Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Some benefits of standardisation for conditional extremes

Christian Rohrbeck[1] | **Jonathan A Tawn**[2]

[1]Department of Mathematical Sciences, University of Bath, Bath, U.K.

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, U.K.

Correspondence
Christian Rohrbeck
Email: cr777@bath.ac.uk.

Present address
Department of Mathematical Sciences
University of Bath
Claverton Down
BA2 7AY
UNITED KINGDOM.

Abstract

A key aspect where extreme values methods differ from standard statistical models is through having asymptotic theory to provide a theoretical justification for the nature of the models used for extrapolation. In multivariate extremes many different asymptotic theories have been proposed, partly as a consequence of the lack of ordering property with vector random variables. One class of multivariate models, based on conditional limit theory as one variable becomes extreme, developed by Heffernan and Tawn (2004), has received wide practical usage. The underpinning value of this approach has been supported by further theoretical characterisations of the limiting relationships by Heffernan and Resnick (2007) and Resnick and Zeber (2014). However, Drees and Janßen (2017) provide a number of counterexamples to their results. This paper studies these counterexamples in the Keef, Papastathopoulos, and Tawn (2013) framework, which involves marginal standardisation to a common exponentially decaying tailed marginal distribution. Our calculations show that some of the issues identified by Drees and Janßen (2017) can be addressed in this way.

KEYWORDS

Conditional multivariate extreme value theory, Copulas, Laplace marginal distribution

1 | INTRODUCTION

Multivariate extreme value problems are important across a range of subject domains, such as sea levels (Coles & Tawn, 1994), air pollution (Heffernan & Tawn, 2004), rainfall (Davison, Padoan, & Ribatet, 2012) and river flow (Engelke & Hitz, 2020), which all feature in influential discussion papers. The typical formulation is to have n independent and identically distributed replicate observations $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, from a d-dimensional vector random variable $\mathbf{X}$ with unknown joint distribution $F_{\mathbf{X}}$. Here the aim is to estimate $\Pr(\mathbf{X} \in A)$ where $A \subset \mathbb{R}^d$, such that all the elements in A are in the upper tail of at least one of the marginal distributions of $\mathbf{X}$, with A being determined by the characteristics of the problem of interest. The usual approach to inference is to estimate the marginal distributions and dependence structure (copula) with a focus on their behaviour in their upper extremes. Univariate extreme value methods are well established (Coles, 2001; Davison & Smith, 1990), with the dependence modelling being the key challenge. One established inference framework, which we also employ, is to transform the margins of $\mathbf{X}$ to a common distribution based on the univariate extreme value models fitted to the individual components. If transformed to Uniform(0, 1) margins then this is the widely-used copula modelling approach, but in extreme value analysis other marginal choices are common, e.g., including Fréchet, Gumbel, Exponential and Laplace. Critically, the uncertainty in the marginal distributions is not ignored in the subsequent dependence analysis, so the choice of margins on which to study dependence is made only to simplify the mathematical presentation.

In the bivariate case, that we will focus on for variables (X, Y), there are two distinct types of extremal dependence, which are easiest explained via the coefficient of asymptotic dependence $\chi = \lim_{p\uparrow 1} \Pr\left[F_Y(Y) > p \mid F_X(X) > p\right]$, where F_X and F_Y are the marginal distributions of X and Y respectively. Having $\chi > 0$ coincides with asymptotic dependence between X and Y, a situation in which both variables can take their largest values simultaneously; while when $\chi = 0$, termed asymptotic independence, such limiting dependence is impossible. Many models for bivariate extremes are only suitable in one of these situations: multivariate max-stable distributions (Gudendorf & Segers, 2012) and multivariate generalised Pareto distributions (Kiriliouk, Rootzén, Segers, & Wadsworth, 2019) only allow $\chi > 0$ or independent variables. Therefore, distinguishing between these cases, or having a model that incorporates both in a flexible way, can play a crucial role in model selection.

One class of multivariate extreme models, based on conditional limit theory as one variable becomes extreme, developed by Heffernan and Tawn (2004), has received wide practical usage, with applications linked to widespread river flooding (Keef, Tawn, & Lamb, 2013), time series dependence in heatwaves (Winter & Tawn, 2017), spatial air temperature extremes (Wadsworth & Tawn, 2022), spatio-temporal sea-surface temperatures (Simpson & Wadsworth, 2021), offshore metocean environmental design contours (Ewans & Jonathan, 2014), coastal flooding (Gouldby et al., 2017), food chemicals (Paulo, van der Voet, Wood, Marion, & van Klaveren, 2006), and laboratory trials (Southworth & Heffernan, 2012).

This Heffernan and Tawn (2004) class of models has considerable flexibility covering both asymptotic dependence and asymptotic independence classes. In the multivariate case it allows for different extremal dependence classes between separate subsets of the variables, unlike the models of Wadsworth, Tawn, Davison, and Elton (2017) and Huser and Wadsworth (2019). Since its initial presentation, the model proposed by Heffernan and Tawn (2004) has been extended by Keef, Papastathopoulos, and Tawn (2013) to its current widely adopted form. Specifically, for (X, Y) marginally transformed to have Laplace marginals, denoted (X_L, Y_L), it is assumed that there exists values $(\alpha_{|X}, \beta_{|X}) \in [-1, 1] \times (-\infty, 1)$ such that for $x > 0$ and $z \in \mathbb{R}$

$$\Pr\left\{\frac{Y_L - \alpha_{|X} X_L}{X_L^{\beta_{|X}}} \leq z, X_L - t > x \mid X_L > t\right\} \to G_{|X}(z)\exp(-x) \qquad \text{as } t \to \infty, \tag{1}$$

where $G_{|X}$ is the distribution function of a non-degenerate random variable with values in $[-\infty, \infty)$; the restriction that $G_{|X}$ places no mass at $\{+\infty\}$ ensures that $\alpha_{|X}$ is uniquely identifiable. Limit (1) gives that the normalised Y_L is conditionally independent of X_L in the limit. To characterise the full joint tail of (X_L, Y_L) in addition to limit (1) we need the equivalent limit for the reverse conditional distribution of X_L given Y_L is large.

Despite the strong applied value of the conditional modelling framework, some concerns exist about breadth of validity of the limiting assumptions. Heffernan and Resnick (2007) and Resnick and Zeber (2014) weaken some of these assumptions producing results involving the both random and non-random norming, which in the context of limit (1) corresponds to norming Y_L by a function of X_L or t respectively. However, Drees and Janßen (2017) provided a number of counterexamples of their results.

We explore these counterexamples to see if they undermine the asymptotic justifications for the statistical methods stemming from the Heffernan and Tawn (2004) framework and their practical adoption. There is a critical difference between the framework studied in Heffernan and Resnick (2007), Resnick and Zeber (2014), Drees and Janßen (2017) from the Heffernan and Tawn (2004) framework, namely the latter requires an initial marginal standardisation, so that after transformation (X, Y) have identical marginal distributions before studying the conditional extremes behaviour. This distribution was Gumbel in Heffernan and Tawn (2004) and Laplace (as above) in Keef, Papastathopoulos, and Tawn (2013). Such standardisation of variables to common margins is quite usual in the study of dependence structure, e.g., Nelsen (1999) and Beirlant, Goegebeur, Segers, and Teugels (2004), as this makes relationships more easy to model through linearity, with exponential margins being particularly desirable for this, see Papastathopoulos, Strokorb, Tawn, and Butler (2017). Our intuition is that having marginal variables on completely different marginal tail behaviours (i.e., different shape parameters/tail indices) imposes a major restriction on a conditional approach using affine transformations, as in the norming of Y_L in limit (1). Also, we believe that there are advantages of random over non-random norming, as it is the exact value of the conditioning variable that affects the response variable, which leads to simpler limiting distributions, vital for the study of graphical structures (Engelke & Hitz, 2020).

The paper is structured as follows: in Section 2 we present the background theory of the different conditional representations. In Section 3 we cover each of the counterexamples given by Drees and Janßen (2017), with simulations to help interpretation, and state which features of Das and Resnick (2011) and Resnick and Zeber (2014) they show are not appropriate. In each case we illustrate how the problems are overcome through an initial standardisation of the marginal distributions. We then discuss the practical implications of our calculations in Section 4 and conclude with a discussion in Section 5. Some technical details of the calculations for the examples are given in the Supplementary Material.

2 | BACKGROUND THEORY

2.1 | Multivariate and Conditional Extremes

For notational simplicity, we focus on the bivariate case with (X, Y), where X and Y are continuous random variables. Classical multivariate extreme value models assume that the marginal distributions F_X and F_Y of (X, Y) belong to the domain of attraction of some extreme value distribution: F_X is in the domain of attraction of an extreme value distribution if there exist functions $p_X : \mathbb{R} \to \mathbb{R}_+$ and $q_X : \mathbb{R} \to \mathbb{R}$ such that

$$F_X^t\left\{p_X(t)x + q_X(t)\right\} \to \exp\left\{-(1+\gamma_X x)^{-1/\gamma_X}\right\} \quad \text{as } t \to \infty \tag{2}$$

for some $\gamma_X \in \mathbb{R}$ and all $x \in E^{(\gamma_X)} := \{x \in \mathbb{R} \mid 1 + \gamma_X x > 0\}$. Multivariate extreme value distributions arise as the limiting joint distribution of the componentwise maxima of independent and identically distributed random variables (X_i, Y_i), with joint distribution function $F_{X,Y}$ and marginal distribution functions $X_i \sim F_X$ and $Y_i \sim F_Y$. Specifically, it is assumed that functions p_X, q_X exist, as in limit (2), and similarly p_Y, q_Y, such that

$$\Pr\left(\frac{\max_{i=1,\dots,t} X_i - q_X(t)}{p_X(t)} \le x, \frac{\max_{i=1,\dots,t} Y_i - q_Y(t)}{p_Y(t)} \le y\right) = \left[F_{X,Y}(p_X(t)x + q_X(t), p_Y(t)y + q_Y(t))\right]^t \to H(x, y) \qquad \text{as } t \to \infty,$$

where H is a bivariate distribution function with non-degenerate marginal distributions, given by limit form (2), with tail indices of γ_X and γ_Y respectively, and with a copula possessing a specific max-stable property which, amongst other features, excludes the possibility of negative dependence, see Coles, Heffernan, and Tawn (1999) and Beirlant et al. (2004).

Heffernan and Tawn (2004) propose examining the dependence in the tail of (X, Y) by first standardising the marginals via the probability integral transformation to have Gumbel distributions, denoted (X_G, Y_G), with $\Pr(X_G \le x) = \Pr(Y_G \le x) = \exp\{-\exp(-x)\}$ for $x \in \mathbb{R}$, and considering the conditional distribution of $Y_G \mid (X_G = t)$ as $t \to \infty$. The assumption underlying their approach is that there exist normalising functions $\tilde{a}_{|X}(y) : \mathbb{R}_+ \to \mathbb{R}$ and $\tilde{b}_{|X}(y) : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\Pr\left\{\frac{Y_G - \tilde{a}_{|X}(X_G)}{\tilde{b}_{|X}(X_G)} \le z \mid X_G = t\right\} \to \tilde{G}_{|X}(z) \qquad \text{as } t \to \infty, \tag{3}$$

where the limit distribution $\tilde{G}_{|X}$ is non-degenerate. To ensure that $\tilde{a}_{|X}$, $\tilde{b}_{|X}$ and $\tilde{G}_{|X}$ are well-defined, we require $\lim_{z\to\infty} \tilde{G}_{|X}(z) = 1$, i.e., $\tilde{G}_{|X}$ has no mass at $\{+\infty\}$, and $\tilde{b}_{|X}(x)/x \to 0$ as $x \to \infty$ (Keef, Papastathopoulos, & Tawn, 2013). Heffernan and Tawn (2004) find that (up to type) the functions $\tilde{a}_{|X}$ and $\tilde{b}_{|X}$ in (3) have a common parametric form for all copulas in Joe (1997) and Nelsen (1999).

Heffernan and Resnick (2007) modify and extend the Heffernan and Tawn (2004) framework, by replacing the condition $X = t$ in (3) by $X > t$, i.e., they analyse

$$\Pr\left\{\frac{Y - a_{|X}(X)}{b_{|X}(X)} \le z \mid X > t\right\} \to G_{|X}(z) \qquad \text{as } t \to \infty. \tag{4}$$

This is the most widely applied and considered conditional extreme value model framework and we will use it in the remainder of the paper. Heffernan and Resnick (2007) further drop the assumption that X and Y have Gumbel margins and provide theoretical results subject to F_X lying in the domain of attraction of some extreme value distribution.

Keef, Papastathopoulos, and Tawn (2013) focus on X_L and Y_L having standard Laplace margins, i.e.,

$$\Pr(X_L < x) = \Pr(Y_L < x) = \begin{cases} \exp(x)/2 & \text{if } x \le 0, \\ 1 - \exp(-x)/2 & \text{if } x > 0. \end{cases}$$

Under these conditions, the functions in (4) are of the form $a_{|X}(x) = \alpha x$ and $b_{|X}(x) = x^\beta$, with $(\alpha, \beta) \in [-1, 1] \times (-\infty, 1)$, in all of the standard copulas studied by Heffernan and Tawn (2004). When X and Y are positively associated, standardisation to Laplace margins (X_L, Y_L) gives the same limiting behaviour as when the variables were transformed to Gumbel margins. However, the limiting behaviours differ when X and Y are negatively associated, with the symmetry of the Laplace margins giving a simpler form. Estimation of the parameters (α, β) and the distribution function $G_{|X}$ for Laplace margins are considered in Keef, Papastathopoulos, and Tawn (2013) and Keef, Tawn, and Lamb (2013).

2.2 | Linking Multivariate and Conditional Extremes Models

Heffernan and Resnick (2007) considered how the conditional extremes model of Heffernan and Tawn (2004) is linked to established multivariate extreme value models. To address this they define the class of conditional extreme value models (CEVM) which does not require the margins to be standardised to common margins. The limit distribution of $Y \mid (X > t)$ as $t \to \infty$ lies in the CEVM class if

1. The distribution function F_X of X is in a domain of attraction of an extreme value distribution with parameter $\gamma_X \in \mathbb{R}$.
2. There exist normalising functions $c, f : \mathbb{R} \to \mathbb{R}$ and $d, g : \mathbb{R} \to \mathbb{R}_+$, such that

$$t \ \Pr\left\{\frac{X - c(t)}{d(t)} > x, \frac{Y - f(t)}{g(t)} \leq y\right\} \to \mu_{Y|X>}((x, \infty] \times [-\infty, y]) \quad \text{as } t \to \infty, \tag{5}$$

where $\mu_{Y|X>}((x, \infty] \times [-\infty, y])$ is a non-degenerate distribution function in y, and $\mu_{Y|X>}((x, \infty] \times [-\infty, y]) < \infty$.

Resnick and Zeber (2014) add the condition $\mu_{Y|X>}((x, \infty] \times \{\infty\}) = 0$, however Drees and Janßen (2017, Example 2.3) shows that to ensure uniqueness of the limit measure, the limit measure cannot put any mass for Y at $\{-\infty\}$ or $\{+\infty\}$, which requires that

$$\lim_{y\to\infty} \mu_{Y|X>}((x, \infty] \times [y, \infty)) = \lim_{y\to\infty} \mu_{Y|X>}((x, \infty] \times (-\infty, -y]) = 0.$$

Returning to the link between the CEVM framework and multivariate extreme value models, assume that X and Y belong to the domain of attraction of some extreme value distribution, with parameters γ_X and γ_Y respectively. Theorem 2.1 in Das and Resnick (2011) states that (X, Y) lies in the domain of attraction of a multivariate extreme value distribution if $Y \mid (X > t)$ and $X \mid (Y > t)$ both lie in the CEVM class by Heffernan and Resnick (2007). Example 4.4 in Drees and Janßen (2017) illustrates that this result is not true, unless the normalisations of $Y \mid (X > t)$ and $X \mid (Y > t)$ in (5) are identical.

Consider when (X, Y) lies in the domain of attraction of a multivariate extreme value distribution. Suppose that (i) X and Y are asymptotically dependent and (ii) $\gamma_X, \gamma_Y \leq 0$. Drees and Janßen (2017) show that these conditions are sufficient for the limits of $Y \mid (X > t)$ and $X \mid (Y > t)$ to lie in the class of CEVM. The restriction $\gamma_X, \gamma_Y \leq 0$ is required, as demonstrated by Example 4.2 in Drees and Janßen (2017). Note, the conditions (i) and (ii) are not necessary conditions for $Y \mid (X > t)$ and $X \mid (Y > t)$ to lie in the CEVM class; see Section 5 in Heffernan and Resnick (2007) for the case of X and Y being asymptotically independent.

2.3 | Standardisation of Marginals in the CEVM Class

Heffernan and Resnick (2007) examined how the standardisation of X to a standard Pareto distributed random variable X_P, but leaving Y unchanged, affected the limiting measure $\mu_{Y|X>}$ in (5). They show that the limiting behaviour of (X_P, Y) satisfies

$$t \ \Pr\left\{\frac{X_P}{t} > x, \frac{Y - f(t)}{g(t)} \leq y\right\} \to \begin{cases} \mu_{Y|X>}\left(\left(\frac{x^{\gamma_X} - 1}{\gamma_X}, \infty\right] \times [-\infty, y]\right) & \text{if } \gamma_X \neq 0, \\ \mu_{Y|X>}((\log x, \infty] \times [-\infty, y]) & \text{if } \gamma_X = 0, \end{cases}$$

where $\mu_{Y|X>}$ is as in (5), and $c(t) = 0$ and $d(t) = t$ from standardisation to X_P.

The standardisation of Y is more challenging than that of X, because the CEVM (5) does not require F_Y to be in a domain of attraction of an extreme value distribution. Heffernan and Resnick (2007) and Das and Resnick (2011) consider the task of finding a monotone and unbounded function $h : \mathbb{R} \to \mathbb{R}_+$ such that

$$t \ \Pr\{X_P/t > x, h(Y)/t \leq y\} \to \tilde{\mu}_{Y|X>}((x, \infty] \times [-\infty, y]) \quad \text{as } t \to \infty, \tag{6}$$

where $\tilde{\mu}_{Y|X>}$ is finite and non-degenerate, $\tilde{\mu}_{Y|X>}((x, \infty] \times \{\infty\}) = 0$ and $\tilde{\mu}_{Y|X>}(\{\infty\} \times [-\infty, y]) = 0$ for all x, y. Das and Resnick (2011) argue that such a function h exists if, and only if, $\mu_{Y|X>}$ is not a product measure. However, Drees and Janßen (2017, Section 3) illustrates that neither implication is true, and the limit measures $\mu_{Y|X>}$ and $\tilde{\mu}_{Y|X>}$ may convey different information if they exist.

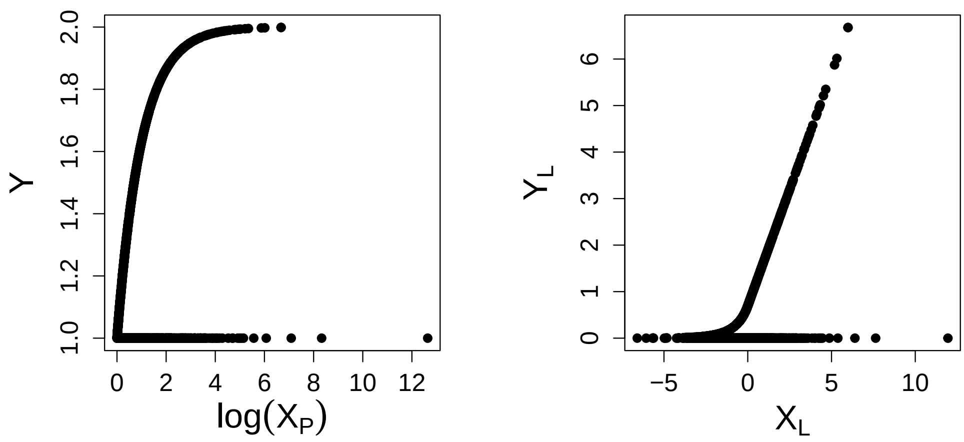

FIGURE 1 Illustration of 2,000 samples for the framework in Example 2.3. The left panel shows the simulated observations (x_p, y) on the original scale, while the right panel corresponds to the transformed samples for (x_L, y_L).

3 | INVESTIGATING DREES & JANßEN (2017) EXAMPLES

3.1 | Strategy

Most examples in Drees and Janßen (2017) work with the joint distribution of (X_P, Y), where $\Pr(X_P > x) = 1 - x^{-1}$ $(x > 1)$, i.e., the conditioning variable has standard Pareto distribution, and the distribution of Y is given indirectly through the distributions of X_P and $Y \mid X_P$. We consider their examples, using their numbering. We work within the framework of Keef, Papastathopoulos, and Tawn (2013), i.e., we consider the limiting behaviour of (X_P, Y) after transformation to Laplace margins, denoted by (X_L, Y_L). The standardisation of X_P to Laplace margins yields the variable X_L and the link between the values x_L of X_L and x of X_P is given by

$$\frac{1}{x} = \begin{cases} 1 - (1/2)\exp(x_L) & \text{if } x_L \leq 0, \\ (1/2)\exp(-x_L) & \text{if } x_L > 0. \end{cases} \tag{7}$$

When transforming Y to Y_L, we derive the distribution function F_Y and the value y_L of Y_L as $y_L = F_L^{-1}\left[F_Y(y)\right]$, where F_L is the distribution function of a Laplace variable.

3.2 | Example 2.3

Let B be a discrete random variable that is uniformly distributed on $\{0, 1\}$ and independent of X_P, and define $Y = B + (1-B)(2-1/X_P)$. Then Y can take any value in the interval $[1, 2)$, with the highest values occurring when $B = 0$ and X_P large, and its marginal distribution is given by $\Pr(Y = 1) = 1/2$ and $\Pr(Y < y) = y/2$ for $1 < y \leq 2$. This example in Drees and Janßen (2017) showed that the condition $\mu_{Y|X}([x, \infty] \times \{\infty\}) = 0$ in Resnick and Zeber (2014) is not sufficient to ensure uniqueness of the limit measure in expression (5), and that the stronger condition $\mu_{Y|X}([x, \infty] \times \{-\infty, \infty\}) = 0$ is required.

As outlined in Section 3.1, we are interested in the limiting behaviour of the transformed variable Y_L given that the transformed variable X_L is large. Transformation of Y to Laplace margins gives $Y = 2 - \exp(-Y_L)$ for $B = 0$, while $Y = 1$ when $B = 1$; this second case implies $Y_L = 0$ irrespective of X_P for $B = 1$ and, thus, the lower tail of Y_L is not Laplace distributed. Figure 1 left panel shows that realised values of Y are close to $y = 1$ or $y = 2$ for large values of X_P, while Y_L, shown in the right panel, has no upper bound. When $B = 0$, substituting the realisations x and y by their transformed values x_L and y_L, gives for $x_L > 0$

$$y = 2 - \frac{1}{x} \quad \Leftrightarrow \quad 2 - \exp(-y_L) = 2 - (1/2)\exp(-x_L) \quad \Leftrightarrow \quad y_L = \log(2) + x_L.$$

This linear relationship between the values, for $x_L > 0$ and $B = 0$, is also visible in Figure 1 right panel, while, for $B = 1$, we have $y_L = 0$ for all possible values x_L.

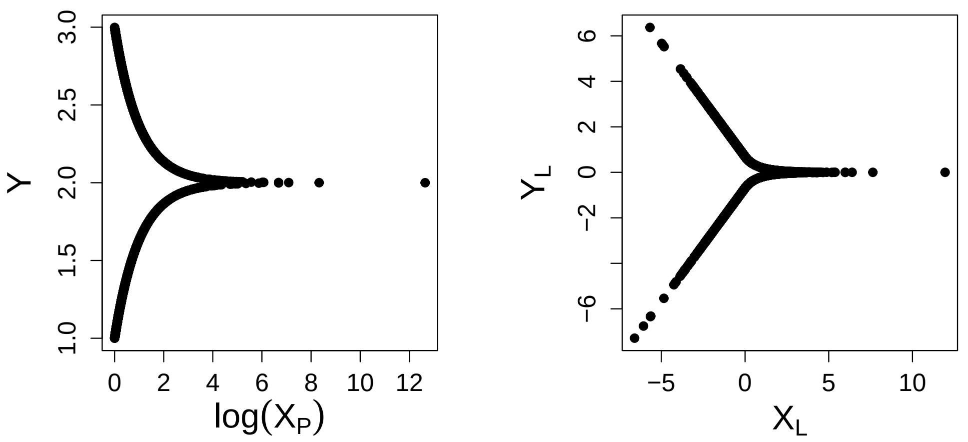

FIGURE 2 Illustration of 2,000 samples for the framework in Example 3.1. The left panel shows the simulated observations (x_p, y) on the original scale, while the right panel corresponds to the transformed samples for (x_L, y_L).

From the calculations above, the functions $a_{|X}(x) = x$ and $b_{|X}(x) = 1$ in expression (4) give the limiting behaviour as

$$\Pr(Y_L - X_L \leq z \mid X_L > x_L) \to (1/2)(1 + \mathbf{I}\{\log 2 \leq z\}) = G_{|X}(z) \qquad \text{as } x_L \to \infty,$$

where $\mathbf{I}$ is the indicator function, and $G_{|X}$ is a non-degenerate distribution function. The result $\lim_{z\to-\infty} G_{|X}(z) = 0.5$ is due to the case $B = 1$ which occurs with probability 0.5. Other choices for $a_{|X}$ and $b_{|X}$ lead to a degenerate limiting distribution $G_{|X}$, contradicting the Heffernan and Tawn (2004) assumption, or yield $\mu_{Y|X>}([x, \infty] \times \{\infty\}) = 0$, violating the constraint $\lim_{z\to\infty} G_{|X}(z) = 1$ by Keef, Papastathopoulos, and Tawn (2013). We have for $x_L > 0$

$$\mu_{Y_L|X_L>}((x_L, \infty] \times [-\infty, y_L]) = (1/2)(1 + \mathbf{I}\{\log 2 \leq y_L\}) \times (1/2)\exp(-x_L).$$

While this is similar to the first limit found by Drees and Janßen (2017), we do not require the constraint $\mu_{Y|X>}([x, \infty] \times \{-\infty\}) = 0$ they introduced to ensure a unique limiting behaviour, because we transformed the variables to common Laplace margins.

3.3 | Example 3.1

Let B be a discrete random variable that is uniformly distributed on $\{-1, 1\}$ and independent of the Pareto distributed random variable X_P. The variable Y is defined as $Y = 2 - B/X_P$. For large X_P, the values of Y are concentrated around 2 (see Figure 2 left panel). The marginal distribution of Y is $Y \sim \text{Uniform}(1, 3)$. Drees and Janßen (2017) present this and the following Example 3.2, to illustrate that the result by Das and Resnick (2011) linked to the standardisation (6) of Y does not hold in general.

We again start by transforming the random variables X_P and Y to Laplace margins. Substitution of the values y by their transformed values y_L gives $y = 1 + \exp(y_L)$ if $y_L \leq 0$ and $y = 3 - \exp(-y_L)$ if $y_L > 0$, and the transformation of X_P to Laplace margins is given in (7). For the case $B = 1$, Y takes values smaller than 2, while only values greater than 2 are observed for Y when $B = -1$. Therefore, we have to consider the transformation with $y_L \leq 0$ for $B = 1$, and $y_L > 0$ for $B = -1$. For the case $B = 1$, we find that

$$y = 2 - \frac{1}{x} \quad \Leftrightarrow \quad 1 + \exp(y_L) = 2 - (1/2)\exp(-x_L) \quad \Leftrightarrow \quad y_L = \log\{1 - \exp(-x_L)/2\}.$$

Hence we find y_L can be approximated by $-\frac{1}{2}\exp(-x_L)$ as $x_L \to \infty$. Calculations, similar to the $B = -1$ case, give that as $x_L \to \infty$

$$y = 2 + \frac{1}{x} \quad \Leftrightarrow \quad y_L = -\log\{1 - \exp(-x_L)/2\} \sim \exp(-x_L)/2.$$

Without norming, $Y_L \mid (X_L > x) \to^P 0$ as $x \to \infty$. To avoid this degeneracy, we need to take the functions in (4) to be $a_{|X}(x) = 0$ and $b_{|X}(x) = \exp(-x)$, the limiting distribution $G_{|X}$ then assigns probability 1/2 to each of the values $z = -0.5$ and $z = 0.5$.

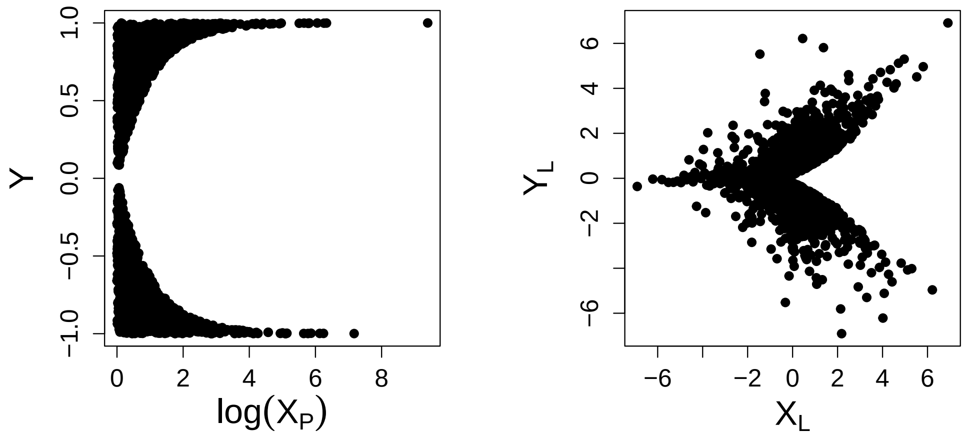

FIGURE 3 Illustration of 2,000 samples for the framework in Example 3.2. The left panel shows the simulated observations (x_p, y) on the original scale, while the right panel corresponds to the transformed samples for (x_L, y_L).

The expression for $b_{|X}(x)$ is not of the form for Laplace margins found by Keef, Papastathopoulos, and Tawn (2013), with $b_{|X}(x)$ tending to zero very rapidly. This form is needed given the speed of convergence of $Y_L \mid (X_L > x)$ towards zero as $x \to \infty$, as seen in Figure 2 right panel. This is not too surprising as it is known that the simple parametric forms of Keef, Papastathopoulos, and Tawn (2013) for the norming functions do not always hold, with Papastathopoulos and Tawn (2016) already identifying that it is possible to have $a_{|X}(x) = x\mathcal{L}_a(x)$ and $b_{|X}(x) = x^{\beta}\mathcal{L}_b(x)$, with $\mathcal{L}_a(x)$ and $\mathcal{L}_b(x)$ being slowly varying functions and $\beta \in (-\infty, 1)$. Here we have an example that is outside that class with $\beta = 0$ and $-\log\{\mathcal{L}_b(x)\}$ being regularly varying. With our norming, the limiting measure $\mu_{Y_L|X_L>}$, as defined in (5), is

$$\mu_{Y_L|X_L>}((x_L, \infty] \times [-\infty, y_L]) = (1/2)\,(\mathbf{I}\{-0.5 \le y_L\} + \mathbf{I}\{0.5 \le y_L\}) \times (1/2)\exp(-x_L).$$

So, a combination of the standardisation of marginals and random norming, by X_L not x_L, gives a simpler product limit measure than found by Drees and Janßen (2017).

3.4 | Example 3.2

Let B be a discrete random variable that is uniformly distributed on $\{-1, 1\}$, $U \sim \text{Uniform}(0, 1)$, and X_P, B and U are all independent. Define $Y = B(1 - U/X_P)$, with the random variable Y taking negative and positive values for $B = -1$ and $B = 1$ respectively. Figure 3 left panel shows that the values of Y are close to $y = -1$ and $y = 1$ for large values of X_P. For $-1 < y < 0$, we calculate the marginal distribution of Y as $\Pr(Y < y) = (y + 1)\{1 - \log(y + 1)\}/2$; see Supplementary Material S1 for details. Using similar calculations, we find $\Pr(Y < y) = (1 + y)/2 + (1 - y)\log(1 - y)/2$ for $0 \le y < 1$.

To transform Y to Laplace margins for $y < 0$, which corresponds to $y_L < 0$, we use the relationship $(1/2)(y+1)\left[1 - \log(y + 1)\right] = (1/2)\exp(y_L)$. Since there is no analytical closed form for y in terms of y_L, we consider approximations in order to derive the link between y_L and y in the limit as $y \to -1$. The calculations in the Supplementary Material S2 give that $(y + 1) \sim -\exp(y_L)/y_L$ for $y \downarrow -1$. Similarly $(1-y) \sim \exp(-y_L)/y_L$ for $y \uparrow 1$. For $B = 1$, the limiting behaviour of Y, as X_P becomes large, is thus described by

$$y = -1 + \frac{u}{x} \quad \Leftrightarrow \quad -\frac{\exp(y_L)}{y_L} - 1 = -1 + \frac{u}{2}\exp(-x_L) \quad \Leftrightarrow \quad y_L - \log(-y_L) = \log\left(\frac{u}{2}\right) - x_L,$$

where u denotes the realisation of the random variable U. For $B = -1$, as $x_L \to \infty$

$$y_L = -x_L + \log x_L + o_P(\log x_L),$$

where the stochasticity is due to U. So, $y_L \xrightarrow{p} -\infty$ as $x_L \to \infty$; this can also be seen in Figure 3 right panel. Similar calculations give $-y_L - \log(y_L) = \log(u/2) - x_L$ when $B = 1$. Hence, for $B = 1$, $y_L = x_L - \log x_L + o_P(\log x_L)$ with $y_L \xrightarrow{p} \infty$ as $x_L \to \infty$.

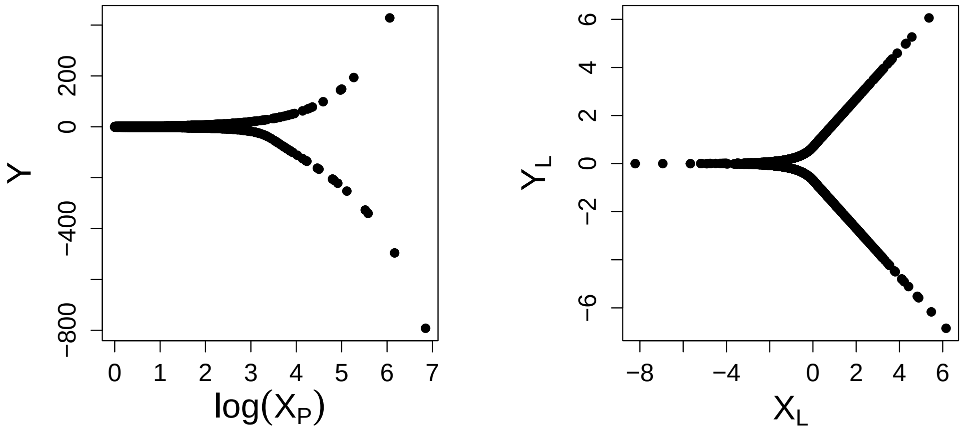

FIGURE 4 Illustration of 2,000 samples for the framework in Example 4.2. The left panel shows the simulated observations (x_p, y) on the original scale, while the right panel corresponds to the transformed samples for (x_L, y_L).

These results appear to correspond to there being non-unique choices for $a_{|X}$ and $b_{|X}$ in (3) that yield a non-degenerate limiting distribution $G_{|X}$. However, there is only one such choice (up to type) with $a_{|X}(x) = x$ and $b_{|X}(x) = \log x$ $(x > 1)$ giving $G_{|X}$ placing mass of 1/2 at $\{-\infty\}$ and $\{-1\}$, i.e., $G_{|X}(x) = 0.5$ for $-\infty < z < -1$ and $G_{|X}(z) = 1$ for $-1 \leq z < \infty$. As in Example 3.1, the norming function $b_{|X}(x)$ is not of the simple power parametric form of Keef, Papastathopoulos, and Tawn (2013). Another possible norming has $b_{|X}(x) = x$ with $a_{|X}(x) = o(b_{|X}(x))$ as $x \to \infty$, giving $G_{|X}$ with mass of 1/2 at $\{-1\}$ and $\{1\}$, but this type of norming is not permitted as $b_{|X}(x)$ cannot grow as fast as x (Keef, Papastathopoulos, & Tawn, 2013).

3.5 | Example 4.2

Let B be a discrete random variable that is uniformly distributed on $\{0, 1\}$. Define the function $g(x) := x(2 + \sin \log x)$ for $x \geq 1$ and we consider $Y = BX_P + (1 - B)\left\{-g^{-1}(2X_P)\right\}$, where g^{-1} is the inverse of g. Figure 4 indicates that Y tends to $-\infty$ and $+\infty$ as X_P becomes large. The purpose of this example in Drees and Janßen (2017) is to illustrate that (X_P, Y) being multivariate extreme value distributed is not a sufficient condition for $Y \mid (X_P > t)$, as $t \to \infty$, to lie in the class of CEVMs of the form (5).

We derive the marginal distribution of Y as

$$\Pr(Y < y) = \begin{cases} 1/g(-y) & \text{if } y < -1, \\ 1/2 & \text{if } -1 \leq y \leq 1 \\ 1 - 1/2y & \text{if } y > 1, \end{cases}$$

in Supplementary Material S3. Transformation of Y to Laplace margins gives $y = -g^{-1}\{2/\exp(y_L)\}$ for $y \leq -1$, and $y = \exp(y_L)$ when $y \geq 1$. The limiting behaviour of the transformed variable Y_L as X_L becomes large, and for $B = 1$, is given by $\exp(y_L) \sim 2\exp(x_L)$, which is equivalent to $y_L = \log 2 + x_L + o(1)$ as $x_L \to \infty$. When $B = 0$, $y_L = -\log 2 - x_L + o(1)$ as $x_L \to \infty$. This symmetry in the limiting behaviour on Laplace marginal distributions is also visible in Figure 4 right panel. So, as $x_L \to \infty$,

$$Y_L = \begin{cases} \log 2 + X_L + o(1) & \text{if } B = 1, \\ -\log 2 - X_L + o(1) & \text{if } B = 0. \end{cases}$$

Defining $a_{|X}(x) = x$ and $b_{|X}(x) = 1$ yields a non-degenerate limiting distribution $G_{|X}$ with $G_{|X}(z) = 0.5$ for $-\infty < z < \log 2$ and $G_{|X}(z) = 1$ for $\log 2 \leq z < \infty$. While the normalising functions $a_{|X}(x) = -x$ and $b_{|X}(x) = 1$ also yield a non-degenerate limiting distribution $G_{|X}$, this choice is not permissible because $G_{|X}$ would have mass at $\{+\infty\}$ (Keef, Tawn, & Lamb, 2013). Consequently, the normalising functions $a_{|X}$ and $b_{|X}$ are well-defined (up to type). Furthermore, this example shows that a transformation to Laplace margins can result in the distribution of $Y_L \mid (X_L > t)$ as $t \to \infty$ being in the class of conditional extreme models by Keef, Papastathopoulos, and Tawn (2013), although $Y \mid (X_P > t)$ as $t \to \infty$ does not lie in the class of CEVMs introduced by Heffernan and Resnick (2007). When we consider the distributions of $X_L \mid (Y_L = y_L)$, where $y_L > 0$, there is a deterministic relationship between X_L and Y_L, and thus there cannot exist a non-degenerate limiting distribution $G_{|Y}$, but the behaviour is trivial $(X_L = Y_L - \log 2) \mid (Y_L > y_L)$ for any $y_L > 0$.

3.6 | Example 4.4

Define the function $g_c(u) = u(1 + c \sin \log u)$, where $0 < u \leq 1$ and $|c| < 1/\sqrt{2}$, and $\psi_c(z) = g_c^{-1}(1/z)$ with $z \geq 1$. Let Z_P be standard Pareto distributed, $\Pr(Z_P > z) = 1 - z^{-1}$ $(z > 1)$, and B be discrete and uniformly distributed on $\{1, 2, 3, 4\}$ and also independent of Z_P. The random variables X and Y are then defined as

$$(X, Y) := \begin{cases} (2 - \psi_{1/2}(Z_P),\ 2 - 1/Z_P) & \text{if } B = 1, \\ (2 - \psi_{-1/2}(Z_P),\ 2 - 1/\sqrt{Z_P}) & \text{if } B = 2, \\ (1 - 1/Z_P,\ 2 - 1/Z_P) & \text{if } B = 3, \\ (2 - 1/Z_P,\ 1 - 1/Z_P) & \text{if } B = 4. \end{cases} \tag{8}$$

The purpose of this example by Drees and Janßen (2017) is to show that (X, Y) does not lie in the class of multivariate extreme value models despite $Y \mid (X > t)$ and $X \mid (Y > t)$ belonging to the CEVM class by Heffernan and Resnick (2007) as $t \to 2$. This inconsistency of the CEVM class with (X, Y) being in the domain of attractions of a bivariate extreme value distribution indicates that these conditional distributions fall outside the framework of the standard assumptions for bivariate extreme values.

We now investigate this inconsistency for the bivariate distribution (8) after marginal standardisation. We start by calculating the marginal distributions of X and Y to see if they are individually in the domain of attraction of the univariate extreme value distribution (2). Figure 5 left panel shows that the random variable X (Y) respectively can take values between 0 and 1 when $B = 3$ $(B = 4)$, while $B \neq 3$ $(B \neq 4)$ leads to the values of X (Y) lying between 1 and 2. The cumulative distribution function of X is

$$\Pr(X \leq x) = \begin{cases} x/4 & \text{if } 0 \leq x \leq 1, \\ 3x/4 - 1/2 & \text{if } 1 \leq x \leq 2, \end{cases}$$

and for Y we have

$$\Pr(Y \leq y) = \begin{cases} y/4 & \text{if } 0 \leq y \leq 1, \\ y/2 - (2 - y)^2/4 & \text{if } 1 \leq y \leq 2, \end{cases}$$

with the derivations provided in Supplementary Material S4. For these two marginals it is straightforward to show that they are each in the domain of attraction of the univariate extreme value distribution with parameters $\gamma_X = \gamma_Y = -1$.

Transformation of X and Y to Laplace margins gives

$$x = \begin{cases} 2\exp(x_L) & \text{if } x_L \leq -\log 2, \\ 2/3 + (2/3)\exp(x_L) & \text{if } -\log 2 < x_L \leq 0, \\ 2 - (2/3)\exp(-x_L) & \text{if } x_L > 0, \end{cases}$$

and

$$y = \begin{cases} 2\exp(y_L) & \text{if } y_L \leq -\log 2, \\ 3 - \sqrt{5 - 2\exp(y_L)} & \text{if } -\log 2 < y_L \leq 0, \\ 3 - \sqrt{1 + 2\exp(-y_L)} & \text{if } y_L > 0. \end{cases}$$

These transformations do not change the property that the marginal distributions are in the domain of attractions of univariate extremes value distribution, only now $\gamma_{X_L} = \gamma_{Y_L} = 0$, and we still have that (X_L, Y_L) is not in the domain of attraction of a bivariate extreme value distribution. The following calculations show that, with standardisation to Laplace margins, the conditional limiting distribution of $Y_L \mid (X_L > x_L)$ fails to meet the conditions of Heffernan and Tawn (2004) as $x_L \to \infty$, unlike the conditional $Y \mid (X > t)$, as $t \to 2$, that falls in the CEVM class of Heffernan and Resnick (2007).

We explore the conditional distributions by looking at the relations between (X_L, Y_L) for each value of B. For $B = 1$, the expressions $X = 2 - \psi_{1/2}(Z)$ and $Y = 2 - 1/Z$ give $Y = 2 - g_{1/2}(2 - X)$. To study the limiting behaviour, we again replace X and Y

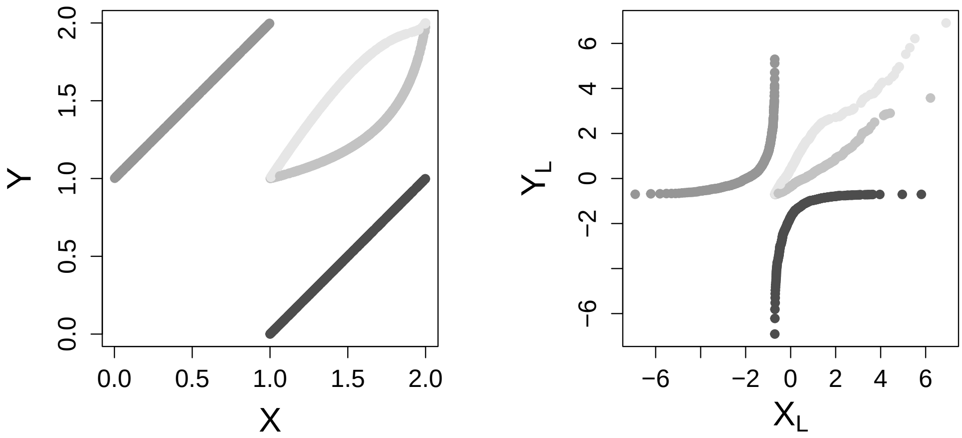

FIGURE 5 Illustration of 2,000 samples for the framework in Example 4.4. The left panel shows the simulated observations (x, y) on the original scale, while the right panel corresponds to the transformed samples (x_L, y_L), with the behaviour for each value of B highlighted: the lightest shade corresponds to $B = 1$, and the points with the darkest shade are the samples for $B = 4$.

by their Laplace distributed transformed expressions. Figure 5 right panel shows that, for $B = 1$, large values of X_L lead to large values of Y_L and we thus consider the equality

$$\begin{aligned} 3 - \sqrt{1 + 2\exp(-y_L)} &= 2 - g_{1/2}\{(2/3)\exp(-x_L)\} \\ &= 2 - (1/3)\exp(-x_L)[2 + \sin\{\log(2/3) - x_L\}]. \end{aligned}$$

For notational brevity, we define $h_1(x_L) = (1/3)\,[2 + \sin\{\log(2/3) - x_L\}]$ and we note $(1/3) \le h_1(x_L) \le 1$. By simplifying the terms and taking squares on both sides, we get

$$1 + 2\exp(-y_L) = 1 + 2\exp(-x_L)h_1(x_L) + \exp(-2x_L)\,[h_1(x_L)]^2\,.$$

Further simplifying the terms and taking logs, we end up with

$$y_L = x_L - \log h_1(x_L) + \log\,[1 + (1/2)\exp(-x_L)h_1(x_L)]\,, \tag{9}$$

which we can write as $y_L = x_L - \log h_1(x_L) + (1/2)\exp(-x_L)h_1(x_L) + O\,(\exp(-2x_L))$.

For $B = 2$, the expressions $X = 2 - \psi_{-1/2}(Z_P)$ and $Y = 2 - 1/Z_P$ give that $Y = 2 - \sqrt{g_{-1/2}(2 - X)}$. On Laplace scale, we then have

$$3 - \sqrt{1 + 2\exp(-y_L)} \;=\; 2 - \sqrt{g_{-1/2}\,\{(2/3)\exp(-x_L)\}} \;=\; 2 - \exp(-x_L/2)h_2(x_L),$$

where $h_2(x_L) = \sqrt{(2/3) - (1/3)\sin\{\log(2/3) - x_L\}}$. By taking squares on both sides,

$$1 + 2\exp(-y_L) = 1 + 2\exp(-x_L/2)h_2(x_L) + \exp(-x_L)\,[h_2(x_L)]^2\,.$$

Following the same steps as for $B = 1$ yields

$$y_L = x_L/2 - \log h_2(x_L) + (1/2)\exp(-x_L/2)h_2(x_L) + O\,(\exp(-x_L))\,.$$

The case $B = 3$ leads to $x_L < -\log 2$, i.e., x_L is not becoming large, and thus this mixture component can be ignored when studying $Y_L \mid (X_L > x_L)$ as $x_L \to \infty$. Finally, $B = 4$ gives $Y = X - 1$ with $2\exp(Y_L) = 2 - (2/3)\exp(-X_L) - 1$. Consequently,

$$y_L = \log\,[1/2 - (1/3)\exp(-x_L)] \;\sim\; -\log(2) \qquad \text{as } x_L \to \infty.$$

Combining the different mixture components and we set $a_{|X}(x) = x - \exp(-x)h_1(x)$ and $b_{|X}(x) = -\log h_1(x) + (1/2)h_1(x)\exp(-x)$ in (4); as $-\log h_1(x) \geq 0$ and $h_1(x) > 0$ this gives $b_{|X}(x) > 0$ as required. For $B = 1$, $Y_L \mid (X_L > x_L)$ as $x_L \to \infty$, behaves as

$$\frac{Y_L - X_L + \exp(-X_L)h_1(X_L)}{-\log h_1(X_L) + (1/2)h_1(X_L)\exp(-X_L)} \sim 1.$$

For $B = 2$ and $B = 4$, $\lim_{x_L \to \infty}(Y_L - a_{|X}(X_L))/b_{|X}(X_L) = -\infty$ for $X_L > x_L$. However, $\Pr(\{Y_L - a_{|X}(x_L)\}/b_{|X}(x_L) \leq z \mid X_L > x_L)$ does not converge as $\Pr(B = 1 \mid X_L > x_L)$ oscillates between 1/6 and 1/2 as $x_L \to \infty$. This oscillation is found by considering $\Pr(B = 1 \mid X > t)$ for $1 < t < 2$. Using similar calculations as in the Supplementary Material S4, we find

$$\Pr(B = 1 \mid X > t) = \frac{\Pr(X > t \mid B = 1)\Pr(B = 1)}{\Pr(X > t)} = \frac{1}{3} + \frac{1}{6}\sin\log(2 - t),$$

which oscillates as $t \to 2$, implying that $\Pr(B = 1 \mid X_L > x_L)$ oscillates between 1/6 and 1/2 as $x_L \to \infty$. Consequently, $Y_L \mid (X_L > x_L)$ as $x_L \to \infty$ does not fall in the Keef, Papastathopoulos, and Tawn (2013) class of conditional extreme value models, despite $Y \mid (X > t)$, as $t \to 2$, being in the CEVM class by Heffernan and Resnick (2007); see Drees and Janßen (2017).

We have focused on the distribution of $Y_L \mid (X_L > x_L)$ for $x_L \to \infty$, but there is also interest in the reverse conditional $X_L \mid (Y_L > y_L)$ as $y_L \to \infty$. Figure 5 provides some insight, with only the mixture terms corresponding to $B = 1$ and $B = 3$ contributing to the tail of Y_L. Further, expression (8) shows that $Y_L \mid (B = 1)$ and $Y_L \mid (B = 3)$ are identical, which gives that the limiting distribution of $X_L \mid (Y_L > y_L)$ as $y_L \to \infty$ must be a mixture distribution with weights 1/2 on each component. When $B = 3$, X_L does not grow with Y_L, so any norming on X_L that is required to handle the growth of X_L with y_L in $X_L \mid (Y_L > y_L)$ will lead to mass tending to $-\infty$. When $B = 1$, the relationship in expression (9) between X_L and Y_L gives $Y_L > X_L$ conditional on X_L being above a sufficiently high threshold, because $-\log h_1(x) \geq 0$ and $\log[1 + (1/2)\exp(-x)h_1(x)] > 0$ for all x. Furthermore, the relation between X_L and Y_L is bijective, since (9) is strictly monotonic increasing. So, we can invert the relation between Y_L and X_L when $B = 1$ to give for $y_L \to \infty$ that $x_L = y_L - Q(y_L)$, where $Q(y_L) > 0$ is an oscillator function that is bounded above, so $(x_L - y_L)/Q(y_L) \sim -1$. Hence, for all $z \in \mathbb{R}$, as $y_L \to \infty$

$$\Pr\left((X_L - Y_L)(Q(Y_L))z \middle| Y_L > y_L\right) \to 0.5[1 + \mathbf{I}(z > -1)].$$

Thus, the reverse conditional has a more straight-forward behaviour.

We further note that the transformation to Laplace margins does not lead to (X_L, Y_L) lying in the class of multivariate extreme value models. Drees and Janßen (2017) showed that (X, Y) is not multivariate extreme value distributed either. Consequently, this is an example for which the limiting behaviour $Y_L \mid (X_L > x_L)$ is not in the class of conditional extremes models by Keef, Papastathopoulos, and Tawn (2013) and (X_L, Y_L) does not lie in the domain of attraction of a multivariate extreme value distribution.

4 | IMPLICATIONS TO THE CONDITIONAL EXTREMES MODEL

In Section 3, we showed that some of the issues, raised by Examples 2.3 to 4.4 in Drees and Janßen (2017), can be resolved by standardisation to common Laplace margins. Here we discuss the practical implications of our results. As stated in Section 1, we believe that working with different marginal tail behaviours imposes a major restriction on a conditional extremes approach using affine transformations. Our calculations demonstrate that standarisation to common margins has benefits. In Example 2.3, this fixed choice of standardisation implied a unique limit measure $G_{|X}$, while the CEVM framework by Heffernan and Resnick (2007) allowed the limit measure $\mu_{Y|X>}$ to vary with the standardisation used. This example also highlighted that it is necessary to allow $G_{|X}$ to have mass at $\{-\infty\}$ in the Keef, Papastathopoulos, and Tawn (2013) framework, because the limit measure might otherwise be degenerate. Given these findings, we believe our results illustrate the versatility of the Heffernan and Tawn (2004) conditional multivariate extremes framework. Nevertheless, the examples of Drees and Janßen (2017) illustrate some statistical limitations of the Keef, Papastathopoulos, and Tawn (2013) framework even with standardised Laplace marginals. Two particular areas relate to handling mixture distributions for $G_{|X}$ and the choice of parametric families for the normalising functions $a_{|X}$ and $b_{|X}$. We discuss these in turn below.

Many of the examples of Drees and Janßen (2017) involved a mixture structure for (X, Y), and hence also for (X_L, Y_L). Although it was possible to identify normalising functions to give a non-degenerate $G_{|X}$ in these cases, it was no surprise that $G_{|X}$ was also a mixture distribution. From a statistical perspective the only complication with $G_{|X}$ having a mixture structure is when

$G_{|X}$ puts an atom of mass at $\{-\infty\}$; with Example 3.1 being the only example where $\{-\infty\}$ has mass zero. The complication with limiting mass at $\{-\infty\}$ is that at non-asymptotic levels of x_L this mass will be at a finite value with its precise value depending on the associated conditioning value, e.g., x_L in this set up. Statistical methods have recently been developed by Tendijck, Eastoe, Tawn, Randell, and Jonathan (2023) which extend the Heffernan–Tawn conditional extreme value model for handling exactly this situation.

Keef, Papastathopoulos, and Tawn (2013) propose parsimonious canonical parametric families for the normalising functions $a_{|X}(x) = \alpha x$ and $b_{|X}(x) = x^{\beta}$ which appears suitable for a wide range of published data applications. The examples in Drees and Janßen (2017) add to the list (first noted by Papastathopoulos and Tawn (2016)) of theoretical joint distributions for (X_L, Y_L) with normalising functions that lie outside the canonical class. The canonical families cannot be extended to cover these examples in a parsimonious way. So future research should identify if it is possible to quantify the errors that can arise from the inappropriate usage of the canonical families in these examples, with the error relating to the bias of estimated probabilities of extremes events for finite extrapolations.

5 | DISCUSSION

By investigating Examples 2.3 to 4.4 by Drees and Janßen (2017) throughout this paper, we found some interesting differences between the conditional extremes frameworks by Keef, Papastathopoulos, and Tawn (2013) and Heffernan and Resnick (2007). In Example 2.3, the standardisation to Laplace margins implied a unique limit measure $G_{|X}$, while the CEVM framework by Heffernan and Resnick (2007) allowed the limit measure $\mu_{Y|X>}$ to vary with the standardisation used. This example also highlighted that it is necessary to allow $G_{|X}$ to have mass at $\{-\infty\}$, because the limit measure might otherwise be degenerate. Consequently, while Drees and Janßen (2017) advocate for the condition $\mu_{Y|X>}\left(\{-\infty,\infty\} \times E^{(\gamma_X)}\right) = 0$ to ensure uniqueness, in our calculations it was sufficient to require $\lim_{z\to\infty} G_X(z) = 1$ for the conditional extremes model by Keef, Papastathopoulos, and Tawn (2013).

A non-degenerate limit measure $G_{|X}$ was found in Examples 3.1 and 3.2, however, the functions $a_{|X}$ and $b_{|X}$ in (5) were not of the simple parametric form of Keef, Papastathopoulos, and Tawn (2013). As mentioned in Section 4, it was already known that the canonical families by Keef, Papastathopoulos, and Tawn (2013) cannot cover all cases in a parsimonious way, and our results add to this set of examples. Example 3.1 further shows that standardisation to Laplace margins can result in a non-degenerate limit measure, despite there not existing a standardisation of the form by Das and Resnick (2011) for the CEVM framework. Conversely, Example 4.4 shows that a model may belong the CEVM class by Heffernan and Resnick (2007), but not fall in the class of conditional extreme models by Keef, Papastathopoulos, and Tawn (2013).

Example 4.2 in Drees and Janßen (2017) showed that (X, Y) being multivariate extreme value distributed is not sufficient for the distributions of $Y \mid (X > t)$ and $X \mid (Y > t)$ to be in the CEVM class as t approaches the upper end of X and Y respectively, while Example 4.4 illustrated that $X \mid (Y > t)$ and $Y \mid (X > t)$ being CEVM does not imply that the distribution of (X, Y) is in the domain of attraction of multivariate extreme value distributions. However, after standardisation to Laplace margins (X_L, Y_L), we show that both examples are ruled out as evidence against this equivalence when margins are standardised. So, the link between the conditional extremes models and the class of multivariate extreme value distributions remains an open research question.

FINANCIAL DISCLOSURE
None reported.

CONFLICT OF INTEREST
The authors declare no potential conflict of interests.

References

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

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag London.

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

Coles, S. G., & 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), 1–49.

Das, B., & Resnick, S. I. (2011). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, *17*(1), 226–252.

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

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

Drees, H., & Janßen, A. (2017). Conditional extreme value models: fallacies and pitfalls. *Extremes*, *20*(4), 777–805.

Engelke, S., & Hitz, A. S. (2020). Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society: Series B*, *82*(4), 871–932.

Ewans, K., & Jonathan, P. (2014). Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. *Journal of Marine Systems*, *130*, 124–130.

Gouldby, B., Wyncoll, D., Panzeri, M., Franklin, M., Hunt, T., Hames, D., … Pullen, T. (2017). Multivariate extreme value modelling of sea conditions around the coast of England. *Proceedings of the Institution of Civil Engineers - Maritime Engineering*, *170*(1), 3–20.

Gudendorf, G., & Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, *142*(12), 3073–3085.

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

Heffernan, J. E., & 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.

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

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC press.

Keef, C., Papastathopoulos, I., & 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.

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, *24*(1), 13–21.

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

Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer Verlag, New York.

Papastathopoulos, I., Strokorb, K., Tawn, J. A., & Butler, A. (2017). Extreme events of Markov chains. *Advances in Applied Probability*, *49*(1), 134–161. doi: 10.1017/apr.2016.82

Papastathopoulos, I., & Tawn, J. A. (2016). Conditioned limit laws for inverted max-stable processes. *Journal of Multivariate Analysis*, *150*, 214-228.

Paulo, M., van der Voet, H., Wood, J., Marion, G., & van Klaveren, J. (2006). Analysis of multivariate extreme intakes of food chemicals. *Food and Chemical Toxicology*, *44*(7), 994-1005.

Resnick, S. I., & Zeber, D. (2014). Transition kernels and the conditional extreme value model. *Extremes*, *17*(2), 263–287.

Simpson, E. S., & Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*, 100482.

Southworth, H., & Heffernan, J. E. (2012). Extreme value modelling of laboratory safety data from clinical studies. *Pharmaceutical Statistics*, *11*(5), 361–366.

Tendijck, S., Eastoe, E., Tawn, J., Randell, D., & Jonathan, P. (2023). Modelling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, *118*, 1373–1384.

Wadsworth, J. L., & Tawn, J. A. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, *51*, 100677.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., & Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B*, *79*(1), 149–175.

Winter, H. C., & Tawn, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes*, *20*(2), 393–415.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website. This material contains the calculations we decided not not to include in the main paper: Sections S1 and S2 provide supplementary material for Example 3, Section S3 concerns Example 4.2, and Section S4 considers the marginal distributions for Example 4.4.