
SIMULTANEOUS TAIL INDEX ESTIMATION

Authors: JAN BEIRLANT

– Department of Mathematics, Katholieke Universiteit Leuven,
Celestijnenlaan 200B, B-3000 Leuven, Belgium
(jan.beirlant@wis.kuleuven.ac.be)

YURI GOEGEBEUR

– Postdoctoral researcher Fund for Scientific Research – Flanders,
Department of Applied Economics, Katholieke Universiteit Leuven,
Naamsestraat 69, B-3000 Leuven,
University Center for Statistics, Katholieke Universiteit Leuven,
W. de Croylaan 54, B-3001 Heverlee, Belgium
(yuri.goegebeur@econ.kuleuven.ac.be)

Received: November 2003

Revised: January 2004

Accepted: January 2004

Abstract:

- The estimation of the extreme-value index γ based on a sample of independent and identically distributed random variables has received considerable attention in the extreme-value literature. However, the problem of combining data from several groups is hardly studied. In this paper we discuss the simultaneous estimation of tail indices when data on several independent data groups are available. The proposed methods are based on regression models linking tail related statistics to the extreme-value index and parameters describing the second order tail behaviour. For heavy-tailed distributions ($\gamma > 0$), estimators are derived from an exponential regression model for rescaled log-spacings of successive order statistics as described in Beirlant *et al.* (1999) and Feuerverger and Hall (1999). Estimators for $\gamma \in \mathbb{R}$ are obtained using the linear model for *UH*-statistics given in Beirlant *et al.* (2000). In both cases, the optimal number of extremes to be used in the estimation is derived from the asymptotic mean squared error matrix.

Key-Words:

- *extreme-value index; regression; quantile-quantile plot.*

AMS Subject Classification:

- 62G30, 62G32.

1. INTRODUCTION

A central topic in extreme-value theory which continues to receive considerable attention is the estimation of the extreme-value index γ . This index is directly related to the tail of a distribution function F with the tail function $1 - F$ becoming more heavy as γ increases. The extreme-value index γ can be estimated from a parametric or a semi-parametric point of view.

Parametric approaches are based on limit theorems which form the core of the extreme-value theory. Consider X_1, \dots, X_n independent and identically distributed random variables and let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the corresponding ascending order statistics. A first possibility is based on the following result of Fisher and Tippett (1928). If for a distribution function F_X there exist sequences of constants $(a_n > 0)_n$ and $(b_n)_n$ such that

$$(1.1) \quad \lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F_X^n(a_n x + b_n) = H(x)$$

at all continuity points of H , with H a nondegenerate distribution, then F_X is said to belong to the domain of (maximum) attraction of H , denoted $F_X \in \mathcal{D}(H)$. Moreover, it is known that if such a nondegenerate limit distribution H exists, it should be of the form

$$(1.2) \quad H_\gamma(x) = \begin{cases} \exp\left(-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right) & 1 + \gamma \frac{x - \mu}{\sigma} > 0, \quad \gamma \neq 0 \\ \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right) & x \in \mathbb{R}, \quad \gamma = 0 \end{cases}$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$. This limit distribution is the so-called generalized extreme-value distribution (GEV). Note that the extreme-value index γ appears as a shape parameter in (1.2). Based on this result, γ can be estimated by fitting (1.2) to sample maxima (Gumbel, 1967). A second possibility is based on the generalized Pareto distribution (GPD) given by

$$(1.3) \quad G_\gamma(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \gamma = 0 \end{cases}$$

with $\sigma > 0$ and with $x > 0$ if $\gamma \geq 0$, $0 < x < -\sigma/\gamma$ if $\gamma < 0$, which is fitted to exceedances over a specified threshold u (Pickands, 1975, Smith, 1985, 1987).

Next to the above described parametric approaches, γ can also be estimated semi-parametrically. Define the tail quantile function U_X as $U_X(x) = \inf\{y: F_X(y) \geq 1 - \frac{1}{x}\}$, $x > 1$. For Pareto-type or heavy tailed distributions

($\gamma > 0$) we have

$$(1.4) \quad \begin{aligned} F_X \in \mathcal{D}(H_\gamma) &\iff 1 - F_X(x) = x^{-\frac{1}{\gamma}} \tilde{l}(x), \quad x > 0 \\ &\iff U_X(x) = x^\gamma l(x), \quad x > 1 \end{aligned}$$

with \tilde{l}, l slowly varying functions at infinity i.e. positive functions g such that

$$(1.5) \quad \frac{g(\lambda x)}{g(x)} \rightarrow 1 \quad \text{as } x \rightarrow \infty \quad \forall \lambda > 0 .$$

The conditions given in (1.4) characterize completely the first order behavior of F_X . The Pareto quantile plot can be found in the literature as the basis for evaluating the goodness-of-fit hypothesis of strict Pareto behavior. For the strict Pareto distribution $\log U(x) = \gamma \log(x)$ so the log theoretical quantiles stand in linear relationship with $\log(x)$. Replacing the theoretical quantiles $\log U(\frac{n+1}{j})$ by their empirical counterparts $\log U_n(\frac{n+1}{j}) = \log X_{n-j+1,n}$, the coordinates of the points on the quantile plot are given by

$$(1.6) \quad \left(\log \left(\frac{n+1}{j} \right), \log X_{n-j+1,n} \right) \quad j = 1, \dots, n .$$

In case of a good fit of the strict Pareto distribution to the data, the points on the Pareto quantile plot should show a straight line pattern. Moreover, the slope of a line through the origin fitted to the Pareto quantile plot will estimate γ . In case the distribution of the data is of Pareto-type, the log-tail quantile function can be written as $\log U(x) = \gamma \log x + \log l(x)$. Since $\log l(x) / \log x \rightarrow 0$ as $x \rightarrow \infty$ we have that $\log U(x) \sim \gamma \log x$ as $x \rightarrow \infty$, so the Pareto quantile plot will be *ultimately* linear. Again, the slope of the linear part will approximate γ . Several well known estimators for γ can be interpreted as estimators of the slope of the linear part of the Pareto quantile plot. For instance, the Hill (1975) estimator given by

$$(1.7) \quad H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n} \quad k = 1, \dots, n-1$$

clearly measures the average increase of the Pareto quantile plot to the right of the anchor point $(\log(\frac{n+1}{k+1}), \log X_{n-k,n})$. Other important estimators for $\gamma > 0$ are the so-called kernel estimators derived by Csörgő *et al.* (1985) and the least squares estimators proposed by Kratz and Resnick (1996) and Schultze and Steinebach (1996) among others.

The estimation of $\gamma \in \mathbb{R}$ has been studied less extensively. In this paper we will concentrate on the approach based on the generalized quantile plot described in Beirlant *et al.* (1996) and Beirlant *et al.* (2000). For a positive random variable X , consider

$$H_X(x) = E \left(\log X - \log U_X(x) \mid X > U_X(x) \right) \quad x > 1 ,$$

the mean residual life function of the log-transformed data, and define the adapted mean excess function UH_X as

$$(1.8) \quad \begin{aligned} UH_X(x) &= U_X(x) H_X(x) \\ &= U_X(x) \int_1^\infty \left(\log U_X(zx) - \log U_X(x) \right) \frac{dz}{z^2} \quad x > 1 . \end{aligned}$$

In Theorem 1 of Beirlant *et al.* (1996) it is shown that $F_X \in \mathcal{D}(H_\gamma)$, $\gamma \in \mathbb{R}$, implies that

$$(1.9) \quad \log UH_X(x) = \gamma \log x + \log \check{l}(x)$$

with \check{l} denoting a slowly varying function at infinity. As a consequence, $\log UH_X(x) \sim \gamma \log x$ for $x \rightarrow \infty$. Consider X_1, \dots, X_n independent and identically distributed positive random variables. Replacing U_X and H_X in (1.8) by their empirical counterparts yields

$$(1.10) \quad UH_{j,n} = \widehat{UH}_X\left(\frac{n}{j}\right) = X_{n-j,n} \left(\frac{1}{j} \sum_{i=1}^j \log X_{n-i+1,n} - \log X_{n-j,n} \right)$$

as sample versions for $UH_X(n/j)$, $j = 1, \dots, n-1$. For $F_X \in \mathcal{D}(H_\gamma)$, the generalized quantile plot, defined by

$$(1.11) \quad \left(\log \frac{n}{j}, \log UH_{j,n} \right) \quad j = 1, \dots, n-1 ,$$

should be ultimately linear in the smaller j -values. Further, the slope of the straight line behind the linear part of the generalized quantile plot is the unknown γ . Applying a Hill-type operation on $UH_{j,n}$, $j = 1, \dots, k$, results in the following estimator for γ , called the adapted Hill estimator $H_{k,n}^{*2}$,

$$(1.12) \quad H_{k,n}^{*2} = \frac{1}{k} \sum_{j=1}^k \log UH_{j,n} - \log UH_{k+1,n} .$$

Other well known estimators for $\gamma \in \mathbb{R}$ have been proposed by Pickands (1975) and Dekkers *et al.* (1989).

As is clear from the above discussion, the literature on extreme-value methods for a sample of independent and identically distributed data is quite elaborate. However, the problem of combining data from different independent data groups is hardly studied. Nevertheless, the problem is important: consider for instance the combination of earthquake data from different geographical regions such as subduction zones and midocean ridge zones. Often the amount of available data is small and hence the combination of different samples is important in order to gain efficiency.

Of course, regression models with dummy explanatory variables describing the groups, can be used in combination with classical extreme-value models such as the generalized extreme-value distribution (1.2), which is fitted to maxima, or the generalized Pareto distribution (1.3), which is fitted to so-called peaks (or excesses) over threshold data. This approach can be found for instance in Davison and Smith (1990). A major difficulty when working with the GPD in a regression setting is the selection of the threshold. Ideally, the threshold should depend on the covariates in order to take the relative extremity of the observations into account. This issue was also noted by Davison and Smith (1990) and Coles and Tawn (1998). Up to now, solutions seem rather ad hoc and especially designed for the data set at hand. Often the threshold is taken equal over the different groups leading to inefficient use of the data if the scale in the different groups is quite different.

In contrast, the semi-parametric approaches where only the k largest data are used for tail estimation can overcome this problem. In this paper we consider in section 2 the estimation problem of γ in case data on several Pareto-type groups are available. Next, in section 3, we extend the procedure to the general case where the extreme-value index can be positive or negative. The performance of the proposed methods will be illustrated using small sample simulations.

2. LINEAR MODEL FORMULATION, $\gamma > 0$

2.1. Description of the model

Consider independent and identically distributed positive random variables $X_1^{(j)}, \dots, X_{n_j}^{(j)}$ with a common distribution function $F_{X^{(j)}}$, $j = 1, \dots, G$, where G denotes the number of groups. Assume further that the G groups are independent of each other and that the response distributions are of Pareto-type i.e. the tail quantile functions $U_{X^{(j)}}$, $j = 1, \dots, G$, satisfy

$$(2.1) \quad U_{X^{(j)}}(x) = x^{\gamma_j} l_j(x) \quad x > 1, \quad \gamma_j > 0$$

where γ_j and l_j denote the extreme-value index respectively the slowly varying function of group j .

In the extreme-value literature one often imposes the so-called slow variation with remainder condition (see section 3.12.1 of Bingham *et al.*, 1987) on the slowly varying function l in (1.4). This second order condition specifies the rate of convergence of the ratio $l(\lambda x)/l(x)$ to its limit as $x \rightarrow \infty$.

Assumption (R_l): There exists a real constant $\rho < 0$ and a rate function b satisfying $b(x) \rightarrow 0$ as $x \rightarrow \infty$, such that for all $\lambda \geq 1$, as $x \rightarrow \infty$,

$$\frac{l(\lambda x)}{l(x)} - 1 \sim b(x) k_\rho(\lambda)$$

$$\text{with } k_\rho(\lambda) = \int_1^\lambda v^{\rho-1} dv.$$

Note that assumption (R_l) is quite general and is satisfied by the Hall (1982) class of Pareto-type distributions given by

$$(2.2) \quad U_X(x) = a x^\gamma \left(1 + d x^\rho + o(x^\rho)\right) \quad x > 1; \quad a, \gamma > 0; \quad \rho < 0; \quad d \in \mathbb{R},$$

with $b(x) \sim \rho d x^\rho$ as $x \rightarrow \infty$.

As in a classical one-way ANOVA situation we introduce the parametrization $\gamma_j = \beta_0 + \beta_j$, $j = 1, \dots, G$, with $\sum_{j=1}^G \beta_j = 0$, so that the parameters β_j denote the difference of the extreme-value index of group j with respect to the global average over all groups. This transformation will now be combined with the following linear model describing the estimation problem of every γ_j , $j = 1, \dots, G$.

Under the slow variation with remainder condition on the l_j , $j = 1, \dots, G$, it can be shown as in Beirlant *et al.* (1999) that the following regression model holds approximately

$$(2.3) \quad i \left(\log X_{n_j-i+1, n_j}^{(j)} - \log X_{n_j-i, n_j}^{(j)} \right) \approx \left(\gamma_j + b_j \left(\frac{n_j+1}{k+1} \right) \left(\frac{i}{k+1} \right)^{-\rho_j} \right) F_i^{(j)} \quad i = 1, \dots, k,$$

with b_j and ρ_j denoting the function b respectively the parameter ρ of group j and the $F_i^{(j)}$, $i = 1, \dots, k$, are independent standard exponential random variables. In Beirlant *et al.* (2002), the approximation error in (2.3) is shown to be $o_P(b_j(\frac{n_j+1}{k+1}))$, $j = 1, \dots, G$. Remark that regression model (2.3) is not identifiable when $\rho = 0$, for then γ_j and $b_j((n_j+1)/(k+1))$ together make up the mean response.

The classical way to estimate the parameters γ_j , $j = 1, \dots, G$, is then given by the Hill (1975) estimates which are obtained as maximum likelihood estimates by omitting the terms $b_j(\frac{n_j+1}{k+1})(\frac{i}{k+1})^{-\rho_j}$ in model (2.3) (these terms tend to 0 as $n_j \rightarrow \infty$ and $k/n_j \rightarrow 0$) leading to a simple average of the scaled log-spacings $i(\log X_{n_j-i+1, n_j}^{(j)} - \log X_{n_j-i, n_j}^{(j)})$, $i = 1, \dots, k$, as an estimator of γ_j , and hence

$$(2.4) \quad \hat{\beta}_0 = \frac{1}{G} \sum_{j=1}^G H_{k, n_j}^{(j)} \quad \text{and} \quad \hat{\beta}_j = H_{k, n_j}^{(j)} - \hat{\beta}_0, \quad j = 1, \dots, G,$$

in which $H_{k,n_j}^{(j)}$ denotes the Hill estimator for group j

$$(2.5) \quad H_{k,n_j}^{(j)} = \frac{1}{k} \sum_{i=1}^k \log X_{n_j-i+1,n_j}^{(j)} - \log X_{n_j-k,n_j}^{(j)} .$$

Introducing $\mathbf{\Lambda} = \text{Block-diag}(\gamma_j^2 I_k; j=1, \dots, G)$ and the $kG \times G$ matrix

$$\mathbf{L} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{1} & -\mathbf{1} & \cdots & -\mathbf{1} \end{bmatrix}$$

with $\mathbf{1}$ denoting a k -vector of ones, we find that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}' = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{G-1})$ is given by

$$(2.6) \quad \text{Acov}(\hat{\boldsymbol{\beta}}) = (\mathbf{L}'\mathbf{\Lambda}^{-1}\mathbf{L})^{-1} .$$

On the other hand the main term of the bias of the estimators (when $n_j \rightarrow \infty$ and $k/n_j \rightarrow 0$) is given by

$$(2.7) \quad \text{Abias}(\hat{\beta}_0) = \frac{1}{G} \sum_{j=1}^G \frac{b_j \left(\frac{n_j+1}{k+1} \right)}{1 - \rho_j} ,$$

$$(2.8) \quad \text{Abias}(\hat{\beta}_j) = \frac{b_j \left(\frac{n_j+1}{k+1} \right)}{1 - \rho_j} - \frac{1}{G} \sum_{l=1}^G \frac{b_l \left(\frac{n_l+1}{k+1} \right)}{1 - \rho_l} \quad j = 1, \dots, G-1 .$$

Application of the estimators defined by (2.4) and (2.5) involves the selection of the number of extreme order statistics k to be used in the estimation. Remark that we take the tail sample fraction k equal for all groups. If k is chosen too small, the resulting estimators will have a high variance. On the other hand, for larger k values the estimators will perform quite well with respect to variance but will be affected by a larger bias as observations are used which are not really informative for the tail of $F_{X^{(j)}}$, $j = 1, \dots, G$. Hence, a good k value should represent a good bias-variance trade-off. Here we will use the trace of the asymptotic mean squared error (AMSE) matrix as optimality criterion.

Defining the AMSE matrix Ω of $\hat{\boldsymbol{\beta}}$ as

$$(2.9) \quad \Omega(k) = (\mathbf{L}'\mathbf{\Lambda}^{-1}\mathbf{L})^{-1} + \boldsymbol{\kappa} \boldsymbol{\kappa}' ,$$

with $\boldsymbol{\kappa}$ denoting the G -vector containing the asymptotic bias expressions given by (2.7) and (2.8), the optimal number of extremes to be used in the estimation, k_{opt} , is defined as

$$k_{\text{opt}} = \arg \min \text{tr} \Omega(k) .$$

Note that $\Omega(k)$ depends on the unknown γ_j , ρ_j , $j=1, \dots, G$, and $b_j \left(\frac{n_j+1}{k+1} \right)$, $k=1, \dots, n_j-1$, $j=1, \dots, G$, which implies that the optimal k has to be derived from an estimate of $\Omega(k)$. The following algorithm is used to estimate k_{opt} and hence γ_j , $j=1, \dots, G$, adaptively:

1. Obtain initial estimates of γ_j , ρ_j , $j = 1, \dots, G$, together with estimates of $b_j(\frac{n_j+1}{k+1})$, $k=1, \dots, n_j-1$, $j=1, \dots, G$,
2. for $k = 2, \dots, \min(n_j; j=1, \dots, G) - 1$:
compute $\text{tr } \hat{\Omega}(k)$ and let

$$\hat{k}_{\text{opt}} = \arg \min \text{tr } \hat{\Omega}(k) ,$$

3. repeat step 2 but with the parameter estimates obtained from using a common k and obtain an update of the parameter estimates.

The initial estimates for the unknown parameters (cf. step 1) are obtained by fitting model (2.3) to the k largest observations of each group using a maximum likelihood method (see Beirlant *et al.*, 1999).

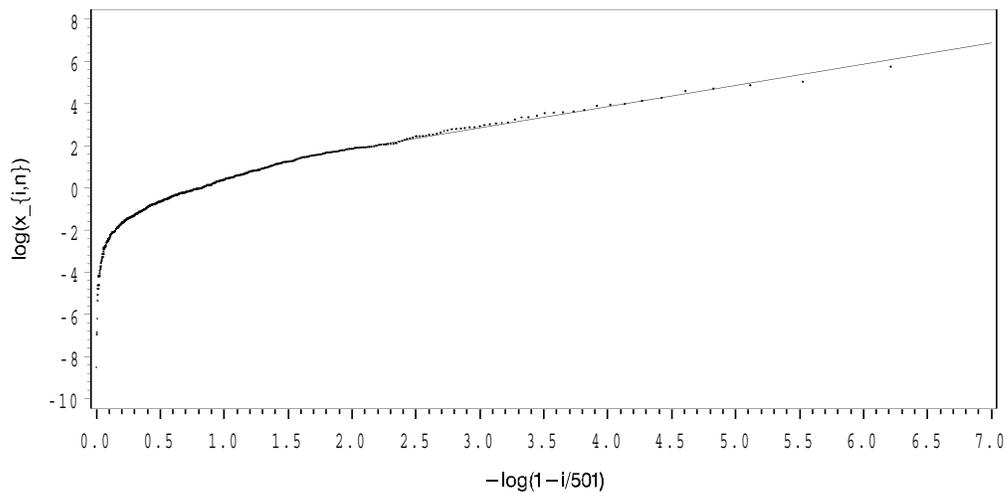
Inference about the regression vector β can be drawn using a likelihood ratio test statistic. For k/n_j , $j=1, \dots, G$, sufficiently small, the slowly varying nuisance part of (2.3) can be ignored and hence inference can be based on the reduced model $j(\log X_{n_j-i+1, n_j}^{(j)} - \log X_{n_j-i, n_j}^{(j)}) \approx (\beta_0 + \beta_j) F_i^{(j)}$, $i=1, \dots, k$, $j=1, \dots, G$. As in a 'classical' one-way ANOVA situation the hypothesis of main interest is $H_0: \beta_1 = \dots = \beta_{G-1} = 0$.

2.2. An illustration

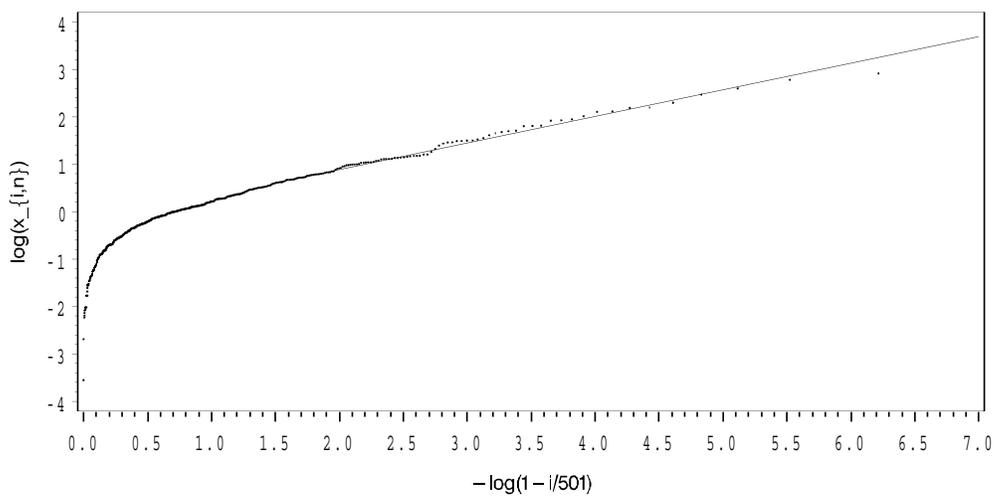
The procedure described above will be illustrated on a simulated dataset. The dataset contains observations from two groups with $n_1 = n_2 = 500$. Data were generated from Burr(η, τ, λ) distributions (Burr, 1942). The Burr(η, τ, λ) distribution function, given by

$$F_X(x) = 1 - \left(\frac{\eta}{\eta + x^\tau} \right)^\lambda \quad x > 0; \quad \eta, \tau, \lambda > 0 ,$$

is clearly of Pareto-type with $\gamma = \frac{1}{\lambda\tau}$ and $\rho = -\frac{1}{\lambda}$. For the simulated data $\lambda = 1$ and $\eta = 1$ for both groups and $\tau_1 = 1$, $\tau_2 = 2$ so $\gamma_1 = 1$, $\gamma_2 = 0.5$ and $\rho_1 = \rho_2 = -1$. Application of the above algorithm results in $\hat{k}_{\text{opt}} = 99$, $H_{99,500}^{(1)} = 1.007$ and $H_{99,500}^{(2)} = 0.560$. In Figure 1 we show the Pareto quantile plot for both groups. On each quantile plot we superimposed the fitted lines passing through the anchor points $(\log(\frac{501}{100}), \log x_{401,500}^{(j)})$, $j=1, 2$, with respective slopes $H_{99,500}^{(j)}$, $j=1, 2$. The lines fit the linear part of the Pareto quantile plot quite well. In Figure 2 we plot the trace of the AMSE-matrix (full line) and the trace of the estimated AMSE-matrix (broken line) versus the number of extremes used in the estimation of the regression coefficients, k . Note that around the optimal k value $\text{tr } \Omega(k)$ is estimated quite well. The $\text{tr } \Omega(k)$ function reaches its minimum at $k = 79$ whereas for the estimate the minimum is reached at $k = 99$. For this simulated dataset, the observed value of the likelihood ratio test statistic to assess the validity of $H_0: \beta_1 = 0$ equals 16.834, leading to a rejection of H_0 .



(a)



(b)

Figure 1: Burr($1, \tau_j, 1$) simulation with $n_1 = n_2 = 500$, $\tau_1 = 1$ and $\tau_2 = 2$.

- (a) Pareto quantile plot for group 1 with line through $(\log(\frac{501}{100}), \log x_{401,500}^{(1)})$ and slope $H_{99,500}^{(1)} = 1.007$ superimposed;
- (b) Pareto quantile plot for group 2 with line through $(\log(\frac{501}{100}), \log x_{401,500}^{(2)})$ and slope $H_{99,500}^{(2)} = 0.560$ superimposed.

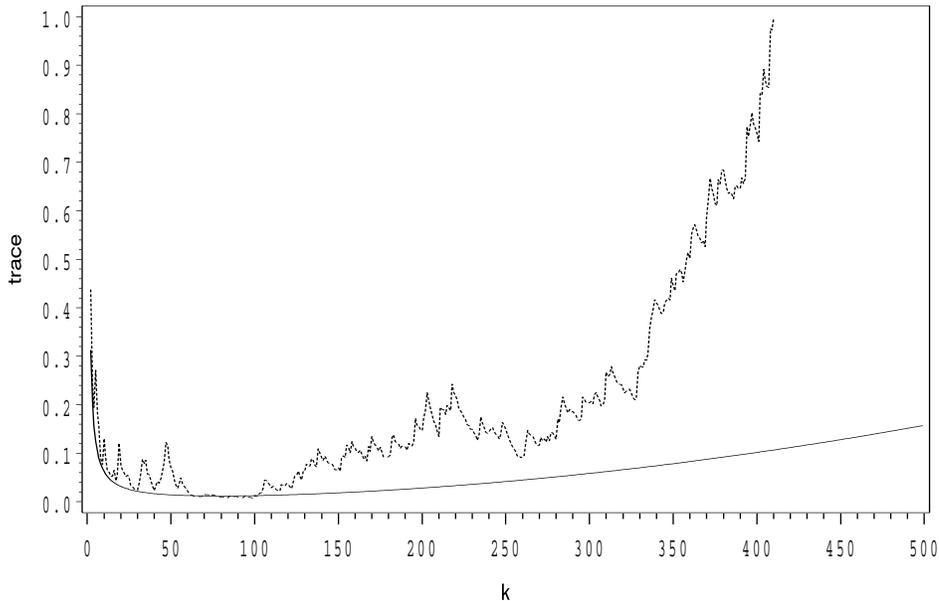


Figure 2: $\text{tr} \Omega(k)$ (full line) and $\text{tr} \hat{\Omega}(k)$ (broken line) vs k .

2.3. Simulation results and practical examples

2.3.1. Simulation results

We illustrate the small sample behaviour of the weighted least squares estimator $\hat{\beta}_{k_{\text{opt}}}$ using a simulation study. Datasets containing observations on 2, 3 and 4 groups were generated from $\text{Burr}(1, \tau_j, \lambda_j)$, $j = 1, \dots, G$, distributions. In Tables 1 and 2 we report the sample mean, sample standard deviation, empirical RMSE and the ratio (empirical RMSE under common optimal k)/(empirical RMSE under optimal k for each group separately) for samples of respectively 200 and 500 observations per group. The blocks $\lambda = 0.5$, $\lambda = 1$ and $\lambda = 2$ of both tables report the results in case of a common λ , and hence a common ρ , over the groups. The last λ -block of both tables reports results for the case $\lambda_j = 1/j$, $j = 1, \dots, G$, and hence $\rho_j = -j$, $j = 1, \dots, G$. Values for the τ -parameters were selected such that $\gamma_j = j$. From the ratio results it is clear that joint estimation of the extreme-value indices with a common k can lead to important gains in empirical MSE compared with a separate analysis. For instance in case $G=3$ and $\lambda=2$,

joint estimation of β_2 leads to a 30% gain in RMSE. Further, inspection of the first three λ -blocks of both tables indicates that the gains tend to increase with λ .

Table 1: Burr data, 200 observations/group, 500 simulation runs.

		$G = 2$		$G = 3$			$G = 4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	-0.5	2	-1	0	2.5	-1.5	-0.5	0.5
$\lambda = 0.5$	mean	1.5556	-0.5039	2.0909	-1.0209	0.0021	2.6197	-1.5434	-0.5131	0.5305
	sd	0.1930	0.1596	0.2281	0.2137	0.2373	0.2544	0.2443	0.2697	0.3567
	RMSE	0.2007	0.1595	0.2454	0.2146	0.2371	0.2809	0.2479	0.2697	0.3577
	ratio	0.9531	0.8323	1.0255	0.8837	0.8361	1.0097	0.8905	0.7855	0.8762
$\lambda = 1$	mean	1.6694	-0.5386	2.2485	-1.0960	-0.0106	2.8057	-1.6346	-0.5522	0.5450
	sd	0.2787	0.2257	0.3316	0.2706	0.3156	0.3906	0.3449	0.3778	0.4614
	RMSE	0.3259	0.2288	0.4141	0.2869	0.3154	0.4957	0.3699	0.3810	0.4632
	ratio	0.9936	0.8345	0.9957	0.8150	0.7776	1.0058	0.8428	0.7733	0.8475
$\lambda = 2$	mean	2.0410	-0.6390	2.6862	-1.2943	-0.0011	3.4326	-1.9971	-0.6735	0.6607
	sd	0.4398	0.3463	0.4951	0.3962	0.4137	0.5749	0.4783	0.5366	0.6491
	RMSE	0.6969	0.3729	0.8459	0.4932	0.4133	1.0952	0.6895	0.5634	0.6681
	ratio	0.9695	0.7929	0.9685	0.7994	0.7061	0.9729	0.8206	0.7830	0.7343
$\lambda_j = 1/j$	mean	1.6079	-0.4277	2.1199	-0.8654	-0.0088	2.6173	-1.3233	-0.4795	0.4147
	sd	0.2094	0.1807	0.2251	0.1931	0.2342	0.2203	0.2039	0.2548	0.3257
	RMSE	0.2354	0.1944	0.2548	0.2352	0.2342	0.2494	0.2696	0.2554	0.3363
	ratio	0.9840	0.8974	1.0892	1.0082	0.8329	1.1020	1.0170	0.8230	0.9015

Table 2: Burr data, 500 observations/group, 500 simulation runs.

		$G = 2$		$G = 3$			$G = 4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	-0.5	2	-1	0	2.5	-1.5	-0.5	0.5
$\lambda = 0.5$	mean	1.5443	-0.5044	2.0412	-1.0094	-0.0043	2.5617	-1.5185	-0.4996	0.5020
	sd	0.1380	0.1147	0.1769	0.1528	0.1868	0.1838	0.1775	0.2069	0.2613
	RMSE	0.1448	0.1147	0.1815	0.1529	0.1867	0.1937	0.1783	0.2067	0.2610
	ratio	1.0223	0.8638	1.1600	0.9581	0.9118	1.1233	0.9360	0.8576	0.8929
$\lambda = 1$	mean	1.5915	-0.5126	2.1547	-1.0720	0.0112	2.7099	-1.6101	-0.5293	0.5192
	sd	0.2361	0.1741	0.2474	0.2115	0.2329	0.2580	0.2405	0.2891	0.3657
	RMSE	0.2530	0.1744	0.2915	0.2232	0.2330	0.3324	0.2643	0.2903	0.3658
	ratio	1.0454	0.8382	0.9870	0.8447	0.7551	1.0001	0.8506	0.8173	0.8216
$\lambda = 2$	mean	1.8470	-0.5969	2.4778	-1.2054	0.0009	3.1297	-1.8401	-0.5969	0.5744
	sd	0.3471	0.2575	0.3909	0.3110	0.3424	0.4022	0.3442	0.3804	0.4884
	RMSE	0.4905	0.2749	0.6171	0.3724	0.3421	0.7470	0.4837	0.3922	0.4936
	ratio	1.0089	0.8265	0.9869	0.8235	0.7558	0.9995	0.8697	0.7440	0.7543
$\lambda_j = 1/j$	mean	1.5648	-0.4519	2.0722	-0.9158	-0.0330	2.5736	-1.3565	-0.5005	0.4668
	sd	0.1578	0.1391	0.1481	0.1367	0.1849	0.1523	0.1533	0.1912	0.2406
	RMSE	0.1704	0.1470	0.1647	0.1604	0.1877	0.1690	0.2099	0.1910	0.2427
	ratio	1.0395	0.9922	1.0576	0.9993	0.9823	1.1274	1.1048	0.8420	0.9789

Also, we performed a small sample simulation study to assess whether the likelihood ratio test for the hypothesis of no factor effects in the reduced model

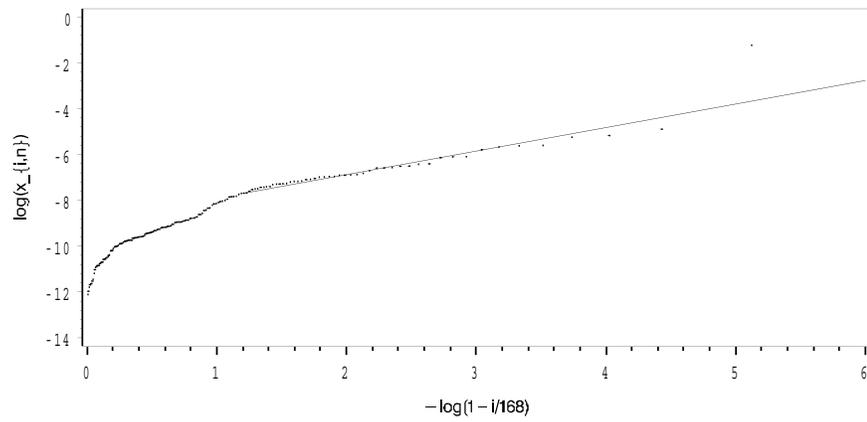
$j(\log X_{n_j-i+1, n_j}^{(j)} - \log X_{n_j-i, n_j}^{(j)}) \approx (\beta_0 + \beta_j)F_i^{(j)}$, $i=1, \dots, k$, $j=1, \dots, G$, satisfies the proposed significance level. Datasets containing 3 groups of 500 observations each were generated from Burr(1, τ_j , λ_j) distributions. Concerning the parameter λ_j (and hence ρ_j), 4 cases were considered: $\lambda_j = 0.5$, $\lambda_j = 1$, $\lambda_j = 2$ and $\lambda_j = 1/j$, $j=1, \dots, G$. The parameters τ_j , $j=1, \dots, G$, were selected such that all $\gamma_j = 1$ (simulation under H_0) and the significance level was set at $\alpha = 0.05$. Table 3 reports the empirical significance level for each setting of ρ for different k -values. As can be seen, the empirical significance levels are slightly below 0.05 for the cases with common ρ . In case the ρ parameter varies over the groups, the test performs only well at the smaller values of k , a result that could be expected.

Table 3: Burr data, 500 observations/group, 500 simulation runs: empirical significance levels.

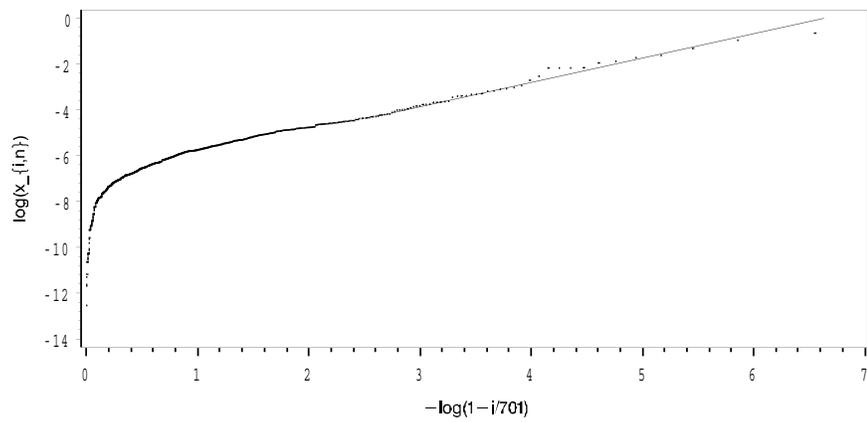
	$k = 10$	$k = 50$	$k = 100$	$k = 200$	$k = 400$
$\rho = -2$	0.034	0.040	0.044	0.044	0.018
$\rho = -1$	0.048	0.036	0.034	0.026	0.044
$\rho = -0.5$	0.040	0.030	0.030	0.020	0.058
$\rho_j = -j$	0.048	0.042	0.094	0.612	1.000

2.3.2. Practical example 1: fire claim data

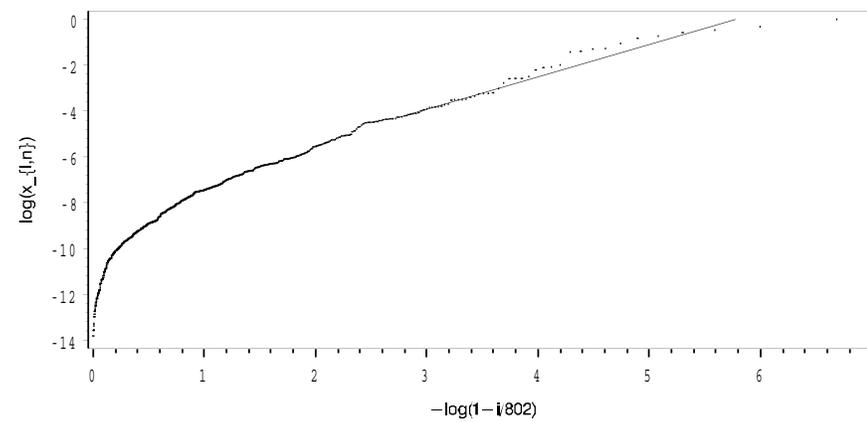
Our first example comes from an actuarial context. The reinsurance broker Aon Re Belgium provided claim data, generated by a fire insurance portfolio, for three types of buildings. The sample sizes are $n_1 = 167$, $n_2 = 700$ and $n_3 = 801$. Application of the proposed procedure results in $\hat{k}_{\text{opt}} = 50$, $H_{50,167}^{(1)} = 1.027$, $H_{50,700}^{(2)} = 1.064$ and $H_{50,801}^{(3)} = 1.413$. The Pareto quantile plots together with the lines passing through $(\log(\frac{n_j+1}{\hat{k}_{\text{opt}}+1}), \log x_{n_j-\hat{k}_{\text{opt}}, n_j}^{(j)})$ and slopes $H_{\hat{k}_{\text{opt}}, n_j}^{(j)}$, $j=1, 2, 3$, are given in Figure 3. As is clear from this figure, the Pareto quantile plots are almost linear in their extreme values indicating a reasonable fit of the Pareto distribution to the tails of the conditional claim size distributions. Concerning the γ estimate for group 3 (see also Figure 3 (c)) actuaries will find the estimate high. Remark however that other characteristics, such as the sum insured, can have an important influence on the tail index estimates but have been ignored in this analysis. Given an observed value for the likelihood ratio test statistic of 3.152, the null hypothesis of no group effects cannot be rejected.



(a)



(b)

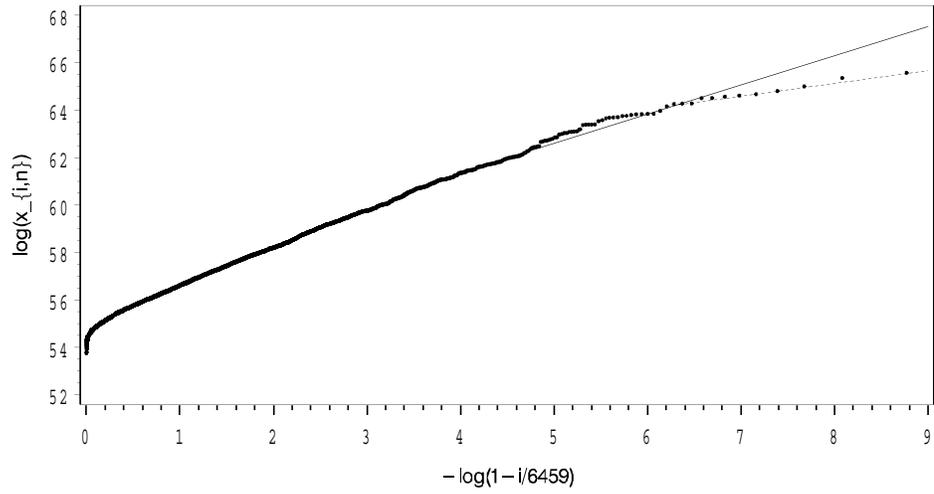


(c)

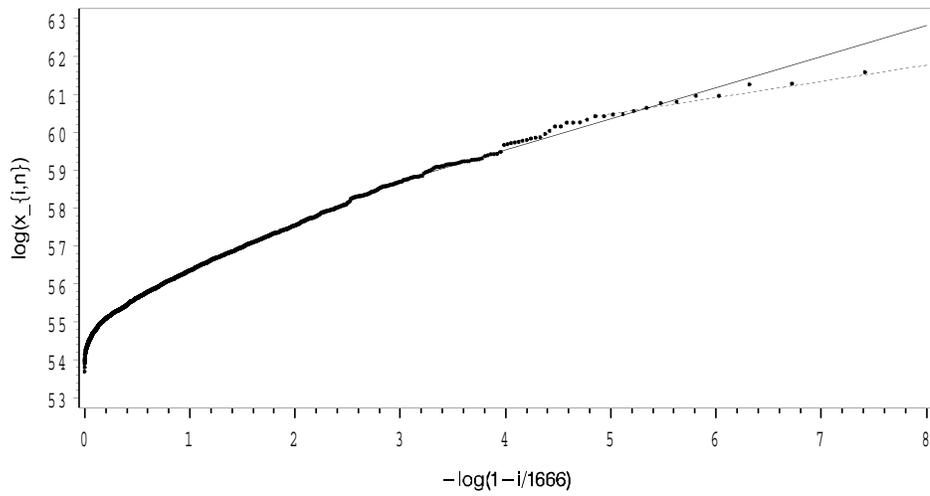
Figure 3: Claim data: Pareto quantile plots.

2.3.3. Practical example 2: earthquake data

As a second example we examined the earthquake data introduced in Pisarenko and Sornette (2001). This dataset is extracted from the Harvard catalog and contains information about the seismic moment (in dyne-cm) of shallow earthquakes (depth < 70 km) over the period 1977–2000. In Pisarenko and Sornette (2001), the tails of the seismic moment distributions for subduction and midocean ridge zones are compared by fitting the generalized Pareto distribution to seismic moment exceedances over 10^{24} dyne-cm. For these data $n_1 = 6458$ (subduction zones) and $n_2 = 1665$ (midocean ridge zones). The procedure described above with $k \geq 20$ yielded $\hat{k}_{\text{opt}} = 97$ with $H_{97,6458}^{(1)} = 1.232$ and $H_{97,1665}^{(2)} = 0.821$. In Figure 4 we show the Pareto quantile plots of the seismic moments for (a) subduction zones and (b) midocean ridge zones on which we superimposed the lines through $(\log(\frac{n_j+1}{\hat{k}_{\text{opt}}+1}), \log x_{n_j-\hat{k}_{\text{opt}},n_j}^{(j)})$ with slope $H_{\hat{k}_{\text{opt}},n_j}^{(j)}$, $j = 1, 2$ (solid lines). For the hypothesis test of no difference between the tail heaviness of the seismic moment distribution of subduction and midocean ridge zones a likelihood ratio statistic of 7.92 was obtained, resulting in a rejection of H_0 . The GPD based approach described in Pisarenko and Sornette (2001) yielded tail index estimates of 1.51 and 1.02 for subduction respectively midocean ridge zones, so our results are slightly more conservative. Likewise these authors found significant differences in the tail heaviness of the seismic moment distributions. Note that the Pareto quantile plots bend down at the largest observations indicating a weaker behaviour of the ultimate tail of the seismic moment distribution. Nevertheless, these largest observations form more or less a straight line pattern. So, also the ultimate tail could be described by a Pareto-type law. This fact is further illustrated in Figure 5 where we plot $\text{tr} \hat{\Omega}(k)$ as a function of $\log(k)$. Relaxation of the constraint that k should be at least 20 results in the global optimum $\hat{k}_{\text{opt}} = 12$ with $\hat{\gamma}_1 = 0.541$ and $\hat{\gamma}_2 = 0.427$. In Figure 4 the resulting optimal fits are plotted with dotted lines. At \hat{k}_{opt} the null hypothesis of no difference in tail behaviour cannot be rejected on basis of the above described likelihood ratio test statistic. Similarly to the results presented here, Pisarenko and Sornette (2001) also found deviations between the GPD and the ultimate tail of the seismic moment distribution. For plausible explanations of this phenomenon we refer to their paper and the references therein.



(a)



(b)

Figure 4: Earthquake data: Pareto quantile plots of seismic moments for
(a) subduction zones and
(b) midocean ridge zones.

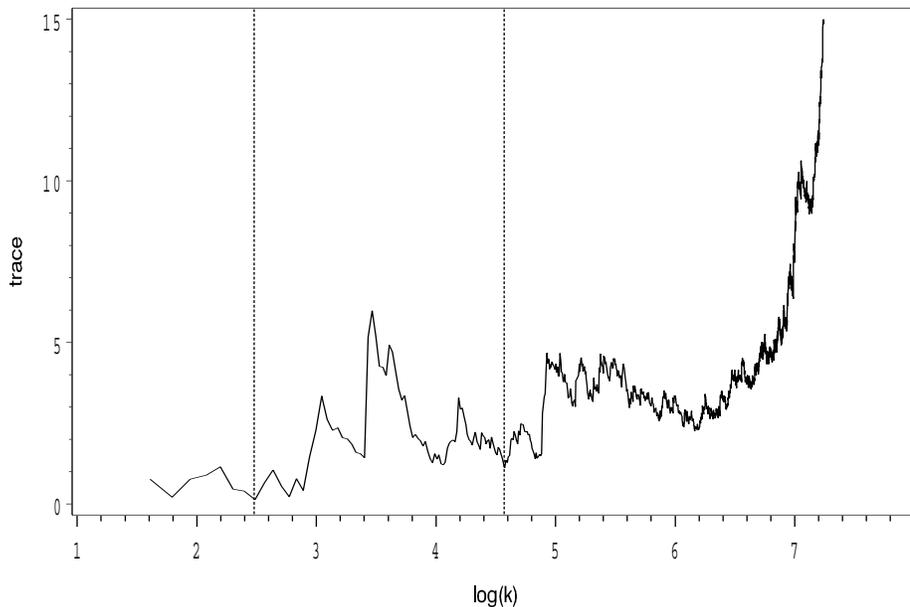


Figure 5: Earthquake data: $\text{tr } \hat{\Omega}(k)$ vs $\log(k)$.

3. LINEAR MODEL FORMULATION, $\gamma \in \mathbb{R}$

3.1. Description of the model

In this section we discuss the simultaneous estimation of several extreme-value indices in the general case $\gamma \in \mathbb{R}$. Consider again a sample of independent and identically distributed positive random variables $X_1^{(j)}, \dots, X_{n_j}^{(j)}$ according to some distribution function $F_{X^{(j)}}$, $j=1, \dots, G$, with G denoting the number of groups. Further, assume that the G groups are independent of each other and that $F_{X^{(j)}} \in \mathcal{D}(H_{\gamma_j})$ for some $\gamma_j \in \mathbb{R}$. In Theorem 1 of Beirlant *et al.* (1996) it is shown that $F_{X^{(j)}} \in \mathcal{D}(H_{\gamma_j})$ implies that

$$UH_{X^{(j)}}(x) = x^{\gamma_j} \check{l}_j(x)$$

with $UH_{X^{(j)}}$ and \check{l}_j denoting the UH function respectively the slowly varying function of group j .

Under the slow variation with remainder condition imposed on \check{l}_j , the following relation holds (see Beirlant *et al.*, 2000)

$$(3.1) \quad (i+1) \log \frac{UH_{i,n_j}^{(j)}}{UH_{i+1,n_j}^{(j)}} - (i+1) \log \frac{i+1}{i} + \frac{i+1}{i} = \gamma_j + g_j \left(\frac{n_j+1}{k+1} \right) \left(\frac{i+1}{k+1} \right)^{-\check{\rho}_j} + \varepsilon_i^{(j)}$$

$i=1, \dots, k, \quad j=1, \dots, G,$

where g_j is some generic notation for a function decreasing to zero for increasing values of the argument and

$$UH_{i,n_j}^{(j)} = X_{n_j-i,n_j}^{(j)} \left(\frac{1}{i} \sum_{m=1}^i \log X_{n_j-m+1,n_j}^{(j)} - \log X_{n_j-i,n_j}^{(j)} \right).$$

The residuals of model (3.1) have a mean approximately equal to zero and a covariance matrix given by

$$\mathbf{\Sigma}_{\varepsilon^{(j)}} = \left[\text{Cov}(\varepsilon_s^{(j)}, \varepsilon_t^{(j)}) \right]_{s,t} \sim \begin{cases} \frac{\gamma_j}{t} & s < t \\ (\gamma_j - 1)^2 + \frac{1+2s}{s^2} & s = t \end{cases} \quad j=1, \dots, G.$$

After introduction of the classical one-way ANOVA parametrization described above and deletion of the terms $g_j \left(\frac{n_j+1}{k+1} \right) \left(\frac{i+1}{k+1} \right)^{-\check{\rho}_j}$ in model (3.1) (these terms tend to 0 as $n_j \rightarrow \infty$ and $k/n_j \rightarrow 0$), the following estimators are obtained:

$$(3.2) \quad \tilde{\beta}_0 = \frac{1}{G} \sum_{j=1}^G \tilde{\gamma}_j \quad \text{and} \quad \tilde{\beta}_j = \tilde{\gamma}_j - \tilde{\beta}_0, \quad j=1, \dots, G,$$

with

$$(3.3) \quad \tilde{\gamma}_j = \frac{1}{k} \sum_{i=1}^k \left((i+1) \log \frac{UH_{i,n_j}^{(j)}}{UH_{i+1,n_j}^{(j)}} - (i+1) \log \frac{i+1}{i} + \frac{i+1}{i} \right).$$

Using least squares computations, the asymptotic covariance matrix of $\tilde{\beta}' = (\tilde{\beta}_0, \tilde{\beta}_1, \dots, \tilde{\beta}_{G-1})$ is given by:

$$(3.4) \quad \text{Acov}(\tilde{\beta}) = (\mathbf{L}'\mathbf{L})^{-1} \mathbf{L}' \mathbf{\Sigma} \mathbf{L} (\mathbf{L}'\mathbf{L})^{-1}$$

with $\mathbf{\Sigma} = \text{Block-diag}(\mathbf{\Sigma}_{\varepsilon^{(j)}}; j=1, \dots, G)$. For the main term of the bias of the estimators we have

$$\text{Abias}(\tilde{\beta}_0) = \frac{1}{G} \sum_{j=1}^G \frac{g_j \left(\frac{n_j+1}{k+1} \right)}{1 - \check{\rho}_j},$$

$$\text{Abias}(\tilde{\beta}_j) = \frac{g_j \left(\frac{n_j+1}{k+1} \right)}{1 - \check{\rho}_j} - \frac{1}{G} \sum_{l=1}^G \frac{g_l \left(\frac{n_l+1}{k+1} \right)}{1 - \check{\rho}_l}, \quad j=1, \dots, G-1.$$

Application of the estimators defined by (3.2) and (3.3) requires the selection of the number of UH statistics k to be used in the estimation. Again the asymptotic variance and bias are combined in an AMSE criterion. Hence, the optimal k -value is defined as

$$k_{\text{opt}} = \arg \min \text{tr} \left[(\mathbf{L}'\mathbf{L})^{-1} \mathbf{L}' \boldsymbol{\Sigma} \mathbf{L} (\mathbf{L}'\mathbf{L})^{-1} + \check{\boldsymbol{\kappa}} \check{\boldsymbol{\kappa}}' \right]$$

with $\check{\boldsymbol{\kappa}}$ the vector containing the asymptotic bias expressions given above.

For k/n_j sufficiently small, hypothesis tests about the regression coefficients can be based on the reduced model

$$(3.5) \quad (i+1) \log \frac{UH_{i,n_j}^{(j)}}{UH_{i+1,n_j}^{(j)}} - (i+1) \log \frac{i+1}{i} + \frac{i+1}{i} \approx \beta_0 + \beta_j + \varepsilon_i^{(j)}, \quad i = 1, \dots, k, \quad j = 1, \dots, G.$$

After transformation of model (3.5) by a matrix \mathbf{C} such that $\text{Cov}(\mathbf{C}\boldsymbol{\varepsilon}) = \mathbf{I}$ (see Beirlant *et al.*, 2000), where $\boldsymbol{\varepsilon}' = (\varepsilon_i^{(j)}; i=1, \dots, k, j=1, \dots, G)$, hypothesis about $\boldsymbol{\beta}$ can be tested using a classical F -test statistic.

3.2. Simulation results and a practical example

3.2.1. Simulation results

First, we apply the above proposed estimation procedure for $\gamma_j \in \mathbb{R}$, $j = 1, \dots, G$, to the simulated Burr datasets described in the previous section. Tables 4 and 5 contain the results for samples with 200 respectively 500 observations per group. Unlike the algorithm for $\gamma_j > 0$, $j = 1, \dots, G$, where the empirical MSE gains obtained from using a common k tend to increase with ρ , here the gains are quite stable with respect to the parameter $\check{\rho}$.

Next we examine the small sample properties of the proposed procedure using datasets with $\gamma_j < 0$, $j = 1, \dots, G$. Datasets containing observations on 2, 3 and 4 groups were generated from reversed Burr distributions. The reversed Burr distribution function, given by

$$1 - F_X(x) = \left(\frac{2}{1 + (1-x)^{-\tau}} \right)^\delta \quad 0 < x < 1; \quad \delta, \tau > 0,$$

belongs to the maximum domain of attraction of the GEV for some $\gamma < 0$, and hence the UH_X function can be written as in (1.9) with $\gamma = -1/(\delta\tau)$.

Table 4: Burr data, 200 observations/group, 500 simulation runs.

		$G = 2$		$G = 3$			$G = 4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	-0.5	2	-1	0	2.5	-1.5	-0.5	0.5
$\lambda = 0.5$	mean	1.5256	-0.5099	2.0626	-1.0214	-0.0032	2.5822	-1.5482	-0.5062	0.5205
	sd	0.1892	0.1640	0.2035	0.2177	0.2399	0.2199	0.2427	0.2820	0.3430
	RMSE	0.1907	0.1641	0.2128	0.2186	0.2397	0.2345	0.2472	0.2818	0.3433
	ratio	1.0344	0.8736	1.0406	0.9479	0.8351	1.0570	0.9699	0.8559	0.8567
$\lambda = 1$	mean	1.6101	-0.5434	2.1938	-1.0937	-0.0098	2.7627	-1.6551	-0.5598	0.5514
	sd	0.2761	0.2220	0.2981	0.2765	0.3298	0.3168	0.3290	0.3654	0.4256
	RMSE	0.2970	0.2260	0.3553	0.2917	0.3297	0.4113	0.3634	0.3699	0.4283
	ratio	1.0450	0.8703	1.0929	0.9609	0.8416	1.1019	0.9746	0.8044	0.8383
$\lambda = 2$	mean	1.9493	-0.6621	2.6153	-1.3241	0.0035	3.3430	-2.0366	-0.6866	0.6582
	sd	0.3754	0.3112	0.4224	0.3845	0.4239	0.5087	0.4792	0.4817	0.6217
	RMSE	0.5853	0.3506	0.7461	0.5026	0.4235	0.9843	0.7191	0.5161	0.6409
	ratio	1.0126	0.8442	1.0240	0.9237	0.7982	1.0679	0.9670	0.8137	0.7989
$\lambda_j = 1/j$	mean	1.5676	-0.4476	2.0855	-0.9109	0.0056	2.5848	-1.3936	-0.4717	0.4442
	sd	0.2193	0.1847	0.1951	0.1872	0.2328	0.1942	0.2151	0.2328	0.3021
	RMSE	0.2293	0.1918	0.2128	0.2072	0.2326	0.2118	0.2398	0.2343	0.3069
	ratio	1.0134	0.8453	1.0516	0.8570	0.8256	1.0663	0.8528	0.7954	0.8536

Table 5: Burr data, 500 observations/group, 500 simulation runs.

		$G = 2$		$G = 3$			$G = 4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	-0.5	2	-1	0	2.5	-1.5	-0.5	0.5
$\lambda = 0.5$	mean	1.5418	-0.5127	2.0358	-1.0206	0.0090	2.5558	-1.5364	-0.4929	0.4996
	sd	0.1224	0.1080	0.1477	0.1434	0.1580	0.1559	0.1647	0.1766	0.2327
	RMSE	0.1293	0.1087	0.1519	0.1447	0.1581	0.1655	0.1685	0.1765	0.2325
	ratio	0.9691	0.8252	1.0148	0.8856	0.7395	1.0485	0.8854	0.7377	0.8045
$\lambda = 1$	mean	1.5714	-0.5267	2.1576	-1.1004	0.0046	2.7033	-1.6413	-0.5303	0.5206
	sd	0.2143	0.1610	0.2171	0.2016	0.2072	0.2236	0.2291	0.2584	0.3106
	RMSE	0.2257	0.1630	0.2681	0.2250	0.2070	0.3020	0.2689	0.2599	0.3110
	ratio	1.0499	0.8362	1.0968	0.9123	0.7304	1.1378	0.9745	0.7719	0.7635
$\lambda = 2$	mean	1.8209	-0.6052	2.4664	-1.2380	0.0095	3.1143	-1.8831	-0.6011	0.5939
	sd	0.2977	0.2396	0.3398	0.3064	0.2972	0.3628	0.3374	0.3519	0.4398
	RMSE	0.4375	0.2615	0.5769	0.3878	0.2971	0.7132	0.5103	0.3658	0.4493
	ratio	1.0465	0.8265	1.0787	0.8909	0.7386	1.1111	0.9947	0.7491	0.7358
$\lambda_j = 1/j$	mean	1.5557	-0.4606	2.0663	-0.9269	-0.0044	2.5786	-1.4014	-0.4880	0.4672
	sd	0.1504	0.1255	0.1336	0.1346	0.1623	0.1226	0.1346	0.1659	0.2057
	RMSE	0.1603	0.1314	0.1490	0.1531	0.1622	0.1455	0.1668	0.1662	0.2081
	ratio	1.0041	0.7875	1.0173	0.8254	0.7941	1.1045	0.8256	0.7483	0.8470

Also, the \check{l} function associated with the UH_X function satisfies the slow variation with remainder condition with $\check{\rho} = -\min(\frac{1}{\delta}, \frac{1}{\delta\tau})$. In this simulation a common δ value was used for all groups. Further, values for the τ parameters were selected such that $\gamma_j = -j$, $j = 1, \dots, G$. Two cases were considered: $\delta = 1$ and $\delta = 2$ giving $\check{\rho} = -1$ and $\check{\rho} = -0.5$. Table 6 summarizes the simulation results. As expected,

the estimators are more biased as the $\check{\rho}$ -parameter increases. The gains obtained by using a common k -value compared to a separate analysis of each group are quite stable with respect to $\check{\rho}$.

Table 6: Reversed Burr data, 500 observations/group, 500 simulation runs.

		$G = 2$		$G = 3$			$G = 4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		-1.5	0.5	-2	1	0	-2.5	1.5	0.5	-0.5
$\delta = 1$	mean	-1.6382	0.5689	-2.1603	1.1005	-0.0094	-2.6820	1.6190	0.5077	-0.5401
	sd	0.1310	0.1124	0.1696	0.1510	0.1445	0.1861	0.1718	0.1832	0.2433
	RMSE	0.1903	0.1317	0.2332	0.1813	0.1447	0.2602	0.2089	0.1832	0.2464
	ratio	0.9650	0.9516	1.0028	1.0213	0.7320	1.0149	1.0585	0.8410	0.8030
$\delta = 2$	mean	-1.8888	0.6507	-2.4375	1.2155	-0.0080	-2.9633	1.7470	0.5565	-0.5651
	sd	0.1646	0.1301	0.2234	0.1969	0.1785	0.2716	0.2466	0.2353	0.2821
	RMSE	0.4221	0.1990	0.4911	0.2918	0.1785	0.5369	0.3488	0.2418	0.2892
	ratio	0.9628	0.9593	0.9294	0.9865	0.7342	0.9667	1.0708	0.7785	0.7152

Finally, the procedure was applied to datasets containing groups for which the γ_j , $j=1, \dots, G$, can have a different sign and/or be equal to zero. Here, datasets containing observations on 2 and 3 groups were generated from the generalized Pareto distribution with distribution function given by (1.3). The slowly varying function \check{l} of the GPD satisfies the slow variation with remainder condition with $\check{\rho} = -|\gamma|$. In this simulation, we took $\sigma = 1$. Table 7 contains the results for these problem sets. Also here we see that using a common optimal k can yield important efficiency gains compared to a separate analysis of each group, except in case β_0, β_1 are both negative.

Table 7: GPD data, 500 observations/group, 500 simulation runs.

		$G = 2$		$G = 2$		$G = 2$		$G = 3$		
		β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_2
value		-0.25	-0.25	0	-0.5	0.25	-0.25	0	-0.5	0
mean		-0.1582	-0.2591	0.0970	-0.5046	0.3442	-0.2518	0.0977	-0.5063	-0.0024
sd		0.0749	0.0734	0.0867	0.0800	0.1031	0.0819	0.0725	0.0789	0.0865
RMSE		0.1184	0.0739	0.1300	0.0800	0.1396	0.0819	0.1216	0.0791	0.0865
ratio		1.0551	0.9719	1.0204	0.8498	1.0317	0.8839	1.1251	0.8635	0.8507

3.2.2. Practical example 3: US wind speed data

The wind speed database, provided by the National Institute of Standards and Technology (NIST), contains information about 49 weather stations in the U.S.. The data have been filed for a period of 15 to 26 years. They are the daily fastest-mile wind speeds, measured by anemometers 10 m above ground. For more information about these data we refer to Simiu *et al.* (1979) and Simiu and Heckert (1995). We restrict our attention to three cities: Des Moines (Iowa), Grand Rapids (Michigan) and Albuquerque (New Mexico). Boxplots of the daily fastest wind speeds (in miles per hour) are given in Figure 6. The generalized quantile plot for each city is given in Figure 7. These plots allow to distinguish between the wind speed tail behavior of the different cities: the Des Moines data (Figure 7(a)) are heavy tailed ($\gamma > 0$), the Grand Rapids data (Figure 7(b)) seem to be moderately tailed with $\gamma \approx 0$ and the Albuquerque data (Figure 7(c)) are weakly tailed ($\gamma < 0$). The line structures in these plots are the result of an inherent grouping of the data due to a loss of accuracy during the data collection process. Consequently many wind speed levels are registered more than once. Application of the above described procedure resulted in $\hat{k}_{\text{opt}} = 357$ with $\tilde{\gamma}_1 = 0.144$, $\tilde{\gamma}_2 = 0.053$ and $\tilde{\gamma}_3 = -0.088$. On each generalized quantile plot we superimposed the line passing through the anchor point $(\log(\frac{n_j}{k+1}), \log UH_{k+1, n_j}^{(j)})$ with slope $\tilde{\gamma}_j$.

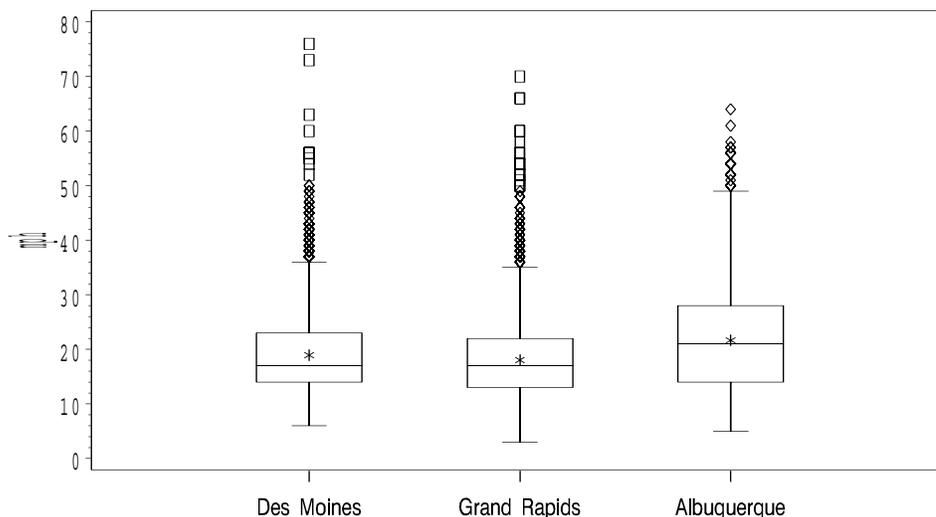


Figure 6: Wind speed data: Boxplots of the daily fastest wind speeds in miles per hour.

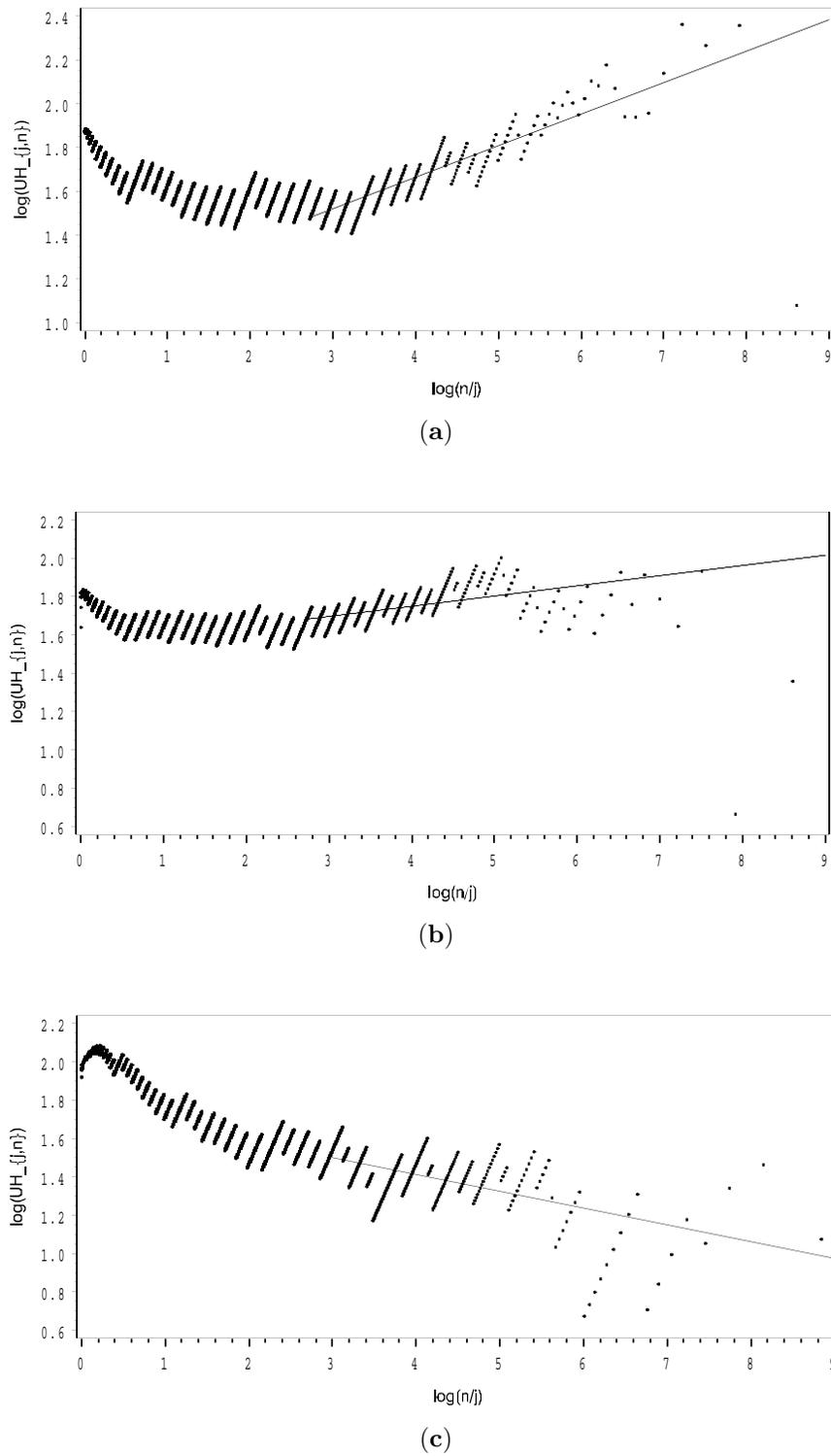


Figure 7: Wind speed data: generalized quantile plots for (a) Des Moines, (b) Grand Rapids and (c) Albuquerque.

4. CONCLUSION

In this paper we discussed the simultaneous estimation of tail indices when data on several independent groups are available. The proposed methods are based on regression models linking statistics related to the tail of the underlying distribution function to the extreme-value index and parameters describing the second order tail behaviour. The optimal number of extremes (in case $\gamma > 0$) or *UH* statistics (in case $\gamma \in \mathbb{R}$) was derived from the trace of the AMSE matrix. It appears from the simulation results that combining data from several groups can lead to significant improvements in the estimation of the extreme-value index. A drawback of using a common k -value is that the procedure can run into difficulties when the design is severely unbalanced. However, this problem is easily remedied by using a common relative tail sample fraction. Future work will concentrate on the further extension of the proposed methods towards the estimation of other tail characteristics such as extreme quantiles or small exceedance probabilities.

ACKNOWLEDGMENTS

The authors would like to thank R. Verlaak and P. Vynckier (Aon Re Belgium) for providing the claim data and V.F. Pisarenko and D. Sornette for providing the earthquake data. The authors also take pleasure in thanking the referee for providing useful and interesting comments.

REFERENCES

- [1] BEIRLANT, J.; DIERCKX, G.; GOEGBEUR, Y. and MATTHYS, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.
- [2] BEIRLANT, J.; DIERCKX, G. and GUILLOU, A. (2000). *Estimation of the extreme value index and regression on generalized quantile plots*, Technical Report, Department of Mathematics, K.U. Leuven.
- [3] BEIRLANT, J.; DIERCKX, G.; GUILLOU, A. and STĂRICĂ, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.
- [4] BEIRLANT, J.; VYNCKIER, P. and TEUGELS, J.L. (1996). Excess functions and estimation of the extreme-value index, *Bernoulli*, **2**, 293–318.

- [5] BINGHAM, N.H.; GOLDIE, C.M. and TEUGELS, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.
- [6] BURR, I.W. (1942). Cumulative frequency functions, *Annals of Mathematical Statistics*, **13**, 215–232.
- [7] COLES, S. and TAWN, J. (1998). *Statistical Methods for Extreme Values*, A course presented at the 1998 RSS conference, Strathclyde, September, 1998.
- [8] CSÖRGŐ, S.; DEHEUVELS, P. and MASON, D.M. (1985). Kernel estimates of the tail index of a distribution, *Annals of Statistics*, **13**, 1050–1077.
- [9] DAVISON, A.C. and SMITH, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society B*, **52**, 393–442.
- [10] DEKKERS, A.L.M.; EINMAHL, J.H.J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**, 1833–1855.
- [11] FEUERVERGER, A. and HALL, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Annals of Statistics*, **27**, 760–781.
- [12] FISHER, R.A. and TIPPETT, L.H.C. (1928). Limiting forms of the frequency distribution in the largest particle size and smallest number of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.
- [13] GUMBEL, E.J. (1967). *Statistics of Extremes*, Columbia University Press.
- [14] HALL, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society B*, **44**, 37–42.
- [15] HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.
- [16] KRATZ, M. and RESNICK, S. (1996). The qq-estimator of the index of regular variation, *Communications in Statistics: Stochastic Models*, **12**, 699–724.
- [17] PICKANDS III, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.
- [18] PISARENKO, V.F. and SORNETTE, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution, *Pure and Applied Geophysics*, **160**, 2343–2364.
- [19] SCHULTZE, J. and STEINEBACH, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.
- [20] SIMIU, E.; CHANGERY, M.J. and FILLIBEN, J.J. (1979). *Extreme Wind Speeds at 129 Stations in the Contiguous United States*, NBS Building Science Series 118, National Bureau of Standards (U.S.), Washington, DC.
- [21] SIMIU, E. and HECKERT, N.A. (1995). *Extreme wind distribution tails: a peak over threshold approach*, NIST, Gaithersburg.
- [22] SMITH, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.
- [23] SMITH, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.