}$.

The smoothing method requires the introduction of further tuning parameters, in the choice of κ and the position of the knots. Tuning parameter selection is discussed in detail in the supplementary material [43, Section B], where a default choice of $\kappa = 7$ is proposed, in addition to the suggested tuning parameters for $\hat{G}^L$ in Section 4.3. Knot-placement; the positions of w_j^* ($j \in J_k$); and sensitivity to the extrapolation quantile q are also discussed in the supplementary material [43, Sections B and E], and we show limited sensitivity to κ for an asymmetric logistic copula [43, Section I].

5 Estimation of extremal dependence properties

5.1 Overview

Further exploiting the new procedure to estimate G, in Section 5.2, we obtain estimates of η, $\lambda(\omega)$, $\tau_1(\delta)$, $\tau_2(\delta)$, α_1 and α_2 in a unified way. Due to the subtlety of result (13) for the conditional extremes parameters β_1 and β_2, we choose not to estimate this as part of our self-consistent approach. Instead, we fix the corresponding α_i parameter to its estimated value, and then carry out inference for β_i; this is discussed in Section 5.3. In Section 5.4, we detail some techniques that are currently available to separately estimate the extremal dependence features that we consider. In Section 5.5, we highlight some desirable properties of our estimators, including (but not limited to) the self-consistency that has been previously mentioned. We compare the performance of these methods to that of our estimation approach in Section 6 and the supplementary material [43, Sections F-H].

5.2 Exploiting the estimate $\hat{G}$ for parameter estimation

For the estimation of the dependence properties, we consider $\hat{G}$ in (19) at a finite number of angles, for ease of implementation. We choose these as identical to the w_j^*, $j \in J_k$, used in the local estimate $\hat{G}^L$, now denoting $\hat{G} = \left\{ (\hat{x}_{1,j}^G, \hat{x}_{2,j}^G) : j \in J_k \right\}$. To estimate η using $\hat{G}$, we first note that result (9) from [34] can be written as

$$\eta = \min \left\{ s \in (0,1] : G \cap [s,\infty]^2 = \emptyset \right\} = \max \left\{ s \in (0,1] : G \cap [s,\infty]^2 \neq \emptyset \right\}. \tag{20}$$

Considering each point in our set $\hat{G}$ to be a candidate for the intersection of G and $[\eta,\infty]^2$ required in (20), we can examine a corresponding set $[s_j^*,\infty]^2$ for each such points by setting $s_j^* = \min(\hat{x}_{1,j}^G, \hat{x}_{2,j}^G)$. Then the resulting estimate of η is given by $\max(s_1^*, \ldots, s_k^*)$, i.e.,

$$\hat{\eta}_G = \max_{j \in J_k} \left\{ \min \left(\hat{x}_{1,j}^G, \hat{x}_{2,j}^G \right) \right\}. \tag{21}$$

14

We estimate $\lambda(\omega)$, $\omega \in [0, 1]$, via estimation of s_ω in (10). We have

$$s_\omega = \min \{s \in [0, 1] : sS_\omega \cap G = \emptyset\} = \max \{s \in [0, 1] : sS_\omega \cap G \neq \emptyset\},$$

with $S_\omega = \{(x_1, x_2) : x_1 \geq \omega/\max(\omega, 1 - \omega), x_2 \geq (1 - \omega)/\max(\omega, 1 - \omega)\}$. Analogously to $\hat{\eta}_G$, we obtain candidate values

$$s_{\omega,j}^* = \max(\omega, 1 - \omega) \cdot \min \{\hat{x}_{1,j}^G/\omega, \hat{x}_{2,j}^G/(1 - \omega)\}$$

for $j \in J_k$, with the resulting estimator

$$\hat{\lambda}_G(\omega) = \left[\max_{j \in J_k} \left\{ \min \left(\frac{\hat{x}_{1,j}^G}{\omega}, \frac{\hat{x}_{2,j}^G}{1 - \omega} \right) \right\} \right]^{-1}, \qquad w \in [0, 1].$$

For the dependence measure $\tau_1(\delta)$, result (11) gives that for $\delta \in [0, 1]$,

$$\tau_1(\delta) = \min \{s \in [0, 1] : [s, \infty] \times [0, \delta s] \cap G = \emptyset\}$$
$$= \max \{s \in [0, 1] : [s, \infty] \times [0, \delta s] \cap G \neq \emptyset\}.$$

Our candidate points for the intersection of the sets G and $(\tau_1(\delta), \infty] \times [0, \delta\tau_1(\delta)]$ are all our estimated points on G where $\hat{x}_2^G \leq \delta\hat{x}_1^G$. Our estimate of $\tau_1(\delta)$ corresponds to the largest such value of $\hat{x}_1^G$. That is,

$$\hat{\tau}_{G,1}(\delta) = \max \left(\hat{x}_{1,j}^G : \hat{x}_{2,j}^G \leq \delta\hat{x}_{1,j}^G, j \in \{1, \dots, k\} \right).$$

Finally, for the conditional extremes parameter α_1,

$$\hat{\alpha}_{G,1} = \max \left(\hat{x}_{2,j}^G : \hat{x}_{1,j}^G = 1, j \in \{1, \dots, k\} \right). \tag{22}$$

Due to our method of scaling onto $[0, 1]^2$, in practice there may only be exactly one value of $\hat{x}_{1,j}^G$ that is equal to one. The estimators of $\hat{\tau}_{G,2}(\delta)$ and $\hat{\alpha}_{G,2}$ are defined analogously. Our approach ensures self-consistency across the conditioning variables in the conditional extremes approach, since $\hat{\alpha}_{G,1} = 1$ if and only if $\hat{\alpha}_{G,2} = 1$; this issue is also discussed in [29]. Moreover, estimating the α_i ($i = 1, 2$) parameters in this way avoids the need to place a distributional assumption on the random variable Z in (8); this benefit over other approaches is discussed further in Section 5.5. A pictorial illustration of the estimation of η, $\lambda(\omega)$, $\tau_1(\delta)$ and α_1 is given in the supplementary material [43, Section C].

5.3 Estimation of the conditional scaling parameter β_i

As discussed in Section 3.2, the parameter β_i, $i = 1, 2$, in the conditional extremes framework is also linked to the asymptotic shape of a scaled sample cloud. However, the required feature of the set G is very subtle, and reliably estimating β_i using our estimate $\hat{G}$ is not really feasible. Instead, we estimate β_i with α_i fixed to its estimated value, $\hat{\alpha}_{G,i}$ of expression (22). Illustrating this for β_1, limit (8) is taken as an equality for some finite, but large, threshold u. Then, for $X_1 = x > u$ and with $a_1(x) = \hat{\alpha}_{G,1}x$ and $b_1(x) = x^{\beta_1}$, this implies

$$X_2|(X_1 = x) = \hat{\alpha}_{G,1}x + x^{\beta_1}Z, \qquad x > u,$$

where Z is a non-degenerate random variable that is independent of X_1. It is common to take $Z \sim \mathcal{N}(\mu, \sigma^2)$ as a working assumption, for $\mu \in \mathbb{R}$, $\sigma > 0$ (e.g., [25]), although other approaches have been considered (e.g., [30]), which implies

$$X_2 | (X_1 = x) \sim \mathcal{N}\left\{ (\hat{\alpha}_{G,1} x + x^{\beta_1} \mu), (x^{\beta_1} \sigma)^2 \right\}, \qquad x > u. \tag{23}$$

From this, it is straightforward to obtain maximum likelihood estimates for the parameters, with $\hat{\beta}_{G,1}$ denoting the corresponding estimate for β_1.

5.4 Existing estimation techniques

In this section, we briefly present some existing estimators for the extremal dependence properties discussed in Section 3.1. We use these for comparison to our proposed estimators in Section 6. More detail on these estimators is provided in the supplementary material [43, Section D].

For η, we consider three options. The first is a Hill-type estimator [24] on the structure variable $M = \min(X_1, X_2)$. Assuming we have $n_{u,H}$ observations of M above some high threshold u_H, denoted $m_1^*, \ldots, m_{n_{u,H}}^*$, the estimator is

$$\hat{\eta}_H = \min\left\{ \frac{1}{n_{u,H}} \sum_{i=1}^{n_{u,H}} (m_i^* - u_H), 1 \right\}. \tag{24}$$

Alternatively, suppose we have n pairs of observations $(x_{1,1}, x_{1,2}), \ldots, (x_{n,1}, x_{n,2})$, and denote by $x_i^{(j)}$ the jth largest value of component $i \in \{1, 2\}$, i.e., $x_i^{(n)} \leq x_i^{(n-1)} \leq \cdots \leq x_i^{(1)}$. Then for each $j \in \{1, \ldots, n\}$, define the quantity $s_n(j) = \sum_{\ell=1}^{n} \mathbf{1}\left\{ x_{\ell,1} \geq x_1^{(j)}, x_{\ell,2} \geq x_2^{(j)} \right\}$. Two non-parametric estimators, $\hat{\eta}_P$ of [37] and $\hat{\eta}_D$ of [16], are

$$\hat{\eta}_P = \min\left[\frac{\log 2}{\log\{s_n(2c)\} - \log\{s_n(c)\}}, 1 \right], \quad \hat{\eta}_D = \min\left\{ \frac{\sum_{j=1}^{c} s_n(j)}{c s_n(c) - \sum_{j=1}^{c} s_n(j)}, 1 \right\},$$

where the forms of these estimators are as presented in [4, p.352], but we have added the truncation at one. The tuning parameter c relates to the number of exceedances above a high threshold.

Hill-type estimators are also available for $\lambda(\omega)$ and $\tau_1(\delta)$. For the former, set $M_\omega = \min\{X_1/\omega, X_2/(1-\omega)\}$, for $\omega \in [0, 1]$, and assume there are $n_{u,\omega}$ observations of M_ω above a high threshold u_ω, denoted $m_{\omega,1}^*, \ldots, m_{\omega,n_{u,\omega}}^*$. The corresponding estimator is

$$\hat{\lambda}_H(\omega) = \min\left[\left\{ \frac{1}{n_{u,\omega}} \sum_{i=1}^{n_{u,\omega}} (m_{\omega,i}^* - u_\omega) \right\}^{-1}, 1 \right].$$

For $\tau_1(\delta)$, $\delta \in (0, 1]$, let $M_\delta = \{X_1 > 0 : X_2 \leq \delta X_1\}$ and suppose we have $n_\delta \leq n$ observations of X_1 where $X_2 \leq \delta X_1$, and that $n_{u,\delta}$ of these, denoted $m_{\delta,1}^*, \ldots, m_{\delta,n_{u,\delta}}^*$, are above some high threshold u_δ. Then the estimator is

$$\hat{\tau}_{H,1}(\delta) = \min\left\{ \frac{1}{n_{u,\delta}} \sum_{i=1}^{n_{u,\delta}} (m_{\delta,i}^* - u_\delta), 1 \right\}.$$

5.5 Properties of our estimators

The main advantage of our approach over estimating each of the extremal dependence features separately, is that by linking each of the estimators to a single estimate of G, we automatically achieve self-consistency across the different dependence features. One way to think about this is in terms of whether the parameter estimates indicate asymptotic dependence or asymptotic independence, since separate estimation could lead to contradictions. Returning to our example from Section 1, we consider the case of asymptotic dependence, where we know that $\eta = 1$ and $\alpha_1 = \alpha_2 = 1$. In our approach, we obtain $\hat{\eta}_G = 1$ and $\hat{\alpha}_{G,1} = \hat{\alpha}_{G,2} = 1$ exactly when our estimate of the set G contains $(1,1)$. This means that $\hat{\eta}_G = 1$ if and only if $\hat{\alpha}_{G,1} = \hat{\alpha}_{G,2} = 1$, which is not guaranteed when carrying out inference separately for these dependence features.

We mentioned in Section 3.1 that the set of measures $\{\eta, \tau_1(\delta), \tau_2(\delta)\}$ can be used to describe the extremal dependence structure of the variables. In particular, if X_1 can be large while X_2 is small, there must be some value of $\delta^* < 1$ such that $\tau_1(\delta^*) = 1$, and a similar result holds for $\tau_2(\delta)$. Moreover, if both variables can take their largest values simultaneously, then $\eta = 1$. Simpson et al. [44] point out that

$$\max\{\eta, \tau_1(\delta) : \delta \in [0,1)\} = \max\{\eta, \tau_2(\delta) : \delta \in [0,1)\} = 1. \tag{25}$$

This is not guaranteed by $\hat{\eta}_H$, $\hat{\tau}_{H,1}(\delta)$ and $\hat{\tau}_{H,2}(\delta)$ (although a possible way to enforce this is discussed in [41]), but is naturally satisfied by our new estimation procedure, since the scaling/truncation steps ensure $\hat{G}$ intersects both the lines $x_1 = 1$ and $x_2 = 1$ at least once. Further, we are guaranteed to have monotonicity in our estimates of $\tau_i(\delta)$ ($i = 1, 2$) as δ increases; this theoretical property is not achieved through application of the Hill estimator at separate values of δ.

Finally, maximum likelihood estimation for conditional extremes models involves placing a distributional assumption on Z. This often gives reasonable performance, but mis-specifying the distribution has the potential to affect the resulting estimates of α_i and β_i. Our new approach has the benefit that we do not need to place a distributional assumption on Z for estimation of α_i, although we do for subsequent estimation of β_i.

6 Results - simulation study and data example

6.1 Simulation study

We now demonstrate the performance of our proposed estimation procedure via a simulation study. Our primary interest lies in estimating the whole of G, which is the focus of Section 6.1.1. However, no other estimators of G exist, but there are estimators available for certain parts of G if we exploit the links to the extremal dependence measures discussed in Sections 3 and 5. In order to judge our performance against these competitors, we provide additional simulation results for individual dependence measures in Section 6.1.2. Our estimation procedure provides point estimates for G and each of the individual dependence measures; to assess uncertainty, we propose that a non-parametric bootstrapping procedure
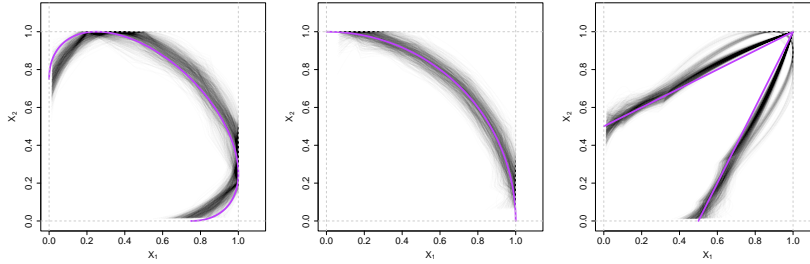
Figure 3: Estimates $\hat{G}$ for the Gaussian (left), inverted logistic (centre) and logistic (right) models, with the ρ or γ parameters set to 0.5. The sample size in each case is $n = 10,000$ and 1000 estimates are shown for each model (grey). The true sets G are shown in purple.

Spline degree	Gaussian	Inverted logistic	Logistic
1	392	355	798
2	482	250	32
3	126	395	170

Table 2: The number of times each spline degree is chosen for the sets $\hat{G}$ shown in Figure 3.

could be used, and show in the supplementary material [43, Section K] that this approach can provide reasonable coverage.

6.1.1 Estimation of G

We begin by considering estimation of the complete set G for the first three copulas from Figure 1, with $\rho, \gamma = 0.5$. For each copula, with Exponential margins, we take a sample of size 10,000 and use the estimator $\hat{G}$ of (19). This is repeated 1000 times for each model, with the resulting estimates $\hat{G}$ shown in Figure 3, alongside the true G. To obtain these results, we select the tuning parameters as stated in Section 4.4. In the supplementary material [43, Section J], we demonstrate the performance of our $\hat{G}$ for samples of size $n = 100,000$ taken from the same three copula models; these results demonstrate that the tuning parameter choices are also reasonable for larger samples.

There is reasonable agreement between $\hat{G}$ and G in Figure 3 for all three copulas. In the inverted logistic simulations, the shape of G is generally well represented by $\hat{G}$. Some small bias occurs in the Gaussian case; this is investigated further in the supplementary material [43, Section J], where we discuss known issues with slow convergence for this copula. For the logistic case, there are some estimates where the pointed shape of G has not been captured, but the estimates are reasonably successful overall.

The degree of the spline chosen to model the log-scale parameter and threshold in the GPD plays an important role in determining the shape of $\hat{G}$; in Table 2, we summarise the selected spline degrees over the 1000 replicated datasets for each copula. Since G is smooth for both the Gaussian and inverted logistic copulas, we would expect to see the procedure favouring quadratic or cubic splines. For both these copulas, there is no overwhelming favourite in terms of the spline degrees chosen, but the non-linear
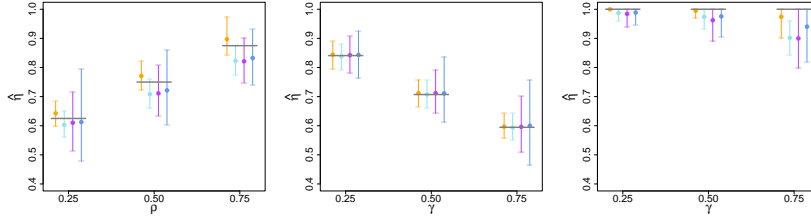
18

Figure 4: Estimates of η for data simulated from Gaussian (left), inverted logistic (centre) and logistic (right) copulas with their corresponding parameters taken as 0.25, 0.5 and 0.75. We show the mean estimates of η (circles) and 0.025 to 0.975 quantile range from 1000 replicates for $\hat{\eta}_G$ (orange), $\hat{\eta}_H$ (light blue), $\hat{\eta}_P$ (purple), $\hat{\eta}_D$ (dark blue), each based on a data sample of 10,000. The true η values, summarised in Table 1, are shown in grey.

options are selected around 60-65% of the time. On the other hand, the pointed shape of G for the logistic copula suggests that linear splines should be preferred here, with this being achieved in almost 80% of cases, and $(1,1) \notin \hat{G}$ corresponding to the selection of quadratic or cubic splines.

6.1.2 Estimation of the dependence measures

As discussed in Section 5.2, once we have $\hat{G}$, we can use this to induce estimates of various extremal dependence properties. For estimating η, we compare our $\hat{\eta}_G$ given in (21), with the existing estimators $\hat{\eta}_H$, $\hat{\eta}_P$ and $\hat{\eta}_D$ described in Section 5.4. We fix the threshold u_H used in the estimator $\hat{\eta}_H$ to coincide with c for the estimators $\hat{\eta}_P$ and $\hat{\eta}_D$, such that all estimators use 500 data points. As with the tuning parameters in our approach, we have not optimised these choices, but have selected values that we found to work reasonably well across a range of examples. We take the set-up of our study as the same as for the results in Figure 3, but now take $\rho, \gamma \in \{0.25, 0.5, 0.75\}$; summaries of the η estimates are shown in Figure 4.

For the Gaussian copula, the existing approaches underestimate the true value of η, while our approach tends to overestimate by a similar magnitude, although our 95% sampling distribution interval does contain the truth in each case and is generally the narrowest. For the inverted logistic copula, both $\hat{\eta}_G$ and $\hat{\eta}_H$ perform consistently well across the different values of γ; all estimators appear to be unbiased but $\hat{\eta}_P$ and $\hat{\eta}_D$ suffer from increasing variability. The similarity of $\hat{\eta}_G$ and $\hat{\eta}_H$ is due to the fact that for this copula, step 2b of our scaling/truncation procedure is generally not required. Where $\hat{\eta}_G$ is particularly successful is in the logistic case, with asymptotic dependence and $\eta = 1$. Across all values of γ, we outperform the existing estimators, with clear success in Figure 4 for $\gamma \in \{0.25, 0.5\}$, where the dependence is stronger. For $\gamma = 0.75$, the picture is less clear, so we find the root mean square error for each of the estimators in this case; these are 0.044, 0.103, 0.113 and 0.080 for $\hat{\eta}_G$, $\hat{\eta}_H$, $\hat{\eta}_P$ and $\hat{\eta}_D$, respectively, confirming that $\hat{\eta}_G$ provides improvement over previous approaches. In terms of self-consistency for the logistic copula, while our approach satisfies the required property that $\eta \geq \max(\alpha_1, \alpha_2)$ in all cases, this is only true for 45%, 46% and 66% of the $\hat{\eta}_H$, $\hat{\eta}_P$ and $\hat{\eta}_D$ estimates in Figure 4, respectively, when compared to the maximum
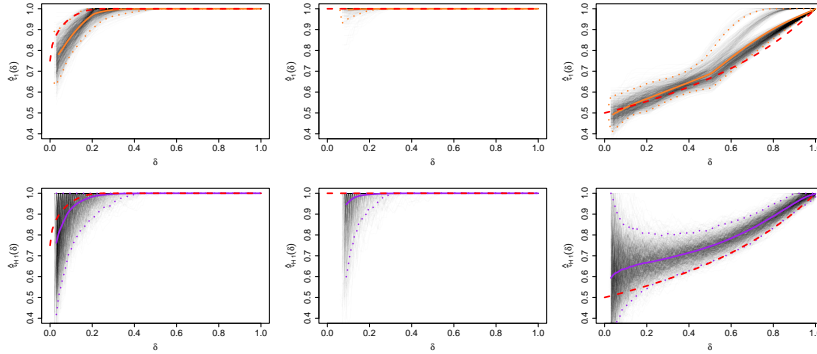
19

Figure 5: Estimates of $\tau_1(\delta)$, $\delta \in [0,1]$, for data simulated from Gaussian (left), inverted logistic (centre) and logistic (right) models with their corresponding parameters set as 0.5. Each grey line represents an estimate obtained using $\hat{\tau}_{G,1}(\delta)$ (top) or $\hat{\tau}_{H,1}(\delta)$ (bottom) over 1000 simulations. The solid and lower/upper dotted lines show the pointwise bootstrapped means and 0.025 and 0.975 quantiles, respectively, for $\hat{\tau}_{G,1}(\delta)$ (orange) and $\hat{\tau}_{H,1}(\delta)$ (purple). The true $\tau_1(\delta)$ values, summarised in Table 1, are shown in red.

likelihood estimator for α_i, $i = 1, 2$ (results for which are given in the supplementary material [43, Section H]). The existing estimators are almost always self-consistent in our study for the Gaussian and inverted logistic copulas, where η is much larger than $\alpha_1 = \alpha_2$, but our approach still has the benefit that this coherence is guaranteed.

In Figure 5, we present estimates of $\tau_1(\delta)$, comparing our procedure to $\hat{\tau}_{H,1}(\delta)$, described in Section 5.4. Following [44], we fix the threshold u_δ for $\hat{\tau}_{H,1}(\delta)$ to the observed 0.85 quantile of the variable M_δ. Since we restrict $\hat{G}$ to values within the range of observed angular values, there are some values of δ for which we cannot compute $\hat{\tau}_{G,1}(\delta)$. To ensure a fair comparison, we present the estimates $\hat{\tau}_{H,1}(\delta)$ over the same range of δ. In terms of bias, the two procedures are relatively similar, but there are two features to highlight in our estimates. First, only 18.8% of the $\hat{\tau}_{H,1}(\delta)$ estimates in Figure 5 (measured over 0.01 increments of δ) satisfy the required monotonicity property of $\tau_1(\delta)$, while our approach guarantees this aspect of self-consistency across different δ values in 100% of cases. Second, for small values of δ, $\hat{\tau}_{H,1}(\delta)$ has much greater variability than $\hat{\tau}_{G,1}(\delta)$ for all three copulas with $\rho, \gamma = 0.5$. This arises as $\hat{\tau}_{G,1}(\delta)$ 'borrows' more information from nearby values of δ than $\hat{\tau}_{H,1}(\delta)$, and therefore suffers less from the lack of data associated with estimates for small δ values.

In the supplementary material [43, Section F], we give further results for $\tau_1(\delta)$ for the copulas used in Figure 5 with $\rho, \gamma \in \{0.25, 0.75\}$; we see similar improvements in the logistic case with $\gamma = 0.25$. We also provide results for estimation of the features $\lambda(\omega)$, α_1 and β_1. Both $\hat{\lambda}_G(\omega)$ and $\hat{\lambda}_H(\omega)$ provide successful estimates for the inverted logistic model across a range of dependence parameters γ, with $\hat{\lambda}_G(\omega)$ providing smoother estimates; our approach exhibits less bias than $\hat{\lambda}_H(\omega)$ in Gaussian cases; and we see less variability in $\hat{\lambda}_G(\omega)$ than $\hat{\lambda}_H(\omega)$ in results for the logistic model. Our estimates of (α_1, β_1) are reasonably similar to those obtained by using maximum likelihood estimation for both parameters α_1 and β_1 simultaneously, i.e., replacing $\hat{\alpha}_{G,1}$ in (23) with α_1 to be estimated. We also provide additional simulation results for the asymmetric logistic model.
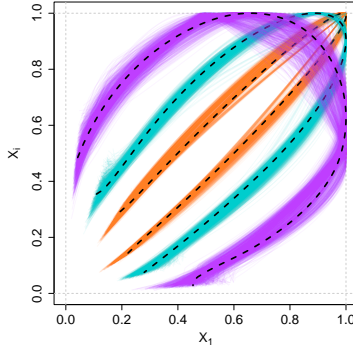
Figure 6: Estimates of G for significant wave heights at three pairs of locations in the North Sea (black dashed lines) with 1000 bootstrapped estimates shown by the orange, turquoise and purple lines for locations at distances of approximately 40km, 130km, 300km, respectively. The variable X_1 represents the transformed and scaled significant wave heights at a location in the north-west of our study region; X_i $(i = 2, 3, 4)$ corresponds to significant wave heights at the other locations.

We acknowledge that, considering estimates of each of the dependence measures separately, our new approach has some relatively small improvements over existing methods. However, we emphasise again the important point that our estimates are jointly self-consistent, which is not generally the case with the other approaches.

6.2 Application to sea wave height data

We now apply our approach to hindcast values of significant wave heights for locations in the North Sea. These are a subset of the data studied by [49] and, as in their paper, we consider only data from the winter months of December to February to limit issues with non-stationarity. We take data from four locations lying on a transect from the north-west to the south-east. The data cover 31 years, with eight observations for each day in our months of interest, resulting in a total of 22,376 observations per location. The transformation (14) is applied to each variable to ensure standard Exponential marginal distributions.

Figure 6 shows $\hat{G}$ at three pairs of locations with varying distances between them, with the tuning parameters as in Sections 4.4 and 6.1. Uncertainty is presented by implementing the stationary bootstrap scheme of [38] over 1000 iterations with the mean block-length set to 16 observations, i.e., two days. The most noticeable feature in these results is the weakening dependence in the significant wave height data as the distance between the locations increases, as is expected in environmental settings. The uncertainty in our estimates increases with the distance between the locations. While previous methods could have captured the weakening dependence for single parameter values, we can now see this change over the whole shape of $\hat{G}$, highlighting the value added by our approach. In terms of risk, these results can be interpreted based on whether pairs of locations may see extreme events simultaneously, compounding the potential impact. For the closest pair of locations, there is a clear risk that the largest wave heights can

occur together. However, as the distance between locations increases, the largest wave heights at one location are associated with increasingly smaller events at the other location, and there becomes less risk of a severe impact at these two sites simultaneously.

For the closest pair of locations, we estimate the coefficient of tail dependence as $\hat{\eta}_G = 0.997$ with a 95% confidence interval of $[0.994, 1.000]$; for comparison, we estimate $\hat{\eta}_H = 0.979\,[0.962, 0.997]$. Our $\hat{\eta}_G$ suggests selecting a model that can capture asymptotic dependence would be most appropriate here, whereas using $\hat{\eta}_H$ the choice is more ambiguous due to a wider confidence interval that does not contain one. For the remaining two pairs of locations, the estimates are $\hat{\eta}_G = 0.974\,[0.964, 0.983]$ and $\hat{\eta}_G = 0.897\,[0.811, 0.932]$, respectively, with corresponding estimates $\hat{\eta}_H = 0.918\,[0.879, 0.954]$ and $\hat{\eta}_H = 0.801\,[0.748, 0.855]$. The $\hat{\eta}_G$ results both indicate that a model for asymptotic independence would be reasonable, although one may prefer the conservative approach of choosing an asymptotically dependent model for the former since the confidence interval is quite close to one. For all three pairs of locations, $\hat{\eta}_G > \hat{\eta}_H$, with narrower confidence intervals obtained for $\hat{\eta}_G$.

7 Discussion

7.1 Summary of our contributions

The aim of this paper was to develop a new inferential technique for bivariate extremes, providing estimation for a limit set and, as a by-product, yielding self-consistent estimation across a range of extremal dependence features. This was motivated by theoretical developments from [34] and [35], who showed that these features can all be linked to the asymptotic shape of a suitably-scaled sample cloud. We have demonstrated that our approach provides estimators which, when considered marginally, are competitive compared to currently available techniques, and provide particularly good improvement for the important case of asymptotic dependence. But critically, in terms of self-consistency, our results are highly superior, as highlighted in Section 5.5, and our method provides additional information about joint extreme events which is not captured by any existing extremal dependence measure, as most clearly illustrated in the discussion of Figure 7 in Section 7.3.

7.2 Links to other approaches

The boundary function may appear at first sight to be essentially like a contour of equal joint density, however, they are different quantities, which is highlighted by the examples in Section 2.2. Also, the existing estimators of density contours in the tail of the joint distribution have major restrictions for practical use, which are not an issue in our consideration of G. Specifically, [7, 18] assume that a multivariate regular variation limit forms hold, which imposes that the marginal distributions must be heavy tailed, and that the variables are either asymptotically dependent or completely independent, meaning that more general forms of asymptotic independence are excluded. Furthermore, they impose conditions such that the spectral measure of the regular variation is non-zero across all interior directional rays.

While our paper has been in the review process, the draft version with working title *"Estimating the limiting shape of bivariate scaled sample clouds for self-consistent inference of extremal dependence properties"* that was uploaded to *arXiv* has been referenced in several subsequent papers also aiming to estimate the limit set, highlighting the value in developing such estimators; see for example [32, 33, 36]. As part of their paper, Murphy-Barltrop et al. [33] carry out an in-depth simulation study comparing estimators of $\lambda(\omega)$. One of their main messages is that "global estimators" such as ours, where simultaneous estimation of $\lambda(\omega)$ is carried out across all values of $\omega \in [0, 1]$, far out-perform local estimates, and the approach we have proposed in particular is shown to perform very well. Moreover, Wadsworth and Campbell [48] have independently developed what they term a "geometric approach" to estimating the probability of multivariate extreme events, with links to the framework we have presented here, and showing benefits over existing methods for multivariate extremes. Part of their inference involves estimation of the gauge function, which they restrict to a class of parametric forms; there is the potential for our proposed method for estimating G to be combined with their approach for enhanced flexibility in this aspect. Overall, statistical inference with the limit set and gauge functions appears to be a promising avenue for future research.

An intermediate step of our estimation procedure considers radial quantiles at high (but non-limiting) levels. Considering these radial quantiles across the range of angles gives a contour that no longer represents a limit set, but which is still a potentially interesting extreme set in its own right. In particular, this can be linked to the idea of environmental contours, often used in engineering applications, and the methodology presented in this paper has now been extended to such a setting [42].

7.3 Potential issues and extensions

We have assumed that our variables have standard Exponential margins. This restricts us to cases where there is independence or positive dependence between the variables, i.e., we cannot handle negatively dependent variables. For conditional extremes modelling, Keef et al. [25] capture this feature by instead using Laplace margins; this possibility is also briefly discussed by [35]. In practice, one could first test for negative dependence between the variables, and proceed with our approach only if it is not present in the data. It would also be possible to extend the methodology presented here for data initially transformed to Laplace margins, with different forms for the (pseudo-)polar coordinates chosen to allow the angular variable to cover all four quadrants, i.e., $W \in [-\pi, \pi)$; this approach is taken for the environmental contours in [42].

As discussed in Section 1, the limit set G could be linked to likely boundary sets for large finite samples, which would take the form G_m in (3) on Exponential margins. However, there may also be cases where a boundary set is required for finite samples on the original scale of the data. Taking the original variables as (Y_1, Y_2) with marginal distributions $F_{Y_i}(y)$ $(i = 1, 2)$, one could achieve this by considering

$$G_m^Y := \{(y_1, y_2) \in \mathbb{R}^2 : g(T_1(y_1)/\log m, T_2(y_2)/\log m) = 1\},$$

where $T_i(y) = -\log\{1 - \hat{F}_{Y_i}(y)\}$ and $\hat{F}_{Y_i}(y)$ is an estimate of $F_{Y_i}(y)$, making use of the probability
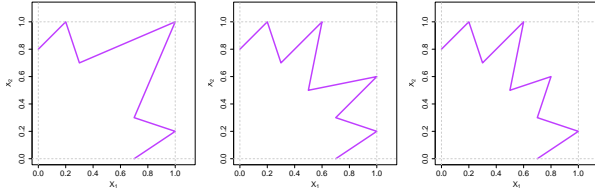
Figure 7: Examples of mixture structures in the set G. Left: mixture structure with asymptotic dependence, but the possibility of X_1 and X_2 both being large while the other is of smaller order. Centre: mixture structure with asymptotic independence, but the largest values of X_1 associated with more than one smaller order value of X_2 (and vice versa). Right: mixture structure with a sub-asymptotic component.

integral transform. This allows for greater flexibility in the possible marginal distributions, including the possibility of non-equal margins.

We have discussed the possibility of having mixture structures in the extremal dependence, and there are different ways that this could arise. The framework of [44] identifies mixture structures where X_1 and X_2 can be simultaneously large but with the possibility of having each variable large while the other is of smaller order. Such a structure is indicated by $\eta = \max_{\delta<1} \tau_1(\delta) = \max_{\delta<1} \tau_2(\delta) = 1$; examples of sets G with this feature are given in Figure 7 (left) and the asymmetric logistic copula in Figure 1. The framework of [47] also allows for a mixture structure in the extremal dependence but with the possibility of $\eta < 1$; this arises if G intersects at least one of the lines $x_1 = 1$ and $x_2 = 1$ at more than one separate location, as in Figure 7 (central). Our estimator $\hat{G}$ can capture such non-convex shapes in the asymmetric logistic case, and therefore has the potential to be further exploited to test for mixture structures. To do this, the scaling/truncation step may need to be altered to ensure that $\hat{G}$ intersects $x_1 = 1$ and $x_2 = 1$ an appropriately many times, perhaps by first testing for convexity of the radial quantile estimates over w. A further interesting mixture possibility considered by [47] is that of sub-asymptotic mixture components, as in Figure 7 (right), where for large X_1, e.g., when $x_1 = 0.7$, there is a mixture of two possible ranges of X_2 values. Here, the largest mixture component has a 'point' at $(0.8, 0.6)$ that does not reach the line $x_1 = 1$; this should also be taken into account if extending our approach as mentioned above.

The approach could be adapted to handle higher dimensional problems. An equivalent set G can be defined in higher dimensions; [34] presents theoretical results on the coefficient of asymptotic independence in any dimension; [35] provides a result equivalent to (11) for calculating the indices of [44] in higher dimensions; and in the conditional approach of [23], pairwise results can determine the relevant normalising functions. Our approach involves transforming to pseudo-polar coordinates, with only a one-dimensional angular component in the bivariate case. In a d-dimensional setting, the angular component would have dimension $d - 1$; this raises questions about how to select the angles used for local estimation to ensure reliability without a high computational cost. Moreover, the selection of the spline functions for $\hat{G}$ comes with added complexity in higher dimensions, as different choices may be needed for different subsets of the variables. Future work that extends our approach to higher dimensions, should involve developing

a strategy for selecting estimation points and spline functions in order to achieve scalable but reliable inference.

A further issue related to higher dimensional problems involves ensuring consistency across different dimensions. Suppose we have variables $\boldsymbol{X} = (X_i : i \in \mathcal{D})$, for $\mathcal{D} = \{1, \ldots, d\}$. Using results on projections of sample clouds, [35] makes links between gauge functions of lower dimensional subsets (for instance, variables indexed by $\mathcal{C} \subset \mathcal{D}$) to the gauge function associated with the full vector $\boldsymbol{X}$. If our method were to be extended to higher dimensions, estimates of extremal dependence features for the variables $\boldsymbol{X}_{\mathcal{C}} = (X_i : i \in \mathcal{C})$ could be obtained by estimating the asymptotic boundary of the scaled sample cloud for any set of variables indexed by $\mathcal{C}^*$ with $\mathcal{C} \subseteq \mathcal{C}^* \subseteq \mathcal{D}$. Ideally, we would ensure consistent estimation from these different boundary estimates.

Code: Code to run the methods presented in this paper is available online in the GitHub repository: https://github.com/essimpson/self-consistent-inference.

References

[1] Balkema, A. A. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.

[2] Balkema, A. A. and Nolde, N. (2020). Samples with a limit shape, multivariate extremes, and risk. *Advances in Applied Probability*, 52(2):491–522.

[3] Barnett, V. (1976). The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.

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

[5] Bofinger, E. and Bofinger, V. J. (1965). The correlation of maxima in samples drawn from a bivariate normal distribution. *Australian Journal of Statistics*, 7(3):57–61.

[6] Brozius, H. and de Haan, L. (1987). On limiting laws for the convex hull of a sample. *Journal of Applied Probability*, 24(4):852–862.

[7] Cai, J.-J., Einmahl, J. H. J., and de Haan, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics*, 39(3):1803–1826.

[8] Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

[9] Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

[10] Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B*, 53(2):377–392.

[11] 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 (Applied Statistics)*, 43(1):1–48.

[12] Davis, R., Mulrow, E., and Resnick, S. I. (1987). The convex hull of a random sample in $\mathbb{R}^2$. *Communications in Statistics. Stochastic Models*, 3(1):1–27.

[13] Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

[14] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B*, 52(3):393–425.

[15] de Valk, C. (2016). Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes*, 19:687–717.

[16] Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280.

[17] Eddy, W. F. and Gale, J. D. (1981). The convex hull of a spherically symmetric sample. *Advances in Applied Probability*, 13(4):751–763.

[18] Einmahl, J. H. J., de Haan, L., and Krajina, A. (2013). Estimating extreme bivariate quantile regions. *Extremes*, 16:121–145.

[19] Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):871–932.

[20] Goix, N., Sabourin, A., and Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. *Journal of Multivariate Analysis*, 161:12–31.

[21] Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

[22] Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

[23] 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.

[24] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

[25] Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

[26] Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

[27] 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.

[28] Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):521–543.

[29] Liu, Y. and Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, 127:19–35.

[30] Lugrin, T., Davison, A. C., and Tawn, J. A. (2016). Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, 19:491–515.

[31] Lugrin, T., Tawn, J. A., and Davison, A. C. (2021). Sub-asymptotic motivation for new conditional multivariate extreme models. *Stat*, 10(1):e401.

[32] Majumder, R., Shaby, B. A., Reich, B. J., and Cooley, D. (2023). Semiparametric estimation of the shape of the limiting multivariate point cloud. *arXiv preprint: arXiv:2306.13257*.

[33] Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2024). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *Extremes*, pages [doi.org/10.1007/s10687--024--00490--4](doi.org/10.1007/s10687--024--00490--4).

[34] Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

[35] Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717.

[36] Papastathopoulos, I., de Monte, L., Campbell, R., and Rue, H. (2023). Statistical inference for radially-stable generalized Pareto distributions and return level-sets in geometric extremes. *arXiv:2310.06130*.

[37] Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters*, 43(4):399–409.

[38] Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

[39] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

[40] Resnick, S. I. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, (5):303–336.

[41] Simpson, E. S. (2019). *Classifying and Exploiting Structure in Multivariate Extremes*. PhD thesis, Lancaster University.

[42] Simpson, E. S. and Tawn, J. A. (2024+a). Inference for new environmental contours using extreme value analysis. *Journal of Agricultural, Biological and Environmental Statistics (to appear)*.

[43] Simpson, E. S. and Tawn, J. A. (2024b). Supplementary material for 'Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties'.

[44] Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.

[45] Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2021). A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis*, 184:104736.

[46] Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

[47] Tendijck, S., Eastoe, E., Tawn, J. A., Randell, D., and Jonathan, P. (2021). Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, 118(542):1373–1384.

[48] Wadsworth, J. L. and Campbell, R. (2024+). Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society: Series B*, page qkae030.

[49] Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

[50] Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

[51] Youngman, B. D. (2019). Generalized additive models for exceedances of high thresholds with an application to return level estimation for U.S. wind gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.

[52] Youngman, B. D. (2020). *evgam: Generalised Additive Extreme Value Models*. R package version 0.1.4.

[53] Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.

Supplementary Material for 'Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties'

Emma S. Simpson[1] and Jonathan A. Tawn[2]

[1]Department of Statistical Science, University College London, WC1E 6BT, U.K.

[2]School of Mathematical Sciences, Lancaster University, LA1 4YF, U.K.

Abstract

The aim of this Supplementary Material is to provide additional details for the paper 'Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties'. In Section A, we provide an illustration of the theoretical links between the set G and the extremal dependence features under consideration. In Section B, we discuss the selection of tuning parameters for our limit set estimates $\hat{G}^L$ and $\hat{G}$, while Section C provides an illustration of our scaling/truncation approach and how we obtain estimates of several bivariate extremal dependence features from $\hat{G}$. In Section D, we provide additional details on the existing estimators for the extremal dependence features we consider. Most of the remaining sections provide additional simulation results for our approach: in Section E, we present estimates of $\hat{G}$ to test sensitivity to the extrapolation quantile q; in Section F, we give results on estimating $\tau_1(\delta)$ for additional copula dependence parameters; Section G covers estimation of $\lambda(\omega)$; Section H provides results on the (α_i, β_i) parameters of the conditional extremes approach; Section I provides additional simulation results for the asymmetric logistic copula; in Section J, we demonstrate the performance of our approach on samples of size $n = 100,000$; and in Section K, we test whether bootstrapping is a reasonable approach to uncertainty assessment for our estimator. A summary of our full approach to estimate G is provided in Section L.

A   Illustration of the link between extremal dependence features and the set G

In Section 3.2 of the main paper, we discussed results from [34] and [35], linking the set G to various bivariate extremal dependence features. Now, we provide further discussion and a pictorial demonstration of the theoretical results.

First, [34] links the coefficient η from [26] using result (9) in the main paper. The result can be thought of as moving the set $[1, \infty]^2$ towards the origin, along the line $x_2 = x_1$, until it just intersects the set G. We demonstrate this feature in the top left panel of Figure A for a Gaussian copula with $\rho = 0.75$. In this case, $\eta = (1 + \rho)/2 = 0.875$, corresponding to the known value of η for this model. We note that although the intersection in this example happens when $x_1 = x_2 = \eta$, this does not have to be the case, and it can occur on one of the edges of the orange rectangle with $x_1 \neq x_2$.
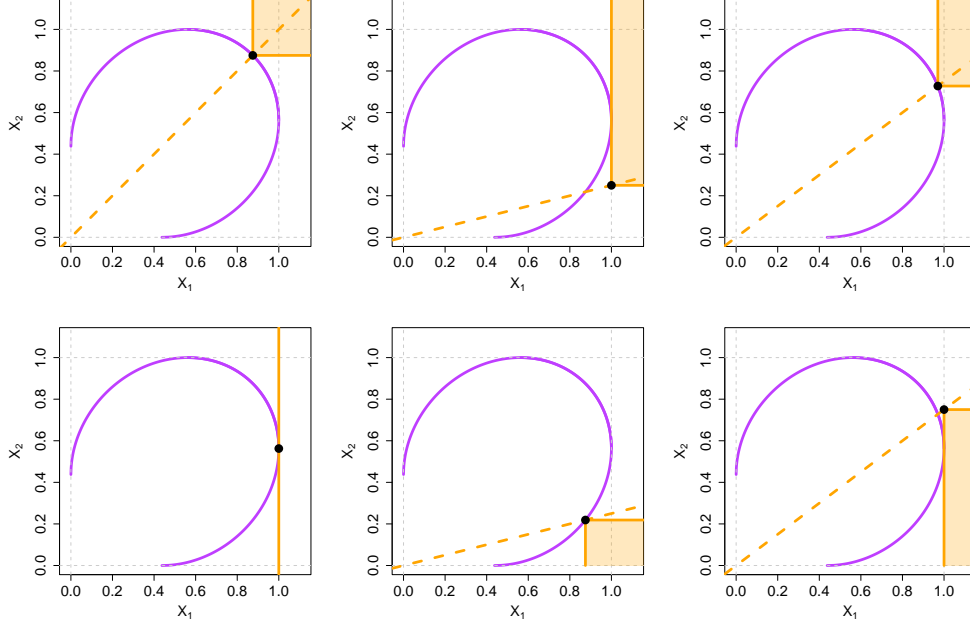
Figure A: Examples of the set G (purple) and demonstration of calculating η, $\lambda(4/5)$, $\lambda(4/7)$, $\tau_1(3/4)$, $\tau_1(1/4)$ and α_1 (clockwise from top left) for a Gaussian model with correlation parameter $\rho = 0.75$. Top left: the set $[\eta, \infty]^2$ (orange region), and the point (η, η) (black circle) which lies on $x_2 = x_1$ (orange dashed line). Top, centre and right: the sets $s_\omega S_\omega$ (orange regions) for $\omega = 4/5, 4/7$, and the points $\left(s_\omega, \frac{1-\omega}{\omega} s_\omega\right)$ (black), lying on $x_2 = \frac{1-\omega}{\omega} x_1$ (orange dashed line). Bottom, right and centre: the sets $\tau_1(\delta) S_{1,\delta}$ (orange regions), for $\delta = 3/4, 1/4$, and the points $(\tau_1(\delta), \delta\tau_1(\delta))$ (black), lying on $x_2 = \delta x_1$ (orange dashed line). Bottom left: the line $x_1 = 1$ (orange) and the point $(1, \alpha_1)$ (black).

The central and right panels of the top row of Figure A demonstrate the sets $s_\omega S_\omega$, used to calculate $\lambda(\omega)$ from [50] through equation (10). These plots are again for the Gaussian copula with $\rho = 0.75$ and with $\omega = 4/5, 4/7$. This example highlights the need to consider the whole set S_ω when deducing $\lambda(\omega)$, not just the vertex $x_2 = \frac{1-\omega}{\omega} x_1$. When $\omega = 4/5$, the intersection of G and $s_\omega S_\omega$ occurs on the edge of the rectangle, with $s_\omega = 1$ and $\lambda(\omega) = \max(\omega, 1 - \omega) = 4/5$. In contrast, for $\omega = 4/7$, the intersection does occur on the line $x_2 = \frac{1-\omega}{\omega} x_1$, with the value of $\lambda(4/7)$ given by the first Gaussian $\lambda(\omega)$ case in Table 1 of the main paper.

In a similar way to the η representation of [34], we can think of $\tau_1(\delta)$ as taking the set $S_{1,\delta}$ and moving it along the line $x_2 = \delta x_1$ towards the origin until it intersects the set G (where, again, the intersection does not necessarily occur on the line $x_2 = \delta x_1$). This is demonstrated in the right and central panels of the bottom row of Figure A for the same Gaussian model, taking $\delta = 0.75, 0.25$, respectively. We have $\tau_1(0.75) = 1$, and $\tau_1(0.25) < 1$, with its exact form given in Table 1.

Finally, we demonstrate the calculation of α_1 for our Gaussian example in the bottom left panel of Figure A. Here, $\alpha_1 = \rho^2 = 0.75^2$.
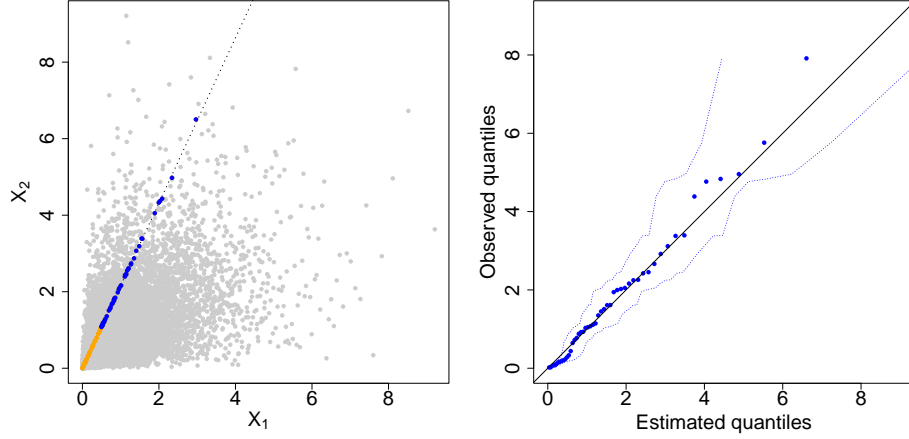
Figure B: Left: a sample of size 10,000 from an inverted logistic distribution with $\gamma = 0.5$ (grey) and an example of a set $\mathcal{R}_{w_j^*}$, with $|\mathcal{R}_{w_j^*}| = m = 100$ (orange: points in $\mathcal{R}_{w_j^*}$ below the empirical 0.5 quantile; blue: points in $\mathcal{R}_{w_j^*}$ above the empirical 0.5 quantile). Right: QQ-plot for the fitted GPD model for exceedances of $\mathcal{R}_{w_j^*}$ above the 0.5 quantile. The blue dashed lines represent approximate 95% confidence intervals for the pointwise empirical quantiles.

B   Tuning parameter selection

Our proposed estimation procedure for the set G, presented in Section 4 of the main paper, involves the choice of several tuning parameters. For $\hat{G}^L$ in Section 4.3, the first step is to choose the value of k and the angular values w_j^*, for $j \in J_k$. The discussion in Section 2 of the main paper highlighted that whether or not the point $(1,1)$ belongs to the set G can play a crucial role in distinguishing between asymptotic dependence and asymptotic independence. As a result, the value of G at $w = 1/2$ is of importance, so $1/2$ should be taken to be one of the estimation angles w_j^*. We only take w_j^* values within the range of observed angles, i.e., setting $w^m = \min_{i=1,\dots,n} w_i$ and $w^M = \max_{i=1,\dots,n} w_i$, we have $w_j^* \in [w^m, w^M]$ for all $j \in J_k$. Natural options for choosing the w_j^* values include splitting the interval $[w^m, w^M]$ into equal segments or using empirical quantiles of $\{w_1, \dots, w_n\}$. The former has the drawback that several w_j^* values could be selected without any nearby values in $\{w_1, \dots, w_n\}$. In such cases, the neighbourhood in (18) would need to be relatively large to allow for inference, but this would come at the cost of obtaining reliable estimates of $r_q(w)$, as that may change considerably over this range of w. We therefore prefer the latter approach, with w_j^* as the $(j-1)/(k-1)$th empirical quantile of $\{w_1, \dots, w_n\}$, for $j = 1, \dots, k-1$, and $w_k^* = 1/2$. We take k to be odd so that if the angular distribution is symmetric, we can have an equal number of w_j^* values either side of $1/2$. Our studies suggest that the choice of k is not particularly crucial to the performance of the estimator, and we set $k = 199$ as a default.

Next consider the sizes of each radial neighbourhood $\epsilon_{w_j^*}$ ($j \in J_k$). We allow these to depend on the angle w_j^*, in order to control the size of each set $\mathcal{R}_{w_j^*}$. In particular, we fix each $\epsilon_{w_j^*}$ such that $|\mathcal{R}_{w_j^*}| = m$, for some choice of m. That is, $\mathcal{R}_{w_j^*}$ corresponds to the radial values of the m nearest angular neighbours of w_j^*. The GPD threshold $u_{w_j^*}$, or equivalently the value $q_u \in [0, q]$, also needs selecting. Figure B demonstrates an example of a single set $\mathcal{R}_{w_j^*}$ for the same data as in Figure 2. The QQ-plot in Figure B

31

(right) suggests that taking $u_{w_j^*}$ as the empirical median of $\mathcal{R}_{w_j^*}$ (i.e., $q_u = 0.5$) is a reasonable choice, and we have observed the same behaviour over a range of different examples during our investigations. For simplicity, we therefore fix $q_u = 0.5$ in the paper. The final tuning parameter to select for the local approach is the value of q in (17). We have found values of q between 0.9 and 0.9999 to work well in practice, and take $q = 0.999$ subsequently. Sensitivity to this parameter is investigated in Section E, where the method appears robust to a reasonable range of choices of q.

When using the GAM framework to produce the smoothed estimates proposed in Section 4.4 of the main paper, we must select the number of knots (κ) and degree of the corresponding spline functions. For the reasons we discussed earlier, $w = 1/2$ is taken as one of the knots, which also suggests that an odd value of κ is reasonable. In practice, we suggest spacing the κ interior spline knots evenly throughout $[w^m, w^M]$, then adjusting the central one to be exactly $1/2$. Any required exterior knots are fixed at 0 and 1. We have found $\kappa = 7$ to be a reasonable choice in practice, illustrating limited sensitivity to this tuning parameter in Section I.

C Illustrations of our scaling/truncation and parameter estimation approaches

The scaling/truncation approach discussed in Section 4.3 of the main paper is demonstrated in Figure C. This allows for estimation of a finite set of points belonging to the set $G \subset [0,1]^2$, with our final estimate given by $\hat{G}^L = \left\{ (\hat{x}_{1,j}^G, \hat{x}_{2,j}^G) : j \in J_k \right\}$. In Figure D, we provide an illustration of our estimation procedure for η, s_ω (for $\lambda(\omega)$), $\tau_1(\delta)$ and α_1. This accompanies the discussion in Section 5.2 of the main paper.
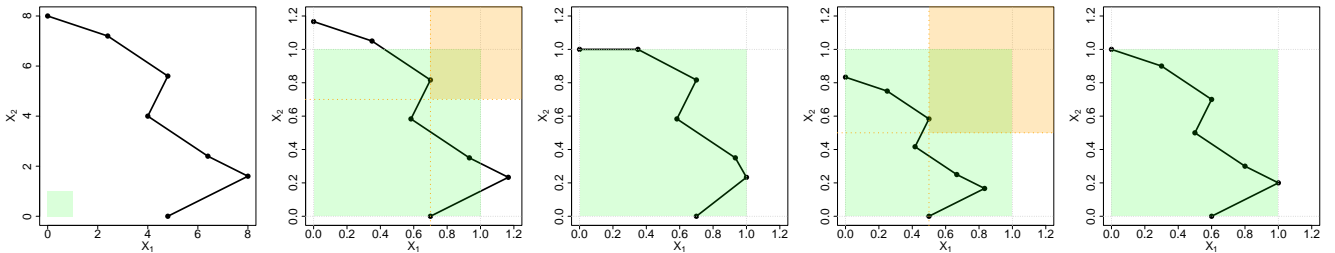

Figure C: An illustration of our scaling approach. From left to right, the first plot shows an example set of pre-scaled points $(\tilde{x}_{1,j}, \tilde{x}_{2,j})$, $j \in J_k$ (black dots), with $k = 7$ for illustrative purposes. The second plot demonstrates step 1 with $\hat{\eta}_H = 0.7$ including the set $[0.7, \infty]^2$ (orange region) and $\hat{G}_\eta^L$ (black points) and the third plot shows the corresponding final set $\hat{G}^L$ with points outside $[0,1]^2$ transformed onto the boundary (step 2a). The fourth plot is equivalent to the second but with $\hat{\eta}_H = 0.5$, and the final plot shows the corresponding final set $\hat{G}^L$ with the final transformation ensuring intersection with $x_1 = 1$ and $x_2 = 1$ (step 2b). The green region in each plot shows the set $[0,1]^2$.
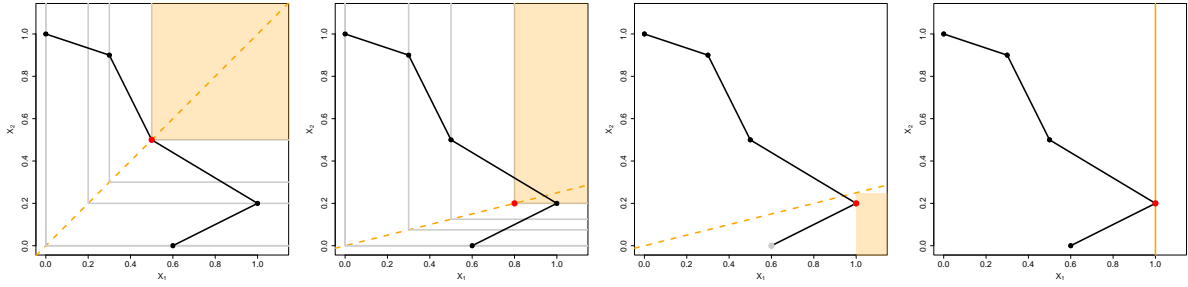
Figure D: A demonstration of how to obtain estimates of η, $s_{0.8}$ (for subsequent estimation of $\lambda(0.8)$), $\tau_1(0.25)$ and α_1 from the estimate of G. Here, $\hat{G}$ uses $k = 5$ for illustrative purposes, but in practice k should be much larger. The black points in each plot demonstrate an estimated set $\hat{G}$. Left: the boundaries of the candidate sets $[s_j^*, \infty]^2$ for each point in $\hat{G}$ (grey); the set $[\hat{\eta}_G, \infty]^2$ (orange region); and the point $(\hat{\eta}_G, \hat{\eta}_G)$ (red) which satisfies $x_1 = x_2$ (orange dashed line). Left-centre: the line $x_2 = \frac{1-0.8}{0.8}x_1 = 0.25x_1$ (orange dashed line); the boundaries of the candidate sets $[s_{0.8,j}^*, \infty] \times [0.25s_{0.8,j}^*, \infty]$ for each point in $\hat{G}$ (grey); the set $[\hat{s}_{0.8}, \infty] \times [0.25\hat{s}_{0.8}, \infty]$ (orange region); the point of intersection of the orange region with $\hat{G}$ (red). Right-centre: the line $x_2 = 0.25x_1$ (orange dashed line); the set $(\tau_1(0.25), \infty] \times [0, 0.25\tau_1(0.25)]$ (orange region); the point of intersection of the orange region with $\hat{G}$ (red) and the other candidate for this intersection point (grey). Right: the line $x_1 = 1$ (orange) and its intersection with the set $\hat{G}$ (red).

D   Existing estimation techniques

In this section, we present some existing estimators for the extremal dependence properties in Section 3.1 of the main paper. We use these for comparison to our proposed estimators in Section 6. [26] estimate η using a maximum-likelihood based approach, which results in a Hill-type estimator [24], but on exponential margins. Setting $M = \min(X_1, X_2)$, and assuming we have $n_{u,H}$ observations of M above some high threshold u_H, denoted $m_1^*, \ldots, m_{n_{u,H}}^*$, this estimator is

$$\tilde{\eta}_H = \frac{1}{n_{u,H}} \sum_{i=1}^{n_{u,H}} (m_i^* - u_H).$$

Alternative estimation procedures for η include the non-parametric approaches of [37] and [16]. These aim to avoid inconsistencies that arise in carrying out estimation using different marginal distributions, which can result in issues when quantifying uncertainty. Suppose we have n pairs of observations $(x_{1,1}, x_{1,2}), \ldots, (x_{n,1}, x_{n,2})$, and denote by $x_i^{(j)}$ the jth largest value of component $i \in \{1, 2\}$, i.e., $x_i^{(n)} \leq x_i^{(n-1)} \leq \cdots \leq x_i^{(1)}$. In both approaches, for each $j \in \{1, \ldots, n\}$, the first step is to define the quantity $s_n(j) = \sum_{\ell=1}^n \mathbf{1}\left\{x_{\ell,1} \geq x_1^{(j)}, x_{\ell,2} \geq x_2^{(j)}\right\}$, with $\mathbf{1}(E)$ the indicator of event E. The estimators $\tilde{\eta}_P$ of [37] and $\tilde{\eta}_D$ of [16] are

$$\tilde{\eta}_P = \frac{\log 2}{\log\{s_n(2c)\} - \log\{s_n(c)\}}, \qquad \tilde{\eta}_D = \frac{\sum_{j=1}^c s_n(j)}{cs_n(c) - \sum_{j=1}^c s_n(j)},$$

where c relates to the number of exceedances above a high threshold. Further details on these estimators can be found in [4]. Although $\eta \in (0, 1]$, all three of these estimators can result in values greater than

one, so the truncated estimators $\hat{\eta}_H = \min(\tilde{\eta}_H, 1)$, $\hat{\eta}_P = \min(\tilde{\eta}_P, 1)$ and $\hat{\eta}_D = \min(\tilde{\eta}_D, 1)$ are preferable. Our estimator from Section 5.2 of the main paper guarantees that $\hat{\eta}_G \in (0, 1]$ due to the scaling procedure proposed in Section 4.3.

[50] also discuss the use of a Hill-type estimator for $\lambda(\omega)$, analogous to $\hat{\eta}_H$, but taking $M_\omega = \min\{X_1/\omega, X_2/(1-\omega)\}$, for $\omega \in [0, 1]$. If we have $n_{u,\omega}$ observations of M_ω above a high threshold u_ω, denoted $m^*_{\omega,1}, \ldots, m^*_{\omega,n_{u,\omega}}$, the estimator is

$$\tilde{\lambda}_H(\omega) = \left\{ \frac{1}{n_{u,\omega}} \sum_{i=1}^{n_{u,\omega}} \left(m^*_{\omega,i} - u_\omega \right) \right\}^{-1},$$

i.e., the reciprocal of the estimated mean excess of M_ω above the threshold u_ω. Again, the truncated estimator $\hat{\lambda}_H(\omega) = \min\left(\tilde{\lambda}_H(\omega), 1 \right)$ is preferred. Similarly, [44] propose a Hill-type estimator for the coefficients $\tau_1(\delta)$, $\delta \in (0, 1]$. The idea is to slightly alter representation (7), and instead consider $\Pr(X_1 > x, X_2 \leq \delta X_1)$, for which estimation is simpler; [35] prove this has index of regular variation $\tau_1(\delta)$. Let $M_\delta = \{X_1 > 0 : X_2 \leq \delta X_1\}$. Suppose we have $n_\delta \leq n$ observations of X_1 where $X_2 \leq \delta X_1$, and that $n_{u,\delta}$ of these, denoted $m^*_{\delta,1}, \ldots, m^*_{\delta,n_{u,\delta}}$, are above some high threshold u_δ. Then,

$$\tilde{\tau}_{H,1}(\delta) = \frac{1}{n_{u,\delta}} \sum_{i=1}^{n_{u,\delta}} \left(m^*_{\delta,i} - u_\delta \right).$$

Again, we could have $\tilde{\tau}_{H,1}(\delta) > 1$ here, so we truncate to give $\hat{\tau}_{H,1}(\delta) = \min\{\tilde{\tau}_{H,1}(\delta), 1\}$.

E   Sensitivity to q

Our method involves the selection of several tuning parameters. Intuitively, one of the most important is the value of q, representing the quantile of interest for the radial distributions before scaling onto $[0, 1]^2$. For all simulations and applications presented in the paper, we set $q = 0.999$, but we now present estimates of the set G for alternative quantile choices. All other tuning parameters are fixed as in Section 6.1 of the main paper. As in Figure 3 of the paper, we present results for datasets of size 10,000 simulated from Gaussian, inverted logistic and logistic models, repeating each test 1000 times. The results are shown in Figure E for $q = 0.99$ and $q = 0.9999$. Here, we choose only values of q close to one in order to appropriately estimate the *limiting* shape of the scaled sample cloud; taking q to be smaller than this risks estimating a sub-asymptotic level set that is different in shape to the limit, which we explore later in this section. Overall, there is very little disparity between the estimates with q set as 0.99, 0.999 or 0.9999, since differences in the $r_q(w)$ estimates are offset by the subsequent scaling/truncation steps. There is some evidence that the estimates for the logistic model deteriorate as we increase q, with the least desirable quadratic spline option being chosen slightly more often for $q = 0.9999$ than $q = 0.99$, although the linear splines are still chosen in around 78% of cases. Hence, the estimates are overall not particularly sensitive to the choice of q, suggesting setting $q = 0.999$ is a reasonable choice.

To highlight the need to take q close to one, we now consider the logistic case in further detail. In Figure F, we present results of a similar simulation study, but where the radial quantile level is set to
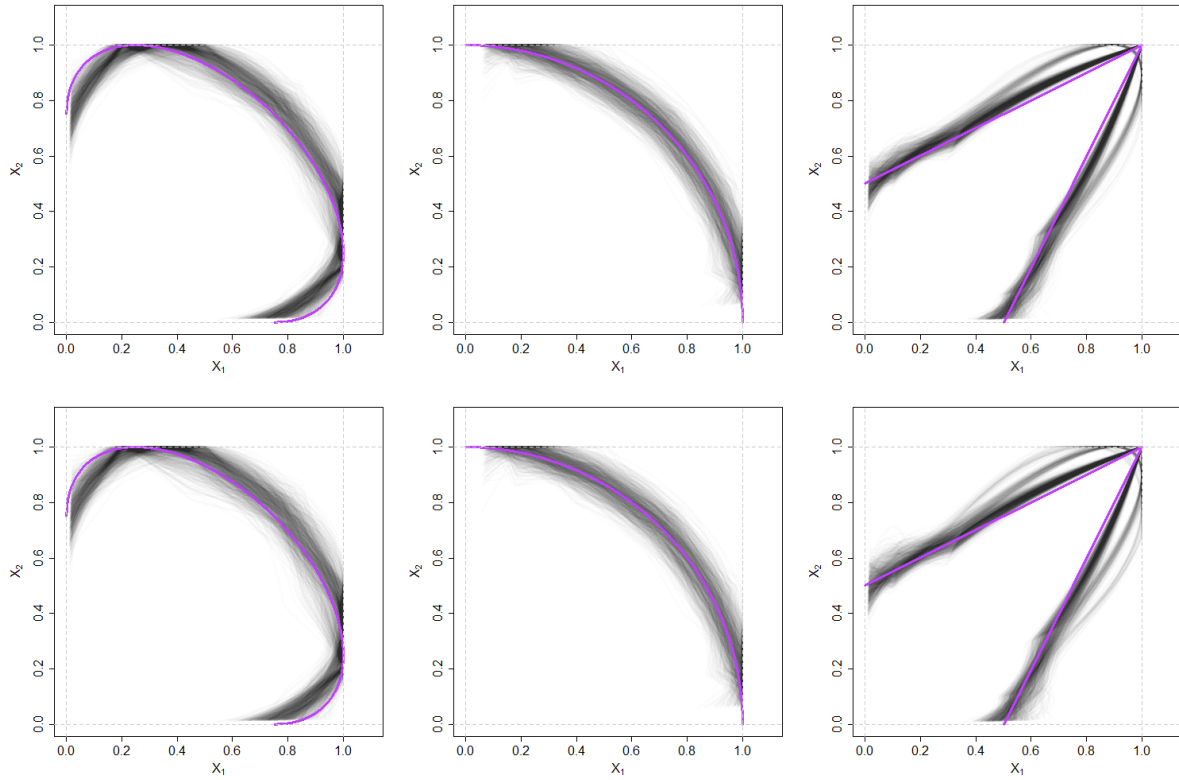
Figure E: Estimates of the sets G for the Gaussian (left), inverted logistic (centre) and logistic (right) copulas, with the corresponding ρ or γ parameters set to 0.5. The sample size in each case is $n = 10,000$ and 1000 estimates are shown for each model (grey). The true sets G are shown in purple. The radial extrapolation quantile is set as $q = 0.99$ (top row) and $q = 0.9999$ (bottom row).
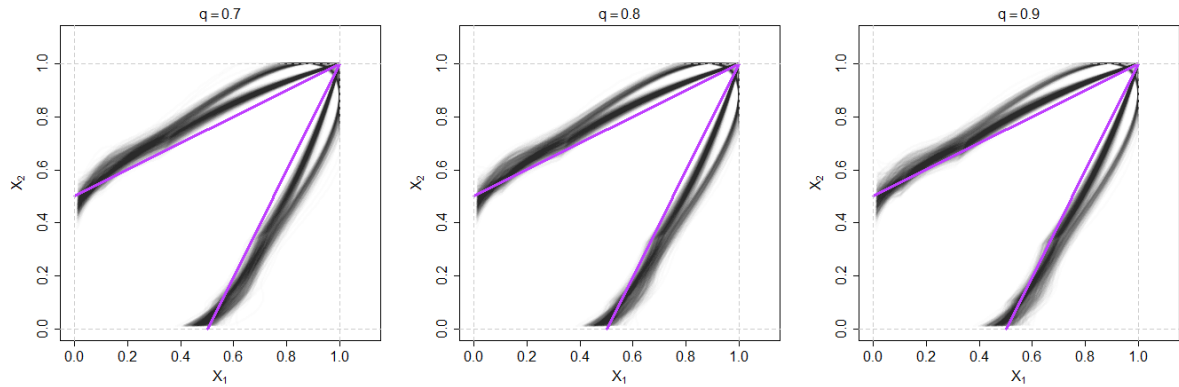

Figure F: Estimates of the sets G for the logistic copula with $\gamma = 0.5$. The sample size in each case is $n = 10,000$ and 1000 estimates are shown for each model (grey). The true sets G are shown in purple. The radial extrapolation quantile is set as $q = 0.7$ (left), $q = 0.8$ (centre) and $q = 0.9$ (right).

$q = 0.7$, 0.8 or 0.9. The percentage of times that the most desirable, linear splines are selected in these cases is 65%, 68% and 76%, respectively, so we are able to correctly identify asymptotic dependence more often as we increase the value of q. This happens since for this copula, the pointed shape we see under asymptotic dependence is a limiting feature, and true level sets at lower quantiles have a smoother, curved shape. Setting q close to one is therefore necessary to improve the estimation of the limiting shape of the scaled sample cloud, G.

F Additional estimates of $\tau_1(\delta)$

In Figure 5 of the main paper, we presented results on estimation of the parameter $\tau_1(\delta)$ [44] for Gaussian, inverted logistic and logistic models with parameters ρ or γ set to 0.5. In Figures G and H, respectively, we now present equivalent results for $\rho, \gamma = 0.25$ and $\rho, \gamma = 0.75$, comparing our new estimator to the Hill-type estimator $\hat{\tau}_{H,1}(\delta)$. We observe similar behaviour to that in the results of the paper. In particular, for smaller values of δ, our $\hat{\tau}_{G,1}(\delta)$ provides improvements in variance over the $\hat{\tau}_{H,1}(\delta)$ estimates. In terms of bias, our results are most successful in the inverted logistic cases and for smaller ρ and γ parameters. For $\rho = 0.75$, we tend to underestimate $\tau_1(\delta)$ for $\delta < 0.5$ in the Gaussian case, which is not a problem for the Hill-type estimator, while for $\gamma = 0.75$ both approaches yield overestimation in the logistic setting.
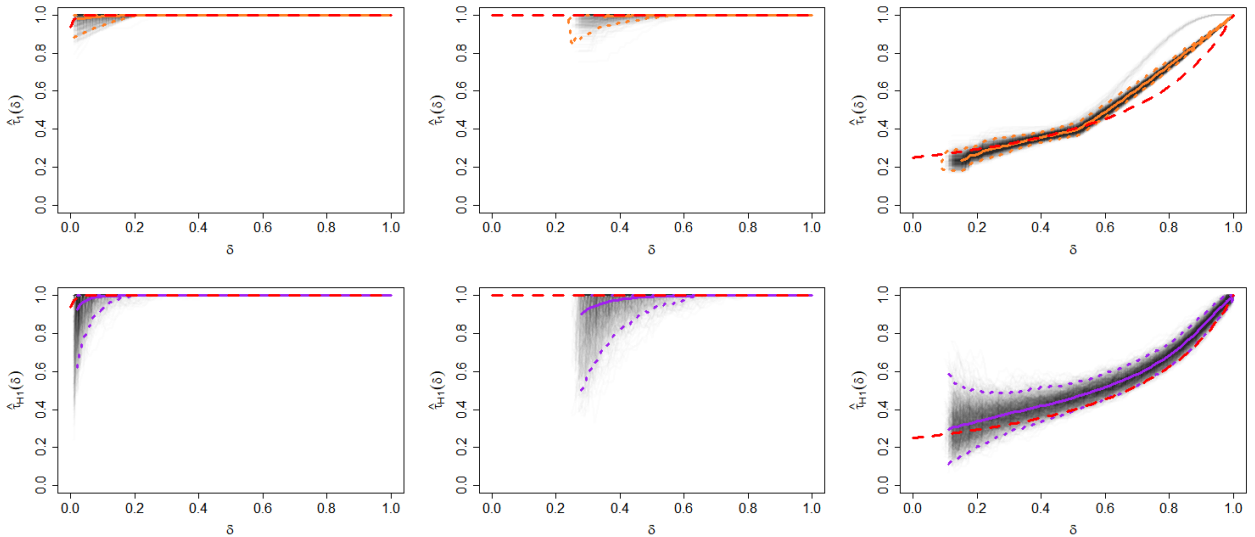

Figure G: Estimates of $\tau_1(\delta)$, $\delta \in [0,1]$, for data simulated from Gaussian (left), inverted logistic (centre) and logistic (right) models with their corresponding parameters set as 0.25. Each grey line represents an estimate obtained using $\hat{\tau}_{G,1}(\delta)$ (top) or $\hat{\tau}_{H,1}(\delta)$ (bottom) over 1000 simulations. The solid and lower/upper dotted lines show the pointwise bootstrapped means and 0.025 and 0.975 quantiles, respectively, for $\hat{\tau}_{G,1}(\delta)$ (orange) and $\hat{\tau}_{H,1}(\delta)$ (purple). The true $\tau_1(\delta)$ values are shown in red.
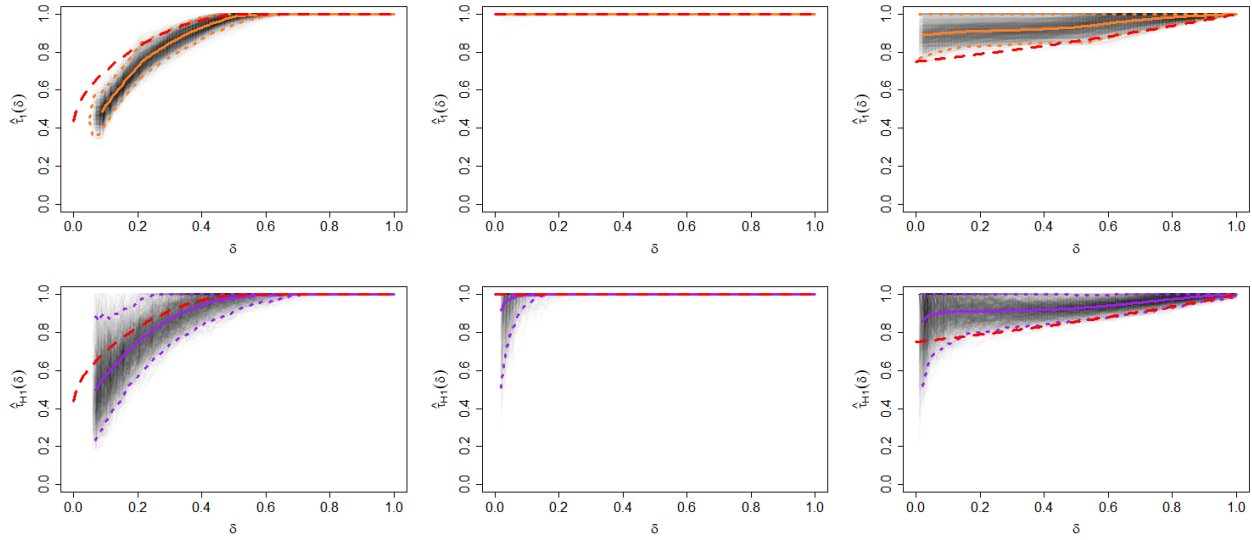
Figure H: Plots equivalent to Figure G, but with the copula parameters set to 0.75.
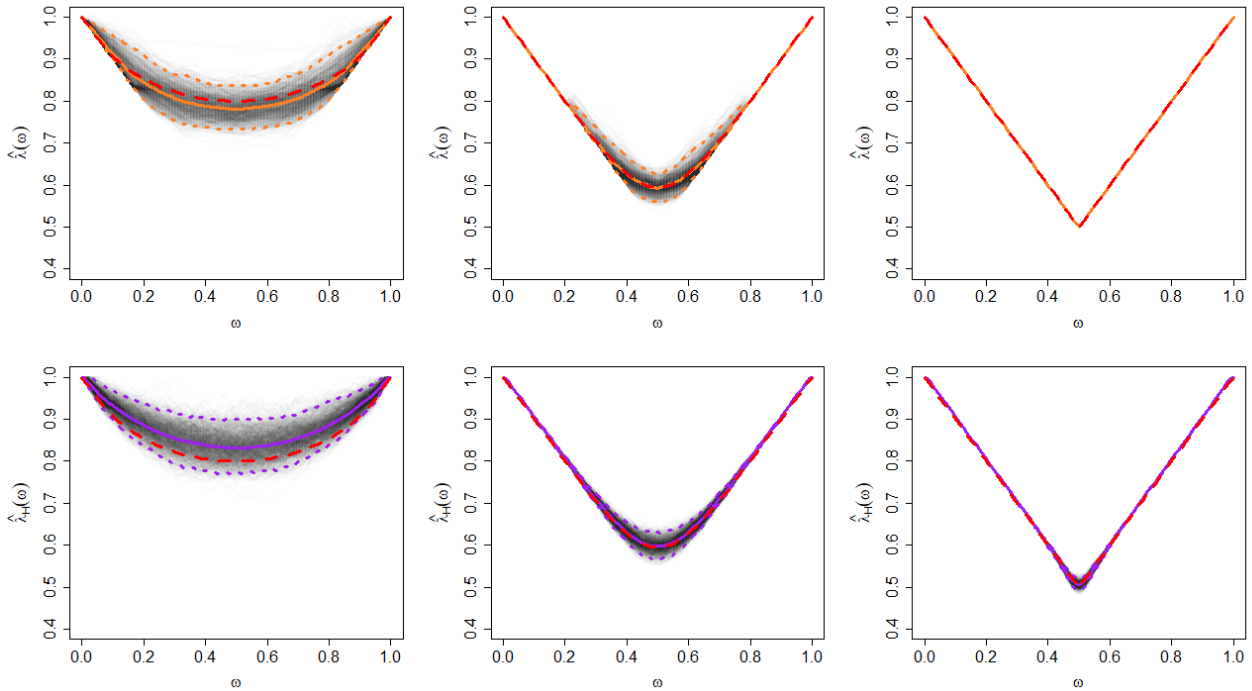

Figure I: Estimates of $\lambda(\omega)$, $\omega \in [0, 1]$, for data simulated from Gaussian (left), inverted logistic (centre) and logistic (right) models with their corresponding parameters set as 0.25. Each grey line represents an estimate obtained using $\hat{\lambda}_G(\omega)$ (top) or $\hat{\lambda}_H(\omega)$ (bottom) over 1000 simulations. The solid and lower/upper dotted lines show the pointwise bootstrapped means and 0.025 and 0.975 quantiles, respectively, for $\hat{\lambda}_G(\omega)$ (orange) and $\hat{\lambda}_H(\omega)$ (purple). The true $\lambda(\omega)$ values are shown in red.

G   Estimates of $\lambda(\omega)$

We now present results on the estimation of $\lambda(\omega)$. The simulation results presented in Figures I, J and K are analogous to those presented in Figures G, 5 (in the main paper) and H in the $\tau_1(\delta)$ case, respectively, i.e., for Gaussian, inverted logistic and logistic data with $\rho, \gamma \in \{0.25, 0.5, 0.75\}$. For the estimator $\hat{\lambda}_H(\omega)$, we set the threshold u_ω as the observed 0.95 quantile of the structure variable M_ω in each case. For the Gaussian examples, $\hat{\lambda}_H(\omega)$ has a tendency towards overestimation; this bias is significantly reduced by our proposed estimator. In the case of the inverted logistic data, the two approaches give comparable, unbiased results with similar variability. Finally, for the logistic examples, with $\gamma = 0.25, 0.50$, our approach is particularly successful, giving almost perfect estimates of $\lambda(\omega)$ across the full range of $\omega \in [0, 1]$. For the logistic model with $\gamma = 0.75$, we also obtain strong results that are an improvement over the Hill-type estimator.
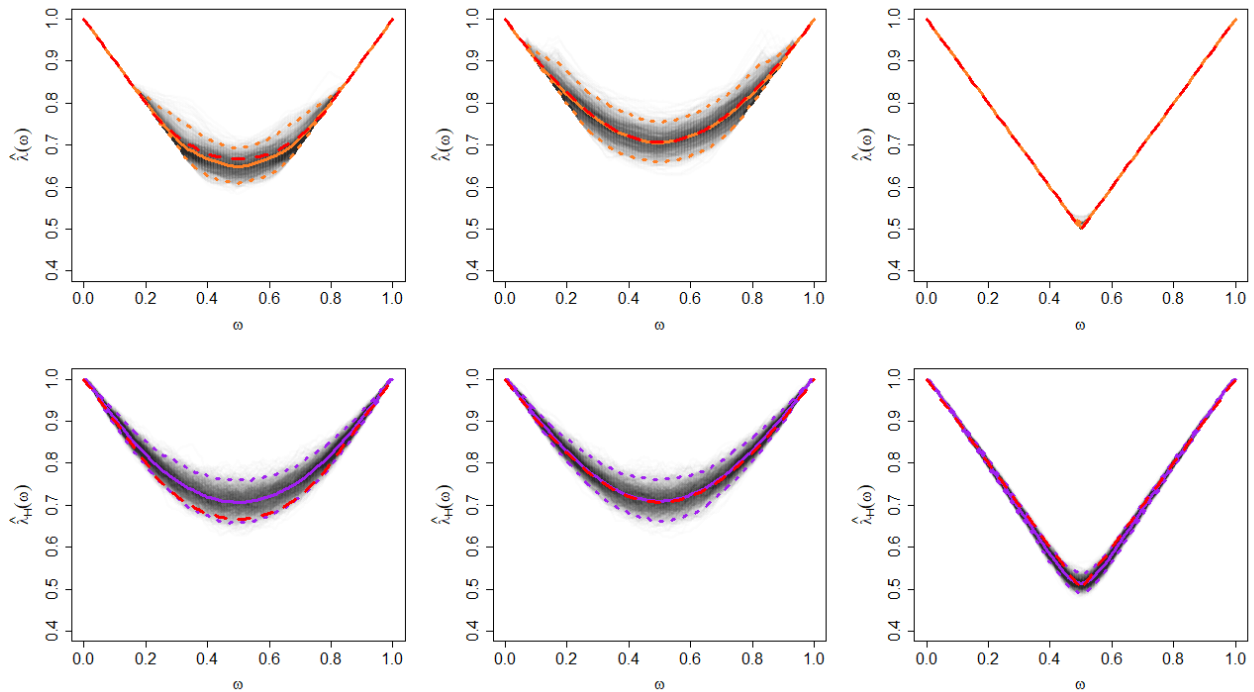

Figure J: Plots equivalent to Figure I, but with the copula parameters set to 0.5.

H   Estimates of (α_1, β_1)

In Figure L, we present results on the estimation of α_1 and β_1 from the conditional extremes model of [23]. These are compared to results obtained using maximum likelihood estimators for model (23) for both parameters simultaneously. Our results are reasonably similar to those obtained using the maximum likelihood estimators, but slightly worse for inverted logistic data, although we do offer some improvement in the logistic case with $\gamma = 0.25, 0.5$. Most noticeably, neither approach is able to provide very successful estimation of the β_1 parameter, but this is known to be difficult to estimate.
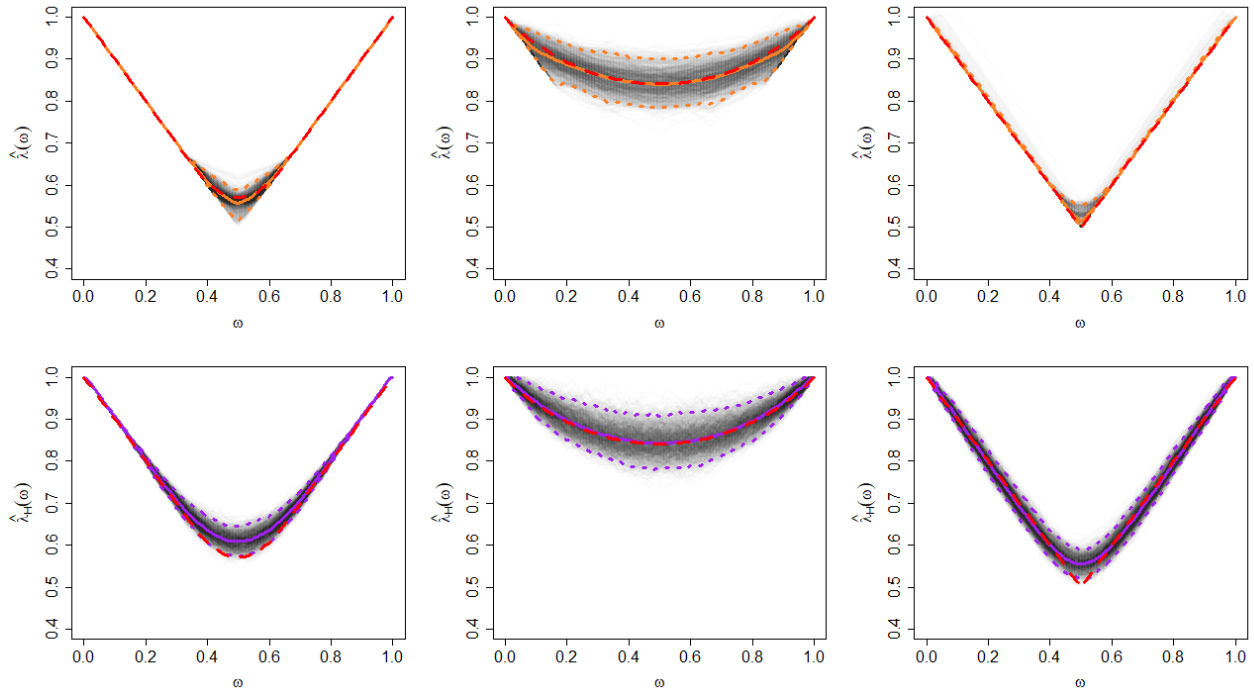
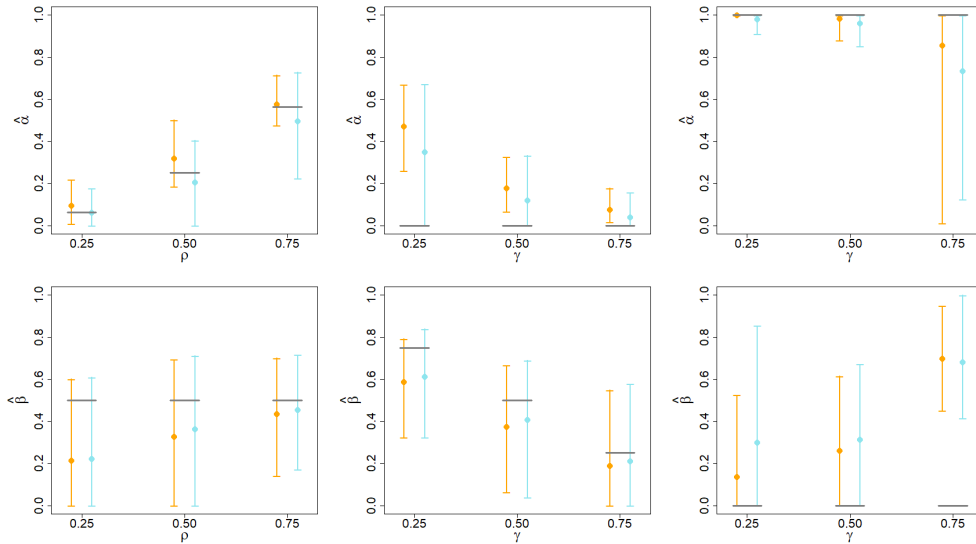Figure K: Plots equivalent to Figure I, but with the copula parameters set to 0.75.

Figure L: Estimates of α_1 (top row) β_1 (bottom row) for data simulated from Gaussian (left), inverted logistic (centre) and logistic (right) models with their corresponding parameters taken as 0.25, 0.5 and 0.75. We show the mean estimates of β_1 (circles) and 0.025 to 0.975 quantile range over 1000 simulations for $\hat{\beta}_{G,1}$ (orange) and maximum likelihood estimates of β_1. The true β_1 values are shown in grey.

I Asymmetric logistic example

The three copulas we considered in Section 6.1 of the main paper exhibit either asymptotic independence, or asymptotic dependence where both variables can only take their largest values simultaneously. We now demonstrate our approach for the asymmetric logistic model [46], which can exhibit a mixture structure in its extremal dependence features, as discussed in Section 3.2.

We first provide more detail on the form of this model. In exponential margins, the bivariate extreme value distribution function has the form

$$F(x_1, x_2) = \exp\left[-V\left\{\frac{-1}{\log(1 - e^{-x_1})}, \frac{-1}{\log(1 - e^{-x_2})}\right\}\right]$$

for $x_1, x_2 \geq 0$ and an exponent measure $V(x, y)$ satisfying certain conditions. For the asymmetric logistic model,

$$V(x, y) = \frac{\theta_1}{x} + \frac{\theta_2}{y} + \left\{\left(\frac{1 - \theta_1}{x}\right)^{1/\gamma} + \left(\frac{1 - \theta_2}{y}\right)^{1/\gamma}\right\}^{\gamma},$$

with $x, y > 0$, $\gamma \in (0, 1]$ and $\theta_1, \theta_2 \in [0, 1]$. For this model, we have

$$\chi = 2 - \theta_1 - \theta_2 - \left\{(1 - \theta_1)^{1/\gamma} + (1 - \theta_2)^{1/\gamma}\right\}^{\gamma}.$$

For a fixed $\gamma < 1$, the parameters (θ_1, θ_2) control the extremal dependence structure of the variables (X_1, X_2) and the strength of asymptotic dependence. If $\theta_1 > 0$, it is possible for X_1 to be large while X_2 is small, and an analogous result holds for $\theta_2 > 0$. If $\max(\theta_1, \theta_2) < 1$, then both variables can take their largest values simultaneously. In the case where $\theta_1, \theta_2 \in (0, 1)$, the asymmetric logistic model can be thought of as a mixture of a logistic model with independence. Asymptotically, proportions $\theta_1/2, \theta_2/2$ and $1 - (\theta_1 + \theta_2)/2$ of the extremal mass are associated with the cases where X_1 is large while X_2 is small, X_2 is large while X_1 is small, and (X_1, X_2) are simultaneously large, respectively. The gauge function of this model does not depend on (θ_1, θ_2), and takes the form (6) in the main paper.

We apply our estimation procedure to three samples of size 10,000 from the asymmetric logistic model with different values of θ_1 and θ_2 (which we take to be equal). If $\theta_1 = \theta_2 = \theta$, we have $\chi = (2 - 2^{\gamma})(1 - \theta)$, i.e., the strength of asymptotic dependence decreases as the value of θ increases. We fix the tuning parameters as in Section 6.1, and the resulting $\hat{G}$ estimates are shown by the red dashed lines in Figure M. This estimated boundary set is reasonably close to the truth for $\theta = 0.25$ and $\theta = 0.5$, where a respective 75% and 50% of the extremal mass is associated with both variables being simultaneously large [44]. When $\theta = 0.75$, so the proportion of mass associated with the simultaneously extreme variables is reduced to 25%, it becomes more difficult for $\hat{G}$ to detect this feature.

We also carry out our estimation procedure using different numbers of knots κ in the spline functions of the GAMs, as here G is not smooth everywhere and increasing the number of knots may enable us to better capture this feature. Results with $\kappa \in \{5, 9, 11\}$ are shown in Figure M. The results for $\kappa = 5$ are indeed slightly worse than for $\kappa \in \{7, 9, 11\}$, but otherwise our estimates of G are not particularly sensitive to this tuning parameter. An issue highlighted in Figure M is that our approach to scaling the boundary estimate onto $[0, 1]^2$ means we are unlikely to have our estimate of G intersecting each of the
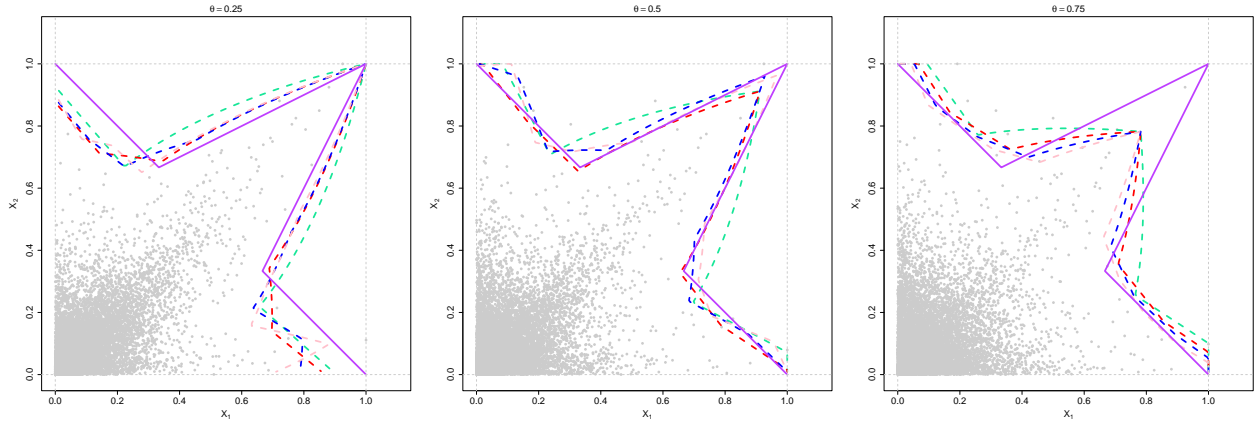
Figure M: An example of our estimation procedure for the asymmetric logistic model with $\gamma = 0.5$ and $\theta \in \{0.25, 0.5, 0.75\}$. The true set G is demonstrated by the purple lines; the simulated data used to obtain the estimates of G are shown in grey; the dashed lines show our estimates of G for different numbers of knots in the spline functions ($\kappa = 5$: green; $\kappa = 7$: red; $\kappa = 9$: blue; $\kappa = 11$: pink).

lines $x_1 = 1$ and $x_2 = 1$ in more than one region. As a result, it is difficult for our approach to detect extremal dependence mixture structures. We expand further on this point in our discussion in Section 7.3 of the main paper.

In Figure N, we present additional results over 1000 repetitions for the default choice of $\kappa = 7$, with the remainder of the tuning parameters fixed as in Section 6.1. The estimates are similar to those given in Figure M, with a similar amount of variability as in the Gaussian, inverted logistic and logistic cases of Section 6.1. In particular, we see that when $\theta = 0.75$, it remains difficult to completely capture the larger mixture component when X_1 is large for each sample replicate.
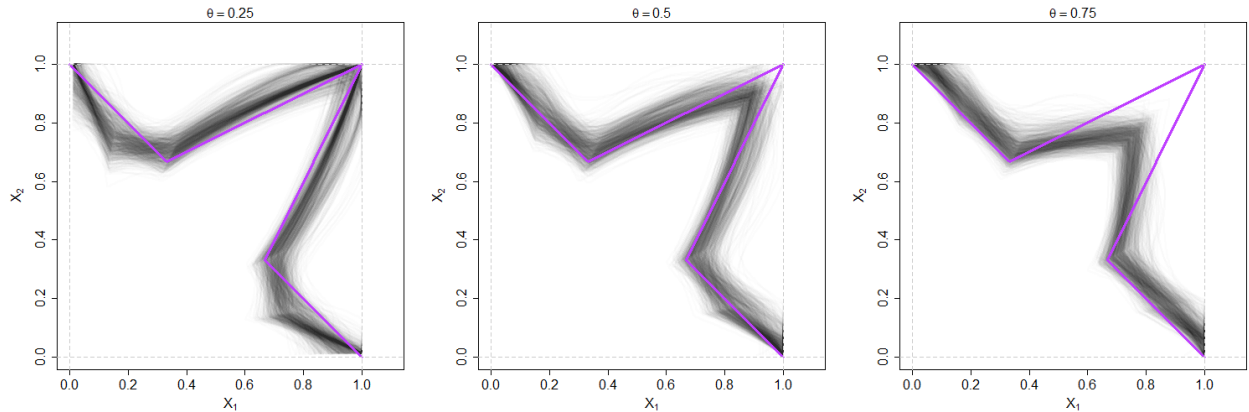

Figure N: Estimates $\hat{G}$ for the asymmetric logistic copula with $\gamma = 0.5$ and $\theta \in \{0.25, 0.5, 0.75\}$ (left to right). The sample size in each case is $n = 10,000$ and 1000 estimates are shown for each model (grey). The true sets G are shown in purple.

J   Results for a larger sample size

In the simulation studies of the main paper, and thus far in this Supplementary Material, we have considered only samples of size $n = 10,000$, both for consistency and to ensure the sample sizes were of a similar order to that of the wave height application in Section 6.2 of the main paper. We now provide results for simulated datasets of size $n = 100,000$ to check the reliability of the procedure for different sample sizes. Estimates of the set G over 100 iterations for the three copula models studied in Section 6.1 of the main paper are shown in Figure O; all tuning parameters are fixed as in Section 6.1. The spline degrees chosen by our algorithm are summarised in Table 3.

The results in Figure O and Table 3 show that the tuning parameters we have selected give reasonable results when the sample size is increased to $100,000$. For all three copula types, the variance of our estimates has decreased compared to the results for samples of size $10,000$. When considering the degree of the splines chosen in the GAM component of the method, it is noticeable that the most appropriate linear splines are chosen more often for the logistic model when the sample size is increased. It is also reassuring that non-linear options are chosen more often for the Gaussian model as the sample size is increased, although the effect of sample size on the choice of spline degree appears negligible in the inverted logistic case.
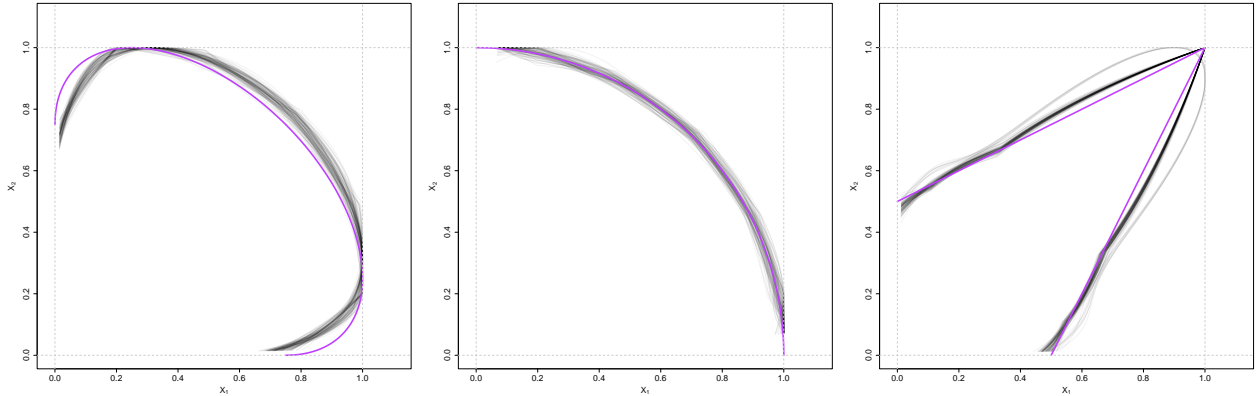

Figure O: Estimates $\hat{G}$ for the Gaussian (left), inverted logistic (centre) and logistic (right) models, with the corresponding ρ or γ parameters set to 0.5. The sample size in each case is $n = 100,000$ and 100 estimates are shown for each model (grey). The true sets G are shown in purple.

Spline degree	Gaussian	Inverted logistic	Logistic
1	22	36	91
2	41	42	0
3	37	22	9

Table 3: The number of times each spline degree is chosen for the estimated sets $\hat{G}$ shown in Figure O.

There remains some bias in the Gaussian examples in Figure O. This raises the question of whether additional spline knots are required as the sample size increases. In Figure P, we provide further estimates
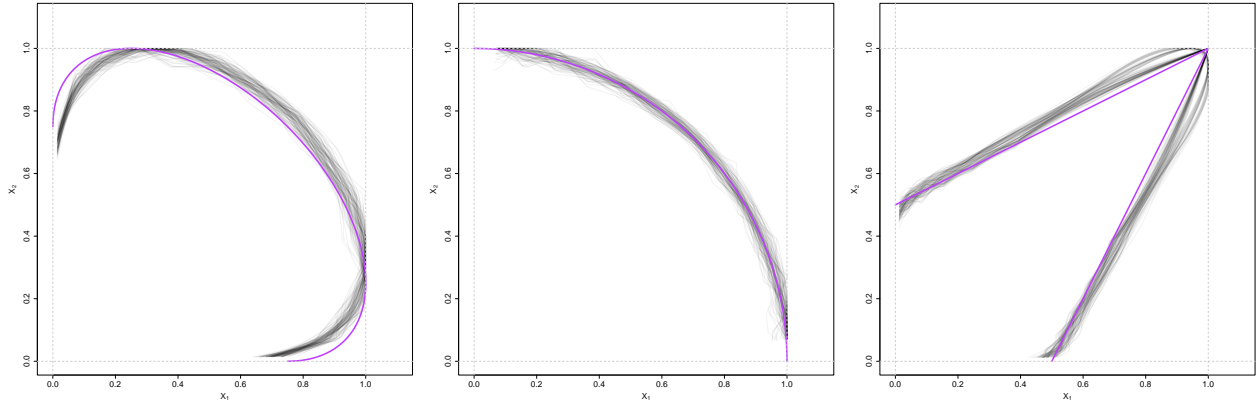
Figure P: Estimates $\hat{G}$ with $\kappa = 15$ for the Gaussian (left), inverted logistic (centre) and logistic (right) models, with the corresponding ρ or γ parameters set to 0.5. The sample size in each case is $n = 100,000$ and 100 estimates are shown for each model (grey). The true sets G are shown in purple.

of the set G with the number of knots increased to $\kappa = 15$. Increasing the number of knots has not improved the Gaussian results, but has led to a deterioration of the estimates in the logistic case, where linear splines are now chosen in only 60% of cases. This suggests having a smaller number of knots, as suggested in the main paper, is reasonable across the range of copula examples and sample sizes we consider.

The bias in the Gaussian example may be due to slow convergence to the limit set for this class of copula. This conclusion is consistent with existing theoretical and empirical studies of convergence in extremal properties for the Gaussian copula. For example, [5] show that convergence of bivariate maxima for the Gaussian copula is very slow; for the coefficient of asymptotic independence, η [26, 27], the penultimate extremal index is slow to converge to its limit value of 1 in the Gaussian case [28]; and there is slow convergence for the conditional extremes distribution [23, 31]. In all cases the convergence rates are controlled by slowly varying functions, which are $O\left[(\log u)^{-\rho/(1+\rho)}\right]$, where ρ is the correlation parameter and u is the threshold. This bias is also evident in our application of existing estimators of the extremal dependence properties to the Gaussian copula in the main paper. Given that each of these extremal characteristics is determined by the set G, it is not at all surprising that the convergence in this case is very slow. This is not to say that the bias in estimating G is too problematic for statistical inference/extrapolation purposes, as at any typical level of extrapolation the bias will persist and the resulting inference will provide reliable estimates for finite quantities of interest.

K   Coverage analysis for the non-parametric bootstrap

In this section, we demonstrate the performance of the non-parametric bootstrap for the assessing uncertainty in our estimates. We focus on one copula, namely the inverted logistic model with $\gamma = 0.5$. All tuning parameters are fixed as in Section 6.1 of the main paper. We focus on the coverage of 95% confidence intervals with respect to the dependence measure $\lambda(\omega)$ for $\omega = 0.25, 0.5, 0.75$. The set up of

our study is as follows:

1. We simulate a dataset of size $n = 10,000$ from our inverted logistic model, then estimate G and extract the corresponding estimates of $\lambda(\omega)$ for $\omega = 0.25, 0.5, 0.75$.

2. We generate 250 bootstrapped samples of size $10,000$, by sampling with replacement from the dataset in step 1. For each of these samples, we again estimate $\lambda(\omega)$ for $\omega = 0.25, 0.5, 0.75$.

3. We extract the 0.025 and 0.975 percentiles from our bootstrapped estimates of $\lambda(\omega)$, for each $\omega = 0.25, 0.5, 0.75$, providing a 95% bootstrapped confidence interval in each case.

4. We repeat steps 1-3 for 300 times, so we have 300 95% bootstrapped confidence intervals for each $\lambda(\omega)$.

In Figure Q, we show an example of the 95% confidence intervals obtained for one sample from our inverted logistic model (i.e., after carrying out steps 1-3 above just once). In this case, all three confidence intervals for $\lambda(\omega)$ ($\omega = 0.25, 0.5, 0.75$) contain the truth. Considering the results across all 300 samples, 99% of the $\lambda(0.25)$ confidence intervals, 92% of the $\lambda(0.5)$ confidence intervals and 99% of the $\lambda(0.75)$ confidence intervals contain the truth, suggesting that we have reasonable coverage with respect to the target of 95%.
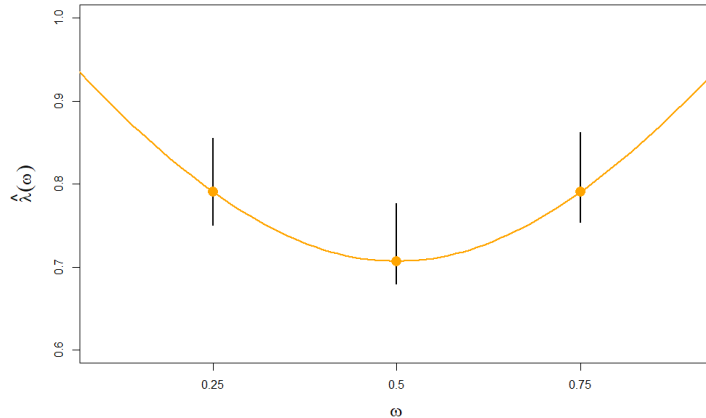

Figure Q: An example of the 95% bootstrapped confidence intervals for $\lambda(0.25)$, $\lambda(0.5)$ and $\lambda(0.75)$ (black lines) for a sample of size $n = 10,000$ taken from an inverted logistic model with $\gamma = 0.5$. The true $\lambda(\omega)$ values for this model are shown in orange.

L   The algorithm for estimating G

In Section 4 of the main paper, we present a detailed discussion of our approach for estimating the set G. Here, we provide a summarised version which may be more convenient for the reader wishing to implement our procedure. Our R code can also be found online in the GitHub repository https://github.com/essimpson/self-consistent-inference. The approach is as follows:

1. Apply the rank transformation to obtain exponential margins as in equation (14).

2. Transform the data to pseudo-polar coordinates as in (15), with observations denoted $(r_1, w_1), \ldots, (r_n, w_n)$.

3. Select the angles w_j^*, $j \in J_k := \{1, \ldots, k\}$, at which to estimate high radial quantiles.

   - **Default tuning parameter:** $k = 199$.
   - **Default angle selection:** set w_j^* as the $(j-1)/(k-1)$th empirical quantile of $\{w_1, \ldots, w_n\}$, for $j = 1, \ldots, k-1$, and $w_k^* = 1/2$. (Note that this labelling leads to unordered w_j^* values.)

4. For each w_j^*:

   (a) Select $\epsilon_{w_j^*}$ in (18) such that $|\mathcal{R}_{w_j^*}| = m$.
       - **Default tuning parameter:** $m = 100$.

   (b) Select the GPD threshold $u_{w_j^*}$ in (16) as the empirical q_uth quantile of the values in $\mathcal{R}_{w_j^*}$. Estimate the parameters $\sigma(w_j^*)$ and $\xi(w_j^*)$ in (16) by applying maximum likelihood estimation to the observed radial values in $\mathcal{R}_{w_j^*}$ exceeding $u_{w_j^*}$.
       - **Default tuning parameter:** $q_u = 0.5$.

   (c) Estimate the high radial quantile $r_q^L(w_j^*)$ via equation (17).
       - **Default tuning parameter:** $q = 0.999$.

5. Choose the position of the κ spline knots to be used in the smooth GPD parameter estimates.

   - **Default tuning parameter:** $\kappa = 7$.
   - **Default knot selection:** take knots evenly spaced within the range of observed angles, but with the central one adjusted to be exactly $1/2$.

6. For each $d \in \{1, 2, 3\}$:

   (a) Use an asymmetric Laplace distribution with a spline of degree d to estimate smooth GPD thresholds u_w at the quantile level q_u (selected above).
       - **Note:** this distribution is fitted to the log-radial values, with the required quantile extracted and back-transformed onto the radial scale.

   (b) Fit a GPD-GAM model to the radial threshold excesses, using a spline of degree d for the log-scale parameter.
       - **Default:** constant shape parameter.

   (c) Estimate the high radial quantiles $r_q^{S_{d*}}(w_j^*)$ ($j \in J_k$) (with q selected as above).

   (d) Obtain an estimate $\hat{G}^{S_d}$ of G by scaling onto $[0, 1]^2$ using the procedure outlined in Section C.

7. Choose G by minimising the mean absolute difference of the radial quantile estimates as in (19).