J. R. Statist. Soc. B (2004)
66, *Part* 3, *pp.* 497–546

A conditional approach for multivariate extreme values

Janet E. Heffernan and Jonathan A. Tawn

Lancaster University, UK

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 15th, 2003*, Professor J. T. Kent *in the Chair*]

Summary. Multivariate extreme value theory and methods concern the characterization, estimation and extrapolation of the joint tail of the distribution of a d-dimensional random variable. Existing approaches are based on limiting arguments in which all components of the variable become large at the same rate. This limit approach is inappropriate when the extreme values of all the variables are unlikely to occur together or when interest is in regions of the support of the joint distribution where only a subset of components is extreme. In practice this restricts existing methods to applications where d is typically 2 or 3. Under an assumption about the asymptotic form of the joint distribution of a d-dimensional random variable conditional on its having an extreme component, we develop an entirely new semiparametric approach which overcomes these existing restrictions and can be applied to problems of any dimension. We demonstrate the performance of our approach and its advantages over existing methods by using theoretical examples and simulation studies. The approach is used to analyse air pollution data and reveals complex extremal dependence behaviour that is consistent with scientific understanding of the process. We find that the dependence structure exhibits marked seasonality, with extremal dependence between some pollutants being significantly greater than the dependence at non-extreme levels.

Keywords: Air pollution; Asymptotic independence; Bootstrap; Conditional distribution; Gaussian estimation; Multivariate extreme value theory; Semiparametric modelling

1. Introduction and background

Multivariate extreme value theory and methods concern the characterization, estimation and extrapolation of the joint tails of multidimensional distributions. Accurate assessments of the probabilities of extreme events are sought in a diversity of applications from environmental impact assessment (Coles and Tawn, 1994; Joe, 1994; de Haan and de Ronde, 1998; Schlather and Tawn, 2003) to financial risk management (Embrechts *et al.*, 1997; Longin, 2000; Stărică, 2000; Poon *et al.*, 2004) and Internet traffic modelling (Maulik *et al.*, 2002; Resnick and Rootzén, 2000). The application that is considered in this paper is environmental. We examine five-dimensional air quality monitoring data comprising a series of measurements of ground level *ozone* (O_3), *nitrogen dioxide* (NO_2), *nitrogen oxide* (NO), *sulphur dioxide* (SO_2) and *particulate matter* (PM_{10}), in Leeds city centre, UK, during the years 1994–1998 inclusively.

Regulation of air pollutants is undertaken because of their well-established deleterious effects on human health, vegetation and materials. Government objectives for concentrations of air pollutants are given in terms of single variables, rather than combinations of variables (Department of the Environment, Transport and the Regions, 2000). However, atmospheric chemists are

Address for correspondence: Jonathan A. Tawn, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.
E-mail: j.tawn@lancaster.ac.uk

 1369–7412/04/66497

increasingly aware of the importance of understanding the dependence between different air pollutants. Recent atmospheric chemistry research (Photochemical Oxidants Review Group, 1997; Colls, 2002; Housley and Richards, 2001) has highlighted issues concerning extremal dependence between air pollutants. In particular, the Photochemical Oxidants Review Group (1997) suggested that the dependence between O_3 and some other atmospheric pollutants strengthens as the level of O_3 increases. This is of concern since it is known that O_3 has synergistic corrosive effects in combination with other sulphur- and nitrogen-based pollutants. The adverse health effects of particulate matter are also believed to be exacerbated by the excessive presence of other gaseous pollutants.

The gases are recorded in parts per billion, and the particulate matter in micrograms per cubic metre. The data are available from

`http://www.blackwellpublishing.com/rss`

We compare data from *winter* (from November to February inclusively) and early *summer* (from April to July inclusively).

Fig. 1 shows the daily maxima of the hourly means of the O_3 and NO_2 variables for each of these seasons. The highest values of O_3 are observed in the summer, as O_3 is formed by a series of reactions that are driven by sunlight (Brimblecombe, 2001). The reactions involve hydrocarbons and NO_2; large values of the latter occur with large O_3 values as shown Fig. 1. This positive dependence between O_3 and NO_2 in summer is not observed during the winter when the sunlight is weaker. Dependence between the air pollution variables influences the combinations which can occur when any one of the pollutants is large. In Section 7 we estimate several functionals of the extreme values of the joint distribution of the air pollution variables. One such functional is the probability that these variables occur in an extreme set $C \subset \mathbb{R}^d$, an example of such a set being shown in the *summer* data plot of Fig. 1(a). The precise specification of this set is discussed in Section 7. Pairs of (O_3, NO_2) could occur in the set that is shown in Fig. 1 by being extreme in a single component, or by being simultaneously (but possibly less) extreme in both components.

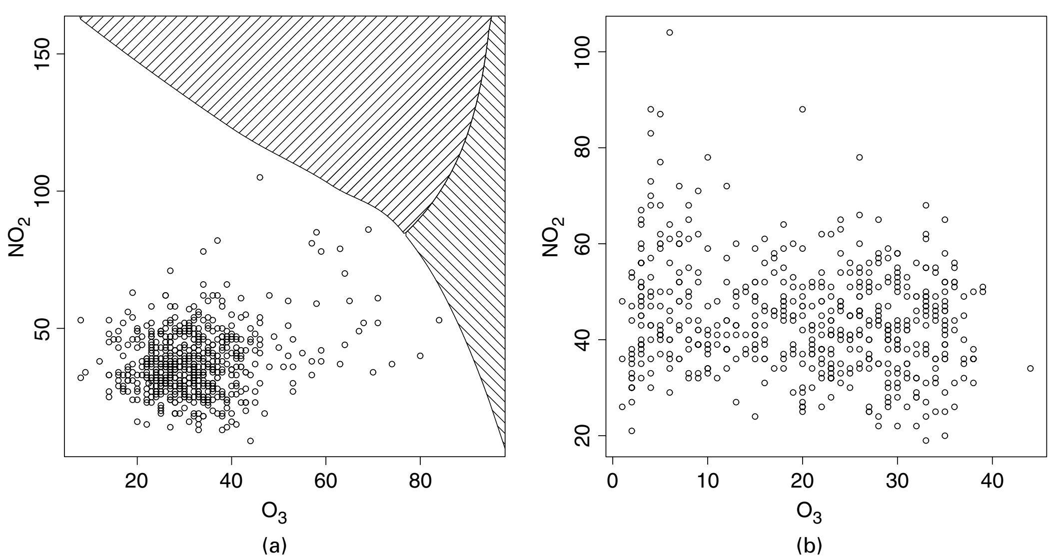


Fig. 1. Daily maxima of O_3 and NO_2 variables during (a) summer and (b) winter periods, 1994–1998 inclusively: the shaded set in (a) indicates an extreme set C which is split into two subsets C_1 (▧) and C_2 (▨)

The air pollution problem is a typical example of multivariate extreme value problems, summarized as follows. Consider a continuous vector variable $\mathbf{X}=(X_1,\ldots,X_d)$ with unknown distribution function $F(\mathbf{x})$. From a sample of n independent and identically distributed observations from F we wish to estimate functionals of the distribution of $\mathbf{X}$ when $\mathbf{X}$ is extreme in at least one component. The methods that are developed in this paper allow any such functional to be considered. However, to simplify the presentation we shall focus much of our discussion on estimating $\Pr(\mathbf{X}\in C)$ where C is an extreme set such that for all $\mathbf{x}\in C$ at least one component of $\mathbf{x}$ is extreme. Typically no observations will have occurred in C. The structure of C motivates the following natural partition of C into d subsets $C=\cup_{i=1}^{d} C_i$. Here, C_i is that part of C for which X_i is the largest component of $\mathbf{X}$, as measured by the quantiles of the marginal distributions. Specifically, for each $i=1,\ldots,d$, let F_{X_i} denote the marginal distribution of X_i; then

$$C_i=C\cap\{\mathbf{x}\in\mathbb{R}^d : F_{X_i}(x_i)>F_{X_j}(x_j);\ j=1,\ldots,d;\ j\neq i\},\qquad \text{for } i=1,\ldots,d.$$

We assume that subsets of C of the form $C\cap\{\mathbf{x}\in\mathbb{R}^d : F_{X_i}(x_i)=F_{X_j}(x_j) \text{ for some } i\neq j\}$ can be ignored; these are null sets provided that on these subsets there are no singular components in the dependence structure of $\mathbf{X}$. The partition of C into C_1 and C_2 for (O_3, NO_2) is shown in Fig. 1; the curved boundary between the sets is due to the inequality of the two marginal distributions.

With the partition of C defined in this way, C is an extreme set if all x_i-values in a non-empty C_i fall in the upper tail of F_{X_i}, i.e., if $v_{X_i}=\inf_{\mathbf{x}\in C_i}(x_i)$, then $F_{X_i}(v_{X_i})$ is close to 1 for $i=1,\ldots,d$. So

$$\Pr(\mathbf{X}\in C)=\sum_{i=1}^{d}\Pr(\mathbf{X}\in C_i)=\sum_{i=1}^{d}\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})\Pr(X_i>v_{X_i}). \tag{1.1}$$

Consider the estimation of $\Pr(\mathbf{X}\in C)$ by using decomposition (1.1). We need to estimate $\Pr(X_i>v_{X_i})$ and $\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})$, the former requiring a marginal extreme value model and the latter additionally needing an extreme value model for the dependence structure. We focus on these two terms in turn.

Methods for marginal extremes are now relatively standard; see Davison and Smith (1990), Smith (1989) and Dekkers *et al.* (1989). Univariate extreme value theory provides an asymptotic justification for the generalized Pareto distribution to be an appropriate model for the distribution of excesses over a suitably chosen high threshold; see Pickands (1975). Thus, we model the marginal tail of X_i for $i=1,\ldots,d$ by

$$\Pr(X_i>x+u_{X_i}|X_i>u_{X_i})=(1+\xi_i x/\beta_i)_+^{-1/\xi_i}\qquad \text{where } x>0. \tag{1.2}$$

Here u_{X_i} is a high threshold for variable X_i, β_i and ξ_i are scale and shape parameters respectively with $\beta_i>0$ and $s_+=\max(s,0)$ for any $s\in\mathbb{R}$. We require a model for the complete marginal distribution F_{X_i} of X_i for each $i=1,\ldots,d$, since to estimate $\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})$ we need to describe all X_j-values that can occur with any large X_i. We adopt the semiparametric model $\hat{F}_{X_i}$ for F_{X_i} of Coles and Tawn (1991, 1994), i.e.

$$\hat{F}_{X_i}(x)=\begin{cases}1-\{1-\tilde{F}_{X_i}(u_{X_i})\}\{1+\xi_i(x-u_{X_i})/\beta_i\}_+^{-1/\xi_i} & \text{for } x>u_{X_i},\\ \tilde{F}_{X_i}(x) & \text{for } x\leqslant u_{X_i},\end{cases} \tag{1.3}$$

where $\tilde{F}_{X_i}$ is the empirical distribution of the X_i-values. We denote the upper end point of the distribution by x^{F_i}, which is ∞ if $\xi_i\geqslant 0$ and $u_{X_i}-\beta_i/\xi_i$ if $\xi_i<0$. Model (1.3) provides the basis for estimating the $\Pr(X_i>v_{X_i})$ term of decomposition (1.1).

Both the marginal and the dependence structures of $\mathbf{X}$ are needed to determine $\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})$. We disentangle these two contributions and focus on the dependence modelling by working with margins that are assumed known for much of the following. We transform all the univariate marginal distributions to be of standard Gumbel form by using the probability integral transform, which for our marginal model (1.3) is

$$\begin{aligned} Y_i &= -\log[-\log\{\hat{F}_{X_i}(X_i)\}] \qquad \text{for } i=1,\ldots,d \\ &= t_i(X_i;\psi_i,\tilde{F}_{X_i}) \\ &= t_i(X_i), \end{aligned} \tag{1.4}$$

where $\psi_i=(\beta_i,\xi_i)$ are the marginal parameters. This transformation gives $\Pr(Y_i \leqslant y)=\exp\{-\exp(-y)\}$ for each i, so $\Pr(Y_i>y)\sim\exp(-y)$ as $y\to\infty$, and Y_i has an exponential upper tail. To clarify which marginal variable we are using, we use $\mathbf{X}$ and $\mathbf{Y}$ throughout to denote the variable with its original marginal distributions and with Gumbel margins respectively.

We now focus on extremal dependence modelling of variables with Gumbel marginal distributions. Modelling dependence for extreme values is more complex than modelling univariate extreme values and despite there already being various proposals the methodologies are still evolving. When interest is in the upper extremes of each component of $\mathbf{Y}$, the dependence structures fall into two categories: asymptotically dependent and asymptotically independent. Variable $\mathbf{Y}_{-i}$ is termed asymptotically dependent on and asymptotically independent of variable Y_i when the limit

$$\lim_{y\to\infty}\{\Pr(\mathbf{Y}_{-i}>\mathbf{y}|Y_i>y)\}$$

is non-zero and zero respectively. Here $\mathbf{Y}_{-i}$ denotes the vector $\mathbf{Y}$ excluding component Y_i and $\mathbf{y}$ a vector of y-values. All the existing methods for multivariate extreme values (outlined in Section 2) are appropriate for estimating $\Pr(\mathbf{X}\in C)$ under asymptotic dependence of the associated $\mathbf{Y}$, or for asymptotically independent variables provided that all $\mathbf{x}\in C$ are large in all components.

Fig. 2 shows the winter air pollution data transformed, by using transformations (1.4), to have identical Gumbel marginal distributions. It is clear from Fig. 2 that the extremal dependence between the NO variable and each of the other variables varies from pair to pair, with asymptotic dependence a feasible assumption only for (NO, NO_2) and (NO, PM_{10}). Thus the range of sets for which existing methods can be used to estimate $\Pr(\mathbf{X}\in C)$ is restricted.

We present an approach to multivariate extreme values that constitutes a change of direction from previous extreme value methods. Our modelling strategy is based on an assumption about the asymptotic form of the conditional distribution of the variable given that it has an extreme component, i.e. the distribution of $\mathbf{Y}_{-i}|Y_i=y_i$ as y_i becomes large. This conditional approach provides a natural extension of the univariate conditional generalized Pareto distribution model (1.2) to the multivariate case as $\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})$ can be expressed as

$$\Pr(\mathbf{X}\in C_i|X_i>v_{X_i})=\int_{v_{X_i}}^{x^{F_i}}\Pr(\mathbf{X}\in C_i|X_i=x)\,\mathrm{d}\hat{F}_{X_i}(x)/\{1-\hat{F}_{X_i}(v_{X_i})\}, \tag{1.5}$$

where the integrand is evaluated by using the distribution of $\mathbf{Y}_{-i}|Y_i=y_i$ after marginal transformation. When $v_{X_i}>u_{X_i}$ the derivative of $\hat{F}_{X_i}(x)/\{1-\hat{F}_{X_i}(v_{X_i})\}$ is the generalized Pareto density function with scale and shape parameters $\beta_i+\xi_i(v_{X_i}-u_{X_i})$ and ξ_i respectively.

Our conditional approach applies whether the variables are asymptotically dependent or asymptotically independent; it can be used to estimate $\Pr(\mathbf{X}\in C)$ for any extreme set C,

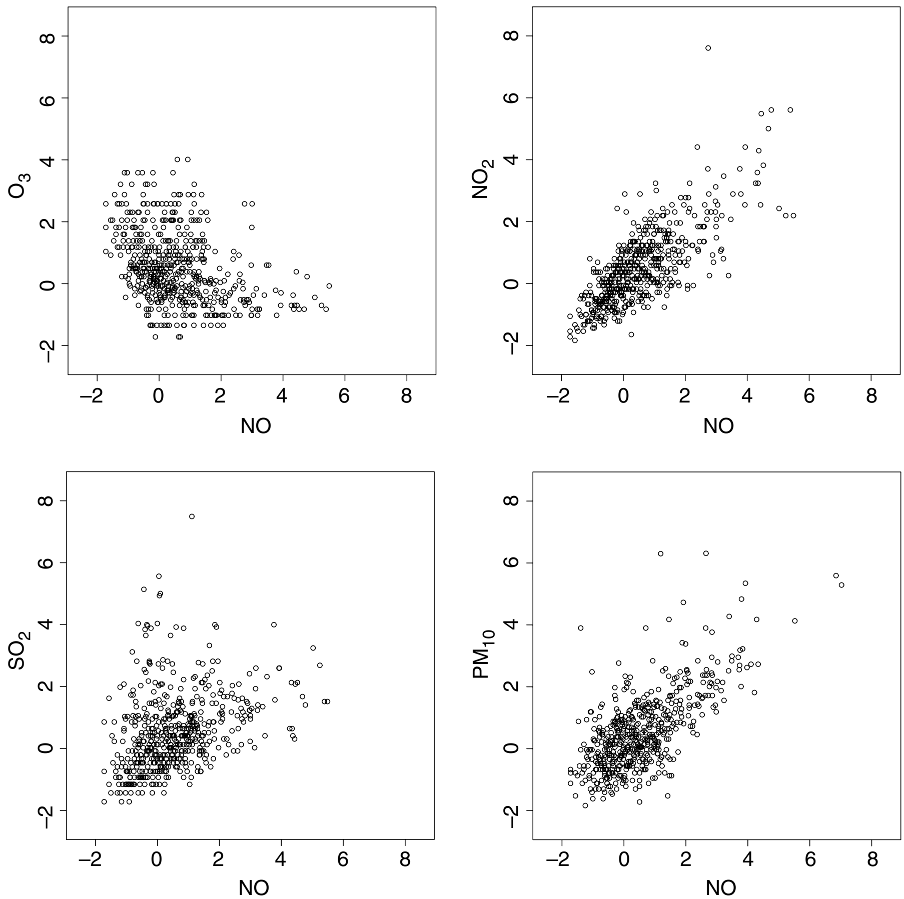


Fig. 2. Winter air pollution data transformed to have Gumbel margins by using transformations (1.4)

and it is applicable in any number of dimensions. The model that we use for the conditional distribution is motivated by an asymptotic distributional assumption and is supported by a range of theoretical examples. The model is semiparametric; parametric regression is used to estimate the location and scale parameters of the marginals of the joint conditional distribution and nonparametric methods are used to estimate the multivariate residual structure. Though our approach lacks a complete asymptotic characterization of the probabilistic structure, such as those which underpin existing extreme value methods, we show that strong mathematical and practical advantages are given by our approach in comparison with existing multivariate extreme value methods.

Existing methods are presented in Section 2. In Section 3 we state the new asymptotic assumption on which our conditional model is based, present some theoretical examples and draw links between the proposed and current methods. The examples motivate the modelling strategy that is introduced in Section 4. In Section 5 inference for the model is discussed. The methods are compared by using simulated data in Section 6. In Section 7 we illustrate the application of the techniques by analysing the extreme values of the air pollution data. Finally, in Section 8 we give the detailed working for the theoretical examples that are presented in Section 3.

2. Existing methods

We present a brief overview of the current methods for variables with Gumbel marginal distributions only. The extension to variables with arbitrary marginal distributions is obtained by incorporating marginal transformation (1.4).

Many multivariate extreme value analyses are based on models which assume implicitly that in some joint tail region each component of $\mathbf{Y}$ is either independent of or asymptotically dependent on the other components. Approaches which rely on these assumptions include the models for the multivariate extreme value distribution to describe componentwise maxima of Tawn (1988, 1990), Joe (1994), Capéraà *et al.* (1997) and Hall and Tajvidi (2000) and the multivariate threshold methods of Coles and Tawn (1991, 1994), Joe *et al.* (1992), de Haan and Resnick (1993), Sinha (1997), de Haan and de Ronde (1998), Draisma (2000) and Stărică (2000). Ledford and Tawn (1996, 1997, 1998) showed that these multivariate threshold methods are inappropriate for extrapolation of a variable $\mathbf{Y}$ with components that are dependent but asymptotically independent, when estimation is carried out by using a single selected threshold. Ledford and Tawn (1996, 1997) proposed a bivariate threshold model to overcome this limitation, which has been explored and developed by Bortot and Tawn (1998), Peng (1999), Coles *et al.* (1999), Bortot *et al.* (2000), Heffernan (2000), Draisma *et al.* (2003) and Ledford and Tawn (2003).

Behind all these existing approaches is the assumption of multivariate regular variation in Fréchet margins. For statistical purposes this asymptotic assumption is taken to hold exactly over a joint tail region. For Gumbel margins, these modelling assumptions combine to give a joint distributional model with the property

$$\Pr(\mathbf{Y} \in t + A) = \exp(-t/\eta_{\mathbf{Y}}) \Pr(\mathbf{Y} \in A), \tag{2.1}$$

where $t + A$ is a componentwise translation of every element of set A by a scalar $t > 0$, A is a set in which every element is large in all its components and $\eta_{\mathbf{Y}}$, termed the *coefficient of tail dependence*, satisfies $0 < \eta_{\mathbf{Y}} \leqslant 1$. When $\eta_{\mathbf{Y}} = 1$ the asymptotic theory behind property (2.1) extends to any set A in which every element is large in at least one of its components.

Ledford and Tawn (1996) identified four classes of extremal dependence. The first class is that of asymptotically dependent distributions, for which $\eta_{\mathbf{Y}} = 1$. The other three classes comprise distributions with asymptotically independent dependence structures exhibiting positive extremal dependence ($d^{-1} < \eta_{\mathbf{Y}} < 1$), near extremal independence ($\eta_{\mathbf{Y}} = d^{-1}$) and negative extremal dependence ($0 < \eta_{\mathbf{Y}} < d^{-1}$) for a d-dimensional variable. These three classes correspond respectively to joint extremes of $\mathbf{Y}$ occurring more often than, approximately as often as or less often than joint extremes if all components of the variable were independent.

Relationship (2.1) forms the basis for the estimation of probabilities of extreme multivariate events for all the existing methods. Specifically, for an extreme set D, which will typically contain no observations in a large sample, the approach is to choose a constant $t > 0$ and to identify a set A such that $D = t + A$ and that A is an extreme set in the joint tail that contains sufficient observations for the empirical estimate of $\Pr(\mathbf{Y} \in A)$ to be reliable. Thus the choice of t is equivalent to selecting a threshold. Estimates of $\Pr(\mathbf{Y} \in D)$ follow from property (2.1). Estimates of the parameter $\eta_{\mathbf{Y}}$ are obtained by exploiting the property that $\Pr\{\min(\mathbf{Y}) > y\} \sim \exp(-y/\eta_{\mathbf{Y}})$ for $y \to \infty$. Estimates of $\Pr(\mathbf{Y} \in A)$, or equivalently $\Pr(\mathbf{Y} + t \in D)$, are obtained empirically.

Extrapolation based on relationship (2.1) cannot provide estimates of probabilities for sets D that are not simultaneously extreme in each component. The reason for this is that, for such D, the empirical estimate of $\Pr(\mathbf{Y} + t \in D)$ is likely to be 0 since the translated data $\mathbf{Y} + t$ are unlikely to fall in D. For asymptotically independent variables such sets are of most interest. This weakness of existing methods illustrates the need for a new approach, as it is due to the inadequacy of the asymptotic framework of the existing methods rather than a paucity of available models within this framework.

3. Theoretical motivation

In this section we present a range of theoretical results which motivate our choice of statistical model. In Section 3.1, we make an assumption about the asymptotic form of the conditional distribution and examine the consequences of this assumption. Then, in Section 3.2, we identify the conditions that must be satisfied by the normalizing functions underlying this assumption for the limiting representation to hold. In Section 3.3 we discuss some theoretical examples which suggest that the asymptotic assumption is appropriate for a wide range of distributions, and that the class of normalizing functions is narrow, whereas the range of limit distributions is broad. Finally, in Section 3.4, we draw links between the proposed and existing methods.

3.1. Assumption of a limit representation and its properties

Consider the asymptotic structure of the conditional distributions arising from a d-dimensional random variable $\mathbf{Y}=(Y_1,\ldots,Y_d)$ with Gumbel marginal distributions. For each $i=1,\ldots,d$, we examine the conditional distribution $\Pr(\mathbf{Y}_{-i}\leqslant \mathbf{y}_{-i}|Y_i=y_i)$, where here, and throughout, vector algebra is applied componentwise. To examine the limiting behaviour of these distributions as $y_i\to\infty$ we require the limiting distribution to be non-degenerate in all margins, so we must control the growth of $\mathbf{y}_{-i}$ according to the dependence of $\mathbf{Y}_{-i}$ on Y_i.

Specifically we assume that for a given i there are vector normalizing functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$, both $\mathbb{R}\to\mathbb{R}^{(d-1)}$, which can be chosen such that, for all fixed $\mathbf{z}_{|i}$ and for any sequence of y_i-values such that $y_i\to\infty$,

$$\lim_{y_i\to\infty}[\Pr\{\mathbf{Y}_{-i}\leqslant \mathbf{a}_{|i}(y_i)+\mathbf{b}_{|i}(y_i)\mathbf{z}_{|i}|Y_i=y_i\}]=G_{|i}(\mathbf{z}_{|i}), \tag{3.1}$$

where all the margins of the limit distribution $G_{|i}$ are non-degenerate. An alternative expression of this assumption, which has an easier statistical interpretation, is that the standardized variables

$$\mathbf{Z}_{|i}=\frac{\mathbf{Y}_{-i}-\mathbf{a}_{|i}(y_i)}{\mathbf{b}_{|i}(y_i)} \tag{3.2}$$

have the property that

$$\lim_{y_i\to\infty}\{\Pr(\mathbf{Z}_{|i}\leqslant \mathbf{z}_{|i}|Y_i=y_i)\}=G_{|i}(\mathbf{z}_{|i}), \tag{3.3}$$

where the limit distribution $G_{|i}$ has non-degenerate marginal distributions.

Under assumption (3.1), or equivalently assumption (3.3), we have that, conditionally on $Y_i>u_i$, as $u_i\to\infty$ the variables Y_i-u_i and $\mathbf{Z}_{|i}$ are independent in the limit with limiting marginal distributions being exponential and $G_{|i}(\mathbf{z}_{|i})$ respectively. To see that this result holds, let $y_i=u_i+y$ with $y>0$ fixed; then

$$\begin{aligned}\Pr(\mathbf{Z}_{|i}\leqslant \mathbf{z}_{|i},Y_i-u_i=y|Y_i>u_i)&=\Pr\{\mathbf{Y}_{-i}\leqslant \mathbf{a}_{|i}(u_i+y)+\mathbf{b}_{|i}(u_i+y)\mathbf{z}_{|i}|Y_i=u_i+y\}\frac{f_{Y_i}(u_i+y)}{\Pr(Y_i>u_i)}\\&\to G_{|i}(\mathbf{z}_{|i})\exp(-y),\qquad \text{as } u_i\to\infty,\end{aligned} \tag{3.4}$$

where f_{Y_i} is the marginal density function of Y_i. The final convergence in this derivation is implied by the exponential tail of the Gumbel variables and the property that the conditional limit (3.1) holds irrespectively of how $y_i \to \infty$.

We now consider the marginal and dependence characteristics of $G_{|i}(\mathbf{z}_{|i})$. For each $j \neq i$, we define $G_{j|i}(z_{j|i})$ to be the limiting conditional distribution of

$$Z_{j|i} = \frac{Y_j - a_{j|i}(y_i)}{b_{j|i}(y_i)} \qquad \text{given } Y_i = y_i \text{ as } y_i \to \infty,$$

where $a_{j|i}(y_i)$ and $b_{j|i}(y_i)$ are the component functions of $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ associated with variable Y_j. Thus $G_{j|i}$ is the marginal distribution of $G_{|i}$ associated with variable Y_j. If

$$G_{|i}(\mathbf{z}_{|i}) = \prod_{j \neq i} G_{j|i}(z_{j|i}),$$

then we say that the elements of $\mathbf{Y}_{-i}$ are mutually asymptotically conditionally independent given Y_i.

3.2. Choice of normalization

We now identify the normalizing functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ in terms of characteristics of the conditional distribution of $\mathbf{Y}_{-i}|Y_i$, thus enabling these functions to be identified for theoretical examples. The normalizing functions and limit distribution are not unique in the sense that, if the normalizing functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ give a non-degenerate limit distribution $G_{|i}(\mathbf{z}_{|i})$, using the normalizing functions

$$\begin{aligned} \mathbf{a}^*_{|i}(y_i) &= \mathbf{a}_{|i}(y_i) + \mathbf{A}\mathbf{b}_{|i}(y_i), \\ \mathbf{b}^*_{|i}(y_i) &= \mathbf{B}\mathbf{b}_{|i}(y_i) \end{aligned} \tag{3.5}$$

for arbitrary vector constants $\mathbf{A}$ and $\mathbf{B}$, with $\mathbf{B} > \mathbf{0}$, gives the non-degenerate limit $G_{|i}(\mathbf{B}\mathbf{z}_{|i} + \mathbf{A})$. However, following standard arguments such as used in Leadbetter *et al.* (1983), page 7, this is the only way that two different limits with no mass at ∞ can arise, so the class of limit distributions is unique up to type, and the normalizing functions can be identified up to the constants $\mathbf{A}$ and $\mathbf{B}$ in expression (3.5).

For fixed i, the choice of the vector functions can be broken into $d-1$ separate conditions based on the limiting behaviour of $Y_j|Y_i = y_i$, for each $j \neq i$, since assumption (3.1) specifies that each marginal distribution of $G_{|i}$ must be non-degenerate. Thus we are interested in the conditional distribution function of $Y_j|Y_i = y_i$ which is denoted by $F_{j|i}(y_j|y_i)$. The associated conditional hazard function $h_{j|i}$ is defined as

$$h_{j|i}(y_j|y_i) = \frac{f_{j|i}(y_j|y_i)}{1 - F_{j|i}(y_j|y_i)} \qquad \text{for } -\infty < y_j < \infty,$$

where $f_{j|i}(y_j|y_i)$ is the conditional density function of $Y_j|Y_i = y_i$.

Theorem 1. Suppose that the vector random variable $\mathbf{Y}$ has an absolutely continuous joint density. If, for a given i, the vector functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i) > 0$ satisfy the limiting property (3.1), or equivalently property (3.3), then the components of these vector functions corresponding to variable Y_j, for each $j \neq i$, satisfy, up to type, properties (3.6) and (3.7):

$$\lim_{y_i \to \infty} [F_{j|i}\{a_{j|i}(y_i)|y_i\}] = p_{j|i}, \tag{3.6}$$

where $p_{j|i}$ is a constant in the range $(0, 1)$, and

$$b_{j|i}(y_i) = h_{j|i}\{a_{j|i}(y_i)|y_i\}^{-1}. \tag{3.7}$$

The proof of theorem 1 is given in Appendix A. Owing to the flexibility in the form of normalizing function given by expression (3.5), a simplification of the structure of the normalizing functions can be achieved, as illustrated by corollary 1.

Corollary 1. If functions $a_{j|i}(y_i)$ and $b_{j|i}(y_i) > 0$ satisfy the conditions of theorem 1, and there is a constant $s_{j|i} < \infty$ such that

$$\lim_{y_i \to \infty} \left\{ \frac{a_{j|i}(y_i)}{b_{j|i}(y_i)} \right\} = s_{j|i},$$

then limit relationship (3.1) holds with $a_{j|i}(y_i) = 0$. Furthermore, if $b_{j|i}(y_i) = t_{j|i}\, k_{j|i}(y_i)$ for $t_{j|i} > 0$ any constant independent of y_i, and $k_{j|i}(y_i)$ any function of y_i, then the limit relationship (3.1) holds with $b_{j|i}(y_i)$ replaced by $k_{j|i}(y_i)$.

3.3. *Theoretical examples*

We present the normalizing functions $\mathbf{a}_{|i}(y)$ and $\mathbf{b}_{|i}(y)$, given by theorem 1 and corollary 1, and some properties of the associated non-degenerate limiting conditional distribution $G_{|i}$ for a range of multivariate distributions with Gumbel marginal distributions. The examples are selected to provide a coverage of the four classes of extremal dependence that were identified in Section 2.

As pairwise dependence determines each of the components of the normalizing functions, we present the results categorized by the pairwise coefficient of tail dependence for (Y_i, Y_j), denoted by η_{ij}, with $\eta_{ij} = \frac{1}{2}$ indicating near extremal independence for the pair. Table 1 shows two examples from each of the four classes. The special cases of perfect positive and negative dependence (cases i and viii respectively) are included here to identify upper and lower bounds on the behaviour of the normalizing functions, although strictly the methods of Section 3.2

Table 1. Examples of multivariate dependence structures classified by extremal dependence behaviour†

Extremal dependence structure	η_{ij}	*Normalization*		*Limit distribution* $G_{\|i}$	
		$a_{j\|i}(y)$	$b_{j\|i}(y)$	$G_{j\|i}$	*ACI*‡
1, i	1	y	1	Degenerate	NA
1, ii	1	y	1	§	No
2, iii	$(1+\rho_{ij})/2$	$\rho_{ij}^2 y$	$y^{1/2}$	Normal	No
2, iv	$2^{-\alpha}$	0	$y^{1-\alpha}$	Weibull	Yes
3, v	0.5	0	1	Gumbel	Yes
3, vi	0.5	0	1	Gumbel	No
4, vii	$(1+\rho_{ij})/2$	$-\log(\rho_{ij}^2 y)$	$y^{-1/2}$	Normal	No
4, viii	0	$-\log(y)$	1	Degenerate	NA

†1, asymptotic dependence; 2, asymptotic independence with positive association; 3, near independence; 4, negative dependence. The dependence structures are i, perfect positive dependence, ii, multivariate extreme value distribution, iii, multivariate normal ($\rho_{ij} > 0$), iv, inverted multivariate extreme value distribution with symmetric logistic dependence structure and parameter $0 < \alpha \leqslant 1$, v, independence, vi, multivariate Morgenstern, vii, multivariate normal ($\rho_{ij} < 0$), and viii, perfect negative dependence.
‡ACI, asymptotic conditional independence, which is not applicable (NA) if the variable is degenerate.
§The limiting distribution is complicated and its exact form is given in Section 8.

do not apply to these two distributions as, for each, the associated conditional distribution is degenerate. At this stage, interest is only in the structure of the normalizing functions and the limiting distributions, so discussion of the precise specification of distributions ii–vii is postponed until Section 8, where additional examples are presented. Furthermore, as the limit distribution $G_{|i}$ is often complicated, here we identify only the marginal distribution $G_{j|i}$ and state whether or not the margins of $G_{|i}$ are independent.

The examples that are listed in Table 1, and those given in Section 8, all satisfy the asymptotic assumption (3.1), have a simple structure for the normalizing functions and give a range of limiting distributions $G_{|i}$ that are not contained in any simple distributional family. This finding about $G_{|i}$ is in contrast with the limiting representation for multivariate extreme value distributions (de Haan and Resnick, 1977; Resnick, 1987) but is a consequence of the lack of structure that is imposed on $G_{|i}$ by the limiting operation. The normalizing functions are all special cases of the parametric family

$$\begin{aligned}\mathbf{a}_{|i}(y) &= \mathbf{a}_{|i}\,y + I_{\{\mathbf{a}_{|i}=\mathbf{0},\mathbf{b}_{|i}<\mathbf{0}\}}\{\mathbf{c}_{|i} - \mathbf{d}_{|i}\log(y)\},\\ \mathbf{b}_{|i}(y) &= y^{\mathbf{b}_{|i}}\end{aligned} \tag{3.8}$$

where, on the right-hand side, $\mathbf{a}_{|i}$, $\mathbf{b}_{|i}$, $\mathbf{c}_{|i}$ and $\mathbf{d}_{|i}$ are vector constants and I is an indicator function. The vectors of constants have components such that $0 \leqslant a_{j|i} \leqslant 1$, $-\infty < b_{j|i} < 1$, $-\infty < c_{j|i} < \infty$ and $0 \leqslant d_{j|i} \leqslant 1$ for all $j \neq i$. Parametric family (3.8) has different structural formulations for $\mathbf{a}_{|i}(y)$ for positively and negatively associated pairs, owing to the asymmetry of the Gumbel marginal distribution, for which the upper tail is heavier than the lower tail.

The construction of the limiting operations that give the normalizing functions and limit distribution does not ensure continuity in these functions or distributions as the parameters of the original distribution are changed. Two particular examples illustrate this point as the parameters of the underlying distributions approach values corresponding to independence. A special case of distribution ii is the bivariate extreme value distribution with logistic dependence structure, which is asymptotically dependent when the dependence parameter $0 < \alpha < 1$ (see Section 8 for details). When $\alpha = 1$ the variables are independent. Consequently the normalization that is required is discontinuous in α at $\alpha = 1$. However, as $\alpha \uparrow 1$ the limit distribution $G_{j|i}$ puts all of its mass increasingly close to $-\infty$, indicating that the location normalization is becoming too powerful. Similarly, the multivariate normal distribution iii gives $G_{j|i}$ as normal with variance $2\rho_{ij}^2(1-\rho_{ij}^2)$, so as $\rho_{ij} \downarrow 0$ the limit is degenerate as the scale normalization becomes too strong. Similar inconsistencies are found for η_{ij} (see Heffernan (2000)) and for a range of asymptotically derived probability models.

We obtained the rate of convergence of each margin of the limiting conditional joint distribution, i.e. the order of convergence to 0 of

$$\Pr[\{Y_j - a_{j|i}(y_i)\}/b_{j|i}(y_i) \leqslant z_{j|i} | Y_i = y_i] - G_{j|i}(z_{j|i}) \tag{3.9}$$

as a function of n, where $\Pr(Y_i > y_i) = n^{-1}$ so that n determines how extreme the conditioning variable is in a manner that is invariant to the marginal distribution. Thus specified, the rate of convergence depends only on the underlying dependence structure. Expression (3.9) equals 0 for all $z_{j|i}$ for distributions i, v and viii in Table 1; the convergence rate is $O(n^{-1})$ for distributions ii and vi and $O\{1/\log(n)\}$ for distribution iv whereas for distributions iii and vii it is $O[\log\{\log(n)\}/\log(n)^{1/2}]$. These rates are typical of those that are seen in other extreme value problems.

3.4. Links with existing methods

To clarify the connections with existing methods, we examine the limiting conditional distribution under the existing framework for multivariate extreme values. Let $A=\Pi_{i=1}^{d}(y_i,\infty)$, for fixed large values of each y_i, $i=1,\ldots,d$, in expression (2.1). Differentiating expression (2.1) with respect to y_i and dividing by $f_{Y_i}(y_i+t)$ gives that for all $t>0$

$$\Pr(\mathbf{Y}_{-i}>\mathbf{y}_{-i}+t|Y_i=y_i+t)=\{1-\delta(y_i,t)\}\exp\{-t(1-\eta_{\mathbf{Y}})/\eta_{\mathbf{Y}}\}\Pr(\mathbf{Y}_{-i}>\mathbf{y}_{-i}|Y_i=y_i), \tag{3.10}$$

where $\delta(y_i,t)=1-\exp[-\exp(-y_i)\{1-\exp(-t)\}]\rightarrow 0$ as $y_i\rightarrow\infty$. Hence, for large y_i, to first order, expression (3.10) is invariant to changes in t when $\eta_{\mathbf{Y}}=1$, so the limit distribution of $\mathbf{Y}_{-i}-Y_i$ is non-degenerate for $Y_i=y_i$ as $y_i\rightarrow\infty$. This result is identical to the structure that we find under asymptotic dependence between all the components ($\mathbf{a}_{|i}=1$ and $\mathbf{b}_{|i}=0$). Despite strong connections between the approaches, the statistical model that is developed in Section 4 leads to a new estimator of $\Pr(\mathbf{X}\in C)$ when the variables are asymptotically dependent. When $\eta_{\mathbf{Y}}<1$, expression (3.10) shows that the normalization $\mathbf{Y}_{-i}-Y_i$ leads to a degenerate limit given $Y_i=y_i$ as $y_i\rightarrow\infty$, demonstrating the need for more sophisticated normalizations than those considered previously.

4. Model structure and properties

In Section 4.1 we present a semiparametric dependence model for describing extreme values in multivariate problems. This model is presented for variables with univariate marginal Gumbel distributions. Combined with our marginal model, described in Section 1, this dependence model gives a complete joint model for the extreme values of the random variable $\mathbf{X}$. Issues concerning the self-consistency of the various conditional models are discussed in Section 4.2. Methods for extrapolation for the $\mathbf{X}$-variable under the joint model are described in Section 4.3. Finally, in Section 4.4, we propose diagnostics to aid model selection.

4.1. Conditional dependence model

The model structure is motivated by the findings in Section 3. Using the same approach as in other univariate and multivariate extreme value methods, we take an asymptotic assumption which holds under weak conditions to hold exactly provided that the limiting variable is sufficiently extreme. Here we use the formulation of the limiting conditional distribution (3.1), and its implied limiting independence property (3.4), to capture the behaviour of variable $\mathbf{Y}_{-i}$ occurring with large Y_i. We assume that for each $i=1,\ldots,d$ there is a high threshold u_{Y_i} for which we model

$$\begin{aligned}\Pr\{\mathbf{Y}_{-i}<\mathbf{a}_{|i}(y_i)+\mathbf{b}_{|i}(y_i)\mathbf{z}_{|i}|Y_i=y_i\}&=\Pr(\mathbf{Z}_{|i}<\mathbf{z}_{|i}|Y_i=y_i)\\&=G_{|i}(\mathbf{z}_{|i}),\qquad\text{for all } y_i>u_{Y_i},\end{aligned}$$

where $\mathbf{Z}_{|i}$ is the standardized residual defined by expression (3.2), with distribution function $G_{|i}$, and $\mathbf{Z}_{|i}$ is independent of Y_i for $Y_i>u_{Y_i}$. The extremal dependence behaviour is then characterized by location and scale functions $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ and the distribution function $G_{|i}$.

First consider the specification of the individual conditional models, i.e. $\mathbf{a}_{|i}(y_i)$, $\mathbf{b}_{|i}(y_i)$ and $G_{|i}(\mathbf{z}_{|i})$ for a given i. We adopt the parametric model (3.8) as it is a single parametric family of normalizing functions which is appropriate for the wide range of theoretical examples that are shown in Table 1 and Section 8. We denote the parameters of $\mathbf{a}_{|i}(y)$ and $\mathbf{b}_{|i}(y)$ by $\boldsymbol{\theta}_{|i}=(\mathbf{a}_{|i},\mathbf{b}_{|i},\mathbf{c}_{|i},\mathbf{d}_{|i})$ and adopt the convention that $c_{j|i}=d_{j|i}=0$ unless $a_{j|i}=0$ and $b_{j|i}<0$. We discuss the estimation of $\boldsymbol{\theta}_{|i}$ in Section 5, denoting the estimator of $\boldsymbol{\theta}_{|i}$ by $\hat{\boldsymbol{\theta}}_{|i}$, and the associated estimators of the normalizing functions by $\hat{\mathbf{a}}_{|i}(y)$ and $\hat{\mathbf{b}}_{|i}(y)$. As the limiting operation (3.1) imposes no specific structure on $G_{|i}$, we adopt a nonparametric model for $G_{|i}$. We estimate this distribution by using the empirical distribution of replicates of the random variable $\hat{\mathbf{Z}}_{|i}$, defined by

$$\hat{\mathbf{Z}}_{|i} = \frac{\mathbf{Y}_{-i} - \hat{\mathbf{a}}_{|i}(y_i)}{\hat{\mathbf{b}}_{|i}(y_i)} \qquad \text{for } Y_i = y_i > u_{Y_i}.$$

The theoretical examples suggest that the $\mathbf{Z}_{|i}$ are often asymptotically conditionally independent, so if supported by diagnostic tests it may be advisable to model the components of $\hat{\mathbf{Z}}_{|i}$ as being independent, i.e. $\hat{G}_{|i}(\mathbf{z}_{|i}) = \Pi_{j \neq i} \hat{G}_{j|i}(z_{j|i})$, where $\hat{G}_{j|i}$ is the empirical distribution function of the $\hat{Z}_{j|i}$.

In summary, for $i = 1, \ldots, d$ our dependence model is a multivariate semiparametric regression model of the form

$$\mathbf{Y}_{-i} = \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{Z}_{|i} \qquad \text{for } Y_i = y_i > u_{Y_i}, \tag{4.1}$$

where $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ are given by the parametric model (3.8), and the distribution of the standardized residuals is modelled nonparametrically. The parameters of the overall model are $\boldsymbol{\theta} = (\boldsymbol{\theta}_{|1}, \ldots, \boldsymbol{\theta}_{|d})$. Each regression model applies only above the threshold u_{Y_i} for which the dependence structure is viewed to be well described by model (4.1). There is no necessity for the dependence threshold u_{Y_i} (on the Gumbel scale) and the marginal threshold u_{X_i} (on the original scale) to agree in the sense that $u_{Y_i} = t_i(u_{X_i})$, where transformation t_i is given in equation (1.4).

We categorize the dependence structure that is implied by model (4.1) by using four classes which identify the behaviour of quantiles of the distribution of $Y_j | Y_i = y_i$ as $y_i \to \infty$. If the quantiles of the conditional distribution grow at the same rate as y_i, i.e. $a_{j|i} = 1$ and $b_{j|i} = 0$, the variables (Y_i, Y_j) are asymptotically dependent; otherwise they are asymptotically independent. For asymptotically independent distributions, the conditional quantiles tend to ∞, a finite limit or $-\infty$ as $y_i \to \infty$ if (Y_i, Y_j) exhibit positive extremal dependence, extremal near independence or negative extremal dependence respectively. Thus the variables exhibit positive extremal dependence when at least one of $0 < a_{j|i} < 1$ or $b_{j|i} > 0$ holds, extremal near independence when $a_{j|i} = d_{j|i} = 0$ and $b_{j|i} \leqslant 0$, and negative extremal dependence when $a_{j|i} = 0, d_{j|i} > 0$ and $b_{j|i} < 0$.

Though the examples of Section 3.3 illustrate that the limit operations on the parameters of the original distribution and the conditioning variable cannot be interchanged, we do not see that this poses any problems in practice for model (4.1). The theoretical examples motivate a subclass of the general limiting structure imposed by asymptotic assumption (3.1); the family (3.8) that we have identified varies smoothly over the four classes of dependence. Furthermore, for statistical applications the underlying distribution is fixed and so the issue of interchanging limits does not arise in practice.

Treating the d conditional models separately gives the most general version of our model with parameter $\boldsymbol{\theta}$ an unconstrained vector of length $4d(d-1)$, though, for each ordered pair, $c_{j|i}$ and $d_{j|i}$ are only non-zero if there is no positive association. Dependence submodels may be of interest for identifying scientifically relevant structure in the joint distribution or for parsimony. For example, there are many multivariate distributions whose dependence structure is exchangeable in some way. The most common form of exchangeability is pairwise, i.e. Y_i depends on Y_j in the same way as Y_j depends on Y_i. We say that variables Y_i and Y_j exhibit *weak pairwise extremal exchangeability* if $\boldsymbol{\theta}_{i|j} = \boldsymbol{\theta}_{j|i}$ and *strong pairwise extremal exchangeability* if in addition $G_{i|j} = G_{j|i}$. In Section 8 we show examples of distributions which exhibit each of these forms of exchangeability.

4.2. Self-consistency of separate conditional models

Now consider the self-consistency of the d individual models for the conditional distributions of $\mathbf{Y}_{-i}|Y_i$ for each i and large values of thresholds $(u_{Y_1},\ldots,u_{Y_d})$. Problems of this general type are discussed by Besag (1974) and Arnold *et al.* (1999). As all d conditional distributions are determined by the joint distribution of $\mathbf{Y}$, there are some theoretical constraints on the possible combinations of values taken by the parameters $\boldsymbol{\theta}$ and the distributions $G_{|i}$ for $i=1,\ldots,d$. However, as the individual models are applied to different subsets of the support of the joint distribution, the self-consistency is important only on the intersection of these subsets. Generally the intersections take the form $\{\mathbf{y}: y_i \geqslant u_{Y_i}\ \forall i \in J\}$ where J is a subset of at least two elements of $\{1,\ldots,d\}$. First consider the case where $J=\{i,j\}$; then self-consistency requires that

$$\frac{\mathrm{d}}{\mathrm{d}y_j}\,\mathrm{Pr}(Y_j \leqslant y_j|Y_i=y_i)\ f_{Y_i}(y_i)=\frac{\mathrm{d}}{\mathrm{d}y_i}\,\mathrm{Pr}(Y_i \leqslant y_i|Y_j=y_j) f_{Y_j}(y_j) \tag{4.2}$$

where $y_j=a_{j|i}(y_i)+b_{j|i}(y_i)z_{j|i}$ and $y_i=a_{i|j}(y_j)+b_{i|j}(y_j)z_{i|j}$ for $y_i>u_{Y_i}$ and $y_j>u_{Y_j}$. In general condition (4.2) is too complex to impose. However, unless at least one of $a_{j|i}=1$ and $a_{i|j}=1$ holds, condition (4.2) becomes null since $\mathrm{Pr}\{\min(Y_i,Y_j)\geqslant u|\max(Y_i,Y_j)\geqslant u\}\to 0$ as $u\to\infty$. When $a_{j|i}=1$ and $b_{j|i}=0$, as $\min(u_{Y_i},u_{Y_j})\to\infty$, condition (4.2) imposes that $a_{i|j}=1$ and $b_{i|j}=0$ and, subject to the appropriate convergence of conditional density results, that

$$\frac{\mathrm{d}}{\mathrm{d}z}G_{i|j}(z)=\exp(-z)\frac{\mathrm{d}}{\mathrm{d}z}G_{j|i}(-z).$$

Now suppose that $J=\{1,\ldots,d\}$ and that all the variables are asymptotically dependent. Self-consistency then requires that, for all i and j, $a_{i|j}=1$ and $b_{i|j}=0$ and that

$$\frac{\mathrm{d}}{\mathrm{d}z_{j|i}}G_{|i}(\mathbf{z}_{|i})=\frac{\mathrm{d}}{\mathrm{d}z_{i|j}}G_{|j}(\mathbf{z}_{|j})\bigg|_{\mathbf{z}_{|j}=\mathbf{z}_{|j}^{(i)}}\exp(z_{j|i})$$

where $\mathbf{z}_{|j}^{(i)}$ denotes a $(d-1)$-vector with element associated with variable k $(k\neq j)$ being $z_{k|i}-z_{j|i}$ for $k\neq i$ and $z_{j|i}$ for $k=i$. Analogous conditions apply when only a subset of the variables is asymptotically dependent.

Though we have made progress in characterizing the self-consistency properties for the special case of asymptotic dependence we have no solution for ensuring self-consistency of the conditional distributions more generally. Our general approach is to estimate the d different conditional distributions separately and not to impose further structure in addition to model (4.1). The first defence of this approach is that the data arise from a valid joint distribution and so estimates that are based on the data should not depart greatly from self-consistency. Secondly, we recommend assessing the effect of using different conditionals to estimate probabilities of events in which more than one variable is extreme. Averaging estimates over the different conditionals reduces any problems of inconsistency, and in essence this is what our partitioning of C into $C_1,\ldots,C_d$ ensures. Thirdly, in many applications when submodels are fitted, the $\boldsymbol{\theta}$-component of the model is automatically restricted to be self-consistent. Finally, we might expect that ensuring self-consistency should improve the general performance of the method. Contrary to this expectation, in Section 6 we illustrate the use of models which are not self-consistent and show that imposing self-consistency of the $\boldsymbol{\theta}$-parameters substantially reduces the performance.

4.3. Extrapolation

We generate random samples from the conditional distributions of $\mathbf{X}|X_i>v_{X_i}$ for each i, using the estimated conditional models. These samples are used to obtain Monte Carlo approximations of functionals of the joint tails of the distribution of $\mathbf{X}$. Since we use the estimated model, the parameters are replaced by their estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$ which are obtained by using methods described in Section 5. We employ the following sampling algorithm.

Step 1: simulate Y_i from a Gumbel distribution conditional on its exceeding $t_i(v_{X_i})$.
Step 2: sample $\mathbf{Z}_{|i}$ from $\hat{G}_{|i}$ independently of Y_i.
Step 3: obtain $\mathbf{Y}_{-i} = \hat{\mathbf{a}}_{|i}(Y_i) + \hat{\mathbf{b}}_{|i}(Y_i)\mathbf{Z}_{|i}$.
Step 4: transform $\mathbf{Y} = (\mathbf{Y}_{-i}, Y_i)$ to the original scale by using the inverse of transformation (1.4).
Step 5: the resulting transformed vector $\mathbf{X}$ constitutes a simulated value from the conditional distribution of $\mathbf{X}|X_i > v_{X_i}$.

For example, we evaluate $\Pr(\mathbf{X} \in C_i | X_i > v_{X_i})$ by using a Monte Carlo approximation of integral (1.5) by repeating steps 1–5 and evaluating $\Pr(\mathbf{X} \in C_i | X_i > v_{X_i})$ as the long run proportion of the generated sample that falls in C_i. When C is not contained entirely in the joint tail region on which the dependence component of the conditional model is defined, we first partition C into C_* and $C \backslash C_*$ where

$$C_* = \{\mathbf{x} \in C : x_i < t_i^{-1}(u_{Y_i});\ i = 1, \ldots, d\}.$$

By definition of the u_{Y_i}, the empirical estimator of $\Pr(\mathbf{X} \in C_*)$ will be reliable. In contrast $\Pr(\mathbf{X} \in C \backslash C_*)$ requires model-based estimation, for which we use the conditional model as follows. We partition $C \backslash C_*$ into sets $C_1, \ldots, C_d$ as in Section 1. Using this construction, $v_{X_i} \geqslant t_i^{-1}(u_{Y_i})$ for all $i = 1, \ldots, d$, and the above approximation can be used to evaluate $\Pr(\mathbf{X} \in C_i | X_i > v_{X_i})$.

4.4. Diagnostics

The examples in Section 3.3 indicate that the rate of convergence of the conditional distribution of $\mathbf{Y}_{-i}|Y_i = y$, as $y \to \infty$, to its limiting form can be slow. However, the limiting form of the conditional distribution is used only to motivate our model structure and we are not interested in the true limit values of $\boldsymbol{\theta}_{|i}$ and $G_{|i}$. What is of practical importance is whether the conditional distribution of the normalized variable $\mathbf{Z}_{|i}$ is stable over the range of Y_i- (or equivalently X_i-) values that is used for estimation and extrapolation.

This requirement suggests that diagnostics for our model structure should be based on assessing the stability of the extrapolations that are achieved when fitting the model above a range of thresholds. For marginal estimation, we use diagnostics that are based on the mean residual life plot and the stability in the marginal shape parameter estimates; see Smith (1989) and Davison and Smith (1990). For dependence estimation, a fundamental modelling assumption is that $\mathbf{Z}_{|i}$ is independent of Y_i given $Y_i > u_{Y_i}$, for a high threshold u_{Y_i}, for each i. By fitting the conditional model over a range of high thresholds, the stability of the estimates of $\boldsymbol{\theta}_{|i}$ and the resulting extrapolations can be assessed. Then, for a selected threshold, independence of $\mathbf{Z}_{|i}$ and Y_i is examined. Furthermore, a range of standard tests for independence can be applied to the observed $\mathbf{Z}_{|i}$ to identify whether the variables can be treated as being asymptotically conditionally independent.

5. Inference

Our model comprises the marginal distributional model (1.3) and the dependence model (4.1). Both of these models are semiparametric, consisting of components that are specified parametrically and components for which no parametric model is appropriate. Our strategy for inference is driven by three features: a lack of parametrically specified joint distributions for each conditional distribution, the absence of practical constraints to impose self-consistency between different conditional distributions and a need for simplicity. This leads us to use an algorithm for point estimation which makes simplifying assumptions and a semiparametric bootstrap algorithm for evaluating uncertainty which does not rely on these assumptions.

Inference for marginal and dependence structures is undertaken stepwise: first the marginal parameters ψ are estimated and then the dependence parameters $\boldsymbol{\theta}$ are estimated assuming that the marginal parameters are known. Stepwise estimation is much simpler than joint estimation of all the parameters and findings in Shi *et al.* (1992) suggest that the loss of efficiency relative to joint estimation is likely to be small unless the values of ξ_i, $i=1,\ldots,d$, differ greatly.

Brief details of the marginal estimation step are given in Section 5.1. Following marginal estimation, the data are transformed to have Gumbel marginals by using transformations (1.4), with ψ replaced by their estimates $\hat{\psi}$. In Section 5.2 we describe why we use Gaussian estimation for the normalizing function parameters $\boldsymbol{\theta}_{|i}$ for each separate conditional distribution under the assumption that there are no constraints between $\boldsymbol{\theta}_{|i}$ and $\boldsymbol{\theta}_{|j}$ for any i and j. The fitting of submodels requires the joint estimation of all the conditional model parameters $\boldsymbol{\theta}$. In Section 5.3 we discuss an approach for this joint estimation which has similarities to the pseudolikelihood of Besag (1975). In Section 5.4 we present techniques for evaluating the uncertainty in estimation for the overall model and the resulting extrapolations. Throughout, we assume that the data are realizations of independent and identically distributed random variables $\mathbf{X}_1,\ldots,\mathbf{X}_n$.

5.1. Marginal estimation

We estimate the d univariate marginal distributions jointly, ignoring the dependence between components. Specifically, we assume independence between components of the variable in constructing the log-likelihood function

$$\log\{L(\psi)\} = \sum_{i=1}^{d} \sum_{k=1}^{n_{u_{X_i}}} \log\{\hat{f}_{X_i}(x_{i|i,k})\} \tag{5.1}$$

where $\hat{f}_{X_i}$ is the density that is associated with distribution (1.3), $n_{u_{X_i}}$ is the number of observations with ith component exceeding the marginal threshold u_{X_i} and the jth component of the kth such observation is denoted by $x_{j|i,k}: j=1,\ldots,d;\ k=1,\ldots,n_{u_{X_i}}$. If there are no functional links between the parameters of the various components then maximizing log-likelihood (5.1) is equivalent to fitting the generalized Pareto distribution to the excesses over the marginal thresholds separately for each margin. When there are constraints between marginal parameters, jointly maximizing the log-likelihood function (5.1) enables inferential efficiency to be gained.

5.2. Single conditional

For each i, we wish to estimate $\boldsymbol{\theta}_{|i}$ under minimal assumptions about $G_{|i}$. If we assume that $\mathbf{Z}_{|i}$ has two finite marginal moments, then $\boldsymbol{\theta}_{|i}$ determines the marginal means and variances of the conditional variable $\mathbf{Y}_{-i}|Y_i=y_i$ when $y_i>u_{Y_i}$. Specifically, if the $\mathbf{Z}_{|i}$ have marginal means and standard deviations denoted by vectors $\boldsymbol{\mu}_{|i}$ and $\boldsymbol{\sigma}_{|i}$ respectively, then the random variables $\mathbf{Y}_{-i}|Y_i=y$, for $y>u_{Y_i}$, have vector mean and standard deviation respectively given by

$$\boldsymbol{\mu}_{|i}(y) = \mathbf{a}_{|i}(y) + \boldsymbol{\mu}_{|i}\,\mathbf{b}_{|i}(y),$$

$$\boldsymbol{\sigma}_{|i}(y) = \boldsymbol{\sigma}_{|i}\,\mathbf{b}_{|i}(y),$$

which are functions of y, $\boldsymbol{\theta}_{|i}$ and of the constants $\boldsymbol{\lambda}_{|i}=(\boldsymbol{\mu}_{|i},\boldsymbol{\sigma}_{|i})$. Thus $(\boldsymbol{\theta}_{|i},\boldsymbol{\lambda}_{|i})$ are the parameters of a multivariate regression model with non-constant variance and unspecified error distribution. We exploit the consistency of maximum likelihood estimates of $\boldsymbol{\theta}_{|i}$ achieved by using a parametric model for $G_{|i}$ which is liable to be misspecified. Specifically, we maximize the associated objective function over the parameter space to produce a consistent and valid point estimator for $\boldsymbol{\theta}_{|i}$. For a general discussion of this approach see Hand and Crowder (1996), chapter 7. The parametric model for $G_{|i}$ is chosen for convenience and computational simplicity. We take the components of $\mathbf{Z}_{|i}$ to be mutually independent and Gaussian and hence our inference for $\boldsymbol{\theta}_{|i}$ is based on Gaussian estimation (Hand and Crowder, 1996; Crowder, 2001). The independence simplification appears reasonable as $\boldsymbol{\theta}_{|i}$ determines only the marginal characteristics of the conditional distribution. We considered a range of parametric distributions for the marginals of $\mathbf{Z}_{|i}$ and selected the Gaussian distribution for its simplicity, superior performance in a simulation study and links to generalized estimating equations that arise from this choice of model for $G_{|i}$.

Therefore, the objective function that we use for point estimation of $\boldsymbol{\theta}_{|i}$ and $\boldsymbol{\lambda}_{|i}$ is

$$Q_{|i}(\boldsymbol{\theta}_{|i},\boldsymbol{\lambda}_{|i})=-\sum_{j\neq i}\sum_{k=1}^{n_{u_{Y_i}}}\left[\log\{\sigma_{j|i}(y_{i|i,k})\}+\frac{1}{2}\left\{\frac{y_{j|i,k}-\mu_{j|i}(y_{i|i,k})}{\sigma_{j|i}(y_{i|i,k})}\right\}^2\right], \tag{5.2}$$

where the notation follows the conventions that are adopted in Section 3 and for log-likelihood (5.1). We maximize $Q_{|i}$ jointly with respect to $\boldsymbol{\theta}_{|i}$ and $\boldsymbol{\lambda}_{|i}$ to obtain our point estimate $\hat{\boldsymbol{\theta}}_{|i}$, with $\boldsymbol{\lambda}_{|i}$ being nuisance parameters. To overcome the structural discontinuity in $\mathbf{a}_{|i}(y)$, we fit the dependence model in two stages: first fixing $c_{j|i}=d_{j|i}=0$; then only estimating $c_{j|i}$ and $d_{j|i}$ if $\hat{a}_{j|i}=0$ and $\hat{b}_{j|i}<0$.

5.3. *All conditionals*

We now consider joint estimation of the conditional model parameters $\boldsymbol{\theta}$. For reasons that are similar to those discussed in Section 5.2, we falsely assume independence between different conditional distributions to give the objective function

$$Q(\boldsymbol{\theta},\boldsymbol{\lambda})=\sum_{i=1}^{d}Q_{|i}(\boldsymbol{\theta}_{|i},\boldsymbol{\lambda}_{|i}), \tag{5.3}$$

where $Q_{|i}(\boldsymbol{\theta}_{|i},\boldsymbol{\lambda}_{|i})$ is as in expression (5.2) and $\boldsymbol{\lambda}=(\boldsymbol{\lambda}_{|1},\ldots,\boldsymbol{\lambda}_{|d})$. For Gaussian error distributions it can be shown that objective function (5.3) is an approximation to the pseudolikelihood, which Besag (1975) introduced as an approximation to the joint likelihood function. The approximation of equation (5.3) to the pseudolikelihood follows from Bayes's theorem and the property that the marginal density of $\mathbf{Y}_{-i}$ and the conditional density of $\mathbf{Y}_{-i}|Y_i=y_i$ when $y_i<u_{Y_i}$ influence the shape of the pseudolikelihood negligibly. Further, if the variables are all mutually asymptotically independent then, for sufficiently large thresholds u_{Y_i}, each datum will exceed at most one threshold so the independence assumption underlying the construction of objective function (5.3) will be satisfied.

5.4. *Uncertainty*

Uncertainty arises from the estimation of the semiparametric marginal models, the parametric normalization functions of the conditional dependence structure and the nonparametric models of the distributions of the standardized residuals. To account for all these sources of uncertainty, we use standard semiparametric bootstrap methods to evaluate standard errors of model parameter estimates and of other estimated parameters such as $\Pr(\mathbf{X} \in C)$ (see Davison and Hinkley (1997)). Throughout we assume that the marginal and dependence thresholds are fixed and so the uncertainty that is linked to threshold selection is not accounted for by the bootstrap methods.

Our bootstrap procedure has three stages: data generation under the fitted model, estimation of model parameters and the derivation of an estimate of any derived parameters linked to extrapolation. These stages are repeated independently to generate independent bootstrap estimates. The novel aspect of our algorithm is the data generation. To ensure that the bootstrap samples that are obtained replicate both the marginal and the dependence features of the data, we use a two-step sampling algorithm for data generation. A nonparametric bootstrap is employed first, ensuring the preservation of the dependence structure; then a parametric step is carried out so that uncertainty in the estimation of the parametric models for the marginal tails can be assessed. The precise procedure is as follows. The original data are first transformed to have Gumbel margins, using the marginal model (1.3) which is estimated by using these original data. A nonparametric bootstrap sample is then obtained by sampling with replacement from the transformed data. We then change the marginal values of this bootstrap sample, ensuring that the marginal distributions are all Gumbel and preserving the associations between the ranked points in each component. Specifically, for each i, $i=1,\ldots,d$, we replace the ordered sample of component Y_i with an ordered sample of the same size from the standard Gumbel distribution. The resulting sample is then transformed back to the original margins by using the marginal model that was estimated from the original data. The data that are generated by using this approach have univariate marginal distributions with upper tails simulated from the fitted generalized Pareto model and dependence structure entirely consistent with the data as determined by the associations between the ranks of the components of the variables.

6. Simulation study

Throughout this section we use simulated data with known Gumbel margins to illustrate the application of the methods proposed. In Section 6.1 we present a detailed analysis of a single data set to highlight inference and extrapolation issues. Section 6.2 reports results of simulation studies comparing the performance of the existing and conditional methods for bivariate and multivariate replicated data sets. To allow a comparison with existing methods, we consider only positively dependent variables and hence work with the submodel $\mathbf{a}_{|i}(y)=\mathbf{a}_{|i}y$ with $\mathbf{0}\leqslant \mathbf{a}_{|i}\leqslant \mathbf{1}$. We focus on return level estimation. Specifically, when the multivariate set C is described by a single parameter v say, i.e. $C=C(v)$, then the return level v_p for an event with probability p is defined implicitly by

$$\Pr\{\mathbf{Y}\in C(v_p)\}=p. \tag{6.1}$$

We assess the performance of an estimator $\hat{v}_p$ of v_p by using the relative error $(\hat{v}_p - v_p)/v_p$.

6.1. *Simulated case-study*

We analyse the simulated data set of 5000 points shown in Fig. 3. The underlying distribution is the bivariate extreme value distribution with asymmetric logistic dependence structure; see Section 8.1 and Tawn (1988, 1990) for details. The parameters of this distribution are $\theta_{1,\{1\}}=1-\theta_{1,\{1,2\}}=0.1$, $\theta_{2,\{2\}}=1-\theta_{2,\{1,2\}}=0.75$ and $\alpha_{\{1,2\}}=0.2$, so the limiting parameters for the conditional distributions are $a_{2|1}=a_{1|2}=1$ and $b_{2|1}=b_{1|2}=0$. The simulated data have a complicated structure as, for large Y_1, variable Y_2 behaves as though it were asymptotically

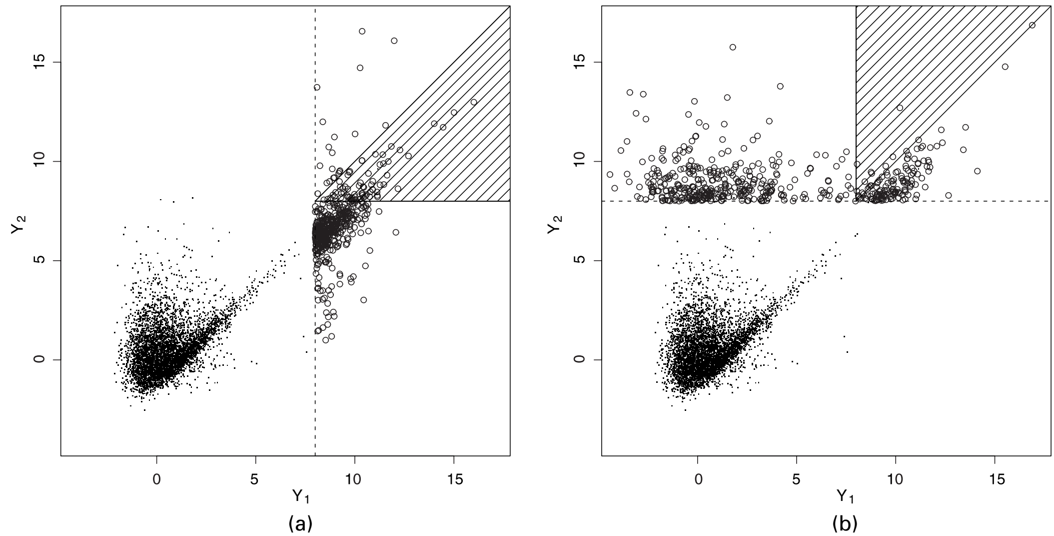


Fig. 3. Observations from the bivariate extreme value distribution with asymmetric logistic dependence structure (•) and pseudosamples (∘) generated under the asymmetric model above a threshold (¦, -------) ($\hat{a}_{2|1}=0.97$, $\hat{a}_{1|2}=0.11$, $\hat{b}_{2|1}=0.40$ and $\hat{b}_{1|2}=0.78$) and sets C_i (▨)

dependent on Y_1 but, for large Y_2, Y_1 arises from a mixture distribution with one component that is independent of Y_2 and the other asymptotically dependent on Y_2. As the normalization stabilizes the growth of the asymptotically dependent component only, the limiting distribution of $Y_1|Y_2$ has substantial mass at $-\infty$, corresponding to the independent component of the mixture distribution. At finite levels the independent points are likely to contaminate the parameter estimates of any asymptotically motivated model. Although the limiting values for the normalization parameters are symmetric, the clear asymmetry in the data suggests that we should compare two models: one with weak pairwise exchangeability ($a_{2|1}=a_{1|2}$ and $b_{2|1}=b_{1|2}$) and the other relaxing this assumption to allow for any form of asymmetry. Both models are fitted by using objective function (5.3) and thresholds corresponding to the 0.9 marginal quantiles.

Diagnostic procedures that are outlined in Section 4.4 aid model selection. Fig. 4 shows scatterplots of residuals $\hat{Z}_{2|1}$ for large Y_1 for each of the models proposed. Fig. 4(a) shows that the estimated distribution of $\hat{Z}_{2|1}$ has a trend in mean value with Y_1 for the weak pairwise exchangeable model, whereas this trend is much diminished in Fig. 4(b), which shows residuals from the fitted asymmetric conditional model. Equivalent plots for $\hat{Z}_{1|2}$ (not shown) indicate approximate independence of these residuals and Y_2 for both models.

Fig. 3 shows the pseudosamples that are obtained by using the fitted asymmetric conditional model with Fig. 3(a) and Fig. 3(b) showing the samples that are obtained conditioning on Y_1 and Y_2 respectively, and revealing the different forms of the conditional distributions. For set $C(v)=(v,\infty)^2$, Fig. 3 shows C_1 and C_2. Empirical estimates of $\Pr(\mathbf{Y}\in C_i|Y_i>v)$ are obtained as the proportion of the respective pseudosamples falling in these sets; $\Pr(\mathbf{Y}\in C)$ is then estimated by using decomposition (1.1). We investigated the effect of inconsistencies of the conditional models for $Y_2|Y_1$ and $Y_1|Y_2$ on the estimation of $\Pr(\mathbf{Y}\in C)$ by comparing approaches using pseudosamples generated under the following models: $Y_2|Y_1$ only, $Y_1|Y_2$ only and the intermediate approach based on decomposition (1.1). Despite the very different forms of the two conditional distributions, the differences between the three estimates are small relative to the uncertainties in estimation.

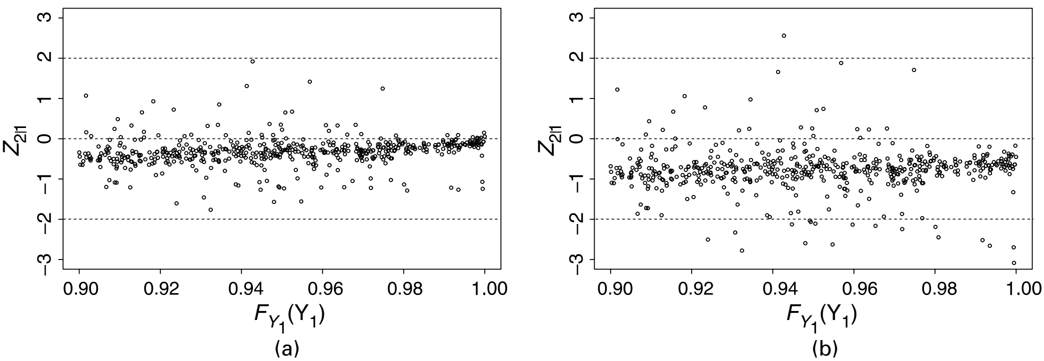


Fig. 4. Diagnostic plots for data from the bivariate extreme value distribution with asymmetric logistic dependence structure—scatterplots of residuals $Z_{2|1}$ against the conditioning variable Y_1, after transforming Y_1 to uniform margins: (a) residuals from fitting the weakly pairwise exchangeable model; (b) residuals from fitting the asymmetric conditional model

6.2. *Multivariate examples*

We consider the following four distributions, all with standard Gumbel margins:

(a) a multivariate extreme value distribution with symmetric logistic dependence structure (8.4) and parameter $\alpha = 0.5$ (distribution A);
(b) a bivariate extreme value distribution with asymmetric logistic dependence structure (8.5) with parameters given in Section 6.1 (distribution B);
(c) an inverted multivariate extreme value distribution with symmetric logistic dependence structure (8.4) with parameter α, for which η_{ij} for any pair of variables is $2^{-\alpha}$ (distribution C);
(d) a bivariate normal distribution with correlation coefficient ρ_{ij}, for which $\eta_{ij} = (1 + \rho_{ij})/2$ (distribution D).

Section 8 shows the theoretical derivation of extremal properties of these distributions. Distributions A and B are asymptotically dependent whereas distributions C and D are asymptotically independent. We select the parameters of distributions C and D so that $\eta_{ij} = 0.75$ for all bivariate pairs. For each distribution we simulated 200 replicate data sets each of size 5000. We applied a range of existing and conditional methods, selecting thresholds so that 10% of each data set was used for estimation by each method. We compare the performance of the methods for a range of forms of extreme event. Preliminary studies showed that the relative errors varied little with p and so we show results for $p = 10^{-4}, 10^{-6}, 10^{-8}$ only.

6.2.1. *Simultaneously extreme bivariate events*

Table 2 shows the median, 2.5 and 97.5 percentiles of the estimated sampling distribution of the relative errors of v_p when $C(v) = (v, \infty)^2$. First consider distributions A and B for which the existing method based on property (2.1) with $\eta_{\mathbf{Y}} = 1$ is asymptotically the correct form of model. The existing method with $\eta_{\mathbf{Y}} = 1$ has small relative errors centred on zero for distribution A, but for distribution B the method overestimates by a small but significant amount. The conditional method and the existing method with estimated $\eta_{\mathbf{Y}}$ are unbiased but have variable relative errors for distribution A. For distribution B the conditional model with weak pairwise exchangeability significantly underestimates whereas the existing method with $\eta_{\mathbf{Y}}$ estimated and the asymmetric conditional model are equally variable and unbiased. For the asymptotically independent distributions C and D, the estimators perform differently from one another but similarly over

Table 2. Median (and 2.5 and 97.5 percentiles) of the estimated sampling distribution of relative errors of v_p for simultaneously extreme bivariate events†

Distribution	*Method*	*Medians (and 2.5 and 97.5 percentiles) (× 100) for the following values of p:*		
		$p=10^{-4}$	$p=10^{-6}$	$p=10^{-8}$
A	Existing, $\eta_{\mathbf{Y}}=1$	−0.1 (−1.0, 0.8)	−0.1 (−0.7, 0.5)	−0.0 (−0.5, 0.4)
	Existing, $\hat{\eta}_{\mathbf{Y}}$	−0.8 (−15.0, 0.6)	−0.8 (−16.0, 0.4)	−0.6 (−17.0, 0.3)
	Conditional (weak pairwise exchangeability)	−1.4 (−4.0, 0.8)	−1.6 (−4.1, 0.5)	−1.6 (−5.0, 0.4)
B	Existing, $\eta_{\mathbf{Y}}=1$	4.6 (3.7, 5.6)	3.0 (2.4, 3.6)	2.1 (1.7, 2.5)
	Existing, $\hat{\eta}_{\mathbf{Y}}$	−1.0 (−14.0, 5.3)	−2.9 (−17.0, 3.4)	−3.9 (−19.0, 2.4)
	Conditional (weak pairwise exchangeability)	−15.0 (−21.0, −8.8)	−14.0 (−21.0, −7.4)	−12.0 (−19.0, −6.3)
	Conditional (asymmetry)	−4.0 (−12.0, 4.2)	−5.7 (−15.0, 0.5)	−6.1 (−17.0, 0.0)
C	Existing, $\eta_{\mathbf{Y}}=1$	23.0 (22.0, 24.0)	26.0 (26.0, 28.0)	29.0 (28.0, 29.0)
	Existing, $\hat{\eta}_{\mathbf{Y}}$	−0.6 (−16.0, 14.0)	−0.1 (−18.0, 17.0)	0.2 (−18.0, 18.0)
	Conditional (weak pairwise exchangeability)	−0.6 (−8.6, 5.3)	0.6 (−13.0, 8.2)	0.8 (−18.0, 9.8)
D	Existing, $\eta_{\mathbf{Y}}=1$	28.0 (27.0, 29.0)	31.0 (30.0, 32.0)	32.0 (32.0, 33.0)
	Existing, $\hat{\eta}_{\mathbf{Y}}$	−1.9 (−15.0, 14.0)	−1.9 (−17.0, 16.0)	−2.2 (−18.0, 17.0)
	Conditional (weak pairwise exchangeability)	−0.6 (−10.0, 7.3)	−0.1 (−15.0, 9.2)	−0.1 (−25.0, 12.0)

†The four distributions are listed in Section 6.2. The true return levels are, for distribution A, $v_p=8.6$, 13.0, 17.9, for distribution B, $v_p=7.8$, 12.4, 17.0, for distribution C, $v_p=6.5$, 9.7, 13.2, and, for distribution D, $v_p=6.9$, 10.2, 13.8, for $p=10^{-4}$, 10^{-6}, 10^{-8} respectively. Four methods of estimation are used: the existing method with $\eta_{\mathbf{Y}}=1$ and with $\eta_{\mathbf{Y}}$ estimated, and the conditional method with weak pairwise exchangeability and asymmetry.

distributions. The existing method with $\eta_{\mathbf{Y}}=1$ grossly overestimates. The other two methods are unbiased with the conditional approach generally having less variability. In Section 3.3 we noted a discontinuity in the normalizing parameters as independence is approached. We extended the above simulation study to assess the performance of the methods as these discontinuities are approached. For distributions A and C, as $\alpha \uparrow 1$, all methods perform similarly to the behaviour shown in Table 2 with a small bias observed for the $\eta_{\mathbf{Y}}=1$ approach, and both of the other two methods being unbiased with similar variances. In summary, these results suggest that the general performance of the conditional method is good but that, when asymmetry is present, the diagnostic procedures of Section 6.1 are vital for model selection.

6.2.2. *Non-simultaneously extreme bivariate events*

Now consider estimating quantiles of the distribution of $Y_2|Y_1>r$ for a given r, i.e. for a given q we estimate v satisfying $\Pr(Y_2<v|Y_1>r)=q$. Equivalently, we wish to estimate v where $C(v)=(r,\infty)\times(-\infty,v)$, and $\Pr(Y_1>r)=p/q$ where p and q are given. Table 3 shows summary characteristics of the sampling distribution of the relative error of the conditional method for combinations of p and q. For $p=10^{-4}$ and $q=0.2, 0.5, 0.8$ the respective true values of $r=7.6, 8.5, 9.0$, and the corresponding values for v are 6.7, 8.8 and 10.5 for distribution A, 6.2, 7.8 and 9.8 for distribution B, 2.5, 4.4 and 6.7 for distribution C, and 2.1, 3.6 and 5.7 for distribution D. This illustrates that if $q<1$ and the variables are asymptotically independent the existing methods are inappropriate for estimating v as all elements of $C(v)$ are not simultaneously extreme in each component. For each distribution, the estimators based on the conditional approach have a larger variance than in Table 2, with the variability increasing as

Table 3. Median (and 2.5 and 97.5 percentiles) of the estimated sampling distribution of relative errors of v_p for non-simultaneously extreme bivariate events†

Distribution (method)	*q*	*Medians (and 2.5 and 97.5 percentiles) (× 100) for the following values of p:*		
		$p=10^{-4}$	$p=10^{-6}$	$p=10^{-8}$
A (weak pairwise exchangeability)	0.2	−3.1 (−13, 2.7)	−4.7 (−15, 1.6)	−5.1 (−16, 1.0)
	0.5	−2.0 (−9.4, 1.2)	−2.5 (−11, 0.8)	−2.6 (−12, 0.5)
	0.8	−0.8 (−6.7, 3.8)	−0.9 (−7.8, 3.5)	−1.0 (−9.2, 2.9)
B (asymmetry)	0.2	−15 (−36, 0.7)	−17 (−43, −1.1)	−16 (−47, −2.0)
	0.5	−11 (−25, −0.9)	−12 (−29, −1.9)	−12 (−32, −2.2)
	0.8	−8.4 (−19, −0.3)	−9.1 (−21, −1.6)	−9.2 (−23, −1.6)
C (weak pairwise exchangeability)	0.2	6.7 (−17, 35)	16 (−18, 58)	25 (−19, 81)
	0.5	3.1 (−15, 23)	7.9 (−17, 36)	13 (−19, 51)
	0.8	0.5 (−18, 22)	2.6 (−21, 32)	5.5 (−20, 42)
D (weak pairwise exchangeability)	0.2	−4.4 (−37, 22)	8.5 (−34, 45)	17 (−32, 60)
	0.5	−1.5 (−22, 22)	4.4 (−23, 35)	6.3 (−27, 44)
	0.8	−1.3 (−25, 27)	−0.7 (−27, 33)	0.2 (−29, 41)

†The four distributions are listed in Section 6.2. The conditional method is used with weak pairwise exchangeability or asymmetry.

q is decreased. Only for long-range extrapolations for distribution B is there a significant bias, but even this is small. The relative errors are much the smallest for distribution A and grow as we extrapolate for distributions C and D.

6.2.3. *Multivariate events*

To illustrate the performance of the conditional method in higher dimensional problems, we consider the estimation of v when $C(v)=\{\mathbf{y}\in\mathbb{R}^5 : \Sigma_{i=1}^5 y_i > v\}$ for distributions A and C. For such sets, values of all the variables are equally influential and the set comprises regions of both simultaneous and non-simultaneous extreme values of the components. Distribution A exhibits asymptotic dependence without being asymptotically conditionally independent, whereas distribution C is asymptotically independent and asymptotically conditionally independent. Because of the symmetry of both dependence structures, in each case we fitted a model with $a_{j|i}=a$ and $b_{j|i}=b$ for all i and j. The limit values of (a,b) for distributions A and C are $(1,0)$ and $(0,0.585)$ respectively. We examined the observed components of the standardized residuals $\mathbf{Z}_{|i}$ for each $Y_i > u_{Y_i}$ to see whether asymptotic conditional independence was a reasonable assumption. Our findings agreed with the limiting properties, so we proceed to estimate v_p assuming asymptotic conditional independence for distribution C only. For distribution A we find that the median (and 2.5 and 97.5 percentiles) ($\times 100$) of the estimated sampling distribution of relative errors of v_p are 4.4 (1.7, 7.6) and 1.8 (−4.5, 6.8) for $p=10^{-4}$ and $p=10^{-6}$ respectively. The same quantities for distribution C are 0.1 (−6.6, 4.9) and −0.1 (−10.0, 7.4). The estimates are close to the true values with increasing variability in relative error for longer-range extrapolation.

7. Air quality monitoring application

We now analyse the extremes of the five-dimensional air pollutant variable that was presented in Section 1. The primary aim of this analysis is to study the underlying extremal dependence structure of the variables. By identifying this structure we can assess whether the relationships between the extreme values of these variables conform with scientific understanding of the production of and interaction between the pollutants and the climatic factor represented by season. We measure the extremal dependence by estimating the individual model parameters and by examining functionals of extremes of the joint distribution.

First we select the data to be analysed. The pollutants exhibit regular seasonal variation, which we account for by focusing separately on two periods: winter (from November to February inclusively) and early summer (from April to July inclusively) and treating the joint distribution of the pollutants as stationary in each period. This proposal is supported by empirical evidence and by knowledge of the seasonal behaviour of the variable (Photochemical Oxidants Review Group, 1997). The measurements follow a diurnal cycle and exhibit marked short-term dependence. By focusing on componentwise daily maxima of hourly means we remove this short-term non-stationarity and substantially reduce temporal dependence. The residual serial dependence is due, among other things, to short-term persistence of local atmospheric pressure systems. We do not attempt to take this temporal dependence into account in this analysis. The data set contains some large values on or around November 5th each year (fireworks night), which were removed for the subsequent analysis. An exploratory analysis also revealed six data points with excessive PM_{10}-values (in excess of 200 μg m^{-3}) during April 1997 and three winter points with unusually large values of some functionals of (NO_2, SO_2, PM_{10}). We performed the modelling and inference stages of the following analysis including and excluding these large points, to assess the sensitivity to their presence. The estimated dependence structures were not affected by the removal of these outliers; however, marginal estimates were more physically self-consistent when the points were left out. We report the analysis that was undertaken with the outliers excluded; these outliers are omitted from the data plots in Figs 1 and 2.

We fit the marginal model (1.3), for each component and for each season. Table 4 shows the resulting values of the threshold u_{X_i}, the threshold non-exceedance probability $\hat{F}_{X_i}(u_{X_i})$, the estimated generalized Pareto distribution parameters $\hat{\beta}_i$ and $\hat{\xi}_i$, and the estimated marginal 0.99 quantiles $\hat{x}_i(0.99) = \hat{F}_{X_i}^{-1}(0.99)$ for each component and each season. The values of $\hat{x}_i(0.99)$ highlight differences in the marginal distributions of separate components within each season

Table 4. Summary of generalized Pareto models fitted to the marginal distributions of the air pollution data†

Season	*Parameter*	*Results for the following pollutants:*				
		O_3	NO_2	NO	SO_2	PM_{10}
Summer	u_{X_i}	43.0	66.1	43.0	22.0	46.0
	$\hat{F}_{X_i}(u_{X_i})$	0.9	0.7	0.7	0.85	0.7
	$\hat{\beta}_i$	15.8 (3.1)	9.1 (1.0)	32.2 (3.5)	42.9 (7.0)	22.8 (2.5)
	$\hat{\xi}_i$	−0.29 (0.14)	0.01 (0.08)	0.02 (0.07)	0.08 (0.12)	0.02 (0.08)
	$\hat{x}_i(0.99)$	70 (2)	75 (3)	180 (10)	152 (16)	127 (8)
Winter	u_{X_i}	28.0	151.6	49.0	23.0	53.0
	$\hat{F}_{X_i}(u_{X_i})$	0.7	0.7	0.7	0.7	0.7
	$\hat{\beta}_i$	6.2 (0.7)	9.3 (0.9)	117.4 (13.1)	19.7 (2.4)	37.5 (4.2)
	$\hat{\xi}_i$	−0.37 (0.06)	−0.03 (0.08)	−0.09 (0.08)	0.11 (0.09)	−0.20 (0.07)
	$\hat{x}_i(0.99)$	40 (1)	80 (3)	494 (30)	104 (10)	145 (6)

†Thresholds used for marginal modelling are denoted u_{X_i}; the associated non-exceedance probabilities are $\hat{F}_{X_i}(u_{X_i})$; estimated scale $\hat{\beta}_i$ and shape $\hat{\xi}_i$ parameters; estimated 0.99 quantiles $\hat{x}_i(0.99)$. Bootstrap-based standard errors are given in parentheses.

and of the same component over seasons, with O_3 and NO exhibiting the largest statistically significant variation over seasons. Though $\hat{\xi}_i$ differ over components, the stability of the $\hat{\xi}_i$-values for each component over season suggests that seasonality primarily affects the variance of the components rather than the shape of their distributions.

Let us now consider the dependence model, applied to the data after transformation to Gumbel margins by using transformation (1.4). We model each season separately, initially considering the most general model consisting of a set of five conditional models with no constraints between the different conditional distributions, so $\boldsymbol{\theta}$ consists of 80 unconstrained dependence parameters.

The first modelling choice to be made is that of the dependence threshold to be used to fit the conditional model (4.1). For simplicity we restrict the search for the dependence thresholds to values of $u_{Y_i} = u$ for all i. Of the diagnostics that are discussed in Section 4.4, we found that those assessing the stability of the $\hat{\boldsymbol{\theta}}$-values and the independence tests were the most revealing in this application. A dependence threshold u such that $\Pr(Y_i < u) = 0.7$ was supported by the diagnostics, although there appeared to be limited sensitivity to this choice. The resulting $(\hat{\mathbf{a}}_{|i}, \hat{\mathbf{b}}_{|i})$ values, and the sampling distributions of pairs $(\hat{a}_{j|i}, \hat{b}_{j|i})$ for all $i \neq j$, are shown for the summer and winter seasons in Fig. 5. In particular, the pairwise sampling distributions are shown by the convex hull of 100 bootstrap realizations from the sampling distribution of $\hat{\boldsymbol{\theta}}$. Plots of this type were used to assess the stability of $\hat{\boldsymbol{\theta}}$ to the choice of threshold. Significant shifts in the region that is encompassed by the convex hull indicate sensitivity of parameter estimates to the choice of threshold. An appropriate threshold should have the property that raising the threshold higher does not result in any significant shifts once the increased variability of estimates made by using higher thresholds is accounted for. The minimum such appropriate threshold is selected for efficiency purposes.

Having decided on a dependence threshold, we consider possible simplifications to the estimated dependence structure. From Fig. 5 and plots of the data with Gumbel marginals (Fig. 2), it is clear that there are significant differences in levels of extremal dependence between different pairs of variables. Fig. 5 shows that, for each season, there are pairs of variables for which the bivariate sampling distributions of $(\hat{a}_{i|j}, \hat{b}_{i|j})$ and $(\hat{a}_{j|i}, \hat{b}_{j|i})$ differ significantly, as the convex hulls do not intersect. For example, Fig. 5 shows that in summer (PM_{10}, O_3) and (SO_2, NO) and in winter (SO_2, NO_2) and (SO_2, NO) do not exhibit weak pairwise exchangeability. This finding indicates that a global weakly pairwise exchangeable dependence structure is inappropriate for these data, a conclusion which is supported in the winter period by a complete lack of stability in $\hat{\boldsymbol{\theta}}$ over all dependence threshold choices for the global weakly pairwise exchangeable model. Though for some pairs of pollutants there is no evidence to reject weak pairwise exchangeability, in the absence of more detailed knowledge about the process we do not attempt to identify subsets of pairs for which we may assume a simplified pairwise dependence model. Finally, we consider whether the $\mathbf{Z}_{|i}$ are independent for any i, i.e. whether we can assume asymptotic conditional independence between the margins of the residual distribution $G_{|i}$. Scatterplots of pairs of components of $\mathbf{Z}_{|i}$ for each i confirm that this assumption is inappropriate. Testing for asymptotic conditional independence between pairs of variables revealed that, for the summer data, SO_2 and O_3 are asymptotically conditionally independent given any other variable, although these two variables are not unconditionally independent. The same conclusion can be drawn for winter NO_2 and O_3 levels.

Fig. 5 also shows substantial differences between the dependence parameter estimates that are obtained for the summer and winter data sets. All pairs which have O_3 as one component exhibit stronger dependence in the summer period than in the winter period, whereas for other pairs the dependence is either of similar strength or weaker in summer than in winter. The strongest

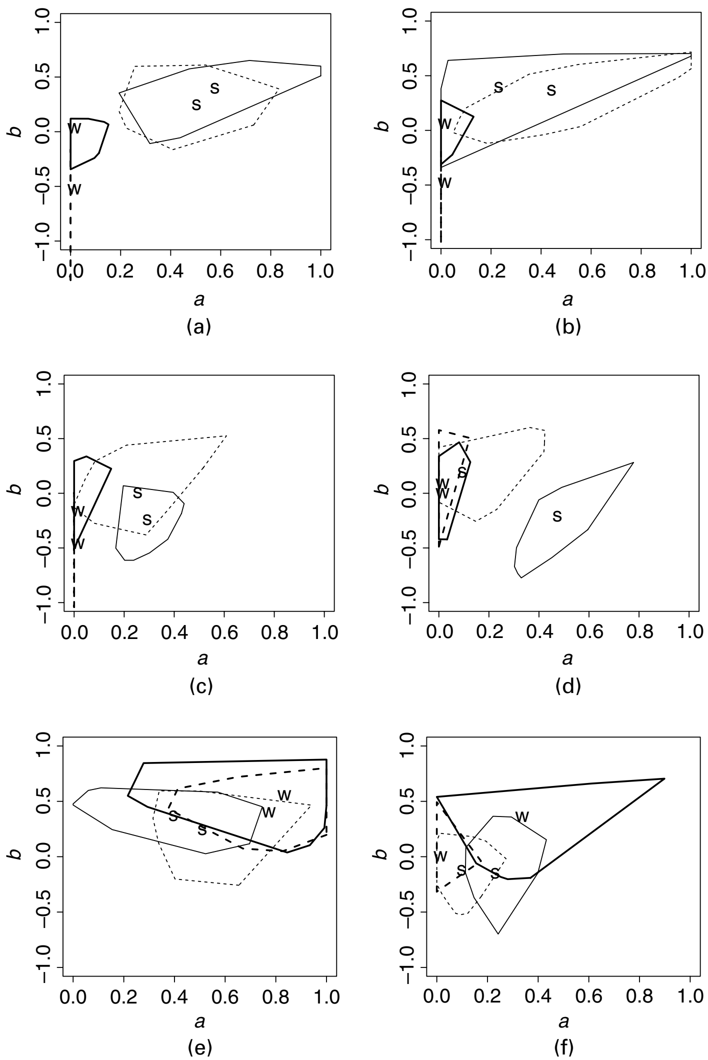


Fig. 5. Comparison of dependence parameter estimates $(\hat{a}_{i|j}, \hat{b}_{i|j})$ for (a) (NO_2, O_3), (b) (NO, O_3), (c) (SO_2, O_3), (d) (PM_{10}, O_3), (e) (NO, NO_2), (f) (SO_2, NO_2), (g) (PM_{10}, NO_2), (h) (SO_2, NO), (i) (PM_{10}, NO) and (j) (PM_{10}, SO_2), using a dependence threshold equal to the 70% marginal quantile: for i and j in the same order as the variables in the descriptor for each part of the figure, bootstrap convex hulls were used for $(\hat{a}_{i|j}, \hat{b}_{i|j})$ (———, summer; **———**, winter) and for $(\hat{a}_{j|i}, \hat{b}_{j|i})$ (· · · · · ·, summer; - - - - - - -, winter) (associated point estimates: s, summer; w, winter)

dependence between any pair occurs in the winter between all the pairs of the triple (NO, NO_2, PM_{10}), with reasonable evidence that these variables are asymptotically dependent. No other pairs of variables exhibit asymptotic dependence in either season.

For non-positively associated variables, estimates of $(c_{j|i}, d_{j|i})$ (not shown) reveal the degree of dependence. In the summer, the only such conditional distribution is that of NO given extreme SO_2, although with the conditioning reversed these variables are clearly positively dependent. In winter, SO_2 and PM_{10} are both negatively dependent on high O_3 values, whereas NO_2 and NO appear to be independent of O_3 when O_3 is extreme. Conversely, in winter O_3 is negatively associated with extreme NO_2, NO and SO_2. Negative dependence is also identified for all winter

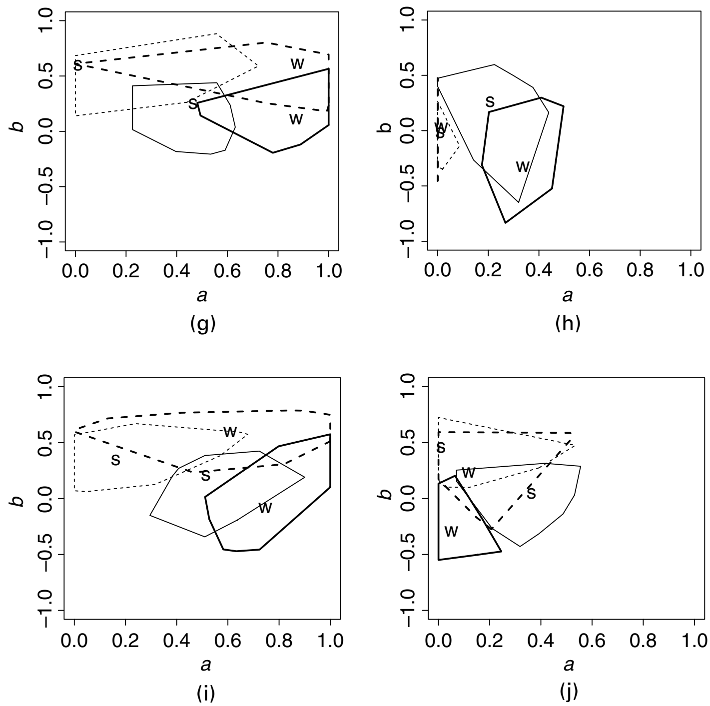


Fig. 5 (*continued*)

variables given that SO_2 is extreme, with the exception of PM_{10} which appears to be independent of extreme SO_2.

These findings are consistent with the current understanding of urban pollution patterns. In winter, air pollution episodes typically occur when cold, stable weather conditions trap pollutants, allowing levels to build. Since the majority of such pollution derives from vehicle emissions, winter episodes consist of simultaneously elevated levels of nitrogen and sulphur compounds and particulate matter. Conversely, since the production of excessive O_3 needs strong sunlight, O_3 levels generally remain at relatively low levels during the winter months regardless of the presence of other pollutants. In the absence of strong sunlight, O_3 levels are negatively associated with high presences of nitrogen compounds as O_3 reacts destructively with NO. The stronger dependence that is observed between O_3 and the other variables during the summer supports the existing understanding of the photochemical processes that produce excessive O_3 levels during summer smog. Temperature inversions and low winds that accompany high pressure systems trap vehicle emissions, which are then exposed to long hours of sunshine. Thus high levels of O_3 accompany elevated levels of the other pollutants (Photochemical Oxidants Review Group, 1997; Colls, 2002; Housley and Richards, 2001).

To illustrate the implications of both the different levels of dependence between the pairs and the different marginal distributions, in Fig. 6 we show pseudosamples, on the measured scale, from the conditional distribution of the remaining variables given that NO exceeds a high threshold. On each pairwise plot the curve corresponds to equal marginal quantiles. The near asymptotic dependence of both NO_2 and PM_{10} on NO is clearly seen by the grouping of simulated points around this curve. The NO_2 points are more scattered than the PM_{10} points for large NO values as $\hat{b}_{j|i}$ are positive and negative respectively. Similarly, O_3 is seen to be

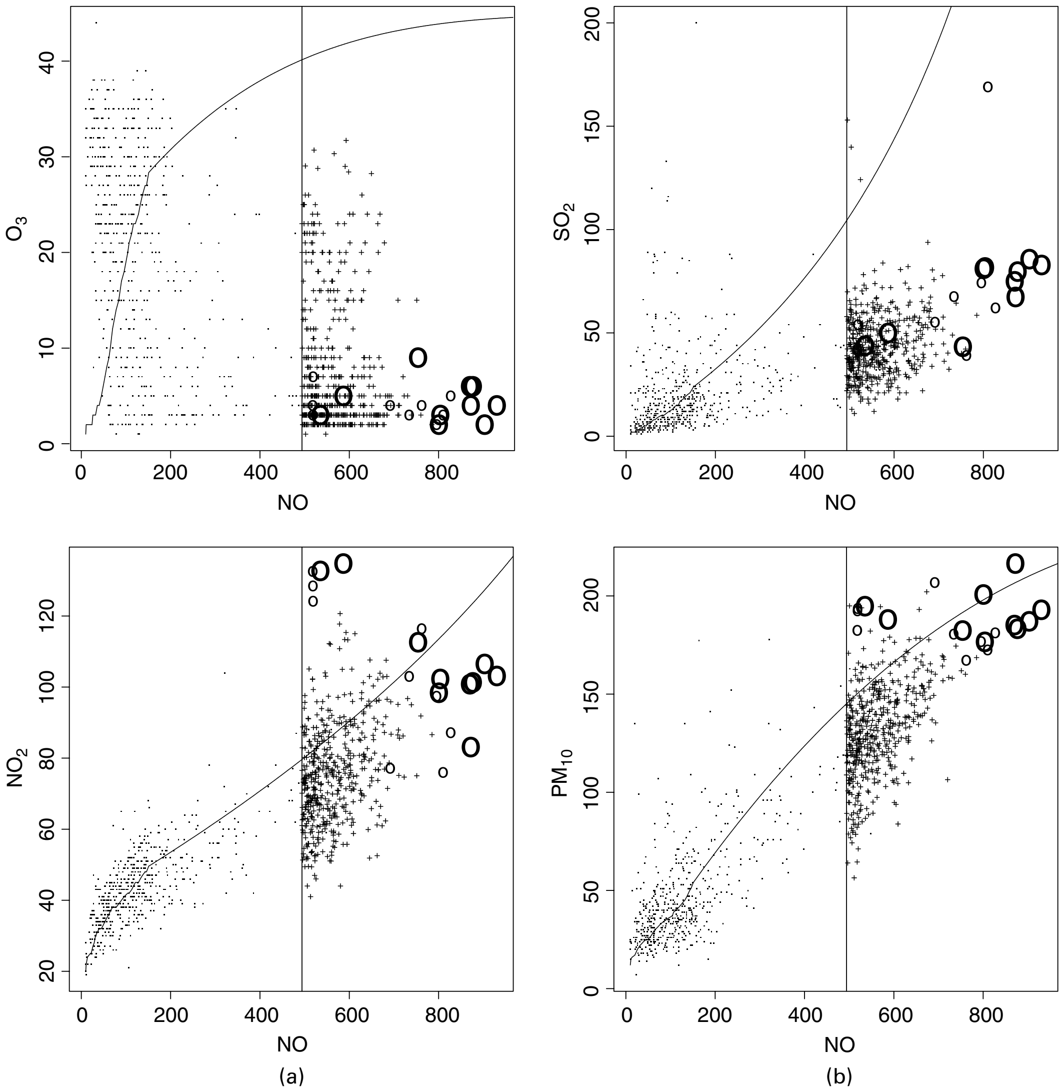


Fig. 6. Simulated winter air pollution points conditional on the NO component exceeding v_{X_i}, the 0.99 marginal quantile of this variable: |, threshold v_{X_i} (points below and above this threshold are the original data and data simulated under the fitted model respectively); +, points that do not fall in the set $C^5(23)$; ∘, points that fall in the set $C^5(23)$; ○, 10 points with the largest values of $\Sigma_{i=1}^5 y_i$; ———, equal marginal quantiles

negatively dependent on NO whereas SO_2 is dependent but asymptotically independent of NO. The effect of the negative $\hat{b}_{j|i}$ for SO_2|NO is the increasing concentration of this conditional distribution for larger NO values.

We now focus on estimating a range of functionals of the joint tails of **X**. Coles and Tawn (1994) discussed several benefits of the multivariate approach (the joint probability method) over the univariate approach (the structure variable method). We see the major advantage of the former being the self-consistency of the resulting estimates of any such functionals; this is particularly important here where we illustrate a range of functionals for which no single structure variable approach could have been used.

Table 5. Empirical and model-based estimates of conditional expectations of the air pollution variables given values of NO in excess of a range of quantiles of that variable†

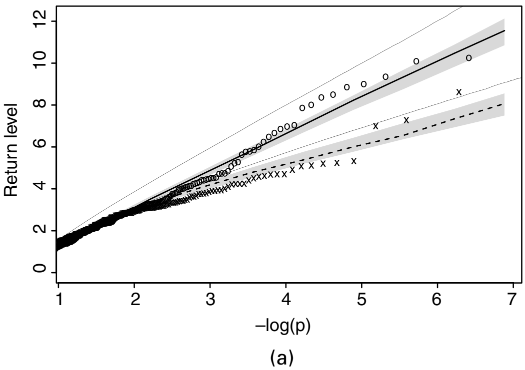
X_j	*Season*	$E(X_j)$, *empirical*	$E\{X_j \mid X_i > x_i(0.95)\}$		$E\{X_j \mid X_i > x_i(0.99)\}$, *model based*
			Empirical	*Model based*	
O_3	Winter	20.0 (0.5)	8.8 (1.4)	10.3 (1.1)	8.3 (1.2)
	Summer	32.0 (0.4)	35.9 (3.0)	34.4 (2.4)	39.6 (4.3)
NO_2	Winter	44.2 (0.5)	67.2 (2.5)	65.1 (2.2)	75.4 (4.4)
	Summer	37.6 (0.5)	57.5 (2.6)	54.6 (2.4)	62.2 (4.3)
NO	Winter	135.5 (4.4)	454.0 (13.0)	431.5 (23.2)	569.9 (45.2)
	Summer	55.2 (1.5)	161.2 (7.2)	157.6 (8.2)	213.5 (17.5)
SO_2	Winter	21.0 (0.9)	38.4 (3.7)	35.6 (4.0)	44.6 (6.7)
	Summer	17.4 (1.2)	36.6 (11.3)	36.9 (5.4)	48.5 (11.8)
PM_{10}	Winter	48.4 (1.2)	105.8 (5.2)	105.0 (4.7)	132.3 (8.2)
	Summer	41.1 (1.0)	72.9 (5.2)	66.3 (4.5)	83.7 (7.9)

†Standard errors are given in parentheses. Variable X_i is NO throughout.

We first turn to the estimation of the conditional expectation of each component given that NO exceeds a particular level. These estimates reflect both the marginal and the dependence features of the air pollution variables. Fig. 6 shows pseudosamples from the conditional distribution of each variable given that NO exceeds its 0.99 marginal quantile. Table 5 shows estimated expectations for each variable conditional on the NO level exceeding various thresholds. When we condition on NO exceeding its 0.95 quantile, empirical estimates of this functional are sufficiently reliable to be compared with the model-based estimates and these are seen to be consistent. Conditional expectations of each variable increase as we move to higher quantiles of NO, with the exception of winter O_3, the only variable to exhibit negative association with large NO values.

We now concentrate on the estimation of return levels of linear combinations of variables on the Gumbel marginal scale. This choice of functional is made to emphasize the effect of dependence on extreme combinations. We focus on subvectors of $\mathbf{Y}$ of size $m = 2, \ldots, d$, indexed by $\mathcal{M} \subseteq \{1, \ldots, d\}$ with associated multidimensional sets $C^m(v) = \{\mathbf{y} \in \mathbb{R}^m : \Sigma_{i \in \mathcal{M}} y_i > v\}$ and we report estimated return levels v_p as defined in equation (6.1). This choice of set allows an exploration of extremal dependence in parts of the space in which not all the variables are simultaneously extreme. To gain insight about combinations of the pollutants that fall in the set $C^5(v)$ for large v, in Fig. 6 we highlight the simulated points with NO exceeding its 0.99 quantile that fall in $C^5(23)$ and indicate which of these have the largest values of $\Sigma_{i=1}^5 y_i$. Simulated points in $C^5(23)$ tend not to have particularly large values of O_3 but do occur with moderate SO_2 values and extreme values of NO_2 and PM_{10}. The strong dependence between (NO, NO_2, PM_{10}) leads to the largest values in $C^5(23)$ occurring when any one of these variables is extreme.

Fig. 7 shows empirical and model-based return level estimates for $C^m(v)$ for $\mathcal{M}$ corresponding to (O_3, NO_2) and (NO_2, SO_2, PM_{10}). Return levels calculated under independence and perfect dependence are also marked. For the pair (O_3, NO_2) the $C^2(11)$ set is shown after transformation to the original margins in Fig. 1. High levels of O_3 and NO_2 are associated with summer photochemical smog. Empirical return level estimates show that stronger dependence between these variables during the summer leads to elevated return levels. Model-based return levels agree closely with the empirical values and show that this seasonal difference is statistically significant as the confidence intervals for the return levels are separated. The estimated return levels for



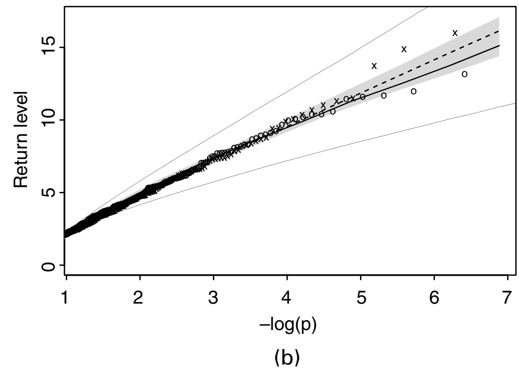


Fig. 7. Return level estimates for the set $C^m(v)$, for $\mathcal{M}$ corresponding to (a) (O_3, NO_2) and (b) (NO_2, SO_2, PM_{10}): ———, point estimates for summer; - - - - - - -, point estimates for winter variables; ■, pointwise 95% confidence intervals (which overlap in (b)); ∘, empirical points for summer; ×, empirical points for winter; ———, return levels calculated under perfect dependence (upper) and exact independence (lower)

the winter (O_3, NO_2) lie significantly below the independence curve, highlighting the negative dependence between these variables during the winter. Elevated levels of all three components of (NO_2, SO_2, PM_{10}) are associated with winter urban air pollution episodes and correspondingly we see larger return levels in winter, indicating stronger dependence between these three variables in this season, although this effect is not significant as the confidence intervals overlap. Both Fig. 7(a) and Fig. 7(b) show excellent agreement between the model-based return level estimates and the empirical estimates, illustrating the good fit of our dependence model.

8. Theoretical examples

We now derive the limiting conditional characteristics that were identified in Section 3 for a range of theoretical examples including those summarized in Table 1. Where possible, results are given for a d-dimensional random variable $\mathbf{Y}$, and in a few special cases for bivariate $\mathbf{Y}$ only.

First we give the precise form of the multivariate extreme value distribution which plays a key role in the examples that are given in this section. A d-dimensional random variable $\mathbf{Y}$ with standard Gumbel margins has a multivariate extreme value distribution if its joint distribution function can be expressed as (Pickands, 1981)

$$\Pr(\mathbf{Y}<\mathbf{y}) = \exp[-V^{(d)}\{\exp(\mathbf{y})\}] \tag{8.1}$$

where $V^{(d)}$, termed the *exponent measure*, is given by

$$V^{(d)}(\mathbf{y}) = d\int_{\mathcal{S}_{d-1}} \max_{1\leqslant j\leqslant d}(w_j y_j^{-1})\,\mathrm{d}H^{(d)}(\mathbf{w}), \tag{8.2}$$

where $H^{(d)}$ is the distribution function of an arbitrary random variable on the $(d-1)$-dimensional unit simplex

$$\mathcal{S}_{d-1} = \left\{\mathbf{w} = (w_1,\ldots,w_{d-1}) : \sum_{j=1}^{d-1} w_j \leqslant 1, w_j \geqslant 0 : j = 1,\ldots,d-1\right\}$$

satisfying the marginal moment constraint

$$\int_{\mathcal{S}_{d-1}} w_j\,\mathrm{d}H^{(d)}(\mathbf{w}) = d^{-1} \qquad \text{for each } j = 1,\ldots,d, \tag{8.3}$$

and $w_d = 1-(w_1+\ldots+w_{d-1})$. We refer to a multivariate extreme distribution as having *mass on the boundary* if $H^{(d)}$ places mass on the boundary of $\mathcal{S}_{d-1}$. We denote by $m_{j|i}$ the mass on the boundary of $\mathcal{S}_{d-1}$ for which $w_j=0$ and $w_i>0$ and let $M_i=\{j : m_{j|i}>0; j\neq i\}$ and $\bar{M}_i=\{1,\ldots,d\}\backslash(M_i\cup\{i\})$. The density of $H^{(d)}$ on the interior of $\mathcal{S}_{d-1}$ is denoted by $h^{(d)}$ when it exists. Some parametric examples of $V^{(d)}$ are given now.

8.1. *Multivariate exchangeable logistic distribution*

Gumbel (1960) introduced the multivariate exchangeable logistic distribution with

$$V^{(d)}(\mathbf{y}) = \left(\sum_{j=1}^{d} y_j^{-1/\alpha}\right)^{\alpha}, \tag{8.4}$$

for any $d\geqslant 2$ and $0<\alpha\leqslant 1$. Independence is given by $\alpha=1$ and perfect positive dependence in the limit as $\alpha\to 0$. There is no mass on the boundary of $\mathcal{S}_{d-1}$ for $0<\alpha<1$.

8.2. *Multivariate asymmetric logistic distribution*

The multivariate asymmetric logistic dependence structure given by Tawn (1990) has

$$V^{(d)}(\mathbf{y}) = \sum_{K\in S}\left\{\sum_{j\in K}\left(\frac{\theta_{j,K}}{y_j}\right)^{1/\alpha_K}\right\}^{\alpha_K}, \tag{8.5}$$

where K is an index variable over the power set S of $\{1,\ldots,d\}$, $0<\alpha_K\leqslant 1$ for all $K\in S$ and $0\leqslant\theta_{j,K}\leqslant 1$ for $j=1,\ldots,d$. Further conditions on $\theta_{j,K}$ are that $\theta_{j,K}=0$ if $j\notin K$ or if $\Pi_{k\in K}\,\theta_{k,K}=0$ and that $\Sigma_{K\in S}\,\theta_{j,K}=1$ for all j. Similarly, for identifiability, $\alpha_K=1$ when $|K|=1$. If, for any K, there exists $j\in K$ with $\theta_{j,K}>0$ then there is positive association between the elements of $\mathbf{Y}_K$, where $\mathbf{Y}_K$ is the subvector of $\mathbf{Y}$ made up of the variables that are indexed by the elements of set K. For this example $m_{j|i}=\Sigma_{K\in S^{(j)}\backslash S^{(i)}}\,\theta_{j,K}$ where $S^{(i)}$ denotes the subclass of S, all of whose members contain i.

8.3. *Bivariate discrete measure*

Ledford and Tawn (1998) defined the distribution for which $H^{(2)}$, satisfying constraint (8.3), places m atoms of mass $\lambda_1,\ldots,\lambda_m$ at points $w_1,\ldots,w_m$ on the interior of $\mathcal{S}_1$:

$$H^{(2)}(w) = \sum_{i=1}^{m}\lambda_i\, I(w\geqslant w_i),$$

where I is the indicator function. For such $H^{(2)}$,

$$V^{(2)}(\mathbf{y}) = \frac{1}{y_1}\left(1-2\sum_{i:w_i\leqslant w^*} w_i\lambda_i\right) + \frac{2}{y_2}\sum_{i:w_i\leqslant w^*}(1-w_i)\lambda_i,$$

where $w^* = y_1/(y_1+y_2)$. There is no mass on the boundary of $\mathcal{S}_1$.

8.4. *Multivariate extreme value distribution*

For $\Pr(\mathbf{Y}<\mathbf{y})$ given by equation (8.1), the conditional distribution function of $\mathbf{Y}_{-i}|Y_i=y_i$ is

$$\Pr(\mathbf{Y}_{-i}<\mathbf{y}_{-i}|Y_i=y_i) = -\exp[-V^{(d)}\{\exp(\mathbf{y})\}]V_{:i}^{(d)}\{\exp(\mathbf{y})\}\exp\{2y_i+\exp(-y_i)\},$$

where $V_{:i}^{(d)}(\mathbf{y})$ is the derivative of $V^{(d)}(\mathbf{y})$ with respect to y_i. If $H^{(d)}$ places any mass on the subset of $\mathcal{S}_{d-1}$ for which $w_iw_j>0$ then $\eta_{ij}=1$; otherwise $\eta_{ij}=\frac{1}{2}$. If $m_{j|i}=0$ there is a unique normalization but if $0<m_{j|i}<1$ there are two normalizations that give non-degenerate $G_{j|i}(z_{j|i})$.

Let $Z_{j|i} = Y_j - a_{j|i} y_i$; then there are non-degenerate limits for $a_{j|i} = 0$ if $m_{j|i} > 0$ and $a_{j|i} = 1$ irrespective of $m_{j|i}$, with the combined limit being

$$G_{j|i}(z_{j|i}) = m_{j|i}[\exp\{-\exp(-z_{j|i})\}]^{I(a_{j|i}=0)} + I(a_{j|i}=1) \lim_{s\to\infty} (-V_{:i}^{(d)}[\exp\{\mathbf{z}(0,i)\}]|_{z_{k|i}=s:k\neq i,j})$$

for $-\infty < z_{j|i} < \infty$ and where $\mathbf{z}(0,i) = (z_{1|i}, \ldots, z_{i-1|i}, 0, z_{i+1|i}, \ldots, z_{d|i})$. Thus $G_{j|i}$ has mass $1 - m_{j|i}$ at $z_{j|i} = \infty$ if $a_{j|i} = 0$ or mass $m_{j|i}$ at $z_{j|i} = -\infty$ if $a_{j|i} = 1$. This may appear to contradict the uniqueness properties, up to type, of the normalization and limit law that was discussed in Section 3.2. However, only one of these non-degenerate limits has no mass at ∞, so we are interested in only one of these limits. When $m_{j|i} = 0$ for all $j \neq i$, the limiting joint conditional distribution of $\mathbf{Z}_{|i}$ is

$$G_{|i}(\mathbf{z}_{|i}) = -V_{:i}^{(d)}[\exp\{\mathbf{z}(0,i)\}].$$

We illustrate these limit properties with the three examples above. For the exchangeable logistic distribution, normalization $\mathbf{Z}_{|i} = \mathbf{Y}_{-i} - y_i$ gives

$$G_{|i}(\mathbf{z}_{|i}) = \left\{1 + \sum_{j\neq i} \exp\left(-\frac{z_{j|i}}{\alpha}\right)\right\}^{\alpha-1}.$$

For the asymmetric logistic distribution, $m_{j|i}$ is given above and

$$\lim_{s\to\infty} (-V_{:i}^{(d)}[\exp\{\mathbf{z}(0,i)\}]\big|_{z_{k|i}=s:k\neq i,j}) = \sum_{K\in S^{(i)}\cap S^{(j)}} \left[\left\{\theta_{j,K}^{1/\alpha_K} \exp\left(-\frac{z_{j|i}}{\alpha_K}\right)\right\} + \theta_{i,K}^{1/\alpha_K}\right]^{\alpha_K-1} \theta_{i,K}^{1/\alpha_K}.$$

For the bivariate discrete measure dependence structure, setting $\mathbf{Z}_{|i} = \mathbf{Y}_{-i} - y_i$ yields

$$G_{j|i}(z_{j|i}) = \begin{cases} 1 - 2 \sum_{w_k \leqslant w_{j|i}^*} \lambda_k w_k, & \text{if } i=1, \\ 2 \sum_{w_k \leqslant w_{j|i}^*} \lambda_k (1-w_k), & \text{if } i=2, \end{cases}$$

where $w_{j|i}^* = 1/\{1+\exp(z_{j|i})\}$ when $i=1$ and $w_{j|i}^* = 1/\{1+\exp(-z_{j|i})\}$ when $i=2$.

8.5. *Inverted multivariate extreme value distribution*

Ledford and Tawn (1997) examined the inverted bivariate extreme value distribution, which we extend here to the multivariate case. For $V^{(d)}$ defined in equation (8.2), the survivor function of this multivariate distribution is given by

$$\Pr(\mathbf{Y} > \mathbf{y}) = \exp\{-V^{(d)}(-1/\log[1-\exp\{-\exp(-\mathbf{y})\}])\}.$$

For this distribution $\eta_{ij} = 1/\lim_{s\to\infty}\{V^{(d)}(\mathbf{y})\big|_{y_i=y_j=1; y_k=s \forall k\neq i,j}\}$. Assuming that the Y_j grow with Y_i, then, as $y_i \to \infty$,

$$\Pr(\mathbf{Y}_{-i} > \mathbf{y}_{-i} | Y_i = y_i) \approx -\exp\{y_i - V^{(d)}(\mathbf{y}^{-1})\} V_{:i}^{(d)}(\mathbf{y}^{-1})/y_i^2.$$

Further simplification is not possible without more information about the shape of $H^{(d)}$ around $w_i = 0$. We first consider the bivariate case where all the mass of $H^{(2)}$ is in the interior of $\mathcal{S}_1$ and the measure density satisfies $h^{(2)}(w_i) \sim s_i w_i^{t_i}$ as $w_1 \to 0$ and $w_2 \to 1$ for $0 < s_i$ and $-1 < t_i$ for $i = 1, 2$. The transformation $Z_{j|i} = Y_j / y_i^{b_{j|i}}$ where $b_{j|i} = (t_i+1)/(t_i+2)$ gives the following limiting survivor function of variable $Z_{j|i}$:

$$\bar{G}_{j|i}(z_{j|i}) = \exp\left\{-\frac{s_i z_{j|i}^{t_i+2}}{(t_i+1)(t_i+2)}\right\}.$$

Thus the limiting distribution of $Z_{j|i}$ is Weibull.

We now consider the logistic examples. When $V^{(d)}$ is of exchangeable logistic form, $\eta_{ij} = 2^{-\alpha}$. Using normalization $\mathbf{Z}_{|i} = \mathbf{Y}_{-i}/y_i^{\mathbf{b}_{|i}}$, where $b_{j|i} = 1-\alpha$ for all $j \neq i$, gives the limiting survivor function of the variable $\mathbf{Z}_{|i}$:

$$\bar{G}_{|i}(\mathbf{z}_{|i}) = \prod_{j\neq i} \exp(-\alpha z_{j|i}^{1/\alpha}); \tag{8.6}$$

hence the $\mathbf{Z}_{|i}$ are asymptotically conditionally independent Weibull variables. Although not in the multivariate extreme class of distributions, the inverted multivariate Crowder distribution (Crowder, 1989) has the same η_{ij} as the inverted multivariate extreme value distribution with exchangeable logistic dependence structure and the same values of $\mathbf{a}_{|i}$ and $\mathbf{b}_{|i}$ with $\bar{G}_{|i}(\mathbf{z}_{|i})$ as in equation (8.6).

When $V^{(d)}$ is of asymmetric logistic form, $\eta_{ij} = 1/\Sigma_{K\in S}(\theta_{i,K}^{1/\alpha_K} + \theta_{j,K}^{1/\alpha_K})^{\alpha_K}$. Let $Z_{j|i} = Y_j/y_i^{b_{j|i}}$; then $b_{j|i} = I(j \in \bar{M}_i)(1-\alpha^{(ij)})$ gives non-degenerate $G_{j|i}(z_{j|i})$ where $\alpha^{(ij)} = \max(\alpha_K : K \in S^{(i)} \cap S^{(j)})$. Let $K^{(ij)}$ be the set of $\{K : K \in S^{(i)} \cap S^{(j)}\ \&\ \alpha_K = \alpha^{(ij)}\}$. Under this normalization, the joint survivor function for the $\mathbf{Z}_{|i}$ is

$$\begin{aligned}\bar{G}_{|i}(\mathbf{z}_{|i}) = &\prod_{j\in\bar{M}_i} \exp(-\alpha^{(ij)} A_{ij} z_{j|i}^{1/\alpha^{(ij)}})\\ &\times \exp\Big(\sum_{K\in S\setminus C^{(i)}}\Big[\sum_{j\in K\cap M_i}\{\theta_{j,K}(-\log[1-\exp\{\exp(-z_{j|i})\}])\}^{1/\alpha_K}\Big]^{\alpha_K}\Big),\end{aligned}$$

where $A_{ij} = \Sigma_{K\in K^{(ij)}} \theta_{i,K}(\theta_{j,K}/\theta_{i,K})^{1/\alpha^{(ij)}}$. Thus the variables in set $\overline{M}_i$ are asymptotically conditionally independent whereas the variables in M_i are not. Variables in set M_i are asymptotically conditionally independent of those in set $\overline{M}_i$.

8.6. Multivariate normal distribution

Let $\mathbf{V}$ be a d-dimensional random variable, distributed as a standard multivariate normal random variable, with correlation matrix Σ. Let $\mathbf{Y}$ represent $\mathbf{V}$ after transformation to Gumbel marginal distributions, via marginal transformations:

$$\mathbf{Y} = -\log[-\log\{\Phi(\mathbf{V})\}], \tag{8.7}$$

where Φ is the standard normal distribution function. The pairwise coefficient of tail dependence for this distribution is $\eta_{ij} = (1+\rho_{ij})/2$. We use Mill's ratio to approximate transformation (8.7) for large positive (or negative) components v and y of $\mathbf{v}$ and $\mathbf{y}$ to give

$$v \approx \begin{cases} \surd(2y) - \{\log(y) + \log(4\pi)\}/2(2y)^{1/2} & \text{for large positive } y,\\ -\surd 2\exp(-y/2) + \exp(y/2)\{\log(4\pi) - y\}/\surd 2 & \text{for large negative } y, \end{cases} \tag{8.8}$$

and

$$y \approx \begin{cases} \log(v) + \log(2\pi)/2 + v^2/2 & \text{for large positive } v,\\ -\log\{\log(2\pi)/2 + v^2/2 + \log(-v)\} & \text{for large negative } v. \end{cases} \tag{8.9}$$

The normalization that is used to give a non-degenerate limit for $\mathbf{Z}_{|i}$ is

$$Z_{j|i} = \begin{cases} (Y_j - \rho_{ij}^2 y_i)/y_i^{1/2} & \text{for } \rho_{ij} > 0, \\ \{Y_j + \log(\rho_{ij}^2 y_i)\}/y_i^{-1/2} & \text{for } \rho_{ij} < 0. \end{cases}$$

To determine the limiting distribution of $\mathbf{Z}_{|i}$ we use the property that the event $\mathbf{Z}_{|i} \leqslant \mathbf{z}_{|i} | Y_i = y_i$ for large y_i can be approximated by the event $\mathbf{V}_{-i} \leqslant \mathbf{v}_{-i} | V_i = v_i$ for large v_i where $v_i = \Phi^{-1}[\exp\{-\exp(-y_i)\}]$ and $\mathbf{v}_{-i}$ has elements $v_{j|i}$ which using expressions (8.8) and (8.9) are found to satisfy

$$v_{j|i} \sim \rho_{ij} v_i + z_{j|i} |\rho_{ij}|^{-\operatorname{sgn}(\rho_{ij})} / \sqrt{2}.$$

The conditional distribution of $\mathbf{V}_{-i} | V_i = v_i$ is $(d-1)$-dimensional multivariate normal with mean vector $\boldsymbol{\rho}_{-i} v_i$ and covariance matrix $\Sigma_{-i} - \boldsymbol{\rho}_{-i}^{\mathrm{T}} \boldsymbol{\rho}_{-i}$, where $\boldsymbol{\rho}_{-i}$ is the ith column of Σ with ith element omitted and Σ_{-i} is Σ with ith row and ith column omitted. Hence it follows that the $\mathbf{Z}_{|i}$ are jointly $(d-1)$-dimensional multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{S}(\Sigma_{-i} - \boldsymbol{\rho}_{-i}^{\mathrm{T}} \boldsymbol{\rho}_{-i})\mathbf{S}$, where $\mathbf{S}$ is the diagonal matrix with diagonal $\sqrt{2} |\boldsymbol{\rho}_{-i}|^{\operatorname{sgn}(\boldsymbol{\rho}_{-i})}$.

8.7. Multivariate Morgenstern distribution

The bivariate Morgenstern distribution is stated in Joe (1997), page 149. A multivariate extension of this distribution is given by

$$\Pr(\mathbf{Y} < \mathbf{y}) = \prod_{j=1}^{d} \exp\{-\exp(-y_j)\} \left(1 + \alpha \prod_{j=1}^{d} [1 - \exp\{-\exp(-y_j)\}]\right),$$

for $-1 \leqslant \alpha \leqslant 1$. Independence is given by $\alpha = 0$. Negative and positive dependence are respectively given by $\alpha < 0$ and $\alpha > 0$. Perfect positive or negative dependence is not attainable under this model. For this distribution, $\eta_{ij} = \frac{1}{2}$. Taking $\mathbf{Z}_{|i} = \mathbf{Y}_{-i}$ gives

$$G_{|i}(\mathbf{z}_{|i}) = \prod_{j \neq i} \exp\{-\exp(-z_{j|i})\} \left(1 - \alpha \prod_{j \neq i} [1 - \exp\{-\exp(-z_{j|i})\}]\right),$$

so $\mathbf{Z}_{|i}$ is distributed as a $(d-1)$-dimensional Morgenstern random variable with the sign of parameter α reversed. For positively and negatively dependent $\mathbf{Y}$, the $\mathbf{Z}_{|i}$ are respectively negatively and positively dependent. For $d \geqslant 3$ the marginal distributions of the $\mathbf{Z}_{|i}$ are Gumbel and all margins of dimension less than $d-1$ are mutually independent. In contrast, for $d = 2$, $G_{j|i}$ is a mixture of Gumbel distributions.

Acknowledgements

We are grateful to the Engineering and Physical Sciences Research Council for funding, and Stuart Coles, Anthony Ledford, Gareth Roberts and the referees for very helpful contributions. The air pollution data set originates from the National Air Quality Archive `www.airquality.co.uk`. This and the measurement programmes are funded by the Department for the Environment, Food and Rural Affairs, the Scottish Executive, the Welsh Assembly Government and the Department of the Environment for Northern Ireland.

Appendix A: Proof of theorem 1

Non-degeneracy of each marginal distribution of the limiting conditional distribution (3.1) requires that

$$F_{j|i}\{a_{j|i}(y_i) + b_{j|i}(y_i) z_{j|i} | y_i\} \to G_{j|i}(z_{j|i}) \qquad \text{as } y_i \to \infty, \tag{A.1}$$

where $G_{j|i}$ is the jth marginal distribution of $G_{|i}$. Putting $z_{j|i}=0$ in equation (A.1) gives the required condition for $a_{j|i}(y_i)$. The limit relationship (A.1) holds for all $z_{j|i}$, because **Y** has an absolutely continuous density, so the limit relationship continues to hold when differentiated with respect to $z_{j|i}$. Dividing the resulting limit relationship by $1-F_{j|i}$ gives

$$b_{j|i}(y_i)\frac{f_{j|i}\{a_{j|i}(y_i)+b_{j|i}(y_i)z_{j|i}|y_i\}}{1-F_{j|i}\{a_{j|i}(y_i)+b_{j|i}(y_i)z_{j|i}|y_i\}}\rightarrow\frac{g_{j|i}(z_{j|i})}{1-G_{j|i}(z_{j|i})},\tag{A.2}$$

where

$$g_{j|i}(z_{j|i})=\frac{\partial}{\partial z_{j|i}}G_{j|i}(z_{j|i}).$$

Putting $z_{j|i}=0$ in equation (A.2) we see that up to proportionality

$$b_{j|i}(y_i)=h_{j|i}\{a_{j|i}(y_i)|y_i\}^{-1},$$

which gives the required result up to type.

References

Arnold, B., Castillo, E. and Sarabia, J. M. (1999) *Conditional Specification of Statistical Models*. New York: Springer.
Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.
Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.
Bortot, P., Coles, S. G. and Tawn, J. A. (2000) The multivariate Gaussian tail model: an application to oceanographic data. *Appl. Statist.*, **49**, 31–49.
Bortot, P. and Tawn, J. A. (1998) Models for the extremes of Markov chains. *Biometrika*, **85**, 851–867.
Brimblecombe, P. (2001) Urban air pollution. In *Air Pollution Reviews*, vol. 1, *The Urban Atmosphere and Its Effects* (eds P. Brimblecombe and R. L. Maynard), pp. 1–19. London: Imperial College Press.
Capéraà, P., Fougères, A.-L. and Genest, C. (1997) Estimation of bivariate extreme value copulas. *Biometrika*, **84**, 567–577.
Coles, S. G., Heffernan, J. E. and Tawn, J. A. (1999) Dependence measures for extreme value analyses. *Extremes*, **2**, 339–365.
Coles, S. G. and Tawn, J. A. (1991) Modelling extreme multivariate events. *J. R. Statist. Soc.* B, **53**, 377–392.
Coles, S. G. and Tawn, J. A. (1994) Statistical methods for multivariate extremes: an application to structural design (with discussion). *Appl. Statist.*, **43**, 1–48.
Colls, J. (2002) *Air Pollution*, 2nd edn. London: Spon.
Crowder, M. J. (1989) A multivariate distribution with Weibull connections. *J. R. Statist. Soc.* B, **51**, 93–108.
Crowder, M. J. (2001) On repeated measures analysis with misspecified covariance structure. *J. R. Statist. Soc.* B, **63**, 55–62.
Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc.* B, **52**, 393–442.
Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989) A moment estimator for the index of an extreme value distribution. *Ann. Statist.*, **17**, 1833–1855.
Department of the Environment, Transport and the Regions (2000) *The Air Quality Strategy for England, Scotland, Wales and Northern Ireland*. London: Department of the Environment, Transport and the Regions.
Draisma, G. (2000) Parametric and semi-parametric methods in extreme value theory. *Thesis*. Erasmus University Rotterdam, Rotterdam.
Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004) Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, **10**, no. 2, in the press.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. London: Springer.
Gumbel, E. J. (1960) Bivariate exponential distributions. *J. Am. Statist. Ass.*, **55**, 698–707.
de Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Z. Wahrsch. Theor.*, **40**, 317–337.
de Haan, L. and Resnick, S. I. (1993) Estimating the limit distribution of multivariate extremes. *Stochast. Models*, **9**, 275–309.
de Haan, L. and de Ronde, J. (1998) Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7–45.
Hall, P. and Tajvidi, N. (2000) Distribution dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, **6**, 835–844.

Hand, D. and Crowder, M. J. (1996) *Practical Longitudinal Data Analysis*. London: Chapman and Hall.
Heffernan, J. E. (2000) A directory of coefficients of tail dependence. *Extremes*, **3**, 279–290.
Housley, D. and Richards, R. (2001) Mechanism of toxicity of gaseous air pollutants. In *Air Pollution Reviews*, vol. 1, *The Urban Atmosphere and Its Effects* (eds P. Brimblecombe and R. L. Maynard), pp. 247–272. London: Imperial College Press.
Joe, H. (1994) Multivariate extreme-value distributions with applications to environmental data. *Can. J. Statist.*, **22**, 47–64.
Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.
Joe, H., Smith, R. L. and Weissman, I. (1992) Bivariate threshold methods for extremes. *J. R. Statist. Soc.* B, **54**, 171–183.
Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Series*. New York: Springer.
Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.
Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *J. R. Statist. Soc.* B, **59**, 475–499.
Ledford, A. W. and Tawn, J. A. (1998) Concomitant tail behaviour for extremes. *Adv. Appl. Probab.*, **30**, 197–215.
Ledford, A. W. and Tawn, J. A. (2003) Diagnostics for dependence within time series extremes. *J. R. Statist. Soc.* B, **65**, 521–543.
Longin, F. M. (2000) From value at risk to stress testing: the extreme value approach. *J. Bnkng Finan.*, **24**, 1097–1130.
Maulik, K., Resnick, S. I. and Rootzén, H. (2002) Asymptotic independence and a network traffic model. *J. Appl. Probab.*, **39**, 671–699.
Peng, L. (1999) Estimation of the coefficient of asymptotic independence in bivariate extremes. *Statist. Probab. Lett.*, **43**, 399–409.
Photochemical Oxidants Review Group (1997) Ozone in the United Kingdom. *Fourth Report of the Photochemical Oxidants Review Group*. Photochemical Oxidants Review Group.
Pickands, J. (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.
Pickands, J. (1981) Multivariate extreme value distributions. *Proc. 43rd Sess. Int. Statist. Inst.*, 859–878.
Poon, S.-H., Rockinger, M. and Tawn, J. A. (2004) Extreme-value dependence in financial markets: diagnostics, models and financial implications. *Rev. Finan. Stud.*, **17**, 581–610.
Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.
Resnick, S. I. and Rootzén, H. (2000) Self-similar communication models and very heavy tails. *Ann. Appl. Probab.*, **10**, 753–778.
Schlather, M. and Tawn, J. A. (2003) A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika*, **90**, 139–156.
Shi, D., Smith, R. L. and Coles, S. G. (1992) Joint versus marginal estimation for bivariate extremes. *Technical Report 2074*. Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill.
Sinha, A. K. (1997) Estimating failure probability when failure is rare: multidimensional case. *Thesis*. Erasmus University Rotterdam, Rotterdam.
Smith, R. L. (1989) Extreme value analysis of environmental time series: an application to trend detection in ground level ozone (with discussion). *Statist. Sci.*, **4**, 367–393.
Stărică, C. (2000) Multivariate extremes for models with constant conditional correlations. *J. Emp. Finan.*, **6**, 513–553.
Tawn, J. A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

Discussion on the paper by Heffernan and Tawn

Richard L. Smith (*University of North Carolina, Chapel Hill*)
The original formulation of multivariate extreme value theory was

$$\lim_{n\to\infty}\left(\Pr\left[\frac{\max\{X_{11},\ldots,X_{n1}\}-b_{n1}}{a_{n1}}\leqslant x_1,\ldots,\frac{\max\{X_{1d},\ldots,X_{nd}\}-b_{nd}}{a_{nd}}\leqslant x_d\right]\right)=G(x_1,\ldots,x_d), \tag{1}$$

where $(X_{i1},\ldots,X_{id}), i=1,2,\ldots,$ are independent identically distributed d-dimensional random vectors, $a_{n1},\ldots,a_{nd}$ and $b_{n1},\ldots,b_{nd}$ are normalizing constants and G is a non-degenerate d-dimensional distribution function.

I do not know who first proposed this definition, but it emerged in several papers in the 1950s, during what might be called the golden age of asymptotic distributions in probability. As such, it seems to have been motivated more by considerations of mathematical elegance than by messy practical problems such as controlling air pollution. In retrospect, it seems surprising that this original formulation survived so long.