Peaks over thresholds modelling with multivariate generalized Pareto distributions

Anna Kiriliouk  Johan Segers
Université catholique de Louvain
Institut de Statistique, Biostatistique et Sciences Actuarielles
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
E-mail: anna.kiriliouk@uclouvain.be, johan.segers@uclouvain.be

Holger Rootzén
Chalmers University of Technology
Department of Mathematical Sciences
SE-412 96 Gothenburg, Sweden.
E-mail: hrootzen@chalmers.se

Jennifer L. Wadsworth
Lancaster University
Department of Mathematics and Statistics
Fylde College LA1 4YF, Lancaster, England.
E-mail: j.wadsworth@lancaster.ac.uk

Abstract

The multivariate generalized Pareto distribution arises as the limit of a normalized vector conditioned upon at least one component of that vector being extreme. Statistical modelling using multivariate generalized Pareto distributions constitutes the multivariate analogue of univariate peaks over thresholds modelling. We exhibit a construction device which allows us to develop a variety of new and existing parametric tail dependence models. A censored likelihood procedure is proposed to make inference on these models, together with a threshold selection procedure and several goodness-of-fit diagnostics. The models are fitted to returns of four UK-based banks and to rainfall data in the context of landslide risk estimation.

Keywords: financial risk; landslides; multivariate extremes; tail dependence.


1 Introduction

Peaks over threshold modelling of univariate time series has been common practice since the seminal paper of Davison and Smith (1990), who advocated the use of the asymptotically motivated generalized Pareto (GP) distribution (Balkema and de Haan, 1974; Pickands, 1975) as a model for exceedances over high thresholds. The multivariate generalized Pareto distribution was introduced in Tajvidi (1996), Beirlant et al. (2004), and Rootzén and Tajvidi (2006). There is a growing body of probabilistic literature devoted to multivariate GP distributions (Rootzén and Tajvidi (2006); Falk and Guillou (2008); Ferreira and de Haan (2014); Rootzén et al. (2016)). To our knowledge, however, there are only a few papers that use these as a statistical model (Thibaud and Opitz, 2015; Huser et al., 2016).

In this paper we advance and exhibit methods needed for practical use of multivariate GP statistical models. We use these to derive risk estimates for returns of four UK-based banks (Section 6), and show that these can be more useful for portfolio risk handling than currently available one-dimensional estimates. Environmental risks often involve physical constraints not taken into account by available methods. We estimate landslide risks using new methods which handle such constraints, thereby providing more realistic estimates (Section 7).

The results build on development of many new parametric multivariate GP models (Sections 3 and 4); a computationally tractable strategy for model selection, model fitting and validation (Section 5.2); and new threshold selection and goodness-of-fit diagnostics (Section 5.3). We throughout use censored likelihood-based estimation (Section 5.1). An important feature is that we estimate marginal and dependence parameters simultaneously, so that confidence intervals include the full estimation uncertainty. We also briefly discuss the probabilistic background (Section 2).

The “point process method” (Coles and Tawn, 1991) provides an alternative approach. However, the multivariate GP distribution has practical and conceptual advantages, in so much as it is a proper multivariate distribution. It also separates modelling of the times of threshold exceedances and the distribution of the threshold excesses in a useful way.

Multivariate GP distributions can place mass on lower-dimensional subspaces. Such cases constitute a challenging area for future research. However, in the remainder of the
paper we consider the simplified situation where there is no mass on any lower-dimensional subspace. This assumption implies so-called asymptotic dependence of all components, and we term it full asymptotic dependence.

2 Background on GP distributions

Let \( Y \) be a random vector in \( \mathbb{R}^d \) with distribution function \( F \). A common assumption on \( Y \) is that it is in the so-called max-domain of attraction of a multivariate max-stable distribution, \( G \). This means that if \( Y_1, \ldots, Y_n \) are independent and identically distributed copies of \( Y \), then one can find sequences \( a_n \in (0, \infty)^d \) and \( b_n \in \mathbb{R}^d \) such that

\[
\Pr[\left\{ \max_{1 \leq i \leq n} Y_i / a_n \leq x \right\}] \rightarrow G(x),
\]

with \( G \) having non-degenerate margins. In (2.1) and throughout, operations involving vectors are to be interpreted componentwise. If convergence (2.1) holds, then

\[
\max \left\{ \frac{Y - b_n}{a_n}, \eta \right\} \mid Y \not\leq b_n \overset{d}{\to} X,
\]

where \( X \) follows a multivariate GP distribution (Rootzén et al., 2016), and where \( \eta \) is the vector of lower endpoints of the GP distribution. We let \( H \) denote the distribution function of \( X \), and \( H_1, \ldots, H_d \) its marginal distributions. Typically the margins \( H_j \) are not univariate GP, due to the difference between the conditioning events \( \{Y_j > b_{n,j}\} \) and \( \{Y \not\leq b_n\} \) in the one-dimensional and \( d \)-dimensional limits. However, the marginal distributions conditioned to be positive are GP distributions. That is, writing \( a_+ = \max(a, 0) \), we have

\[
\overline{H}_j^+(x) := \Pr[X_j > x \mid X_j > 0] = (1 + \gamma_j x / \sigma_j)^{-1/\gamma_j},
\]

where \( \sigma_j \) and \( \gamma_j \) are marginal scale and shape parameters, and \( \overline{H}_j^+ \) has lower endpoints \( \eta_j = -\sigma_j / \gamma_j \) if \( \gamma_j > 0 \) and \( \eta_j = -\infty \) otherwise. The link between \( H \) and \( G \) is \( H(x) = (\log G(\min(x, 0)) - \log G(x)) / \{\log G(0)\} \), and we say that \( H \) and \( G \) are associated.

Following common practice in the statistical modelling of extremes, \( H \) may be used as a model for data which arise as multivariate excesses of high thresholds. Hence, if \( u \in \mathbb{R}^d \) is a threshold that is “sufficiently high” in each margin, then we approximate \( Y - u \mid Y \not\leq u \) by a member \( X \) of the class of multivariate GP distributions, with \( \sigma, \gamma \), the marginal
exceedance probabilities $\mathbb{P}(Y_j > u_j)$, and the dependence structure to be estimated. In practice the truncation by the vector $\eta$ in (2.2) is only relevant when dealing with mass on lower-dimensional subspaces, and is outside the scope of the present paper. The following are useful properties of the GP distributions; for further details and proofs we refer to Rootzén et al. (2016).

**Threshold stability:** GP distributions are *threshold stable*, meaning that if $X \sim H$ follows a GP distribution and if $w \geq 0$, with $H(w) < 1$ and $\sigma + \gamma w > 0$, then

$$X - w \mid X \not\leq w$$

is GP with parameters $\sigma + \gamma w$ and $\gamma$.

Hence if the thresholds are increased, then the distribution of conditional excesses is still GP, with a new set of scale parameters, but retaining the same vector of shape parameters.

A special role is played by the levels $w = w_t := \sigma(t^n - 1)/\gamma$: these have the stability property that for any set $A \subset \{x \in \mathbb{R}^d : x \not\leq 0\}$ it holds that, for $t \geq 1$,

$$\mathbb{P}[X \in w_t + t^n A] = \mathbb{P}[X \in A]/t,$$

(2.4)

where $w_t + t^n A = \{w_t + t^n x : x \in A\}$. This follows from equation (3.1) along with the representation of $X_0$ to be given in equation (3.2). The $j$-th component of $w_t$, $\sigma_j(t^n - 1)/\gamma_j$, is the $1 - 1/t$ quantile of $H_j^+$. Equation (2.4) provides one possible tool for checking if a multivariate GP is appropriate; see Section 5.3.

**Lower dimensional conditional margins:** Lower dimensional margins of GP distributions are typically not GP. Let $J \subset \{1, \ldots, d\}$, with $X_J = (x_j : j \in J)$, and similarly for other vectors. Then $X_J \mid X_J \not\leq 0_J$ does follow a GP distribution. Combined with the threshold stability property above, we also have that if $w_J \in \mathbb{R}^{|J|}$ is such that $w_J \geq 0$, $H_J(w_J) < 1$ and $\sigma_J + \gamma_J w_J > 0$ then $X_J - w_J \mid X_J \not\leq w_J$ follows a GP distribution.

**Sum-stability under shape constraints:** If $X$ follows a multivariate GP distribution, with scale parameter $\sigma$ and shape parameter $\gamma = \gamma 1$, then for weights $a_j > 0$ such that $\sum_{j=1}^d a_j X_j > 0$ with positive probability, we have

$$\sum_{j=1}^d a_j X_j \mid \sum_{j=1}^d a_j X_j > 0 \sim \text{GP}(\sum_{j=1}^d a_j \sigma_j, \gamma).$$

(2.5)

Thus weighted sums of components of a multivariate GP distribution with equal shape parameters, conditioned to be positive, follow a univariate GP distribution with the same
shape parameter and with scale parameter equal to the weighted sum of the marginal scale parameters. The dependence structure does not affect this result, but it does affect the probability of the conditioning event that the sum of components is positive.

3 Model construction

The parametric GP densities derived in Sections 4 below are obtained from three general forms $h_T, h_U,$ and $h_R$ of GP densities. For the first two one first constructs a standard form density for a variable $X_0$ with $\sigma = 1, \gamma = 0$, and then obtains a density on the observed scale through the standard transformation

$$X = \frac{\sigma e^{x_0\gamma} - 1}{\gamma},$$

(3.1)

with the distribution $X$ supported on $\{x \in \mathbb{R}^d : x \not< 0\}$. For $\gamma_j = 0$, the corresponding component of the right-hand side of equation (3.1) is simply $\sigma_j x_{0,j}$. The first density, $h_T$, comes from a convenient reformulation of the the Ferreira and de Haan (2014) spectral representation. More details, alternative constructions, and intuition for the three forms are given in Rootzén et al. (2016, 2017).

**Standard form densities:** We first focus on how to construct suitable densities for the random vector $X_0$, which, through equation (3.1), lead to densities for the multivariate GP distribution with marginal parameters $\sigma$ and $\gamma$. Let $E$ be a unit exponential random variable and let $T$ be a $d$-dimensional random vector, independent of $E$. Define $\max(T) = \max_{1 \leq j \leq d} T_j$. Then the random vector

$$X_0 = E + T - \max(T)$$

(3.2)

is a GP vector with support included in the set $\{x \in \mathbb{R}^d : x \not< 0\}$ and with $\sigma = 1$ and $\gamma = 0$ (interpreted as the limit for $\gamma_j \to 0$ for all $j$). Moreover, *every* such GP vector can be expressed in this way (Ferreira and de Haan, 2014; Rootzén et al., 2016). The probability of the $j$-th component being positive is $P[X_{0,j} > 0] = E[e^{T_j - \max(T)}]$, which, in terms of the original data vector $Y$, corresponds to the probability $P[Y_j > u_j \mid Y \not< u]$, i.e., the probability that the $j$-th component exceeds its corresponding threshold given that one of the $d$ components does.
Suppose $T$ has a density $f_T$ on $(-\infty, \infty)^d$. By Theorem 5.1 of Rootzén et al. (2016), the density of $X_0$ is given by

$$h_T(x; 1, 0) = \frac{1}{e^{\max(x)}} \int_0^\infty f_T(x + \log t) t^{-1} \, dt. \tag{3.3}$$

One way to construct models therefore is to assume distributions for $T$ which provide flexible forms for $h_T$, and for which ideally the integral in (3.3) can be evaluated analytically.

One further construction of GP random vectors is given in Rootzén et al. (2016). If $U$ is a $d$-dimensional random vector with density $f_U$ and such that $E[e^{U_j}] < \infty$ for all $j = 1, \ldots, d$, then the following function also defines the density of a GP distribution:

$$h_U(x; 1, 0) = \frac{1}{E[e^{\max(U)}} \int_0^\infty f_U(x + \log t) \, dt. \tag{3.4}$$

The marginal exceedance probabilities are now $P[X_{0,j} > 0] = E[e^{U_j}] / E[e^{\max(U)}]$. Formulas (3.3) and (3.4) can be obtained from one another via a change of measure.

Where $f_T$ and $f_U$ take the same form, then the similarity in integrals between (3.3) and (3.4) means that if one can be evaluated, then typically so can the other; several instances of this are given in the models presented in Section 4. What is sometimes more challenging is calculation of the normalization constant $E[e^{\max(U)}] = \int_0^\infty P[\max(U) > \log t] \, dt$ in (3.4). Nonetheless, the model in (3.4) has the particular advantage over that of (3.3) that it behaves better across various dimensions: if the density of the GP vector $X$ is $h_U$ and if $J \subset \{1, \ldots, d\}$, then the density of the GP subvector $X_J | X_J \not\geq 0_J$ is simply $h_{U_J}$.

This property is advantageous when moving to the spatial setting, since the model retains the same form when numbers of sites change, which is useful for spatial prediction.

**Densities after transformation to the observed scale.** The densities above are in the standardized form $\sigma = 1$, $\gamma = 0$. Using (3.1), we obtain general densities which is are approximations to the conditional density of $Y - u$ given that $Y \not\geq u$, for the original data $Y$:

$$h(x; \sigma, \gamma) = h\left(\frac{1}{\gamma} \log(1 + \gamma x/\sigma); 1, 0\right) \prod_{j=1}^d \frac{1}{\sigma_j + \gamma_j x_j}. \tag{3.5}$$

In (3.5), $h$ may be either $h_T$ or $h_U$. Observe that it is always possible to transform this density to be suitable for large values of $Y$ itself, i.e., to a density for the conditional distribution of $Y$ given that $Y \not\geq u$ by replacing $x$ by $x - u$ in (3.5).
Densities constructed on observed scale. The models (3.5) are built on a standardized scale, and then transformed to the observed, or “real” scale. Alternatively, models can be constructed directly on the real scale, which gives the possibility of respecting structures, say additive structures, in a way which is not possible with the other two models; this approach will be used to model ordered data in Section 7. One way of presenting this is to define the random vector \( R \) in terms of \( U \) in (3.4) through the componentwise transformation

\[
R_j = \begin{cases} 
(\sigma_j/\gamma_j) \exp(\gamma_j U_j), & \gamma_j \neq 0, \\
\sigma_j U_j, & \gamma_j = 0,
\end{cases}
\]

and develop suitable models for \( R \). This gives the GP density

\[
h_R(x; \sigma, \gamma) = \frac{1}{E[\exp(\max(U))]} \int_{0}^{\infty} t^{\gamma_j} f_R \left( \left( g(t; x_j, \sigma_j, \gamma_j) \right)_{j=1}^{d} \right) dt,
\]

where \( f_R \) denotes the density of \( R \) and where

\[
g(t; x_j, \sigma_j, \gamma_j) = \begin{cases} 
\frac{t^{\gamma_j}}{(x_j + \sigma_j/\gamma_j)}, & \gamma_j \neq 0, \\
x_j + \sigma_j \log t, & \gamma_j = 0.
\end{cases}
\]

The \( d \) components of \( U \) are found by inverting equation (3.6). For \( \sigma = 1 \) and \( \gamma = 0 \), the densities (3.4) and (3.7) are the same.

4 Parametric models

Here we provide details of certain probability distributions for \( T, U \) and \( R \) that generate tractable multivariate GP distributions. We refer to these as “generators”. Several articles have previously used random vectors to generate dependence structures for extremes, e.g. Segers (2012), Thibaud and Opitz (2015) and Aulbach et al. (2015). The literature on max-stable modelling for spatial extremes also relies heavily on this device (de Haan, 1984; Schlather, 2002; Davison et al., 2012).

To control bias when using a multivariate GP distribution, we often need to use censored likelihood (Section 5) and thus not just to be able to calculate densities, but also integrals of those densities. These considerations guide our choice of models presented below. For
each model we give the uncensored densities in the subsequent subsections, whilst their censored versions are given in the supplementary material.

In Sections 4.1 and 4.2 we consider particular instances of densities \( f_T \) and \( f_U \) to evaluate the corresponding densities \( h_T \) and \( h_U \) in (3.3) and (3.4). As noted in Section 3, even if \( f_T = f_U \), the GP densities \( h_T \) and \( h_U \) are still different in general. Thus we will focus on the density of a random vector \( V \), denoted \( f_V \), and create two GP models per \( f_V \) by setting \( f_T = f_V \) and then \( f_U = f_V \), in the latter case with the restriction \( E[e^{U_j}] < \infty \). The support for each GP density given in Sections 4.1 and 4.2 is \( \{ x \in \mathbb{R}^d : x \not\leq 0 \} \), and for brevity, we omit the indicator \( 1\{ \max(x) > 0 \} \). In Section 4.3 we exhibit a construction of \( h_R \) in (3.7) with support depending on \( \gamma \) and \( \sigma \).

In all models, identifiability issues occur if \( T \) or \( U \) have unconstrained location parameters \( \beta \), or if \( R \) has unconstrained scale parameters \( \lambda \). Indeed, replacing \( \beta \) or \( \lambda \) by \( \beta + k \) or \( c\lambda \), respectively, with \( k \in \mathbb{R} \) and \( c > 0 \), lead to the same GP distribution (Rootzén et al., 2016, Proposition 1). A single constraint, such as fixing the first parameter in the vector, is sufficient to restore identifiability.

### 4.1 Generators with independent components

Let \( V \in \mathbb{R}^d \) be a random vector with independent components and density \( f_V(v) = \prod_{j=1}^d f_j(v_j) \), where \( f_j \) are densities of real-valued random variables. The dependence structure of the associated GP distributions is determined by the relative heaviness of the tails of the \( f_j \): roughly speaking, if one component has a high probability of being “large” compared to the others, then dependence is weaker than if all components have a high probability of taking similar values. Throughout, \( x \in \mathbb{R}^d \) is such that \( \max(x) > 0 \).

**Generators with independent Gumbel components:** Let

\[
f_j(v_j) = \alpha_j \exp\{-\alpha_j(v_j - \beta_j)\} \exp\{-\exp\{-\alpha_j(v_j - \beta_j)\}\}, \quad \alpha_j > 0, \beta_j \in \mathbb{R}.
\]

**Case** \( f_T = f_V \). Density (3.3) is

\[
h_T(x; 1, 0) = e^{-\max(x)} \int_0^\infty t^{-1} \prod_{j=1}^d \alpha_j (te_j - \beta_j)^{-\alpha_j} e^{-(te_j - \beta_j)^{-\alpha_j}} dt. \tag{4.1}
\]
If $\alpha_1 = \ldots = \alpha_d = \alpha$ then the integral can be explicitly evaluated:

$$h_T(x; 1, 0) = e^{-\max(x)} \alpha^{d-1} \frac{\Gamma(d) \prod_{j=1}^d e^{-\alpha(x_j - \beta_j)} \left(\sum_{j=1}^d e^{-\alpha(x_j - \beta_j)}\right)^d}{\prod_{j=1}^d e^{-\alpha(x_j - \beta_j)}}.$$

**Case** $f_U = f_V$. The marginal expectation of the exponentiated variable is $E[e^{U_j}] = e^{\beta_j} \Gamma(1 - 1/\alpha_j)$ for $\alpha_j > 1$ and $E[e^{U_j}] = \infty$ for $\alpha_j \leq 1$. For $\min_{1 \leq j \leq d} \alpha_j > 1$, density (3.4) is

$$h_U(x; 1, 0) = \frac{\int_0^\infty \prod_{j=1}^d \alpha_j \left(t e^{x_j - \beta_j}\right)^{-\alpha_j} e^{-\left(t e^{x_j - \beta_j}\right)^{-\alpha_j}} dt}{\int_0^\infty \left(1 - \prod_{j=1}^d e^{-\left(t e^{x_j - \beta_j}\right)^{-\alpha_j}}\right) dt}. \quad (4.2)$$

If $\alpha_1 = \ldots = \alpha_d = \alpha$ then this simplifies to:

$$h_U(x; 1, 0) = \frac{\alpha^{d-1} \Gamma(d - 1/\alpha) \prod_{j=1}^d e^{-\alpha(x_j - \beta_j)} \left(\sum_{j=1}^d e^{-\alpha(x_j - \beta_j)}\right)^{d-1/\alpha}}{\Gamma(1 - 1/\alpha) \left(\sum_{j=1}^d e^{\beta_j} \alpha_j\right)^{1/\alpha}}.$$

Observe that if in addition to $\alpha_1 = \ldots = \alpha_d = \alpha$, also $\beta_1 = \ldots = \beta_d = 0$, then this is the multivariate GP distribution associated to the well-known *logistic* max-stable distribution.

**Generators with independent reverse Gumbel components:** Let

$$f_j(v_j) = \alpha_j \exp\{\alpha_j(v_j - \beta_j)\} \exp\{-\exp\{\alpha_j(v_j - \beta_j)\}\}, \quad \alpha_j > 0, \beta_j \in \mathbb{R}.$$

As the Gumbel case leads to the multivariate GP distribution associated to the logistic max-stable distribution, the reverse Gumbel leads to the multivariate GP distribution associated to the *negative logistic* max-stable distribution\(^1\). Calculations are very similar to the Gumbel case, and hence omitted.

**Generators with independent reverse exponential components:** Let

$$f_j(v_j) = \exp\{(v_j + \beta_j)/\alpha_j\}/\alpha_j, \quad v_j \in (-\infty, -\beta_j), \alpha_j > 0, \beta_j \in \mathbb{R}.$$

**Case** $f_T = f_V$. Density (3.3) is

$$h_T(x; 1, 0) = e^{-\max(x)} \int_0^{e^{-\max(x+\beta)}} t^{-1} \prod_{j=1}^d \frac{1}{\alpha_j} \left(t e^{x_j + \beta_j}\right)^{1/\alpha_j} dt.$$

$$= \frac{e^{-\max(x) - \max(x+\beta)} \prod_{j=1}^d 1/\alpha_j \left(e^{x_j + \beta_j}\right)^{1/\alpha_j}}{\prod_{j=1}^d 1/\alpha_j \left(e^{x_j + \beta_j}\right)^{1/\alpha_j}, \quad (4.3)}$$

\(^1\)The authors are grateful to Clément Dombry for having pointed out this connection.
Case $f_U = f_V$. The expectation of the exponentiated variable is $\mathbb{E}[e^{U_j}] = 1 / \{e^{\beta_j}(\alpha_j + 1)\}$, which is finite for all permitted parameter values. Density (3.4) is

$$h_U(x; 1, 0) = \frac{1}{\mathbb{E}[e^{\max(U)}]} \int_0^{-\max(x+\beta)} \prod_{j=1}^d \frac{1}{\alpha_j} (te^{x_j+\beta_j})^{1/\alpha_j} dt$$

$$= \frac{(e^{-\max(x+\beta)}\sum_{j=1}^d 1/\alpha_j+1)}{\mathbb{E}[e^{\max(U)}]} \prod_{j=1}^d \frac{1}{\alpha_j} (e^{x_j+\beta_j})^{1/\alpha_j}. \quad (4.4)$$

The normalization constant may be evaluated as

$$\mathbb{E}[e^{\max(U)}] = \int_0^\infty \left(1 - \prod_{j=1}^d \min(e^{\beta_j t}, 1)^{1/\alpha_j}\right) dt$$

$$= e^{-\beta_{(d)}} - \prod_{j=1}^d e^{\beta_j/\alpha_j} \sum_{j=1}^d 1/\alpha_j + 1$$

$$+ \sum_{i=1}^{d-1} \prod_{j=i+1}^d e^{\beta_{(i+j)}/\alpha_{[j]}} \left(e^{-\beta_{(i+1)}(\sum_{j=i+1}^d 1/\alpha_{[j]}+1)} - e^{-\beta_{(i)}(\sum_{j=i+1}^d 1/\alpha_{[j]}+1)}\right),$$

where $\beta_{(1)} > \beta_{(2)} > \cdots > \beta_{(d)}$ and where $\alpha_{[j]}$ is the component of $\alpha$ with the same index as $\beta_{(j)}$ (thus the $\alpha_{[j]}$s are not in general ordered). As far as we are aware the associated max-stable model is not well known.

Generators with independent log-gamma components: if $e^V_j \sim \text{Gamma}(\alpha_j, 1)$ then

$$f_j(v_j) = \exp(\alpha_j v_j) \exp\{-\exp(v_j)\}/\Gamma(\alpha_j), \quad \alpha_j > 0, \ v_j \in (-\infty, \infty).$$

Case $f_T = f_V$. Density (3.3) is

$$h_T(x; 1, 0) = e^{-\max(x)} \prod_{j=1}^d \left(\frac{e^{\alpha_j x_j}}{\Gamma(\alpha_j)}\right) \int_0^\infty t^{\sum_{j=1}^d \alpha_j-1} e^{-t^{\sum_{j=1}^d e^{x_j}}} dt$$

$$= \frac{\Gamma\left(\sum_{j=1}^d \alpha_j\right)}{\prod_{j=1}^d \Gamma(\alpha_j)} \frac{e^{\sum_{j=1}^d x_j - \max(x)}}{(\sum_{j=1}^d e^{x_j})^{\sum_{j=1}^d \alpha_j}}.$$

Case $f_U = f_V$. The marginal expectation of the exponentiated variable is $\mathbb{E}[e^{U_j}] = \alpha_j$, hence finite for all permitted parameter values. Density (3.4) is

$$h_U(x; 1, 0) = \frac{1}{\mathbb{E}[e^{\max(U)}]} \prod_{j=1}^d \left(\frac{e^{\alpha_j x_j}}{\Gamma(\alpha_j)}\right) \int_0^\infty t^{\sum_{j=1}^d \alpha_j} e^{-t^{\sum_{j=1}^d e^{x_j}}} dt$$

$$= \frac{1}{\mathbb{E}[e^{\max(U)}]} \Gamma\left(\sum_{j=1}^d \alpha_j + 1\right) \frac{e^{\sum_{j=1}^d x_j - \max(x)}}{(\sum_{j=1}^d e^{x_j})^{\sum_{j=1}^d \alpha_j+1}}.$$

10
The normalization constant is
\[
E[e^{\max(U)}] = \frac{\Gamma \left( \sum_{j=1}^{d} \alpha_j + 1 \right)}{\prod_{j=1}^{d} \Gamma(\alpha_j)} \int_{\Delta_{d-1}} \max(u_1, \ldots, u_d) \prod_{j=1}^{d} u_j^{\alpha_j-1} du_1 \cdots du_{d-1},
\]
where \(\Delta_{d-1} = \{(u_1, \ldots, u_d) \in [0, 1]^d : u_1 + \cdots + u_d = 1\}\) is the unit simplex, and the integral can be easily computed using the R package \texttt{SimplicialCubature}. This GP distribution is associated to the Dirichlet max-stable distribution (Coles and Tawn, 1991; Segers, 2012).

### 4.2 Generators with multivariate Gaussian components

Let \(f_V(v) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\{- (v - \beta)^T \Sigma^{-1} (v - \beta)/2\}\), where \(\beta \in \mathbb{R}^d\) is the mean parameter and \(\Sigma \in \mathbb{R}^{d \times d}\) is a positive-definite covariance matrix. As before, \(\max(x) > 0\). For calculations, it is simplest to make the change of variables \(s = \log t\) in (3.3) and (3.4).

**Case** \(f_T = f_V\). Density (3.3) is
\[
h_T(x; 1, 0) = e^{-\max(x)} \int_{-\infty}^{\infty} \frac{(2\pi)^{-d/2}}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \beta - s1)^T \Sigma^{-1} (x - \beta - s1) \right\} ds
\]
\[
= \frac{(2\pi)^{(1-d)/2}}{(1^T \Sigma^{-1} 1)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \beta)^T A (x - \beta) - \max(x) \right\}
\]
with
\[
A = \Sigma^{-1} - \frac{\Sigma^{-1} 1 1^T \Sigma^{-1}}{1^T \Sigma^{-1} 1},
\]
a \(d \times d\) matrix of rank \(d - 1\).

**Case** \(f_U = f_V\). As required, the expectation \(E[e^{U}] = e^{\beta_j + \Sigma_{jj}/2}\) is finite for all permitted parameter values, where \(\Sigma_{jj}\) denotes the \(j\)th diagonal element of \(\Sigma\). Density (3.4) is
\[
h_U(x; 1, 0) = \frac{1}{E[e^{\max(U)}]} \int_{-\infty}^{\infty} \frac{(2\pi)^{-d/2}}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \beta - s1)^T \Sigma^{-1} (x - \beta - s1) + s \right\} ds
\]
\[
= \frac{(2\pi)^{(1-d)/2}}{E[e^{\max(U)}] (1^T \Sigma^{-1} 1)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \beta)^T A (x - \beta) + \frac{1}{2} (x - \beta)^T \Sigma^{-1} 1 - 1 \right\}
\]
with \(A\) as in (4.6). This is the GP distribution associated to the Brown–Resnick or Hüsler–Reiss max-stable model (Kabluchko et al., 2009; Hüsler and Reiss, 1989). A variant of the density formula with \(E[e^{U}] = 1\) (equivalently \(\beta = -\text{diag}(\Sigma)/2\)) was given in Wadsworth and Tawn (2014). The normalization constant is \(\int_{0}^{\infty} \left[ 1 - \Phi_d(\log t 1 - \beta; \Sigma) \right] dt\), where \(\Phi_d(\cdot; \Sigma)\) is the zero-mean multivariate normal distribution function with covariance matrix \(\Sigma\). This normalization constant can be expressed as a sum of multivariate normal distribution functions (Huser and Davison, 2013).
4.3 Generators with structured components

We present a model for $\mathbf{R}$ based on cumulative sums of exponential random variables and whose components are ordered; for the components of the corresponding GP vector to be ordered as well, we assume that $\gamma = \gamma_1$ and $\sigma = \sigma_1$. We restrict our attention to $\gamma \in [0, \infty)$ in view of the application we have in mind: this model is used in Section 7 to model cumulative precipitation amounts which may trigger landslides.

Case $\gamma = 0$. By construction, the densities $h_{\mathbf{R}}(\cdot; \mathbf{1}, 0)$ and $h_{\mathbf{U}}(\cdot; \mathbf{1}, 0)$ coincide since $\mathbf{R} = \mathbf{U}$. Let $\mathbf{R} \in (-\infty, \infty)^d$ be the random vector whose components are defined by

$$R_j = \log \left( \sum_{i=1}^d E_i \right), \quad E_j \overset{iid}{\sim} \text{Exp}(\lambda_j), \quad j = 1, \ldots, d,$$

where the $\lambda_j$ are the mean values of the exponential distributions. Its density, $f_{\mathbf{R}}$, is

$$f_{\mathbf{R}}(\mathbf{r}) = \begin{cases} \left( \prod_{j=1}^d \lambda_j e^{r_j} \right) \exp \left\{ - \sum_{j=1}^d (\lambda_j - \lambda_{j+1})e^{r_j} \right\}, & \text{if } r_1 < \ldots < r_d, \\ 0, & \text{otherwise}, \end{cases}$$

where we set $\lambda_{d+1} = 0$. In view of (3.4), $R_1 < \ldots < R_d$ (or equivalently $U_1 < \ldots < U_d$) implies $X_{0,1} < \ldots < X_{0,d}$. The density of $\mathbf{X}_0$ is given as follows: if $x_1 < \ldots < x_d$, then

$$h_{\mathbf{R}}(\mathbf{x}; 1, 0) = \frac{1}{\mathbb{E}[e^{\mathbf{R}_d}]} \left( \prod_{j=1}^d \lambda_j e^{x_j} \right) \int_0^\infty t^d \exp \left\{ -t \left( \sum_{j=1}^d (\lambda_j - \lambda_{j+1})e^{x_j} \right) \right\} \, dt$$

$$= \frac{\mathbb{1}(x_d > 0) \, d! \, \prod_{j=1}^d \lambda_j e^{x_j}}{\left( \sum_{j=1}^d 1/\lambda_j \right) \left( \sum_{j=1}^d (\lambda_j - \lambda_{j+1})e^{x_j} \right)^{d+1}}, \quad (4.7)$$

while $h_{\mathbf{R}}(\mathbf{x}; 1, 0)$ is zero otherwise. The density $h_{\mathbf{R}}(\mathbf{x}; \sigma, 0)$ is obtained from (3.5).

Case $\gamma > 0$. Let $\mathbf{R} \in (0, \infty)^d$ be the random vector whose components are defined by

$$R_j = \sum_{i=1}^j E_i, \quad E_j \overset{iid}{\sim} \text{Exp}(\lambda_j), \quad j = 1, \ldots, d,$$

Its density, $f_{\mathbf{R}}$, is similar to the one for $\gamma = 0$. Then

$$\mathbb{E} \left[ e^{\max(U)} \right] = \mathbb{E} \left[ \max_{1 \leq j \leq d} \left( \frac{\gamma R_j}{\sigma} \right)^{1/\gamma} \right] = \left( \frac{\gamma}{\sigma} \right)^{1/\gamma} \mathbb{E} \left[ R_d^{1/\gamma} \right].$$

The distribution of $R_d$ is called generalized Erlang if $\lambda_i \neq \lambda_j$ for all $i \neq j$ (Neuts, 1974), and, letting $f_{R_d}$ denote its density we get

$$\mathbb{E} \left[ R_d^{1/\gamma} \right] = \int_0^\infty r^{1/\gamma} f_{R_d}(r) \, dr = \Gamma \left( \frac{1}{\gamma} + 1 \right) \sum_{i=1}^d \lambda_i^{-1/\gamma} \left( \prod_{j=1, j \neq i}^d \frac{\lambda_j}{\lambda_j - \lambda_i} \right).$$
If \( \lambda_1 = \ldots = \lambda_d \), then \( R_d \) follows an Erlang distribution. By (3.7), the density of \( X \) becomes, for \( x_d > \ldots > x_1 > -\sigma/\gamma \) and \( x_d > 0 \),

\[
h_{R}(x; \sigma, \gamma) = \left( \prod_{j=1}^{d} \lambda_j \right) \int_{0}^{\infty} t^{d\gamma} \exp \left\{ -t^\gamma \sum_{j=1}^{d} (\lambda_j - \lambda_{j+1})(x_j + \sigma/\gamma) \right\} \, dt
\]

\[
= \frac{\left( \prod_{j=1}^{d} \lambda_j \right) \left( \frac{\sigma}{\gamma} \right)^{-1/\gamma} \Gamma \left( d + \frac{1}{\gamma} \right) / \Gamma \left( \frac{1}{\gamma} \right)}{\left( \sum_{j=1}^{d} (\lambda_j - \lambda_{j+1})x_j + (\sigma/\gamma)\lambda_1 \right)^{d+1/\gamma} \sum_{i=1}^{d} \lambda_i^{-1/\gamma} \left( \prod_{j=1, j \neq i}^{d} \frac{\lambda_j}{\lambda_j - \lambda_i} \right)}.
\]

5 Likelihood-based inference

Working within a likelihood-based framework for inference allows many benefits. Firstly, comparison of nested models can be done using likelihood ratio tests. This is important as the number of parameters can quickly grow large if margins and dependence are fitted simultaneously, allowing us to test for simplifications in a principled manner. Secondly, incorporation of covariate effects is straightforward in principle. For univariate peaks over thresholds, such ideas were introduced by Davison and Smith (1990), but nonstationarity in dependence structure estimation has received comparatively little attention. Thirdly, such likelihoods could also be exploited for a Bayesian approach to inference if desired.

5.1 Censored likelihood

The density (3.5) is the basic ingredient in a likelihood, however, we will use (3.5) as a contribution only when all components of the observed translated vector \( Y - u \) are “large”, in the sense of exceeding a threshold \( v \), with \( v \leq 0 \). The reasoning for this is twofold:

1. For \( \gamma_j > 0 \), the lower endpoint of the multivariate GP distribution is \( -\sigma_j/\gamma_j \). Censored likelihood avoids that small values of a component affect the fit of too strongly.

2. Without censoring, bias in the estimation of parameters controlling the dependence can be larger than for censored estimation, see Huser et al. (2016).

The use of censored likelihood for inference on extreme value models was first used by Smith et al. (1997) and Ledford and Tawn (1997), and is now a standard approach to enable more robust inference. Let \( C \subset D = \{1, \ldots, d\} \) contain the indices for which components of
\( Y - u \) fall below the corresponding component of \( v \), i.e., \( Y_j - u_j \leq v_j \) for \( j \in C \), and \( Y_j - u_j > v_j \) for \( j \in D \setminus C \), with at least one such \( Y_j > u_j \). For each realization of \( Y \), we use the likelihood contribution

\[
h_C(y_{D \setminus C} - u_{D \setminus C}, v_C; \gamma, \sigma) = \int_{\mathbb{R}^{|C|}} h(y - u; \gamma, \sigma) \, dy_C,
\]

(5.1)

with \( y_C = (y_j)_{j \in C} \), which is equal to (3.5) with \( x = y - u \) if \( C \) is empty. The supplementary material contains forms of censored likelihood contributions for the models in Section 3.

For \( n \) independent observations \( y_1, \ldots, y_n \) of \( Y \mid Y \not\leq u \), the censored likelihood function to be optimized is

\[
L(\theta, \sigma, \gamma) = \prod_{i=1}^{n} h_{C_i}(y_{i,D \setminus C_i} - u_{D \setminus C_i}, v_{C_i}; \theta, \sigma, \gamma),
\]

(5.2)

where \( C_i \) denotes the censoring subset for \( y_i \), which may be empty, and \( \theta \) represents parameters related to the model that we assumed for the generator.

### 5.2 Model choice

When fitting multivariate GP distributions to data on the observed scale we have a large variety of potential models and parameterizations. For non-nested models, Akaike’s Information Criterion (AIC = \(-2 \times \) log-likelihood + 2 \times number of parameters) can be used to select a model with a good balance between parsimony and goodness-of-fit. When looking at nested models, e.g., to test for simplifications in parameterization, we can use likelihood ratio tests. Because of the many possibilities for model fitting, to reduce the computational burden we propose the following model-fitting strategy, that we will employ in Section 6.

(i) Standardize the data to common exponential margins, \( Y_E \), using the rank transformation (i.e., the probability integral transform using the empirical distribution function);

(ii) select a multivariate threshold, denoted \( u \) on the scale of the observations, and \( u_E \) on the exponential scale, using the method of Section 5.3;

(iii) fit the most complicated standard form model within each class (i.e., maximum number of possible parameters) to the standardized data \( Y_E - u_E \mid Y_E \not\leq u_E \);

(iv) select as the standard form model class the one which produces the best fit to the standardized data, in the sense of smallest AIC;
(v) use likelihood ratio tests to test for simplification of models within the selected standard form class, and select a final standard form model;
(vi) fit the GP margins simultaneously with this standard form model, to \( Y - u \mid Y \leq u \) by maximizing (5.2);
(vii) Use likelihood ratio tests to find simplifications in the marginal parameterization.

Although this strategy is not guaranteed to result in a final GP model that is globally optimal, in the sense of minimizing an information criterion such as AIC, it should still result in a sensible model whilst avoiding enumeration and fitting of an unfeasibly large number of possibilities. The goodness of fit of the final model can be checked via diagnostics.

5.3 Threshold selection and model diagnostics

An important issue that pervades extreme value statistics — in all dimensions — is the selection of a threshold above which the limit model provides an adequate approximation of the distribution of threshold exceedances. Here this amounts to “how can we select a vector \( u \) such that \( Y - u \mid Y \not\leq u \) is well-approximated by a GP distribution?”. There are two considerations to take into account: \( Y_j - u_j \mid Y_j > u_j \) should be well-approximated by a univariate GP distribution, for \( j = 1, \ldots, d \), and the dependence structure of \( Y - u \mid Y \not\leq u \) should be well-approximated by that of a multivariate GP distribution. Marginal threshold selection has a large body of literature devoted to it; see Scarrott and MacDonald (2012) and Caeiro and Gomes (2016) for recent reviews. Threshold selection for dependence models is a much less well studied problem. Contributions include Lee et al. (2015) who considers threshold selection via Bayesian measures of surprise, and Wadsworth (2016) who examines how to make better use of so-called parameter stability plots, offering a method that can be employed on any parameter, pertaining to the margins or dependence structure. Recently, Wan and Davis (2017) proposed a method based on assessing independence between radial and angular distributions.

Here we propose exploiting the “constant conditional exceedance” property of multivariate GP distributions, and use the measure of asymptotic dependence

\[
\chi_{1:a}(q) := \frac{\mathbb{P}[F_1(Y_1) > q, \ldots, F_d(Y_d) > q]}{1 - q},
\]
where \( Y_j \sim F_j \) and the related quantity for the limiting GP distribution
\[
\chi_H(q) := \frac{\mathbb{P}[H_1(X_1) > q, \ldots, H_d(X_d) > q]}{1 - q}, \quad q \in (0, 1)
\]
to guide threshold selection for the dependence structure. By property (2.4), for a suitable choice of \( A \), it holds that \( \chi_H(q) \) is constant for \( q \) sufficiently large such that \( H_j(X_j) > q \) implies \( X_j > 0 \) for \( j \in \{1, \ldots, d\} \).

If \( Y \sim F \) and \( Y - \mathbf{u} \ | \ Y \nless \mathbf{u} \sim H \), then on the region \( q > \max_j F_j(u_j) \), we have \( \chi_{1:d}(q) = \chi_H(q') \) with \( q' = \{q - F(\mathbf{u})\}/\{1 - F(\mathbf{u})\} \). A consequence of this is that \( \chi_{1:d}(q) \) should be constant on the region \( Y > \mathbf{u} \), if \( \mathbf{u} \) represents a sufficiently high dependence threshold. The empirical version \( \hat{\chi}_{1:d}(q) \) of \( \chi_{1:d}(q) \) is defined by
\[
\hat{\chi}_{1:d}(q) := \frac{\sum_{i=1}^{n} \mathbb{1}\{\hat{F}_1(Y_1) > q, \ldots, \hat{F}_d(Y_d) > q\}}{n(1 - q)}, \quad q \in [0, 1), \quad (5.3)
\]
where \( \hat{F}_1, \ldots, \hat{F}_d \) represent the empirical distribution functions. If we use (5.3) to identify \( q^* = \inf\{0 < \tilde{q} < 1 : \chi_{1:d}(q) \equiv \chi \ \forall \ q > \tilde{q}\} \), then \( \mathbf{u} = (F_1^{-1}(q^*), \ldots, F_d^{-1}(q^*)) \) should provide an adequate threshold for the dependence structure. Once suitable thresholds have been identified for margins, \( \mathbf{u}_m \), and dependence, \( \mathbf{u}_d \), then a threshold vector which is suitable for the entire multivariate model is \( \mathbf{u} = \max(\mathbf{u}_m, \mathbf{u}_d) \).

Having identified a multivariate GP model and a threshold above which to fit it, a key concern is to establish whether the goodness-of-fit is adequate. For the dependence structure, one diagnostic comes from comparing \( \hat{\chi}_{1:d}(q) \) for \( q \to 1 \) to \( \chi_{1:d} \), which for models \( h_T \) in (3.3) has the form \( \chi_{1:d} = \mathbb{E}\left[\min_{1 \leq j \leq d}\{e^{T_j - \max(T)} / \mathbb{E}(e^{T_j - \max(T)})\}\right] \), whilst for models \( h_U \) in (3.4) we get \( \chi_{1:d} = \mathbb{E}\left[\min_{1 \leq j \leq d}\{e^{U_j} / \mathbb{E}(e^{U_j})\}\right] \). The form of \( \chi_{1:d} \) for \( h_R \) models follows through equation (3.6). In some cases these expressions may be obtained analytically, but they can always be evaluated by simulation (Rootzén et al., 2016).

A further diagnostic uses that \( \mathbb{P}[X_j > 0] = \mathbb{E}[e^{T_j - \max(T)}] = \mathbb{E}[e^{U_j}] / \mathbb{E}[e^{\max(U)}] \). Thus one compares \( \mathbb{P}[Y_j > u_j] / \mathbb{P}[Y \nless \mathbf{u}] \) with the relevant model-based probability. If the \( u_j \) are equal marginal quantiles these are the same for each margin.

Equation (2.4) suggests a model-free diagnostic of whether a multivariate GP model may be appropriate. To exploit this, one defines a set of interest \( A \), and compares the number of points of \( Y - \mathbf{u} \ | \ Y \nless \mathbf{u} \) that lie in \( A \) to \( t \) times the number of points of
\((Y - u - w_t)/t^\gamma \mid Y \leq u\) lying in \(A\) for various choices of \(t > 1\). According to (2.4), the ratio of these numbers should be approximately equal to 1. Note that setting \(A = \{x : x > 0\}\) is equivalent to computing \(\chi_H\) with \(H_1, \ldots, H_d\) replaced by \(H_1^+, \ldots, H_d^+\).

Finally, in the event that the margins can be modelled with identical shape parameters, one can test property (2.5) by examining the adequacy of the implied univariate GP distribution from a multivariate fit.

6 UK bank returns

We examine weekly negative raw returns on the prices of the stocks from four large UK banks: HSBC (H), Lloyds (L), RBS (R) and Barclays (B). Data were downloaded from Yahoo Finance. Letting \(Z_{j,t}, j \in \{H, L, R, B\}\), denote the closing stock price (adjusted for stock splits and dividends) in week \(t\) for bank \(j\), the data we examine are \(Y_{j,t} = 1 - \frac{Z_{j,t}}{Z_{j,t-1}}\), so that large positive values of \(Y_{j,t}\) correspond to large relative losses for that stock. The observation period is 10/29/2007 – 10/17/2016, with \(n = 470\) datapoints.

Figure 1 displays pairwise plots of the negative returns. There is evidence of strong extremal dependence from these plots, as the largest value of \(Y_L, Y_R, Y_B\) occurs simultaneously, with positive association amongst other large values. The largest value of \(Y_H\) occurs at a different time, but again there is positive association between other large values. As is common in practice the value of \(\hat{\chi}_{HLRB}(q)\) generally decreases as \(q\) increases (see Figure 1 in the supplementary material), but is plausibly stable and constant from slightly above \(q = 0.8\). Consequently, we proceed with fitting a GP distribution. Ultimately, we wish to fit a parametric GP model to the raw threshold excesses \(\{Y_t - u : Y_t \leq u\}\). In view of the large variety of potential models and parameterizations, we use the model selection strategy detailed in Section 5.2. Throughout we use censored likelihood with \(v = 0\).

Based on the plot of \(\hat{\chi}_{HLRB}(q)\) we select the 0.83 marginal quantile as the threshold in each margin; there are 149 observations with at least one exceedance. We fit the models with densities (4.1), (4.2), (4.3), (4.4) and (4.5) to the standardized data. The smallest AIC is given by model (4.1), i.e., where \(f_T\) is the density of independent Gumbel random variables. We therefore select this class and proceed with item (v) of the procedure in Section 5.2 to test for simplifications within this class. In Table 1, model M1 is the most
Figure 1: Pairwise scatterplots of the negative weekly returns of the stock prices of four UK banks: HSBC (H), Lloyds (L), RBS (R) and Barclays (B), from 10/29/2007 to 10/17/2016.

The complex model with all dependence parameters. Model M2 imposes the restriction $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, whilst M3 imposes $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha$, and M4 imposes both. We observe that both possible sequences of likelihood ratio tests between nested models lead to M4 when adopting a 5% significance level. This model only contains a single parameter, which is a useful simplification.

Table 1: Negative UK bank returns: parameterizations of (4.1) for standardized data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Number</th>
<th>Maximized log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3$</td>
<td>7</td>
<td>-917.0</td>
</tr>
<tr>
<td>M2</td>
<td>$\alpha_1, \alpha_2, \alpha_3, \alpha_4$</td>
<td>4</td>
<td>-918.2</td>
</tr>
<tr>
<td>M3</td>
<td>$\alpha, \beta_1, \beta_2, \beta_3$</td>
<td>4</td>
<td>-920.8</td>
</tr>
<tr>
<td>M4</td>
<td>$\alpha$</td>
<td>1</td>
<td>-921.0</td>
</tr>
</tbody>
</table>

Finally we fit a full GP distribution to Model M4, and test for simplifications in the marginal parameterization. Marginal parameter stability plots suggest that the 0.83 quantile is adequate, which is also supported by diagnostics from the fitted model (supplemen-
Table 2: Negative UK bank returns: maximum likelihood estimates (MLE) and standard errors (SE) of parameters from the final model for the original data.

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>1.29</td>
<td>0.14</td>
</tr>
<tr>
<td>σ₃</td>
<td>0.020</td>
<td>0.0026</td>
</tr>
<tr>
<td>σ₄</td>
<td>0.041</td>
<td>0.0053</td>
</tr>
<tr>
<td>σ₅</td>
<td>0.038</td>
<td>0.0052</td>
</tr>
<tr>
<td>σ₆</td>
<td>0.035</td>
<td>0.0049</td>
</tr>
<tr>
<td>γ</td>
<td>0.43</td>
<td>0.082</td>
</tr>
</tbody>
</table>

At a 5% significance level, a likelihood ratio test for the hypothesis of \( γ_H = γ_L = γ_R = γ_B \) provides no evidence to reject the null hypothesis, whereas a test for \( σ_H = σ_L = σ_R = σ_B \) is rejected. The parameter estimates are displayed in Table 2.

To scrutinize the fit of the model, we examine marginal, dependence, and joint diagnostics. Quantile-quantile (QQ) plots for each of the univariate GP distributions implied for \( Y_{t,j} - u_j \mid Y_{t,j} > u_j \) are displayed in the supplementary material (Figure 2) indicating reasonable fits in each case. Estimates of the pairwise \( χ_{ij}(q) \), \( i ≠ j ∈ \{H, L, R, B\} \), are plotted in Figure 2, with the corresponding fitted value and threshold indicated; triplet-wise plots and the plot of \( \hat{χ}_{HLRB}(q) \) show similarly good agreement. Since the model has a single parameter, all pairs are exchangeable and have the same fitted value of \( χ \) for any fixed dimension.

Other model diagnostics are available; recall Section 5.3. One is sum-stability under shape constraints, Equation (2.5). As a diagnostic, we fit a univariate GP distribution to

\[
\sum_{j∈\{H,L,R,B\}} (Y_{t,j} - u_j) \mid \sum_{j∈\{H,L,R,B\}} (Y_{t,j} - u_j) > 0,
\]

with scale parameter estimate (standard error) obtained as 0.10 (0.021), and shape parameter estimate 0.45 (0.17). QQ plots suggest that the fit is good; see the supplementary material (Figure 3). For comparison, \( \sum_{j∈\{H,L,R,B\}} \hat{σ}_j = 0.13 \) with standard error 0.014 obtained using the delta method, whilst the maximized univariate GP log-likelihood is 63.5, and that for the parameters obtained via the multivariate fit is 62.2, showing that the theory holds well.

Weighted sums of raw stock returns correspond to portfolio performance. We use the final fitted model to compute two commonly-used risk measures, Value at Risk (VaR) and Expected Shortfall (ES), for a time horizon of one week. If the conditional distribution of
Figure 2: Negative UK bank returns: estimates of pairwise $\chi_{ij}(q)$ with fitted pairwise $\chi_{ij}$ (horizontal line), for HSBC (H), Lloyds (L), RBS (R) and Barclays (B). Clockwise from top left: $\chi_{HL}$, $\chi_{HR}$, $\chi_{HB}$, $\chi_{RB}$, $\chi_{LB}$, $\chi_{LR}$. The vertical line is the threshold used. Approximate 95% pointwise confidence intervals are obtained by bootstrapping from $\{Y_t : t = 1, \ldots, n\}$.

$\sum_j a_j(Y_{t,j} - u_j)$ given the event $\sum_j a_j(Y_{t,j} - u_j) > 0$ is $\text{GP}(\sum_j a_j \sigma_j, \gamma)$, then

$$\text{VaR}(p) = \sum_j a_j u_j + \frac{\sum_j a_j \sigma_j}{\gamma} \left\{ \left( \frac{\phi}{p} \right)^\gamma - 1 \right\},$$

where $0 < p < \phi = \mathbb{P}[\sum_j a_j(Y_{t,j} - u_j) > 0]$, so that (6.2) is the unconditional $1-p$ quantile of $\sum_j a_j Y_{t,j}$. We estimate the probability $\phi$ by maximum likelihood using the assumption $\sum_1^1 \{\sum_j a_j(Y_{t,j} - u_j) > 0\} \sim \text{Bin}(n, \phi)$, and in the univariate model, $\phi$ is orthogonal to the parameters of the conditional excess distribution. In the multivariate model

$$\mathbb{P}\left[\sum_j a_j(Y_{t,j} - u_j) > 0\right] = \mathbb{P}\left[\sum_j a_j(Y_{t,j} - u_j) > 0 \mid Y_t \not\subseteq u\right] \mathbb{P}[Y_t \not\subseteq u] = p(\theta) \tilde{\phi},$$

where $p(\theta)$ is an expression involving the parameters of the multivariate GP model, and $\tilde{\phi}$ is the proportion of points for which $Y_t \not\subseteq u$. The expression $p(\theta)$ is not tractable here, thus we continue to estimate $\phi$ as the binominal maximum likelihood estimate, and as a working assumption treat it as orthogonal to the other parameters. However, an estimate of $p(\theta)$ can be obtained by simulation using the estimated $\theta$; the utility of this will be demonstrated in Figure 4.
The expected shortfall is defined as the expected loss given that a particular VaR threshold has been exceeded. Under the GP model, and provided \( \gamma < 1 \), it is given by

\[
ES(p) = E\left[ \sum_j a_j Y_{t,j} \mid \sum_j a_j Y_{t,j} > \text{VaR}(p) \right] = \text{VaR}(p) + \sum_j a_j \sigma_j + \gamma \left[ \frac{\text{VaR}(p) - \sum_j a_j u_j}{1-\gamma} \right].
\]

Asymptotic theory suggests that a univariate GP model fit directly to \( \sum_j a_j (Y_{t,j} - u_j) \) or the implied GP(\( \sum_j a_j \sigma_j, \gamma \)) model obtained from the multivariate fit could be used. An advantage of using the GP(\( \sum_j a_j \sigma_j, \gamma \)) model derived from the multivariate fit is reduced uncertainty, combined with consistent estimates across different portfolio combinations.

Figures 3 displays VaR curves and confidence intervals for two different weight combinations and for both the univariate and multivariate fits, together with empirical counterparts, whilst Figure 4 in the supplementary material shows the corresponding ES curves. For VaR the univariate fit is closer in the body and the multivariate fit is closer to the data in the tails. The reduction in uncertainty is clear and potentially quite useful for smaller \( p \). For ES (supplementary material, Figure 4) the univariate fit estimates smaller values than the multivariate fit in each case, and seems to reflect the observed data better. However, the empirical ES values fall within the 95% confidence intervals obtained from the multivariate model, suggesting that the model is still consistent with the data. Note that the univariate fit is tailored specifically to the data \( \sum_j a_j Y_{t,j} \) and as such, we would always expect the point estimates from Figure 3 to look better for the univariate fit. On the other hand, when interest lies in different functions of the extremes of \( Y_{t,j} \), the multivariate approach is able to deliver self-consistent inference.

Figure 4 illustrates how the multivariate model provides more consistent estimates of VaR across different portfolio combinations compared to the use of multiple univariate models. To produce the figures, we suppose that \( \sum_j a_j = 100 \) represents the total amount available to invest. The value \( a_H = 10 \) is fixed, with other weights varying, but with each \( a_j \geq 1 \). Two estimates making use of the multivariate model are provided: one for which a model-based estimate of \( p(\theta) \) from (6) is used (with estimation based on 100,000 draws from the fitted model), and one where the empirical binomial estimate of \( \phi \) is used, as in Figure 3 and the supplementary material (Figure 4). Both sets of multivariate estimates suggest much more consistent behaviour across portfolio combinations than the use of univariate fits. In particular, behaviour is very smooth once a model-based estimate for
Figure 3: VaR estimates and pointwise 95% delta-method confidence intervals for portfolio losses based on the weights given as percentages invested in HSBC, Lloyds, RBS and Barclays as in the figure title. Estimates based on the multivariate GP fit are on the left; estimates based on the univariate fit are on the right.

\( p(\theta) \) is included.

7 Landslides

Rainfall can cause ground water pressure build-up which, if very high, can trigger a landslide. The cause can be short periods with extreme rain intensities, or longer periods of up to three days of more moderate, but still high rain intensities. Guzzetti et al. (2007) consolidate many previous studies and propose threshold functions which link duration in hours, \( D \), with total rainfall in millimeters, \( P \), such that rainfall below these thresholds are unlikely to cause landslides. For highland climates in Europe this function is

\[
P = 7.56 \times D^{0.52}.
\] (7.1)

Thus, a one-day rainfall below 39.5 mm, a two-day rainfall below 56.6 mm, or a three-day rainfall below 69.9 mm are all unlikely to cause a landslide.

We use a long time series of daily precipitation amounts \( P_1, \ldots, P_N \) collected by the Abisko Scientific Research Station in northern Sweden in the period 1/1/1913 – 12/31/2014, to estimate a lower bound for the probability of the occurrence of rainfall events which may lead to landslides. The total cost of landslides in Sweden is around SEK 200 million/year. There have been several landslides in the Abisko area in the past century, for instance in October 1959, August 1998, and July 2004 (Rapp and Strömquist, 1976; Jonasson and
Figure 4: Maximum likelihood estimates of VaR(0.001) for $\sum_j a_j Y_{t,j}$ with $a_H = 10$ and $a_B = 90 - a_L - a_R$ for four UK banks, HSBC, Lloyds, RBS and Barclays. Left: from multivariate model including simulation to estimate $p(\theta)$ from (6); centre: from multivariate model using the binomial estimate of $\phi$; right: from univariate model fit to each combination separately. Note the different colour scales on each panel.

Nyberg, 1999; Beylich and Sandberg, 2005). The rainfall episodes causing the landslides are clearly visible in the data, with 24.5 mm of rain on October 5, 1959, 21.0 mm of rain on August 24, 1998, and 61.9 mm of rain on July 21, 2004. The 2004 rain amount is well above the 1-day risk threshold, whereas the 1959 and 1998 rain amounts are below the 1-day threshold. The explanation may be that the durations of the latter two rain events were shorter than 24 hours, and that the threshold in (7.1) was still exceeded.

We wish to construct a dataset $Y_1, \ldots, Y_n \in \mathbb{R}^3$, for $n < N$, whose components represent daily, two-day, and three-day extreme rainfall amounts respectively, to account for longer periods of moderate rainfall. Based on a mean residual life plot and parameter stability plots (not shown here) for the daily rainfall amounts $P_1, \ldots, P_N$, we choose the threshold $u = 12$, which corresponds roughly to the 99% quantile. Figure 5 shows the cumulative three-day precipitation amounts $P_i + P_{i+1} + P_{i+2}$ for $i \in \{1, \ldots, N-2\}$. The threshold $u$ is used to extract clusters of data containing extreme episodes; the data $Y_1, \ldots, Y_n$ are constructed as follows:

1. Let $i$ correspond to the first sum $P_i + P_{i+1} + P_{i+2}$ which exceeds the threshold $u$ and set $P_{(1)} = \max(P_i, P_{i+1}, P_{i+2})$.
2. Let the first cluster $C_{(1)}$ consist of $P_{(1)}$ plus the five values preceding it and the five
values following it.

3. Let $Y_{11}$ be the largest value in $C_{(1)}$, $Y_{12}$ the largest sum of two consecutive non-zero values in $C_{(1)}$, and $Y_{13}$ the largest sum of three consecutive non-zero values in $C_{(1)}$.

4. Find the second cluster $C_{(2)}$ and compute $Y_2 = (Y_{21}, Y_{22}, Y_{23})$ in the same way, starting with the first observation after $C_{(1)}$.

Continuing this way, we obtain a dataset $Y_1, \ldots, Y_n$, with $d = 3$ and $n = 580$.

![Three-day rainfall amounts in Abisko](image)

Figure 5: Precipitation data in Abisko: cumulative three-day precipitation amounts $P_i + P_{i+1} + P_{i+2}$ for $i \in \{1, \ldots, N - 2\}$ with threshold $u = 12$ in red.

Annual maxima of a similar data set were analysed in Rudvik (2012), with the conclusion that there was no time trend. We fitted a univariate GP distribution with a fixed shape parameter $\gamma$ but a loglinear trend for the scale parameter to the marginal components $(Y_i)_{i=1}^n$, and also did not find any significant trend; see the supplementary material. The estimated shape parameters obtained from fitting univariate GP distributions to the marginal threshold excesses are close to zero (the hypothesis $\gamma = 0$ is not rejected at a 5% level) and the confidence intervals for the scale parameters overlap (Table 3). Note that a common $\sigma$ and $\gamma$ only implies that the marginal distributions are equal conditional on exceeding the threshold; it does not imply that the unconditional probabilities $P[Y_j > u_j]$ are equal.

In the following analysis, we set $\sigma = \sigma 1$ and $\gamma = \gamma 1$, and we fit the structured models from Section 4.3, both with $\gamma = 0$ and with $\gamma > 0$, using censored likelihood with $v = 0$. To insure identifiability we set $\lambda_1 = 1$ for both models. We choose $u = u 1$ with $u = 24$ since parameter estimates stabilize for thresholds around this value, and continue with the
Table 3: Precipitation data in Abisko: estimates of the parameters of marginal GP models for thresholds \( u = 12, u = 13.5 \) and \( u = 14 \) respectively; standard errors in parentheses.

<table>
<thead>
<tr>
<th>( Y_{i1} )</th>
<th>( Y_{i2} )</th>
<th>( Y_{i3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\gamma} )</td>
<td>-0.06 (0.05)</td>
<td>-0.02 (0.06)</td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>8.26 (0.69)</td>
<td>9.34 (0.74)</td>
</tr>
</tbody>
</table>

142 data points whose third components exceed \( u = 24 \).

The estimates of \( \sigma \) are somewhat higher than in the marginal analysis and again the hypothesis \( \gamma = 0 \) was not rejected (Table 4). The higher estimate of \( \sigma \) is intuitively reasonable since the maximum likelihood estimators for \( \gamma \) and \( \sigma \) are negatively correlated and since \( \hat{\gamma} \) is positive for the second model.

To estimate the risk of a future landslide using formula (7.1), i.e., \( \mathbb{P}[\mathbf{Y} \not\leq \mathbf{y}] \), where \( \mathbf{y} = (39.5, 56.6, 69.9) \), we use the first model, \( \gamma = 0 \), and write

\[
\mathbb{P}[\mathbf{Y} \not\leq \mathbf{y}] = \mathbb{P}[\mathbf{Y} - \mathbf{u} \not\leq \mathbf{y} - \mathbf{u} \mid \mathbf{Y} \not\leq \mathbf{u}] \mathbb{P}[\mathbf{Y} \not\leq \mathbf{u}]
= \{1 - H \left( \frac{\mathbf{y} - \mathbf{u}}{\mathbf{\sigma}}; 1, 0 \right) \} \mathbb{P}[Y_3 > u].
\]

The first term of (7.2) was calculated by integrating the density (4.7), using the parameter estimates \( (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\sigma}) \) from the top row of Table 4. \( \mathbb{P}[Y_3 > u] \) was estimated by the relative number of years with exceedances. For \( \mathbf{y} = (39.5, 56.6, 69.9) \), obtained using formula (7.1), we get

\[
\mathbb{P}[Y_1 > 39.5 \text{ or } Y_2 > 56.6 \text{ or } Y_3 > 69.9] \approx 0.062.
\]

We find that the probability of rain amounts which could lead to a landslide in any given year is 0.062. This is higher than the result in Rudvik (2012) who used data from 1913–2008 and analysed daily, three-day and five-day precipitation amounts.

Marginal QQ-plots show good fits for components 2 and 3, but less so for component 1 for the model with \( \gamma = 0 \) (Figure 5 in the supplementary material). This is due to the restriction \( \mathbf{\sigma} = \mathbf{\sigma}1 \) used to ensure that the components are ordered.

For the dependence structure, using Equation (2.4) (see also Section 5.3) and \( \gamma = 0 \),
Table 4: Precipitation data in Abisko: parameter estimates for the structured components model with $u = 24$; standard errors in parentheses.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\lambda}_1$</th>
<th>$\hat{\lambda}_2$</th>
<th>$\hat{\lambda}_3$</th>
<th>$\tilde{\sigma}$</th>
<th>$\tilde{\gamma}$</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 0$</td>
<td>1.00</td>
<td>0.84 (0.13)</td>
<td>1.08 (0.18)</td>
<td>10.17 (0.80)</td>
<td>0</td>
<td>-870.0</td>
</tr>
<tr>
<td>$\gamma &gt; 0$</td>
<td>1.00</td>
<td>0.83 (0.12)</td>
<td>1.06 (0.18)</td>
<td>9.14 (0.99)</td>
<td>0.11 (0.08)</td>
<td>-868.9</td>
</tr>
</tbody>
</table>

we display the empirical counterpart of the ratio

$$\frac{\mathbb{P}[Y - u \in A \mid y \not\in u]}{t \mathbb{P}[Y - u - \sigma \log t \in A \mid Y \not\in u]} \quad (7.3)$$

where $\sigma$ is the vector of scale parameter estimates of the marginal GP models above $u = 24$ for the sets $A_j = \{x \in \mathbb{R}^3 : x_j > 0\}, j \in \{1, 2, 3\}$ (Figure 6). The plots indicate that the model is appropriate. The plot for $A_1$ uses few observations and hence is more variable. Formulas for pairwise and trivariate $\chi$ and comparisons with their empirical counterpart can be found in Figure 6 of the supplementary material. The model-based estimates of exceedance probabilities are $\mathbb{P}[X_1 > 0] = 0.34$ (0.03), $\mathbb{P}[X_2 > 0] = 0.63$ (0.03) using values from the top row in Table 4 and delta method standard errors. The empirical probabilities are 0.32 and 0.69 respectively. Plots of the empirical probabilities for a range of different thresholds (not shown) confirm the chosen threshold value $u = 24$.

The test statistic in Einmahl et al. (2016, Corollary 2.5) compares the estimates of
$(\chi_{12}, \chi_{13}, \chi_{23}, \chi_{123})$ with an empirical estimator. It depends on a value $k$ which represents
a threshold: a low value of $k$ corresponds to a high threshold. Asymptotically the test
statistic has a chi-square distribution with 2 degrees of freedom whose 95% quantile is
5.99. For $k \in \{50, 75, 100, 125, 150\}$ we obtain the values 1.08, 4.48, 1.17, 5.42, and 0.99,
and hence cannot reject the structured components model for any value of $k$.

Acknowledgements

The authors gratefully acknowledge support from: the Knut and Alice Wallenberg foun-
dation (Kiriliouk, Rootzén, Wadsworth); “Fonds de la Recherche Scientifique - FNRS”
(Kiriliouk); “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Commu-
nauté française de Belgique” (Kiriliouk, Segers); IAP research network grant P7/06 of the
Belgian government (Segers); EPSRC fellowship grant EP/P002838/1 (Wadsworth).

We thank the Abisko Scientific Research Station for access to their rainfall data. To ob-
tain it, see instructions at http://polar.se/en/abisko-naturvetenskapliga-station/
vaderdat. Data for Section 6 can be downloaded from http://finance.yahoo.com. Pro-
grams for the analyses in Sections 6 and 7 are available from the authors on request.

References


Beylich, A. A. and Sandberg, O. (2005). Geomorphic effects of the extreme rainfall event of
20–21 july, 2004 in the Latnjavagge catchment, northern Swedish Lapland. Geografiska


