Journal of Multivariate Analysis 115 (2013) 396–404

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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

Caroline Keef[a,*], Ioannis Papastathopoulos[b], Jonathan A. Tawn[b]

[a] JBA Consulting, South Barn, Broughton Hall, Skipton, North Yorkshire, BD23 3AE, UK
[b] Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK

ARTICLE INFO

Article history:
Received 19 January 2011
Available online 20 November 2012

AMS subject classifications:
62G32
62F30

ABSTRACT

A number of different approaches to study multivariate extremes have been developed. Arguably the most useful and flexible is the theory for the distribution of a vector variable given that one of its components is large. We build on the conditional approach of Heffernan and Tawn (2004) [13] for estimating this type of multivariate extreme property. Specifically we propose additional constraints for, and slight changes in, their model formulation. These changes in the method are aimed at overcoming complications that have been experienced with using the approach in terms of their modelling of negatively associated variables, parameter identifiability problems and drawing conditional inferences which are inconsistent with the marginal distributions. The benefits of the methods are illustrated using river flow data from two tributaries of the River Thames in the UK.

1. Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a continuous d-dimensional vector random variable and suppose that we have n independent and identically distributed realisations from $\boldsymbol{X}$. Multivariate extreme value theory concerns the characterisation and estimation of the extremes of the vector $\boldsymbol{X}$. Methods for univariate extremes are well established with theories and inference methods for the distributions of the maxima and exceedances of high thresholds (Leadbetter et al. [19,3]). Owing to the lack of natural ordering of vectors, a number of different approaches to study multivariate extremes have been developed (Barnett [1]). Arguably the most useful and flexible methods for multivariate extreme values are based on asymptotic theory for the distribution of a vector variable given that one of its components is large. For example in studying the risk of flooding over a river network then it is key to know the status of the rest of the system given that there is flooding at one point in the network. Emergency planners and the insurance industry need knowledge of the chance of multiple flooding events and which sites are likely to be flooded simultaneously.

The initial approaches for multivariate extreme value modelling and estimation were based on the limiting distribution of componentwise maxima, (Hall and Tajvidi [11], Capéraà et al. [2] and Naveau et al. [22]), or were derived under the assumption that the dependence structure is multivariate regular variation in a tail region of $\boldsymbol{X}$ (Einmahl et al. [9], Rootzén and Tajvidi [27] and Coles and Tawn [4,5]). These asymptotically motivated models imply that for every pair of variables they are either independent or satisfy a very strong form of dependence structure termed asymptotic dependence; see definition (1.2) below. However dating back to Sibuya [29], and Tiago de Oliveira [30] it has been known that there are two different

* Corresponding author.
E-mail addresses: carolinekeef@googlemail.com, caroline.keef@yorkshirewater.co.uk (C. Keef), i.papastathopoulos@lancaster.ac.uk (I. Papastathopoulos), j.tawn@lancaster.ac.uk (J.A. Tawn).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.10.012

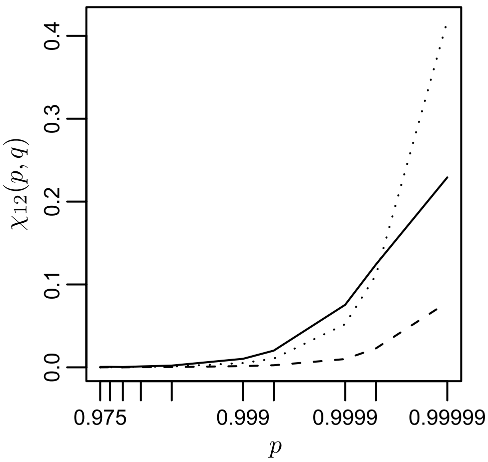

Fig. 1. Estimates of $\chi_{12}(p, q)$ using river flow data for the Pang and Windrush with $q = 0.99999$. The dotted, solid and dashed lines correspond respectively to estimates obtained using asymptotic positive dependence, and unconstrained and constrained Heffernan and Tawn model fit which is an asymptotically independent model.

forms of extremal dependence for a vector variable: asymptotic dependence and asymptotic independence, so these models cover only a very special case of the asymptotically independent class. For example all non-perfectly dependent multivariate normal variables are asymptotically independent and so their extremes are not well modelled by these approaches unless all correlations are equal to zero.

The asymptotically motivated approach of Heffernan and Tawn [13] provides the most flexible current approach for modelling the conditional distribution of $\boldsymbol{X}$ given X_i is large. This model encapsulates a broad range of asymptotic independence and asymptotic dependence forms of extremal dependence structure allowing a different type of dependence between the different pairs of variables. This paper overcomes a number of weaknesses of the Heffernan and Tawn [13] approach that have been identified by exploiting bounds on the quantiles of the conditional distribution for asymptotically independent distributions by imposing an ordering on them relative to the asymptotically dependent cases.

To illustrate this issue first consider the basic measure of dependence between the pair of variables (X_i, X_j) given by

$$\chi_{ij}(p, q) = \Pr\{X_j > F_j^{-1}(q) | X_i > F_i^{-1}(p)\} \quad \text{for } 0 < p < 1, 0 < q < 1 \tag{1.1}$$

where F_i and F_j are the marginal distributions of X_i and X_j respectively. In Fig. 1, for the bivariate river flow data that we analyse in Section 4, the estimated probability $\chi_{12}(p, q)$ is plotted over a range of p for $q = 0.99999$ under asymptotic dependence and the Heffernan and Tawn model fits. We see that the Heffernan and Tawn model estimates this probability to be larger than under the asymptotically dependence model for all $p < q$. A consequence of this lack of ordering is that the original Heffernan and Tawn estimates give simulated realisations from the distribution of $X_j|X_i$ that are inconsistent with the marginal distribution of X_j. The figure also shows our constrained estimate of $\chi_{12}(p, q)$ which has the required ordering for all p.

To understand the different types of dependence we first introduce two measures of extremal dependence $(\chi_{ij}^+, \chi_{ij}^-)$ between the pair (X_i, X_j), $i, j \in \Delta$ with $i \neq j$ and $\Delta = \{1, \ldots, d\}$. The limiting positive dependence between these variables is given by χ_{ij}^+ with

$$\chi_{ij}^+ = \lim_{p \to 1} \chi_{ij}(p, p), \tag{1.2}$$

where (X_i, X_j) are termed asymptotically positive dependent if $\chi_{ij}^+ > 0$ and asymptotically positive independent if $\chi_{ij}^+ = 0$. It is helpful to identify a further class of extremal dependence: let

$$\chi_{ij}^- = \lim_{p \to 1} \Pr\{X_j < F_j^{-1}(1-p) | X_i > F_i^{-1}(p)\} = \lim_{p \to 1} [1 - \chi_{ij}(1-p, p)],$$

then we define (X_i, X_j) to be asymptotically negative dependent if $\chi_{ij}^- > 0$ and asymptotically negative independent if $\chi_{ij}^- = 0$. To help with the interpretation of these tails probabilities Fig. 2 shows the joint events associated with the conditional probabilities $(\chi_{ij}^+, \chi_{ij}^-)$ coupled with the approximate joint probability of these events for large p.

The limiting distribution of componentwise maxima and multivariate regular varying models have the property that $\chi_{ij}^+ > 0$ and $\chi_{ij}^- = 0$ for all cases except when (X_i, X_j) are independent. The first extensions of the multivariate extreme value modelling toolbox to describe dependence structures which could include either asymptotic independence or asymptotic positive dependence date back to Ledford and Tawn [20,21] who looked at extremal dependence in a region of the joint tail where all components of the vector were extreme. Recent work on this approach includes Draisma et al. [7], Ramos

Fig. 2. The joint events associated with the conditional probabilities $(\chi_{ij}^{+}, \chi_{ij}^{-})$ shown by the top and bottom regions respectively. In each case the approximate joint probability of these events are stated for large p.

and Ledford [25], and Frick and Reiss [10]. Furthermore, this forms the basis of the idea of hidden regular variation; see Resnick [26]. However for asymptotically positive independent variables $\chi_{ij}(p, p) \approx 0$ for p near 1 and so these models describe unlikely realisations of X_j given a large X_i value. Therefore these events are not of primary interest when considering extreme events where X_i is large. More generally, events with all components of $\boldsymbol{X}$ simultaneously large are of limited interest for $d > 2$ when the components of $\boldsymbol{X}$ are asymptotically independent as their probability of occurring is an order of magnitude smaller than other multivariate events which are extreme in at least one component.

An alternative approach which focuses on more likely multivariate extreme events has been proposed by Heffernan and Tawn [13], and formalised by Heffernan and Resnick [12]. This method aims to describe the conditional distribution of $\boldsymbol{X}_{-i}|X_i > F_i^{-1}(p)$ as $p \to 1$, where $\boldsymbol{X}_{-i}$ is $\boldsymbol{X}$ with the X_i component removed. Heffernan and Tawn first transform $\boldsymbol{X}$ componentwise to follow Gumbel marginal distributions, i.e.

$$Y_i^{(G)} = -\log\{-\log F_i(X_i)\} \quad \text{for } i \in \Delta \tag{1.3}$$

where F_i is the marginal distribution function X_i. The Heffernan and Tawn model describes the distribution of the vector $\boldsymbol{Y}_{-i}^{(G)}|Y_i^{(G)} = y$ for large y, for $i \in \Delta$ based on an asymptotic representation through vector semi-parametric regression models. The Heffernan and Tawn model has been found to be helpful in large scale applications of multivariate extremes with $d > 100$ for riverflow and rainfall (Keef et al. [16]) and sea levels (Latham [18]), in temporal river flow cases (Eastoe and Tawn [8]), in food safety (Paulo et al. [24]), and in finance (Hilal et al. [14]). Despite such generally positive experiences in using the Heffernan and Tawn method three problems have emerged which limit the method's current usage.

Firstly there are complications with modelling variables with some components positively associated and some negatively associated. This is due to the semi-parametric regression model taking different functional forms in these two cases. We have identified that this problem arises due to the choice of the Gumbel distribution for the marginal distribution in which to apply the model. In this paper we propose transforming the marginal variables to follow Laplace distributions as an alternative to the Gumbel distribution. The Laplace distribution has both exponential tails and symmetry, this captures the exponential upper tail of the Gumbel required for modelling positive dependence but the symmetry also allows for negatively associated variables to be incorporated into the model parsimoniously. In Section 2 we present the Heffernan and Tawn model, based on the transformation to Laplace marginals (limited details are given, see Heffernan and Tawn for full details under Gumbel marginals), and explain the advantages of the use of Laplace marginals.

The other two problems are parameter identifiability and inferences which are inconsistent with the marginal distributions (a feature we term inconsistency). These features arise from the omission by Heffernan and Tawn of joint constraints on the parameters of the semi-parametric regression models and the non-parametric element of the model. Our main focus is introducing additional constraints on the Heffernan and Tawn model parameters to overcome these issues. In Section 3 we develop the additional constraints and in Section 4 we illustrate the benefit of these constraints for river flows on two tributaries of the River Thames in the UK for which the existing Heffernan and Tawn approach leads to inconsistent inferences.

2. The Heffernan and Tawn model with Laplace marginals

2.1. Marginal transformation

Instead of the transformation (1.3) to produce Gumbel marginals we let

$$Y_i = \begin{cases} \log\{2F_i(X_i)\} & \text{for } X_i < F_i^{-1}(0.5) \\ -\log\{2[1-F_i(X_i)]\} & \text{for } X_i \geq F_i^{-1}(0.5) \end{cases} \tag{2.1}$$

for $i \in \Delta$. Then $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ has Laplace marginal distributions with

$$\Pr(Y_i < y) = \begin{cases} \exp(y)/2 & \text{if } y < 0 \\ 1 - \exp(-y)/2 & \text{if } y \geq 0 \end{cases}$$

for all $i \in \Delta$. Thus both lower and upper tails of Y_i are exactly exponentially distributed and so for any $u > 0$ the distributions of $Y_i - u | Y_i > u$ and $-(Y_i + u) | Y_i < -u$ are both exponential with mean 1.

2.2. Dependence model

Here and throughout the vector algebra is to be interpreted as componentwise. The model is motivated by the relatively weak assumption that for each $i \in \Delta$, there exist vector-valued normalising functions, $\boldsymbol{a}_{|i}(y)$, such that $\mathbb{R}_+ \to \mathbb{R}^{d-1}$ and $\boldsymbol{b}_{|i}(y)$ with $\mathbb{R}_+ \to \mathbb{R}_+^{d-1}$, such that for $x > 0$

$$\Pr\left(Y_i - u > x, \frac{\boldsymbol{Y}_{-i} - \boldsymbol{a}_{|i}(Y_i)}{\boldsymbol{b}_{|i}(Y_i)} \leq \boldsymbol{z} \mid Y_i > u\right) \to \exp(-x) G_{|i}(\boldsymbol{z}) \quad \text{as } u \to \infty, \tag{2.2}$$

where the jth marginal distribution $G_{j|i}$ of $G_{|i}$ is a non-degenerate distribution function for all $j \in \Delta$. To ensure that $G_{|i}$ is well-defined the following additional condition is required:

$$\lim_{z \to \infty} G_{j|i}(z) = 1 \quad \text{for all } j \neq i,$$

so there is no mass at $+\infty$ in any margin but some mass is allowed at $-\infty$. A consequence of assumption (2.2) is that conditionally on $Y_i > u$, as $u \to \infty$ then $Y_i - u$ and $\{\boldsymbol{Y}_{-i} - \boldsymbol{a}_{|i}(Y_i)\}/\boldsymbol{b}_{|i}(Y_i)$ converge to independent random variables with the former following an exponential distribution and the latter having joint distribution function $G_{|i}$. No simple class of parametric models exists for $G_{|i}$. The formulation of the Heffernan and Tawn model in expression (2.2) comes from Heffernan and Resnick [12], which represents a reformulation of the model in terms of conditional distribution functions instead of the conditional density functions used in Heffernan and Tawn [13].

Heffernan and Tawn [13] found that for all the standard copula dependence models studied by Joe [15] and Nelsen [23] the form of the equivalent norming functions $\boldsymbol{a}_{|i}^{(G)}(y)$ and $\boldsymbol{b}_{|i}^{(G)}(y)$, found when using Gumbel margins $\boldsymbol{Y}^{(G)}$, fell into a simple class for positively associated variables and a different more complex class for negatively associated variables namely

$$\boldsymbol{a}_{|i}^{(G)}(y) = \boldsymbol{\alpha}_i^{(G)} y + I_{\{\boldsymbol{\alpha}_i^{(G)} = \boldsymbol{0}, \boldsymbol{\beta}_i^{(G)} < 0\}} \left[\boldsymbol{\gamma}_i^{(G)} - \boldsymbol{\delta}_i^{(G)} \log(y)\right] \quad \text{and} \quad \boldsymbol{b}_{|i}^{(G)}(y) = y^{\boldsymbol{\beta}_i^{(G)}}$$

with $\boldsymbol{0} \leq \boldsymbol{\alpha}_i^{(G)} \leq \boldsymbol{1}$, $-\boldsymbol{\infty} < \boldsymbol{\beta}_i^{(G)} < \boldsymbol{1}$, $-\boldsymbol{\infty} < \boldsymbol{\gamma}_i^{(G)} < \boldsymbol{\infty}$, $\boldsymbol{0} \leq \boldsymbol{\delta}_i^{(G)} \leq \boldsymbol{1}$ and I the indicator function. Using our alternative transformation to Laplace marginal distributions a single unified class is found to be

$$\boldsymbol{a}_{|i}(x) = \boldsymbol{\alpha}_i x \quad \text{and} \quad \boldsymbol{b}_{|i}(x) = x^{\boldsymbol{\beta}_i} \tag{2.3}$$

with $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) \in [-1, 1]^{d-1} \times (-\infty, 1)^{d-1}$. With $\alpha_{j|i}$, the component of $\boldsymbol{\alpha}_i$ associated with Y_j, then $0 < \alpha_{j|i} \leq 1$ and $-1 \leq \alpha_{j|i} < 0$ correspond respectively to positive and negative association of Y_j and large Y_i. This unification of the parametric forms simplifies model fitting and inferences, particularly for weakly associated variables. For positively associated variables the use of Laplace margins does not effect the limiting dependence model of Heffernan and Tawn [13] since the upper tail of the Laplace distribution is in the same domain of attraction as the Gumbel distribution; see Heffernan and Resnick [12]. In contrast to the Gumbel distribution the symmetry of the Laplace distribution ensures the limiting dependence model is unchanged simply by inverting the copula to produce a model for negatively associated variables; see Section 2.3.

The statistical model is based on the approximation that limiting relationship (2.2) holds exactly for all values of $Y_i > u$ for a suitably high threshold u. A consequence of this assumption, together with the findings about the possible structure of $\boldsymbol{a}_{|i}(y)$ and $\boldsymbol{b}_{|i}(y)$ functions is that given Y_i, with $Y_i > u$,

$$\boldsymbol{Y}_{-i} = \boldsymbol{\alpha}_i Y_i + (Y_i)^{\boldsymbol{\beta}_i} \boldsymbol{Z}_{|i} \tag{2.4}$$

where $\boldsymbol{Z}_{|i}$ is a $d-1$ dimensional random variable with distribution function $G_{|i}$ and is independent of Y_i. Thus expression (2.4) represents the model as a multivariate regression with $\boldsymbol{Z}_{|i}$ a multivariate residual term, however the marginal variables of $\boldsymbol{Z}_{|i}$ are not required to have zero mean.

2.3. Theoretical examples

To help clarify the interpretation of the model parameters and the parameter space it is helpful to recall five examples derived in detail in Section 8 of Heffernan and Tawn [13] and also to consider the case of asymptotic negative dependence.

Independence. Here $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i = \boldsymbol{0}$ and $G_{|i}$ factorises into $d-1$ Laplace distribution functions.
Asymptotic positive dependence. Here $\boldsymbol{\alpha}_i = \boldsymbol{1}$ and $\boldsymbol{\beta}_i = \boldsymbol{0}$ and $G_{|i}$ takes a range of forms.

Asymptotic negative dependence. Here $\boldsymbol{\alpha}_i = -\mathbf{1}$ and $\boldsymbol{\beta}_i = \mathbf{0}$ and $G_{|i}$ takes a range of forms.

Asymptotic independence. Variable Y_j is asymptotically independent of variable Y_i if $-1 < \alpha_{j|i} < 1$, where $\alpha_{j|i}$ is the element of $\boldsymbol{\alpha}_{|i}$ associated with Y_j.

Gaussian copula. Under this model, with parameter $\rho_{ij} \neq 0$ with $-1 < \rho_{ij} < 1$, corresponding the correlation parameter between variables (Y_i, Y_j) when transformed to having Gaussian marginal, then for Laplace marginal distributions $\alpha_{j|i} = \text{sign}(\rho_{ij})\rho_{ij}^2$ and $\beta_{j|i} = 1/2$. With Gumbel marginals (Heffernan and Tawn [13]) obtained a similar form for $\rho_{ij} > 0$ but for $\rho_{ij} < 0$ they obtained $a_{j|i}^{(G)}(y) = -\log(\rho_{ij}^2 y)\rho_{ij}^2$ and $\beta_{j|i}^G = -1/2$.

Inverted multivariate extreme value copula. Here $\boldsymbol{\alpha}_i = \mathbf{0}$ and $\mathbf{0} < \boldsymbol{\beta}_i < \mathbf{1}$, with the precise value of $\boldsymbol{\beta}_i$ determined by the tail features of the spectral measure of the multivariate extreme value distribution.

These examples illustrate that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ can take any value in their marginal ranges as specified by expression (2.3). However, the strongest form of extremal dependence is asymptotic positive dependence and so this suggests that when $\boldsymbol{\alpha}_i = \mathbf{1}$ no element of $\boldsymbol{\beta}_i$ can be positive. However, this combination of parameters is allowed in the model as specified by Heffernan and Tawn [13] as the parameter space was taken as a cross product over the spaces for $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$.

2.4. Inference

Assume now that we have independent and identically distributed observations on $\boldsymbol{X}$. Like Heffernan and Tawn [13] we estimate each marginal distribution function F_i, $i \in \Delta$ using an empirical distribution function below some high threshold level for X_i and a generalised Pareto distribution (Davison and Smith [6]) above this threshold level. We then transform to have approximately Laplace marginals using transformation (2.1) with F_i replaced by its estimate.

Inference for the parametric part of the model exploits property (2.4). If $\boldsymbol{Z}_{|i}$, has mean vector $\boldsymbol{\mu}_i$ and vector of standard deviations $\boldsymbol{\sigma}_i$ the respective conditional mean and standard deviation vectors of $\boldsymbol{Y}_{-i} \mid Y_i$, for $Y_i > u$, are $\boldsymbol{\alpha}_i Y_i + \boldsymbol{\mu}_i (Y_i)^{\boldsymbol{\beta}_i}$ and $\boldsymbol{\sigma}_i (Y_i)^{\boldsymbol{\beta}_i}$ respectively. Likelihood methods are used to jointly estimate the parameters of interest $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$ and nuisance parameters $(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i)$. We explored making two different false working assumptions in constructing the likelihood: that $\boldsymbol{Z}_{|i}$ are independent Normal random variables and that $\boldsymbol{Z}_{|i}$ are independent Laplace random variables. From a broad range of simulated examples we found that the former was the better and so we adopt that throughout. Numerical maximisation of the likelihood function over the parameter space of the model is required to get parameter estimates $(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\sigma}}_i)$. Finally the distribution $G_{|i}$ is estimated nonparametrically using the empirical joint distribution of observations of

$$\boldsymbol{Z}_{|i} = \frac{\boldsymbol{Y}_{-i} - \hat{\boldsymbol{\alpha}}_i Y_i}{Y_i^{\hat{\boldsymbol{\beta}}_i}} \tag{2.5}$$

for $Y_i > u$. Estimation of probabilities of multivariate extreme events with $Y_i > u$ are obtained from the model by Monte Carlo methods exploiting the conditional independence property (2.2) of $Y_i - u$ and $\boldsymbol{Z}_{|i}$ given $Y_i > u$ to construct samples of $\boldsymbol{Y} = (Y_i, \boldsymbol{Y}_{-i})$ with $Y_i > u$. A nonparametric bootstrap procedure provides an assessment of the uncertainty of the model. Specifically this bootstrap procedure involves resampling from the original data so that the effect of both the marginal transformation to Laplace margins and Heffernan and Tawn conditional dependence model fit are accounted for.

3. New constraints

3.1. Motivation and theory

The constraints we require are determined entirely by consideration of the bivariate marginals, thus throughout this section we focus on the pair (Y_i, Y_j), with the results applying for all $i, j \in \Delta$.

First note that if a pair of variables (Y_i, Y_j) is asymptotically positive dependent then $\chi_{ij}^+ > 0$ and hence larger than if the pair were asymptotically positive independent, as then $\chi_{ij}^+ = 0$. This ordering holds whatever the form of asymptotic independence. Similar arguments hold using the limiting measure χ_{ij}^-, with asymptotically negative independence giving $\chi_{ij}^- = 0$ and asymptotic negative dependence giving $\chi_{ij}^- > 0$.

Statistical methods based on extreme value theory rely on applying asymptotic properties above some high threshold. Therefore these bounds implied by asymptotic dependence need to be imposed on the asymptotically independent class. Considering the general conditional distribution of $Y_j|Y_i = y$ the same properties must hold for all large y. Without imposing the bounds the resulting estimated joint probabilities can exceed the marginal probabilities, *i.e.*,

$$\hat{\Pr}\{X_i > F_i^{-1}(p), X_j > F_j^{-1}(q)\} > \max(1-p, 1-q)$$

for some $0 < p < 1$ and $0 < q < 1$. This implies that we require a stochastic ordering of conditional distributions associated with asymptotic negative dependence, asymptotic independence and asymptotic positive dependence. Consequently conditional quantiles for any form of asymptotic independence cannot be larger than under asymptotic positive dependence nor can it be smaller than under asymptotic negative dependence. Imposing this feature of the conditional distribution leads to our constraints.

To be specific, let the qth ($0 \le q \le 1$) conditional quantile of $Y_j|Y_i = y$ for large y under the Heffernan and Tawn [13] model be $y_{j|i}(q)$ and the associated quantile under the assumptions of asymptotic positive dependence and asymptotic negative dependence be $y^+_{j|i}(q)$ and $y^-_{j|i}(q)$ respectively. Then

$$y^-_{j|i}(q) \le y_{j|i}(q) \le y^+_{j|i}(q). \tag{3.1}$$

Here

$$y^-_{j|i}(q) = -y + z^-_{j|i}(q) \quad \text{and} \quad y_{j|i}(q) = \alpha_{j|i}y + y^{\beta_{j|i}}z_{j|i}(q) \quad \text{and} \quad y^+_{j|i}(q) = y + z^+_{j|i}(q),$$

where $\tilde{G}^-_{j|i}\{z^-_{j|i}(q)\} = \tilde{G}_{j|i}\{z_{j|i}(q)\} = \tilde{G}^+_{j|i}\{z^+_{j|i}(q)\} = q$ with $\tilde{G}^-_{j|i}$, $\tilde{G}_{j|i}$ and $\tilde{G}^+_{j|i}$ are the estimated distributions of $\boldsymbol{Z}_{|i}$ for $Y_i > u$ under the assumption of asymptotic negative dependence, asymptotic independence and asymptotic positive dependence respectively and $(\alpha_{j|i}, \beta_{j|i})$ are the components of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ corresponding to Y_j. We impose the ordering constraint (3.1) for all $y > v$ where the choice of the level v, $v \ge u$, is discussed in Section 3.2.

Theorem 1. *For $v \ge u$, the ordering constraint (3.1) holds for all $y > v$ if and only if both Cases* I *and* II *hold.*

Case I*: either*

$$\alpha_{j|i} \le \min\left\{1,\, 1 - \beta_{j|i}z_{j|i}(q)v^{\beta_{j|i}-1},\, 1 - v^{\beta_{j|i}-1}z_{j|i}(q) + v^{-1}z^+_{j|i}(q)\right\}$$

or

$$1 - \beta_{j|i}z_{j|i}(q)v^{\beta_{j|i}-1} < \alpha_{j|i} \le 1 \quad \textit{and} \quad (1 - \beta^{-1}_{j|i})\{\beta_{j|i}z_{j|i}(q)\}^{1/(1-\beta_{j|i})}(1 - \alpha_{j|i})^{-\beta_{j|i}/(1-\beta_{j|i})} + z^+_{j|i}(q) > 0.$$

Case II*: either*

$$-\alpha_{j|i} \le \min\left\{1,\, 1 + \beta_{j|i}v^{\beta_{j|i}-1}z_{j|i}(q),\, 1 + v^{\beta_{j|i}-1}z_{j|i(q)} - v^{-1}z^-_{j|i}(q)\right\}$$

or

$$1 + \beta_{j|i}v^{\beta_{j|i}-1}z_{j|i}(q) < -\alpha_{j|i} \le 1 \quad \textit{and} \quad (1 - \beta^{-1}_{j|i})(-\beta_{j|i}z_{j|i}(q))^{1/(1-\beta_{j|i})}(1 + \alpha_{j|i})^{-\beta_{j|i}/(1-\beta_{j|i})} - z^-_{j|i}(q) > 0.$$

Proof. Let

$$D_+(y) = y^+_{j|i}(q) - y_{j|i}(q)$$

then the ordering holds for all $y \ge v$ if and only if either $D'_+(v) > 0$ and $D_+(v) > 0$ or if $D'_+(v) < 0$, there exists an $s > v$ such that $D'_+(s) = 0$ and $D_+(s) > 0$. The constraints follow directly from these conditions. Similar issues apply for $D_-(y) = y_{j|i}(q) - y^-_{j|i}(q)$ $\square$

3.2. Practicalities

To apply Theorem 1 to give constraints on a fit of the semi-parametric model of Heffernan and Tawn [13] we need to clarify how we estimate $z^-_{j|i}(q)$, $z_{j|i}(q)$ and $z^+_{j|i}(q)$ and determine which values of q and v to use.

Under the assumption that the asymptotically dependent model is appropriate for (Y_i, Y_j) with $Y_i > u$ we estimate $z^+_{j|i}(q)$ as the empirical qth quantile of the residual variable $Z^+_{j|i} = Y_j - Y_i$ for $Y_i > u$. Similarly, under the assumption that the asymptotically negative model we estimate $z^-_{j|i}(q)$ as the empirical qth quantile of the residual variable $Z^-_{j|i} = Y_j + Y_i$ for $Y_i > u$. Finally, under the Heffernan and Tawn [13] model for parameters $(\alpha_{j|i}, \beta_{j|i})$ we estimate $z_{j|i}(q)$ as the empirical qth quantile of the residual variable $Z_{j|i} = (Y_j - \alpha_{j|i}Y_i)/Y_i^{\beta_{j|i}}$ for $Y_i > u$. In all cases we use the standard R software sample quantile function, R Core Team [28].

The conditions of Theorem 1 are a function of both the quantile q and the level v. For the required stochastic ordering constraint (3.1) through extensive examples we found that both Cases I and II conditions were satisfied for all q if they were each satisfied for both $q = 0$ and 1. The impact of these conditions on the parameter space is explored for the application, with the effect on the parameter space reflected in the line $\alpha_{j|i} = 0$ for positively and negatively associated extremal dependence.

To give the greatest flexibility to the fits we only impose the constraints on extrapolations. Thus we take v to be a value above the maximum observed value of Y_i. We found very little sensitivity to the precise choice of v in this range over a number of unreported examples. In contrast, when v was taken lower, *e.g.* $v = u$, we found that the resulting constrained model fit to be poor.

The constraints are built into the inference as follows. The profile likelihood for $(\alpha_{j|i}, \beta_{j|i})$, obtained by maximising the likelihood referred to in Section 2.4 over the nuisance parameters $(\mu_{j|i}, \sigma_{j|i})$, is now zero if the conditions of Theorem 1

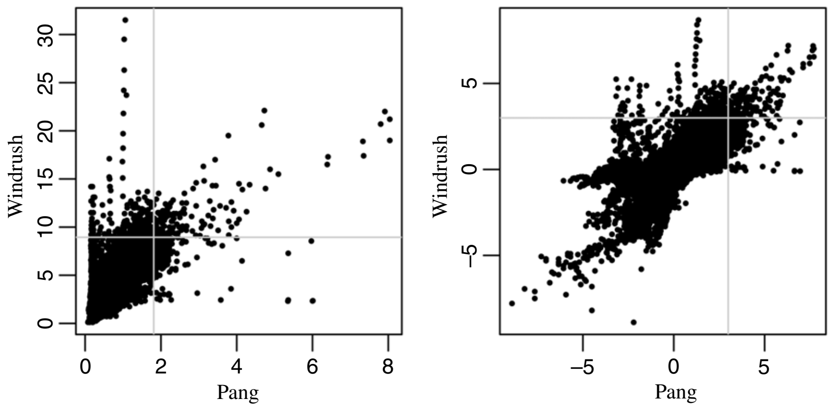

Fig. 3. Scatter plot of river flow data from the Pang and Windrush. The marginal threshold level, corresponding to the 0.975 quantile is shown in grey. Data are shown on original marginals (left plot) and Laplace marginals (right plot).

fail to hold. These conditions restrict the feasible parameter combinations to those which lead to conditional quantiles that are higher and lower than if an asymptotically negative dependent or asymptotically positive dependent model is fitted respectively. This leads to a smaller set of feasible parameters and hence as well as removing inconsistent estimates these constraints also reduce the variance of estimators of any feature derived from the model.

3.3. Simulation study

Previous simulation studies (Keef et al. [17], Heffernan and Tawn [13]) have assessed the ability of the original inference method for estimating the Heffernan and Tawn parameters for data with different forms of extremal dependence. To assess the benefits of imposing joint constraints on the parameter estimates we simply compare estimates made by imposing constraints with those without constraints for data simulated from the Heffernan and Tawn model. The method we use to simulate data is that described in Keef et al. [17] the steps are as follows.

1. Generate a value, Y_i^* for this site conditionally on $Y_i^* > u$, so $Y_i^* - u$ follows a mean one exponential distribution.
2. Simulate $\boldsymbol{Z}_{|i}^*$, from a normal distribution zero mean, standard deviation 1, independently of Y_i^*.
3. Using the dependence parameters $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$ set $Y_j^* = \alpha_{j|i} Y_i^* + (Y_i^*)^{\beta}_{j|i} Z_{j|i}^*$ for $j \in \Delta \setminus \{i\}$.

We repeated this process 10,000 times, and simulated 45 observations for each replication this number of observations approximates the length of record typically observed for environmental data, the threshold, u, in the simulation process was the 0.99 quantile of the Laplace distribution. We repeated the process for $\boldsymbol{\alpha}_i = 0.7$, and $\boldsymbol{\beta}_i = 0.3$ and $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i = 0.1$ representing high and low dependent variables. In estimating the parameters we chose v to be the 0.999 quantile of the Laplace distribution. To assess the benefit of using the constraints we use the efficiency, $RMSE_{new}/RMSE_{old}$ where $RMSE_{old}$, $RMSE_{new}$ denote the root-mean-squared error of the estimates without and with imposing the new constraints. When $\boldsymbol{\alpha}_i = 0.7$, and $\boldsymbol{\beta}_i = 0.3$ the efficiency in the $\boldsymbol{\alpha}_i$ estimate was 0.700 and the efficiency in the $\boldsymbol{\beta}_i$ estimate was 0.985, for $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i = 0.1$ the efficiency in the $\boldsymbol{\alpha}_i$ estimate was 0.699 and the efficiency in the $\boldsymbol{\beta}_i$ estimate was 0.998. These efficiencies are mainly due to reduced variance in the estimates. These results show the improvement in the estimation procedure made by including the joint constraints. This improvement is in addition to that made by removing the possibility of inconsistent parameter estimates.

4. A hydrological application

We analyse daily river flow, measured in $m^3\ s^{-1}$, for a pair sites the Pang and Windrush, which are tributaries of River Thames in the UK. We have 40 years of data from 1968 to 2008. Fig. 3 shows the scatter plot for the data on the original and Laplace marginals. When the flow in the Pang is extreme for most occasions the flow for the Windrush is also extreme but there are a number of events when the Windrush flow is not extreme at all. In Laplace margins (Y_1, Y_2) this suggests that the fitted conditional distribution of $Y_2|Y_1$ should have a large spread for large values of Y_1 which implies that $\beta_{2|1} > 0$.

The modelling threshold u is taken to be the 0.975 marginal quantile. When the original Heffernan and Tawn [13] approach for inference is used and the joint constraints we have obtained on $\alpha_{2|1}$ and $\beta_{2|1}$ are not imposed we obtain $\hat{\alpha}_{2|1} = 0.515$ and $\hat{\beta}_{2|1} = 1.000$. The large value of $\hat{\beta}_{2|1}$, on the boundary of the parameter space, reflects the observed increase in variation in the conditional distribution as the value of Y_1 grows. In the left plot of Fig. 4 we show that although the conditional median is lower than under asymptotic positive dependence the associated estimated conditional 0.975 quantile is larger for this fitted model than when the variables are assumed to be asymptotically positive dependent. This

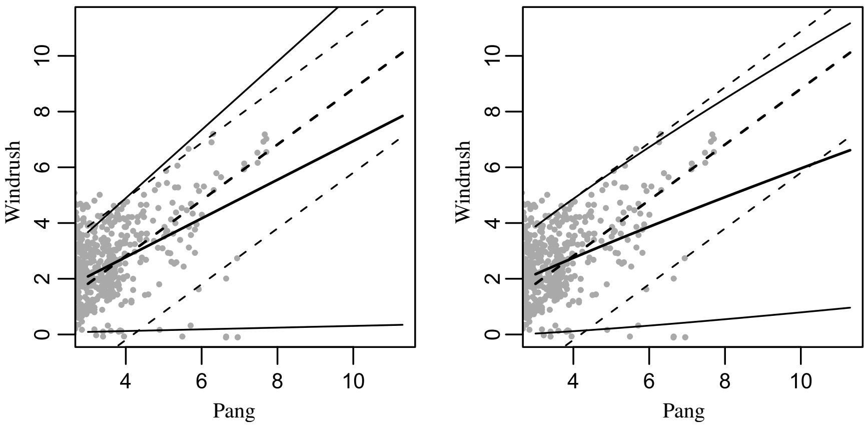

Fig. 4. Estimated 0.025, 0.5 and 0.975 conditional quantiles of $Y_2|Y_1 = y$ using the Heffernan and Tawn model for a range of y above the threshold used for fitting the model. In both plots the dashed and solid lines correspond to the estimates under asymptotic positive dependence and asymptotic independence respectively. The asymptotic independent model fit is unconstrained and constrained in the left and right plots respectively. The grey dots are the data points, on a Laplace marginal scale, which exceed the marginal threshold.

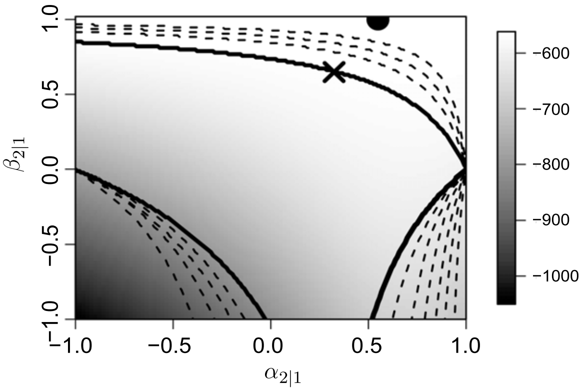

Fig. 5. Profile log-likelihood surface for $(\alpha_{2|1}, \beta_{2|1})$. The solid curves show the boundary of parameter space under the constraints of Theorem 1. Dashed curves show the constraints of Theorem 1 when $0 < q < 1$, showing these constraints are less restrictive than when $q = 0$ and $q = 1$. The dot and cross show estimated parameters for unconstrained and constrained estimation respectively.

leads to the contradiction identified in Fig. 1 with the estimate of probability $\chi_{12}(p, q)$ being larger under asymptotic positive independence than under asymptotic positive dependence.

Fig. 5 shows the profile log-likelihood surface for combinations of $\alpha_{2|1}$ and $\beta_{2|1}$, these were obtained by fixing $\alpha_{2|1}$ and $\beta_{2|1}$ and maximising the likelihood over $\mu_{2|1}$ and $\sigma_{2|1}$. It is clear that the profile log-likelihood surface is relatively flat around $(\hat{\alpha}_{2|1}, \hat{\beta}_{2|1})$. The effect of imposing the constraints of Theorem 1 restricts the feasible set of $(\alpha_{2|1}, \beta_{2|1})$ as shown in Fig. 5. Case I excludes combinations in the parameter space for large $\beta_{j|i}$ and when both $\alpha_{j|i} > 0.5$ and $\beta_{j|i} < 0$: specifically the top and the bottom-right solid curves mark the boundaries of the parameter space implied by $q = 1$ and $q = 0$ respectively. Case II excludes combinations in the parameter space for $\alpha_{j|i} < 0$ and $\beta_{j|i} < 0$: specifically the bottom-left solid curve marks the boundary of the parameter space implied by $q = 1$. The collective bounds to the parameter space exclude the unconstrained estimated parameter combination from the feasible parameter space. As expected, this boundary to the feasible region goes through $(\alpha_{2|1}, \beta_{2|1}) = (0, 1)$ but its other features are non-generic, depending on values on $z^-_{j|i}(q)$, $z_{j|i}(q)$ and $z^+_{j|i}(q)$ which vary with data-set.

Under these constraints the updated parameter estimates are $(\tilde{\alpha}_{2|1}, \tilde{\beta}_{2|1}) = (0.380, 0.656)$. In the right plot of Fig. 4 we show that using this parameter combination has addressed the inconsistent ordering of conditional quantiles identified with the unconstrained estimate, with the conditional quantiles under the fitted Heffernan and Tawn model below the equivalent conditional quantile under asymptotic positive dependence. Specifically although the spread of the distribution of $Y_2|Y_1 = y$ is still large for higher values of Y_1 the upper conditional quantiles are now much lower. Further note that this constrained fit removes the inconsistency in the ordering of estimates of probability $\chi_{12}(p, q)$ under asymptotic independence and asymptotic positive dependence for all $p < q$ as shown in Fig. 1.

Fig. 6 shows confidence intervals for conditional quantiles for the asymptotically independent model fitted with and without constraints. The confidence intervals, obtained using a block bootstrap, show that the upper endpoint of the

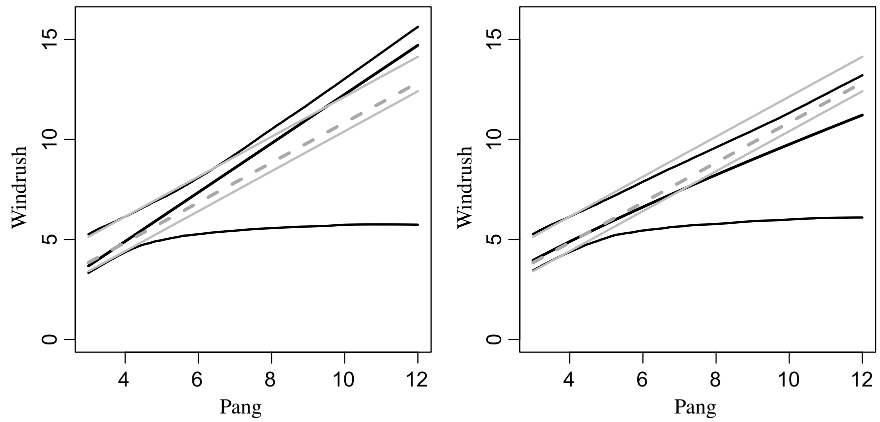

Fig. 6. Estimates and associated 95% confidence intervals for the 0.975 quantile of $Y_2|Y_1 = y$. In each plot grey lines shows the estimate and 95% confidence intervals under asymptotic positive dependence whereas the black lines show estimate and 95% confidence intervals for the fitted Heffernan and Tawn model. Left and right plots show unconstrained and constrained model fits for the Heffernan and Tawn model.

confidence interval for the constrained fit is lower than the point estimate for the unconstrained fit, but the large overlap on confidence intervals suggests that the imposition of the constraints has not substantially changed the fit other than removing the inconsistency issue.

Acknowledgment

We thank a referee for helpful comments that improved the accessibility of the paper.

References

[1] V. Barnett, The ordering of multivariate data (with discussion), J. Roy. Statist. Soc. A 139 (1976) 318–355.
[2] P. Capéraà, A.-L. Fougères, C. Genest, Bivariate distributions with given extreme value attractor, J. Multivariate Anal. 72 (2000) 30–49.
[3] S.G. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, London, 2001.
[4] S.G. Coles, J.A. Tawn, Modelling extreme multivariate events, J. Roy. Statist. Soc. B 53 (1991) 377–392.
[5] S.G. Coles, J.A. Tawn, Statistical methods for multivariate extremes: an application to structural design (with discussion), Appl. Statist. 43 (1994) 1–48.
[6] A.C. Davison, R.L. Smith, Models for exceedances over high thresholds, J. Roy. Statist. Soc. B 52 (1990) 393–442.
[7] G. Draisma, H. Drees, A. Ferreira, L. de Haan, Bivariate tail estimation: dependence in asymptotic independence, Bernoulli 10 (2004) 251–280.
[8] E.F. Eastoe, J.A. Tawn, The distribution for the cluster maxima of exceedances of sub-asymptotic thresholds, Biometrika 99 (2012) 43–55.
[9] J. Einmahl, L. de Haan, V. Piterbarg, Nonparametric estimation of the spectral measure of an extreme value distribution, Ann. Statist. 29 (2001) 1401–1423.
[10] M. Frick, R.-D. Reiss, Expansions of multivariate Pickands densities and testing the tail dependence, J. Multivariate Anal. 100 (2009) 1168–1181.
[11] P. Hall, N. Tajvidi, Distribution and dependence-function estimation for bivariate extreme-value distributions, Bernoulli 6 (2000) 835–844.
[12] J.E. Heffernan, S.I. Resnick, Limit laws for random vectors with an extreme component, Ann. Appl. Probab. 17 (2007) 537–571.
[13] J.E. Heffernan, J.A. Tawn, A conditional approach for multivariate extreme values (with discussion), J. Roy. Statist. Soc. B 66 (2004) 1–34.
[14] S. Hilal, S.-H. Poon, J.A. Tawn, Hedging the Black Swan: conditional heteroskedasticity and tail dependence in S&P500 and VIX, J. Bank. Finan. 35 (2011) 2374–2387.
[15] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, London, 1997.
[16] C. Keef, C. Svenson, J.A. Tawn, Spatial dependence in extreme river flows and precipitation for Great Britain, J. Hydrol. 378 (2009) 240–252.
[17] C. Keef, J.A. Tawn, C. Svenson, Spatial risk assessment for extreme river flows, Appl. Statist. 58 (2009) 601–618.
[18] M. Latham, Statistical methodology for the extreme values of dependent processes, Ph.D. Thesis, Lancaster University, 2006.
[19] M.R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and Related Properties of Random Sequences and Series, Springer-Verlag, New York, 1983.
[20] A.W. Ledford, J.A. Tawn, Statistics for near independence in multivariate extreme values, Biometrika 83 (1996) 169–187.
[21] A.W. Ledford, J.A. Tawn, Modelling dependence within joint tail regions, J. Roy. Statist. Soc. B 59 (1997) 475–499.
[22] P. Naveau, A. Guillou, D. Cooley, J. Diebolt, Modeling pairwise dependence of maxima in space, Biometrika 96 (2009) 1–17.
[23] R.B. Nelsen, An Introduction to Copulas, Springer-Verlag, New York, 1999.
[24] M. Paulo, H. van der Voet, J. Wood, G. Marion, J. van Klaveren, Analysis of multivariate extreme intakes of food chemicals, Food Chem. Toxicol. 44 (2006) 994–1005.
[25] A. Ramos, A.W. Ledford, A new class of models for bivariate joint tails, J. Roy. Statist. Soc. B 71 (2009) 219–241.
[26] S.I. Resnick, Hidden regular variation, second order regular variation and asymptotic independence, Extremes 5 (2002) 303–336.
[27] H. Rootzén, Tajvidi, The multivariate generalized pareto distribution, Bernoulli 12 (2006) 917–930.
[28] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0. URL: http://www.R-project.org/.
[29] M. Sibuya, Bivariate extreme statistics, Ann. Inst. Statist. Math. 11 (1960) 195–210.
[30] J. Tiago de Oliveira, Structure theory of bivariate extremes; extensions, Est. Mat., Estat. e. Econ. 7 (1962/63) 165–295.