Bernoulli **11**(6), 2005, 949–970

Estimation of the extreme-value index and generalized quantile plots

J. BEIRLANT[1], G. DIERCKX[1] and A. GUILLOU[2]

[1]*Katholieke Universiteit Leuven, Department of Mathematics, Celestijnenlaan 200B, 3001 Leuven, Belgium*
[2]*Université Paris VI, LSTA, Boîte 158, 4 Place Jussieu, 75252 Paris Cedex 05, France*

In extreme-value analysis, a central topic is the adaptive estimation of the extreme-value index γ. Hitherto, most of the attention in this area has been devoted to the case $\gamma > 0$, that is, when $\bar{F}$ is a regularly varying function with index $-1/\gamma$. In addition to the well-known Hill estimator, many other estimators are currently available. Among the most important are the kernel-type estimators and the weighted least-squares slope estimators based on the Pareto quantile plot or the Zipf plot, as reviewed by Csörgő and Viharos. Using an exponential regression model (ERM) for spacings between successive extreme order statistics, both Beirlant *et al*. and Feuerverger and Hall introduced bias-reduced estimators.

For the case where γ is real, Hill's estimator has been generalized to a moment-type estimator by Dekkers *et al*. Alternatively, Beirlant *et al*. introduced a Hill-type estimator that is based on the generalized quantile plot. Another popular estimation method follows from maximum likelihood estimation applied to the generalizations of the Pareto distribution. In the present paper, slope estimators for $\gamma > 0$ are generalized to the case where γ is real-valued. This is accomplished by replacing the Zipf plot by a generalized quantile plot. We make an asymptotic comparison of our estimator with the moment estimator and with the maximum likelihood estimator. A case study illustrates our findings. Finally, we offer a regression model that generalizes the ERM in that it allows the construction of bias-reduced estimators. Moreover, the model provides an adaptive selection rule for the number of extremes needed in several of the existing estimators.

Keywords: bias; extreme-value index; least squares; mean squared error; quantile plots

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with distribution function F and tail quantile function $U(x) = \inf\{y;\, F(y) \geqslant 1 - 1/x\}$. We denote the order statistics by $X_{1,n} \leqslant \ldots \leqslant X_{n,n}$. In this paper, the statistical model is given by the maximum domain of attraction condition that governs extreme-value theory: suppose that there exist some $\gamma \in \mathbb{R}$ and sequences of constants $(a_n;\, a_n > 0)$ and (b_n), such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leqslant x\right) = G_\gamma(x) \text{ for all } x, \tag{1}$$

with $G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma})$.

The main aim of this paper is to discuss the problem of estimating the extreme-value index γ under this model. This tail index is known to be the crucial indicator for the decay of the tail of the distribution: distributions with finite endpoint have $\gamma < 0$, exponentially decreasing tails occur when $\gamma = 0$, while $\gamma > 0$ leads to polynomially decreasing Pareto-type tails. The extreme-value index γ should not be confused with the extremal index from stationary time series, which characterizes the change in the distribution of the sample maxima due to dependence in the series.

Most research in extreme-value theory concentrates on the heavy-tailed distributions where $\gamma > 0$. An excellent overview of the relevant literature can be found in Csörgő and Viharos (1998). When γ is strictly positive, it follows from (1) that X is a Pareto-type variable, that is,

$$\frac{\bar{F}(tx)}{\bar{F}(x)} \to t^{-1/\gamma} \qquad \text{as } x \to \infty, \text{ for all } t > 0.$$

The latter condition is equivalent to the regular variation of the tail function $U(x) = Q(1 - 1/x)$ with Q the quantile function of F:

$$U(x) = x^{\gamma} L(x), \tag{2}$$

where L is a slowly varying function, that is, L satisfies the condition $L(tx)/L(x) \to 1$ as $x \to \infty$ for all $t > 0$. For this regular variation model, Hill (1975) first proposed the estimator

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}$$

of γ. For theoretical asymptotic reasons, we consider intermediate sequences $k = k_n$ of positive integers ($1 \leqslant k < n$) satisfying

$$k \to \infty, \qquad \frac{k}{n} \to 0 \quad \text{as } n \to \infty.$$

If L is constant, that is, X has a Pareto distribution, a Pareto quantile plot or a Zipf plot

$$\left(\log\left(\frac{n+1}{j}\right), \log X_{n-j+1,n}\right), \qquad j = 1, \ldots, n, \tag{3}$$

is nearly linear, with its slope approximately equal to γ. If L is not constant, the Pareto quantile plot exhibits this feature only for smaller values of j. Hence, for $\gamma > 0$, a wide variety of estimators of γ emerges from different regression fits to Pareto quantile plots. Beirlant *et al*. (1996a) pointed out that a constrained weighted least-squares fit to an upper part of the Pareto quantile plot (3) leads to the class of kernel estimators

$$\hat{\gamma}_{k,n}^{+,K} = \frac{\sum_{j=1}^{k} jk^{-1}K(jk^{-1})[\log X_{n-j+1,n} - \log X_{n-j,n}]}{k^{-1}\sum_{j=1}^{k} K(jk^{-1})}$$

with a kernel K integrating to 1. For example, Hill's estimator is obtained for $K = I_{(0,1]}$. Kratz and Resnick (1996) and Schultze and Steinebach (1996) introduced the Zipf estimator, an unconstrained least-squares estimator based on (3):

$$\hat{\gamma}^{+,Z}_{k,n} = \frac{\sum_{j=1}^{k} \log(j^{-1}(k+1)) \log X_{n-j+1,n} - k^{-1} \sum_{j=1}^{k} \log(j^{-1}(k+1)) \sum_{j=1}^{k} \log X_{n-j+1,n}}{\sum_{j=1}^{k} \log^2(j^{-1}(k+1)) - k^{-1} \left(\sum_{j=1}^{k} \log(j^{-1}(k+1)) \right)^2}.$$

Due to the asymptotic nature of the definition of the Pareto-type model, any estimator of γ will contain quantities whose selection plays a crucial role in the successful application of the estimator. Adaptive selection rules for the number of extremes k to be used in the estimation procedures have been proposed in Beirlant *et al*. (1996a), Resnick and Stărică (1997), Drees and Kaufmann (1998), Drees *et al*. (2000) and Danielsson *et al*. (2001). The choice of k is important: Hill plots $\{(k, H_{k,n}) : 1 \leqslant k < n\}$ often exhibit strong trends, so guidelines for the choice of k are most helpful. Moreover, the lack of smoothness of these plots results in different estimates for neighbouring values of k. The minimization of the mean squared error of the estimator has been a paradigm in most publications: due to the asymptotic nature of the nuisance part of the model, the bias diminishes with decreasing k, while the variance decreases with increasing k.

However, next to the choice of k, another crucial problem is the appearance of substantial bias. In the Hall (1982) model given by

$$L(x) = M_1(1 + M_2 x^{-\beta}\{1 + o(1)\}), \tag{4}$$

this problem appears when $\beta > 0$ is small. In this model $M_1 > 0$ and $M_2 \in \mathbb{R}$. In Beirlant *et al*. (1999) and Feuerverger and Hall (1999), the introduction of such a second-order slow-variation condition leads to nonlinear regression fits on the upper part of a Pareto quantile plot. With this approach, the bias can be reduced. Moreover, an adaptive estimation procedure for the proper choice of k can be constructed. See Beirlant *et al*. (2002).

The estimation of the more general case where $\gamma \in \mathbb{R}$ has been studied less extensively. There are two main classes of solutions that result from different formulations of the model and that are equivalent to (1).

The first is the peaks over threshold method (see, for instance, Smith 1987; 1989; Davison and Smith 1990). This method is based on results given by Balkema and de Haan (1974) and Pickands (1975) that state that the limit distribution of the exceedances over a threshold u is a generalized Pareto distribution when $u \to \infty$. The fit of the generalized Pareto distribution over a high threshold can, therefore, be performed by a number of alternative procedures: maximum likelihood (Smith 1987), probability-weighted moments (Hosking *et al*. 1985), Bayesian analysis methods (see Coles and Powell 1996), or a percentile method given by Castillo and Hadi (1997). Most of these estimating methods are only valid if additional restrictions are placed on the value of γ.

The second procedure is based on the use of k upper order statistics. The method is motivated by the following asymptotic relation which is equivalent to (1): there exists a positive function a such that, for all $t > 0$,

$$\lim_{x\to\infty} \frac{U(tx) - U(x)}{a(x)} = \begin{cases} \log t & \gamma = 0, \\ \gamma^{-1}(t^{\gamma} - 1) & \gamma \neq 0. \end{cases} \tag{5}$$

Assuming $U(\infty) > 0$, condition (5) implies (see, for instance, Dekkers *et al*. 1989)

$$\lim_{x\to\infty}\frac{\log U(tx)-\log U(x)}{a(x)/U(x)}=\begin{cases}\log t & \gamma\geqslant 0,\\ \gamma^{-1}(t^{\gamma}-1) & \gamma<0.\end{cases} \tag{6}$$

Based on (6), Dekkers *et al*. (1989) proposed the *moment estimator* that can be considered to be an adaptation of the Hill estimator to the case where $\gamma\in\mathbb{R}$:

$$\hat{\gamma}_k^{\mathrm{M}}:=\hat{\gamma}_{k,n}^{\mathrm{M}}=H_{k,n}+1-\frac{1}{2}\left(1-\frac{H_{k,n}^2}{S_{k,n}}\right)^{-1},$$

where in turn

$$S_{k,n}=\frac{1}{k}\sum_{j=1}^{k}(\log X_{n-j+1,n}-\log X_{n-k,n})^2.$$

Also based on (6), Beirlant *et al*. (1996b) proposed an estimator of $\gamma\in\mathbb{R}$ using the slope on a generalized quantile plot. This method puts the Pareto quantile plot in a more general setting. To construct such a plot, observe that, under (6), UH is regularly varying at infinity with index γ where $H(x)=\mathbb{E}[\log X-\log U(x)|X>U(x)]$, i.e.

$$UH(x)=U(x)H(x)=x^{\gamma}L(x), \tag{7}$$

with L again a function slowly varying at infinity. Introduce

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

as an empirical substitute for $UH((n+1)/j)$. It can then be seen that for small j the generalized quantile plot

$$\left(\log\left(\frac{n+1}{j}\right),\ \log UH_{j,n}\right),\qquad j=1,\ldots,n, \tag{8}$$

becomes ultimately linear.

For $\gamma>0$, one can construct regression-based estimators of $\gamma\in\mathbb{R}$. Among them, the generalized Hill estimator

$$\hat{\gamma}_{k,n}^{\mathrm{H}}=\frac{1}{k}\sum_{j=1}^{k}\log UH_{j,n}-\log UH_{k+1,n},$$

is the simplest. Formal replacement of $X_{n-j+1,n}$ by $UH_{j,n}$ leads to kernel estimators $\hat{\gamma}_{k,n}^{K}$ that generalize the Pareto index estimators $\hat{\gamma}_{k,n}^{+,K}$ as shown in Beirlant *et al*. (1996a). Different generalizations of the kernel estimators $\hat{\gamma}_{k,n}^{+,K}$ can be found in Groeneboom *et al*. (2003).

The estimators $\hat{\gamma}_{k,n}^{\mathrm{M}}$, $\hat{\gamma}_{k,n}^{\mathrm{H}}$ as well as the kernel estimators are all based on the logarithms of the observed data, hence they are not shift-invariant. In an attempt to apply these estimators to negative observations, one needs to shift the observations to positive values. But this operation has its influence on the estimates. Furthermore, taking logarithms often introduces further bias. As shown by Drees (1998), this new bias might even dominate the bias of shift-invariant estimators. We will take this up in Section 2.

In Section 2 we study the generalized unconstrained least-squares estimator for the case where $\gamma \in \mathbb{R}$:

$$\hat{\gamma}^{Z}_{k,n} = \frac{\sum_{j=1}^{k} \log(j^{-1}(k+1)) \log UH_{j,n} - k^{-1} \sum_{j=1}^{k} \log(j^{-1}(k+1)) \sum_{j=1}^{k} \log UH_{j,n}}{\sum_{j=1}^{k} \log^2(j^{-1}(k+1)) - k^{-1} \left(\sum_{j=1}^{k} \log(j^{-1}(k+1)) \right)^2}.$$

Because

$$\frac{1}{k} \sum_{j=1}^{k} \log^2 \frac{k+1}{j} - \left(\frac{1}{k} \sum_{j=1}^{k} \log \frac{k+1}{j} \right)^2 \sim 1 \qquad \text{as } k \to \infty \text{ and } \frac{k}{n} \to 0,$$

the above estimator can be approximated by

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

Remark 1. Following Csörgő and Viharos (1998), we can generalize $\hat{\gamma}^{Z}_{k,n}$ to a class of weighted estimators

$$\frac{\sum_{j=1}^{k} [\int_{j-1}^{j/k} J(s)\mathrm{d}s] \log UH_{j,n}}{\sum_{j=1}^{k} [\int_{j-1}^{j/k} J(s)\mathrm{d}s] \log(k/j)}$$

with a weight function J integrating to 0, such as $J_\theta(s) = ((1+\theta)/\theta)(1-(1+\theta)s^\theta)$, $s \in [0, 1]$ with $\theta > 0$. The estimator $\hat{\gamma}^{Z}_{k,n}$ is then retrieved for $\theta \downarrow 0$, leading to the Zipf weight function $J_\theta(s) = -\log(s) - 1$. This potential extension will not be pursued further.

Remark 2. The new estimator $\hat{\gamma}^{Z}_{k,n}$ can, of course, be used for the estimation of extreme tail probabilities and extreme quantiles. Using the general technique presented, for instance, in Dekkers *et al*. (1989), we obtain:

$$\hat{U}(1/p) = \hat{Q}(1-p) = X_{n-k,n} + \hat{a}(n/k) \frac{(k/np)^{\hat{\gamma}^{Z}_{k,n}} - 1}{\hat{\gamma}^{Z}_{k,n}}, \qquad \text{for some } p \in (0, 1/n],$$

and

$$\hat{\bar{F}}(x) = \frac{k}{n} \max\left\{ 0, \left(1 + \hat{\gamma}^{Z}_{k,n} \frac{x - X_{n-k,n}}{\hat{a}(n/k)} \right)^{-1/\hat{\gamma}^{Z}_{k,n}} \right\}, \qquad \text{for some } x \geqslant X_{n-k,n},$$

where

$$\hat{a}(n/k) = X_{n-k,n} H_{k,n} \max(1 - \hat{\gamma}^{Z}_{k,n}, 1).$$

With the help of the asymptotic results obtained in Section 2, asymptotic results for the above estimators can be developed by analogy with the results given, for instance, in de Haan and Rootzén (1993), Ferreira *et al*. (2003) and Ferreira (2002).

In the next section, we give a detailed theoretical asymptotic comparison of the estimators, together with a practical example from insurance. The approach in Beirlant *et al*. (1999) and Feuerverger and Hall (1999) is then extended to the case of a real-valued γ. Starting from the generalized quantile plot, we investigate the induced regression problem in more detail. In Section 3 we derive the regression model that leads to bias-reduced estimators of γ. Furthermore, estimates for the bias of the four estimators considered in Section 2 are derived. The latter are obtained through estimation of the parameters of the regression model. A major result of this is a diagnostic selection procedure for the number of extremes k needed in the estimators. Proofs and technical results are deferred to Appendix B.

2. Asymptotic results and comparisons

In this section, we derive the basic asymptotic results for the estimators $\hat{\gamma}^{\mathrm{H}}_{k,n}$ and $\hat{\gamma}^{\mathrm{Z}}_{k,n}$ that are based on the generalized quantile plot (8). We discuss the asymptotic bias in detail and compare the asymptotic mean squared errors of these estimators with those of the maximum likelihood and moment estimators at their respective asymptotic optimal k-values.

To control the asymptotic bias resulting from the slowly varying parts of the models, one needs a second-order condition on the tail quantile function U. From the theory of generalized regular variation of second order outlined in de Haan and Stadtmüller (1996), one assumes the existence of a positive function a and a second ultimately positive auxiliary function a_2 with $a_2(x) \to 0$ when $x \to \infty$, such that the limit

$$\lim_{x\to\infty} \frac{1}{a_2(x)} \left\{ \frac{U(ux) - U(x)}{a(x)} - h_\gamma(u) \right\} = k(u) \tag{9}$$

exists on $(0, \infty)$.

It follows that there exists a real constant c and a value $\rho < 0$ for which the auxiliary function a satisfies

$$\lim_{x\to\infty} a_2^{-1}(x) \left\{ \frac{a(ux)}{a(x)} - u^\gamma \right\} = cu^\gamma h_\rho(u), \tag{10}$$

with $h_\rho(u) = \int_1^u z^{\rho-1} \mathrm{d}z$. The function k that appears in (9) admits the representation

$$k(u) = c \int_1^u t^{\gamma-1} h_\rho(t) \mathrm{d}t + A h_{\gamma+\rho}(u), \tag{11}$$

where $A \in \mathbb{R}$. We denote the class of generalized second-order regularly varying functions U (satisfying (9)–(11)) by $\mathrm{GRV}_2(\gamma, \rho; a(x), a_2(x); c, A)$.

We restrict ourselves to the case where $\rho < 0$. In this case, a clever choice of the auxiliary function a_2 results in a simplification of the limit function k when $c = 0$.

In Appendix A, we give an overview of possible forms of GRV_2 functions and the corresponding representations for U and $\log U$ as given in Vanroelen (2003). From this list it follows that the second-order rate for $\log U$ in (9) is worse than for U when $\rho < \gamma < 0$ and in some cases when $0 < \gamma < -\rho$. In such cases, the asymptotic relative efficiency for estimators based on log-transformed data compared to shift-invariant estimators (such as the maximum likelihood estimator) reduces to 0, the proviso that all estimators are based on the optimal number of order statistics.

When $0 < \gamma < -\rho$, the above rate problem for $\log U$ is due to the appearance of the constant D in the characterization of U for that case. Indeed, U is then given by

$$U(x) = \ell_+ x^\gamma \left\{ \frac{1}{\gamma} + Dx^{-\gamma} + \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \right\}.$$

When $D = 0$, the original a_2-rate is kept for $\log U$. This is not so when $D \neq 0$, in which case a_2 is replaced by a regularly varying function with index $-\gamma$. Within the Hall class (4) of Pareto-type distributions, the case $D \neq 0$ occurs when $\beta = \gamma$. Examples are the Fisher F and the generalized extreme-value distributions.

In the statement of our results, we use the following notation:

$$b(x) = \begin{cases} \dfrac{A\rho[\rho+\gamma(1-\rho)]}{(\gamma+\rho)(1-\rho)} a_2(x) & \text{if } 0 < -\rho < \gamma \text{ or if } 0 < \gamma < -\rho \text{ with } D = 0, \\ -\dfrac{\gamma^3}{(1+\gamma)} x^{-\gamma} L_2(x) & \text{if } \gamma = -\rho, \\ -\dfrac{\gamma^3 D}{(1+\gamma)} x^{-\gamma} & \text{if } 0 < \gamma < -\rho \text{ with } D \neq 0, \\ \dfrac{1}{\log^2(x)} & \text{if } \gamma = 0, \\ \dfrac{A\rho(1-\gamma)}{(1-\gamma-\rho)} a_2(x) & \text{if } \gamma < \rho, \\ -\dfrac{\gamma}{1-2\gamma} \dfrac{\ell_+}{U(\infty)} x^\gamma & \text{if } \rho < \gamma < 0, \\ \dfrac{\gamma}{1-2\gamma}\left[A(1-\gamma) - \dfrac{\ell_+}{U(\infty)}\right] x^\gamma & \text{if } \gamma = \rho, \end{cases}$$

and

$$\tilde{\rho}^* = \begin{cases} -\gamma & \text{if } 0 < \gamma < -\rho \text{ with } D \neq 0, \\ \rho & \text{if } 0 < -\rho \leqslant \gamma \text{ or if } 0 < \gamma < -\rho \text{ with } D = 0, \\ 0 & \text{if } \gamma = 0, \\ \rho & \text{if } \gamma < \rho, \\ \gamma & \text{if } \rho \leqslant \gamma < 0. \end{cases}$$

We assume here that $k = k_n$ is an intermediate sequence, that is, $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$. Our main asymptotic result is as follows.

Theorem 1. *Suppose that* $\sqrt{k}b(n/k) \to \lambda \in \mathbb{R}$. *Then*

$$\sqrt{k}(\hat{\gamma}^{\mathrm{H}}_{k,n} - \gamma) \to_d N(\mu_{\mathrm{H}}, \sigma^2_{\mathrm{H}})$$

and

$$\sqrt{k}(\hat{\gamma}^{\mathrm{Z}}_{k,n} - \gamma) \rightarrow {}_d N(\mu_{\mathrm{Z}}, \sigma^2_{\mathrm{Z}}),$$

where

$$\sigma^2_{\mathrm{H}} = \begin{cases} 1 + \gamma^2 & \text{if } \gamma \geqslant 0, \\ \dfrac{(1-\gamma)(1+\gamma+2\gamma^2)}{(1-2\gamma)} & \text{if } \gamma < 0, \end{cases}$$

$$\sigma^2_{\mathrm{Z}} = \begin{cases} 2[1+\gamma^2+\gamma] & \text{if } \gamma \geqslant 0, \\ \dfrac{2(1-\gamma)[1+2\gamma+\gamma^2-2\gamma^3]}{(1-2\gamma)(1-\gamma)} & \text{if } \gamma < 0, \end{cases}$$

$$\mu_{\mathrm{H}} = \frac{\lambda}{1-\tilde{\rho}}$$

and

$$\mu_{\mathrm{Z}} = \frac{\lambda}{(1-\tilde{\rho})^2}.$$

The asymptotic variance and bias of the estimators $\hat{\gamma}^{\mathrm{H}}_{k,n}$ and $\hat{\gamma}^{\mathrm{Z}}_{k,n}$ can be derived using the following asymptotic representations for $\log UH_{j,n}$, $j = 1, \ldots, k$, as $k, n \to \infty$, $k/n \to 0$:

$$\log UH_{j,n} = \begin{cases} \gamma \log U^{-1}_{k+1,n} + \log(\ell_+(k/j)^{\gamma}) + \left\{ \dfrac{1}{\sqrt{k}} \dfrac{Z_{0,k,n}(j/k)}{j/k} + \tilde{\rho}^{-1} b\left(\dfrac{n}{k}\right)\left(\dfrac{j}{k}\right)^{|\tilde{\rho}|} \right\}(1+o_p(1)) \\ \hfill \text{if } \gamma > 0, \\ \log \ell_+ + \left\{ \dfrac{1}{\sqrt{k}} \dfrac{P^{(2)}_{k,n}(j/k)}{j/k} - \left(\log \dfrac{n}{j}\right)^{-1} \right\}(1+o_p(1)) \\ \hfill \text{if } \gamma = 0, \\ \log\left(\dfrac{\ell_+}{1-\gamma}(nk)^{\gamma}\right) + \log(k/j)^{\gamma} + \left\{ (1-\gamma)\dfrac{1}{\sqrt{k}} \dfrac{Z_{\gamma,k,n}(j/k)}{(j/k)^{1-\gamma}} + \tilde{\rho}^{-1} b\left(\dfrac{n}{k}\right)\left(\dfrac{j}{k}\right)^{|\tilde{\rho}|} \right\}(1+o_p(1)) \\ \hfill \text{if } \gamma < 0. \end{cases} \tag{12}$$

Here, $U_{1,n} \leqslant \ldots \leqslant U_{n,n}$ denote the order statistics from a random sample of size n from the uniform (0,1) distribution. Further,

$$P^{(2)}_{k,n}(j/k) = \sqrt{k}\left(\frac{j}{k}\right)\left\{\frac{1}{j}\sum_{i=1}^{j}\log\left(\frac{U_{j+1,n}}{U_{i,n}}\right) - 1\right\}, \qquad j = 1, \ldots, k,$$

which can be asymptotically represented by $\int_0^t (W(s)/s)\mathrm{d}s - W(t)$ with W denoting a Wiener process. Moreover, $\{Z_{0,k,n}(t);\ t \in (0, 1)\}$ and $\{Z_{\gamma,k,n}(t);\ t \in (0, 1)\}$ are stochastic processes that are asymptotically represented by the Gaussian processes

$$\int_0^t \frac{W(s)}{s}\mathrm{d}s + (\gamma - 1)W(t) - \gamma t W(1) =: W^{(0)}(t) - \gamma t W(1), \qquad \gamma \geqslant 0$$

and

$$W^{(\gamma)}(t) = \int_0^t \frac{W(s)}{s^{1+\gamma}}\mathrm{d}s - t^{-\gamma}W(t), \qquad \gamma < 0,$$

respectively, with covariances given by

$$\operatorname{cov}(W^{(0)}(s),\ W^{(0)}(t)) = s[1 + \gamma^2 + \gamma\log(t/s)],$$

$$\operatorname{cov}(W^{(\gamma)}(s),\ W^{(\gamma)}(t)) = \frac{s^{1-\gamma}}{(1-\gamma)(1-2\gamma)}[2s^{-\gamma} - t^{-\gamma}(1+\gamma)(1-2\gamma)],$$

with $0 \leqslant s \leqslant t \leqslant 1$.

The above expressions lead to the asymptotic mean squared errors of the different estimators as given in Appendix C. If $\gamma > 0$, the estimator $\hat{\gamma}^{\mathrm{H}}_{k,n}$ and the moment estimator $\hat{\gamma}^{\mathrm{M}}_{k,n}$ have the same asymptotic mean squared error (AMSE). In Appendix E, the minimal AMSE values of the different estimators are given as a function of k (under the assumption that the slowly varying parts of a_2 and L_2 are equivalent to a constant). To facilitate the comparison of these minimal AMSE values, Figure 1 provides plots with contour lines for the ratios of $\mathrm{AMSE}(\hat{\gamma}^{\mathrm{Z}}_{k_{\mathrm{opt}},n})$ to the minimum values of the other estimators together with an indication of the $(\gamma,\ \rho)$ area, taken from $(-2, 2) \times (-2, 0)$, where the Zipf-type estimator is best.

In Figure 1(a)–(c), we compare the Zipf and Hill-type estimators. For small values of γ and ρ, the Zipf approach is better. When $0 < \gamma < -\rho$ (whence $\tilde{\rho} = -\gamma$), the Zipf estimator performs better when $\gamma < 0.22$. When $\gamma < 0$, the AMSE ratio is less than 1 over the whole $(\gamma,\ \rho)$ area. In Figure 1(d) the Zipf and moment estimators are considered for $\gamma < 0$. The AMSE fraction is here always in favour of the Zipf estimator. Finally, in Figure 1(e)–(f) we show the ratios with respect to the maximum likelihood estimator. When $\gamma > 0$, the comparison does not hold if $D \neq 0$ (Figure 1(e)) since then the maximum likelihood estimator always performs better.

An interesting feature of the Zipf estimator is the smoothness of the realizations as a function of k, which alleviates to some extent the problem of choosing k. This is illustrated in Figure 2 for an insurance example. In reinsurance, protection against extreme claims is regularly sought through an excess-of-loss reinsurance contract, where the reinsurance company pays the amount $X - R$ if $X > R$, where R denotes a preset priority level. The example combines 252 claims from a single line of business. All of the claims were for a minimum of €1.1 million and were submitted between 1988 and 2000. They were gathered

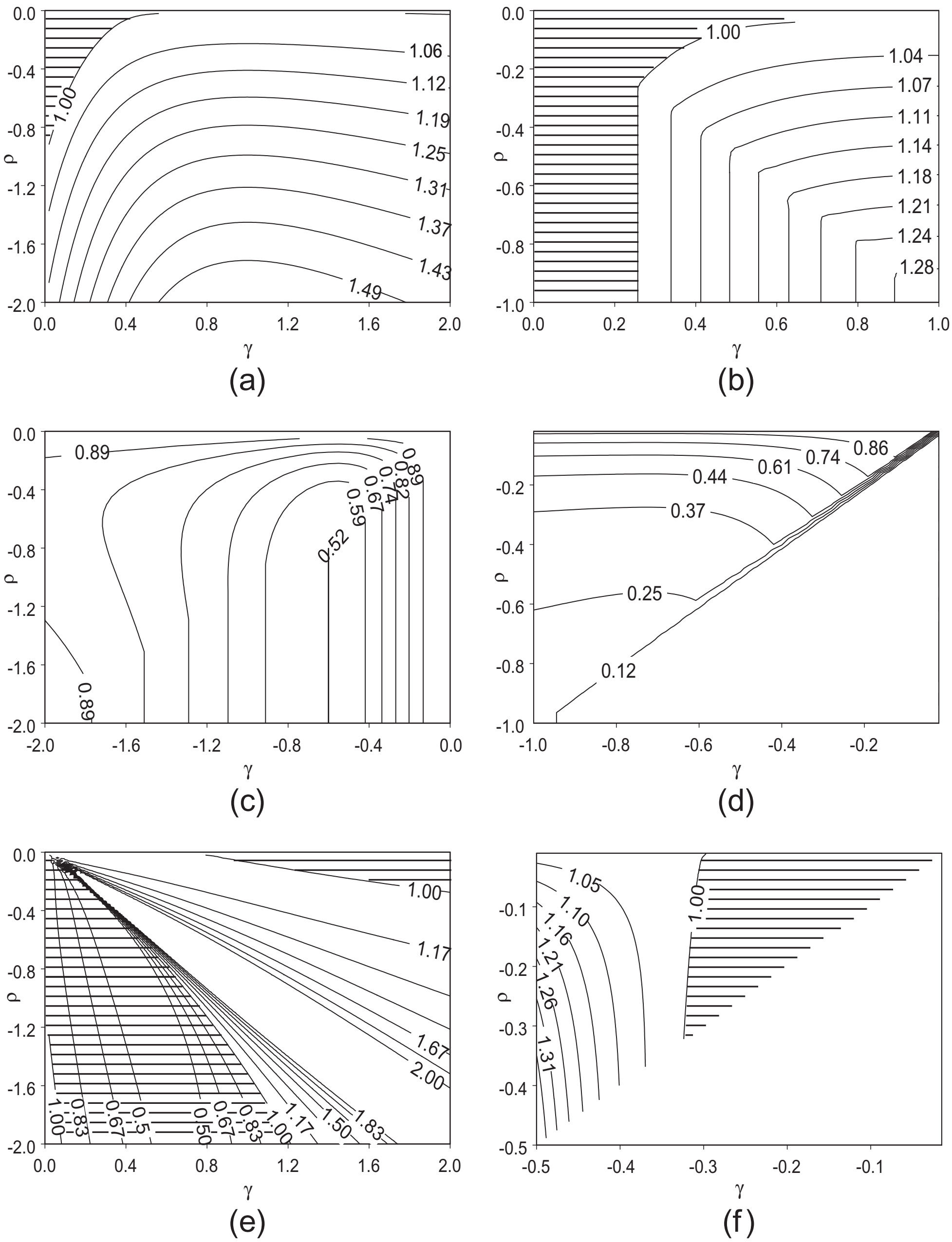

Figure 1. Contour lines of $\mathrm{AMSE}(\hat{\gamma}^{Z}_{k_{\mathrm{opt}},n})/\mathrm{AMSE}(\hat{\gamma}^{H}_{k_{\mathrm{opt}},n})$ for (a) $\gamma > 0$ and $D = 0$, (b) $\gamma > 0$ and $D \neq 0$, (c) $\gamma < 0$; (d) $\mathrm{AMSE}(\hat{\gamma}^{Z}_{k_{\mathrm{opt}},n})/\mathrm{AMSE}(\hat{\gamma}^{M}_{k_{\mathrm{opt}},n})$ for $\gamma < 0$; $\mathrm{AMSE}(\hat{\gamma}^{Z}_{k_{\mathrm{opt}},n})/\mathrm{AMSE}(\hat{\gamma}^{ML}_{k_{\mathrm{opt}},n})$ for (e) $\gamma > 0$ and $D = 0$, (f) $\gamma < 0$. The shaded areas (if available) show the (γ, ρ)-areas where $\hat{\gamma}^{Z}_{k_{\mathrm{opt}},n}$ is the best (except in cases where the Zipf estimator is always best).

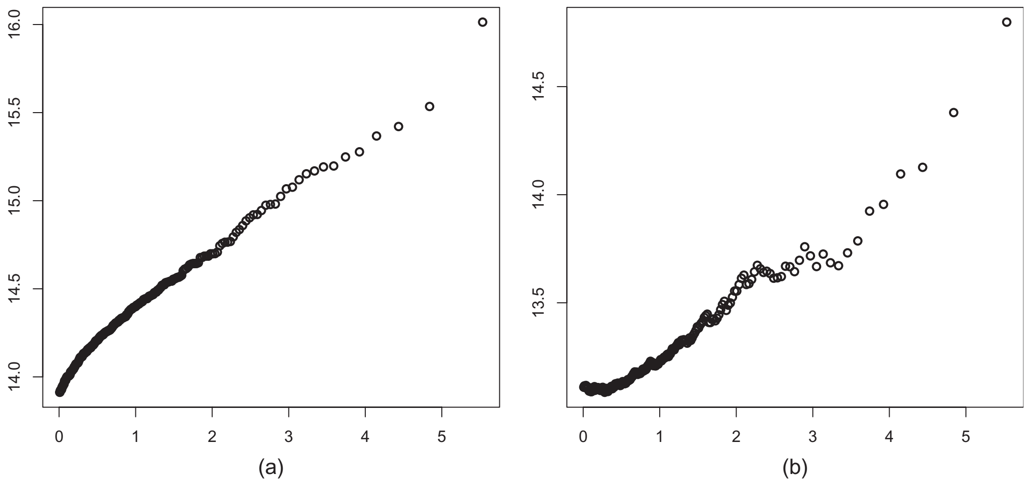

Figure 2. (a) Zipf quantile plot; (b) generalized quantile plot for the insurance data set ($n = 252$).

from different companies. Figure 2 offers the Zipf and generalized quantile plot for these data. Figure 3 compares four estimators of the index. Note that the generalized Zipf estimator is extremely stable from $k = 50$ up to the end.

The authors also performed a simulation study. The results are available from the authors.

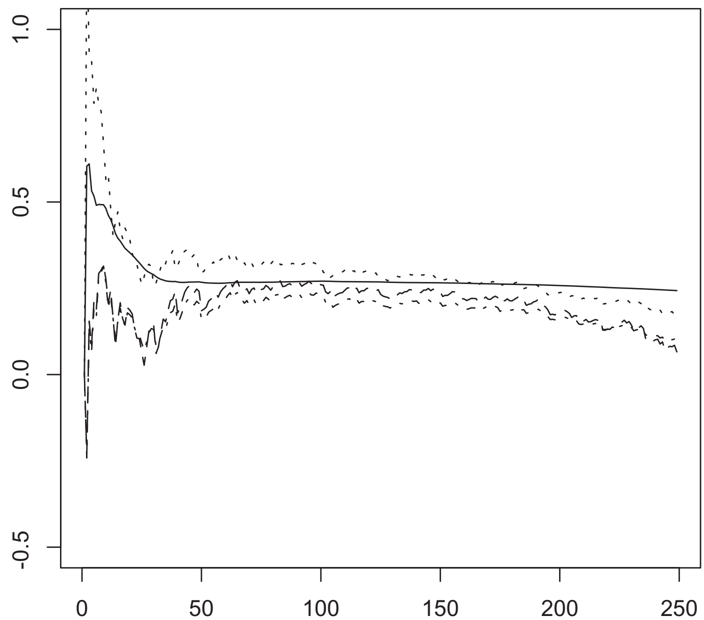

Figure 3. $\hat{\gamma}^{\mathrm{M}}_{k,n}$ (long dashes), $\hat{\gamma}^{\mathrm{H}}_{k,n}$ (dotted line), $\hat{\gamma}^{\mathrm{Z}}_{k,n}$ (solid line) and $\hat{\gamma}^{\mathrm{ML}}_{k,n}$ (dashed-dotted line) as a function of k for the insurance data set ($n = 252$).

In most cases the finite-sample comparisons of the estimators are in line with the asymptotic analyses: $\hat{\gamma}^{Z}_{k,n}$ performs best when $|\rho|$ is small.

3. Regression based on the generalized quantile plot and adaptive threshold selection

In this section, we first construct a regression representation for the tail of a generalized quantile plot (8). Given a value of k, we discuss the regression problem through the points

$$\left(\log\left(\frac{n+1}{j}\right), \log UH_{j,n}\right), \; j = 1, \ldots, k. \tag{13}$$

The main goal is twofold. Once we have such a representation, we can construct other estimators of γ whose bias is reduced compared to, for instance, $\hat{\gamma}^{H}_{k,n}$. This is particularly the case when the limit in (5) is attained at a slow rate. More importantly, we can estimate the bias of a variety of estimators such as those considered in the preceding section. This in turn allows us to construct diagnostics for threshold selection.

From (12) we derive for $\gamma \neq 0$ that

$$\begin{aligned} Z_j &:= (j+1)\log\frac{UH_{j,n}}{UH_{j+1,n}} \\ &= \left(\gamma + b(n/k)\left(\frac{j}{k}\right)^{-\tilde{\rho}}\right) + \varepsilon_j, \qquad 1 \leqslant j \leqslant k-1, \end{aligned} \tag{14}$$

where ε_j are considered as zero-centred error terms.

This representation provides a direct generalization of the regression model for $j\log(X_{n-j+1,n}/X_{n-j,n})$ as studied in Feuerverger and Hall (1999) and Beirlant *et al*. (1999; 2002). If we ignore the term $b(n/k)$ in (14), we retrieve the Hill-type estimator $\hat{\gamma}^{H}_{k,n}$.

By using a least-squares approach, the representation (14) can be further exploited to propose an estimator for γ in which $\tilde{\rho}$ is replaced by an estimator $\hat{\tilde{\rho}}$. We follow the same procedure as in Beirlant *et al*. (2002). For $\lambda \in (0, 1)$, the estimator

$$\hat{\tilde{\rho}}_{k,\lambda,n} = -\frac{1}{\log\lambda}\log\frac{\hat{\gamma}^{H}_{\lfloor\lambda^2 k\rfloor,n} - \hat{\gamma}^{H}_{\lfloor\lambda k\rfloor,n}}{\hat{\gamma}^{H}_{\lfloor\lambda k\rfloor,n} - \hat{\gamma}^{H}_{k,n}}$$

offers a proper consistent estimator if k is chosen such that $\sqrt{k}b(n/k) \to \infty$ (however, with a slow rate of convergence). For practical diagnostic purposes, it might be sufficient to replace $\tilde{\rho}$ by a canonical choice such as -1.

The least-squares estimators of γ and $b(n/k)$ are then given by

$$\begin{aligned} \hat{\gamma}_{\mathrm{LS},k} &= \overline{Z}_k - \hat{b}_{\mathrm{LS},k}(\hat{\tilde{\rho}})/(1-\hat{\tilde{\rho}}), \\ \hat{b}_{\mathrm{LS},k}(\hat{\tilde{\rho}}) &= \frac{(1-\hat{\tilde{\rho}})^2(1-2\hat{\tilde{\rho}})}{\hat{\tilde{\rho}}^2}\frac{1}{k}\sum_{j=1}^{k}\left(\left(\frac{j}{k}\right)^{-\hat{\tilde{\rho}}} - \frac{1}{1-\hat{\tilde{\rho}}}\right)Z_j. \end{aligned}$$

Here we have approximated $\frac{1}{k}\sum_{j=1}^{k}(\frac{j}{k})^{-\hat{\tilde{\rho}}}$ by $1/(1-\hat{\tilde{\rho}})$ and $\frac{1}{k}\sum_{j=1}^{k}((\frac{j}{k})^{-\hat{\tilde{\rho}}}-\frac{1}{1-\hat{\tilde{\rho}}})^2$ by $\frac{\hat{\tilde{\rho}}^2}{(1-\hat{\tilde{\rho}})^2(1-2\hat{\tilde{\rho}})}$.

We propose an adaptive estimation procedure for k_{opt} that is based on the above estimators. The values of k_{opt} for the different estimators are given in Appendix D. For brevity, we only specify the results concerning $\hat{\gamma}^{\mathrm{H}}_{k,n}$ for $\gamma > 0$. It is clear that a similar procedure can be applied without any problem to the other estimators.

From Appendix C, one finds the optimal value of k that minimizes the AMSE of the simple estimator $\hat{\gamma}^{\mathrm{H}}_{k,n}$:

$$k^{\mathrm{H}}_{n,\mathrm{opt}} \sim n\big((1-\tilde{\rho})^2(1+\gamma^2)\big)^{1/(1-2\tilde{\rho})}\left(s^{\leftarrow}\left(\frac{1}{n}\right)\right)^{-1}. \tag{15}$$

Here $s^{\leftarrow}$ is the inverse function of the decreasing function s and satisfies $b^2(t) = (1+o(1))\int_t^{\infty} s(u)\mathrm{d}u$. Note that this expression requires a third-order condition on the tail of the underlying distribution. In the special case where the slowly varying parts of a_2 and L_2 are asymptotically constant, we obtain

$$\begin{aligned} k^{\mathrm{H}}_{n,\mathrm{opt}} &\sim [b(n)]^{-2/(1-2\tilde{\rho})}\left(\frac{(1+\gamma^2)(1-\tilde{\rho})^2}{-2\tilde{\rho}}\right)^{1/(1-2\tilde{\rho})} \\ &\sim \left[b\left(\frac{n}{k_0}\right)\right]^{-2/(1-2\tilde{\rho})} k_0^{-2\tilde{\rho}/(1-2\tilde{\rho})}\left(\frac{(1+\gamma^2)(1-\tilde{\rho})^2}{-2\tilde{\rho}}\right)^{1/(1-2\tilde{\rho})} \end{aligned} \tag{16}$$

for any secondary value $k_0 \in \{1, \ldots, n\}$. Continuing along the same lines, we can replace the quantities $b(n/k_0)$, $\tilde{\rho}$, and γ in (16) by consistent estimators. For example, we can use the least-squares estimators of γ and $b(n/k_0)$ of the regression model (14). For each value of k_0 we obtain an estimator of $k^{\mathrm{H}}_{n,\mathrm{opt}}$ such as

$$\hat{k}^{\mathrm{H}}_{n,k_0} = [\hat{b}^2_{\mathrm{LS},k_0}(\hat{\tilde{\rho}})]^{-1/(1-2\hat{\tilde{\rho}})} k_0^{-2\hat{\tilde{\rho}}/(1-2\hat{\tilde{\rho}})}\left(\frac{(1+\hat{\gamma}^2_{\mathrm{LS},k_0}(\hat{\tilde{\rho}}))(1-\hat{\tilde{\rho}})^2}{-2\hat{\tilde{\rho}}}\right)^{1/(1-2\hat{\tilde{\rho}})}.$$

Theorem 2. *As* $k_0, n \to \infty$, $k_0/n \to 0$ and $\sqrt{k_0}b(n/k_0) \to \infty$ *and if, in the estimation procedure, we substitute for* $\tilde{\rho}$ *a consistent estimator* $\hat{\tilde{\rho}}$ such that $\hat{\tilde{\rho}} - \tilde{\rho} = o_P(1/\log k_0)$, *then*

$$\frac{\hat{k}^{\mathrm{H}}_{n,k_0}}{k^{\mathrm{H}}_{n,\mathrm{opt}}} \to_P 1.$$

In Fraga Alves *et al.* (2003), estimators $\hat{\tilde{\rho}}$ can be found that satisfy

$$\sqrt{k_0}b\left(\frac{n}{k_0}\right)(\hat{\tilde{\rho}} - \tilde{\rho}) = O_P(1).$$

In this expression it is necessary that $\sqrt{k_0}b(n/k_0) \to \infty$ and $\sqrt{k_0}b^2(n/k_0) \to c$ finite, which induces a stronger higher-order condition than given in (9). Consequently, $\hat{\tilde{\rho}} - \tilde{\rho} = o_P(1/\log k_0)$ follows if $\sqrt{k_0}b(n/k_0)/\log k_0 \to \infty$ as $k_0 \to \infty$.

Theorem 2 relies on the following asymptotic expansion, which follows from (12):

$$\hat{b}_{\mathrm{LS},k_0}(\hat{\tilde{\rho}}) = b\left(\frac{n}{k_0}\right) + \frac{N_{k_0,n}}{\sqrt{k_0}} + o_p\left(b\left(\frac{n}{k_0}\right)\right). \tag{17}$$

Here $N_{k_0,n}$ denotes a sequence of random variables that, for k_0, $n \to \infty$ and $k_0/n \to 0$, is asymptotically normal with mean 0 and finite variance depending on $\tilde{\rho}$ and γ. For such a procedure in the case $\gamma > 0$, see Theorem 4 in Beirlant *et al.* (2002).

Of course, the above approach suffers from the practical drawback that, in order to obtain a consistent estimator, one needs to identify the k_0-region for which $\sqrt{k_0}b(n/k_0) \to \infty$. However, when $\sqrt{k_0}b(n/k_0) \to c$ for some $c \in \mathbb{R}$, (17) suggests that $\hat{k}^{\mathrm{H}}_{n,k_0}/k^{\mathrm{H}}_{n,\mathrm{opt}}$ asymptotically behaves as a realization from a normal distribution centred at 1. Consequently, graphs of $\log \hat{k}^{\mathrm{H}}_{n,k_0}$ as a function of k_0 are rather stable except for the k_0-regions where $\sqrt{k_0}b(n/k_0) \to 0$.

Figure 4 illustrates this for the insurance example. We plot $\log\hat{h}^{\mathrm{H}}_{n,k_0}$ against k_0 for k_0 between 3 and n, with $\hat{\tilde{\rho}}$ replaced by -1. Notice how the graph in Figure 4 is stable for k between 150 and n. Using the median value $\log\hat{k}^{\mathrm{H}}_{n,\mathrm{med}}$ of the estimates, we obtain $\hat{k}^{\mathrm{H}}_{n,\mathrm{med}} = \hat{k}^{\mathrm{M}}_{n,\mathrm{med}} = 121$, $\hat{k}^{\mathrm{Z}}_{n,\mathrm{med}} = 263$ and $\hat{k}^{\mathrm{ML}}_{n,\mathrm{med}} = 91$. Also note that, in the case of the generalized Zipf estimator, the k-value is larger than the sample size n, which means that in practice we take n. This should not be surprising in view of the stability of this estimator in Figure 3.

Finally, we refer to Draisma *et al.* (1999) and Groeneboom *et al.* (2003) for other adaptive selection procedures when estimating a real-valued extreme-value index.

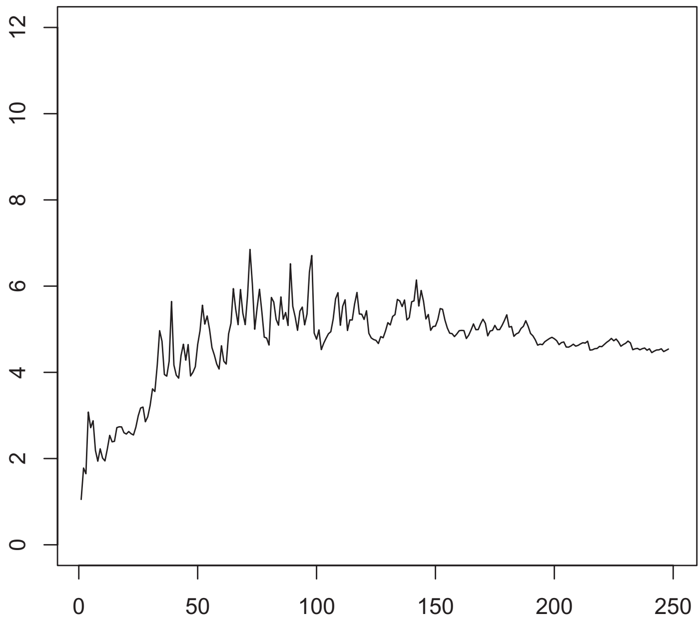

Figure 4. $\log \hat{k}^{\mathrm{H}}_{n,k_0}$ as a function of k_0 for k_0 between 3 and n for the insurance data set.

Appendix A: Overview of all possible kinds of GRV_2 functions with $\rho < 0$

From Vanroelen (2003) we obtain the following representations of U. See also the appendix in Draisma *et al*. (1999).

- $0 < -\rho < \gamma$. For $U \in \mathrm{GRV}_2(\gamma,\ \rho;\ \ell_+ x^\gamma,\ a_2(x);\ 0,\ A)$,

$$U(x) = \ell_+ x^\gamma \left\{ \frac{1}{\gamma} + \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \right\}.$$

- $\gamma = -\rho$. For $U \in \mathrm{GRV}_2(\gamma,\ -\gamma;\ \ell_+ x^\gamma,\ x^{-\gamma}\ell_2(x);\ 0,\ A)$ with ℓ_2 some slowly varying function,

$$U(x) = \ell_+ x^\gamma \left\{ \frac{1}{\gamma} + x^{-\gamma} L_2(x) \right\},$$

  with

$$L_2(x) = B + \int_1^x (A+o(1)) \frac{\ell_2(t)}{t}\,\mathrm{d}t + o(l_2(t)) \qquad \text{for some constant } B.$$

- $0 < \gamma < -\rho$. For $U \in \mathrm{GRV}_2(\gamma,\ \rho;\ \ell_+ x^\gamma,\ a_2(x);\ 0,\ A)$,

$$U(x) = \ell_+ x^\gamma \left\{ \frac{1}{\gamma} + Dx^{-\gamma} + \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \right\}.$$

- $\gamma = 0$. For $U \in \mathrm{GRV}_2(0,\ \rho;\ \ell_+,\ a_2(x);\ 0,\ A)$,

$$U(x) = \ell_+ \log x + D + \frac{A}{\rho} a_2(x)(1+o(1)).$$

- $\gamma < 0$. For $U \in \mathrm{GRV}_2(\gamma,\ \rho;\ \ell_+ x^\gamma,\ a_2(x);\ 0,\ A)$,

$$U(x) = U(\infty) - \ell_+ x^\gamma \left\{ \frac{1}{-\gamma} - \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \right\},$$

where $\ell_+ > 0,\ A \neq 0,\ D \in \mathbb{R}$.

Concerning $\log U$, the following results are available under these representations:

- If $0 < -\rho < \gamma$, then $\log U \in \mathrm{GRV}_2(0,\ \rho;\ \gamma,\ a_2(x);\ 0,\ \rho A/(\gamma+\rho))$.
- If $\gamma = -\rho$, then $\log U \in \mathrm{GRV}_2(0,\ -\gamma;\ \gamma,\ x^{-\gamma} L_2(x);\ 0,\ -\gamma)$.
- If $0 < \gamma < -\rho$, then $\log U \in \mathrm{GRV}_2(0,\ -\gamma;\ \gamma,\ x^{-\gamma};\ 0,\ -\gamma D)$ if $D \neq 0$, and $\log U \in \mathrm{GRV}_2(0,\ \rho;\ \gamma,\ a_2(x);\ 0,\ \rho A/(\gamma+\rho))$ if $D = 0$.
- If $\gamma = 0$, then $\log U \in \mathrm{GRV}_2(0,\ 0;\ a(x)/U(x),\ a(x)/U(x);\ -1,\ 0)$.
- If $\gamma < \rho$, then $\log U \in \mathrm{GRV}_2(\gamma,\ \rho;\ [U(\infty)]^{-1}\ell_+ x^\gamma,\ a_2(x);\ 0,\ A)$.
- If $\rho < \gamma < 0$, then $\log U \in \mathrm{GRV}_2(\gamma,\ \gamma;\ [U(\infty)]^{-1}\ell_+ x^\gamma,\ \ell_+ x^\gamma;\ 0,\ -1/(\gamma U(\infty)))$.
- If $\gamma = \rho$, then $\log U \in \mathrm{GRV}_2(\gamma,\ \gamma;\ [U(\infty)]^{-1}\ell_+ x^\gamma,\ a_2(x);\ 0,\ A - \ell_+/(\gamma U(\infty)))$.

Appendix B: Details of proofs

Here we show how to derive the asymptotic representation (12). Denote by $U_{1,n} \leqslant \ldots \leqslant U_{n,n}$ the order statistics of a pure random sample of size n from the uniform (0,1) distribution.

Throughout, we use the fact that, when $k, n \to \infty$ and $k/n \to 0$, $U^{\beta}_{j+1,n} = (j/n)^{\beta}(1 + o_p(1))$ uniformly in $j = 1, \ldots, k$, for any constant $\beta > 0$.

We first deal with the case $\gamma > 0$. As the other cases are similar, we only give the proof for the subcase $\gamma + \rho > 0$. From Appendix A, we have

$$UH_{j,n} =_d \ell_+ U^{-\gamma}_{j+1,n}\left\{\frac{1}{\gamma} + \frac{A}{\gamma+\rho} a_2({U_{j+1,n}}^{-1})(1 + o_p(1))\right\}$$

$$\times \left\{\frac{\gamma}{j}\sum_{i=1}^{j} \log \frac{U_{j+1,n}}{U_{i,n}} + \frac{\gamma A}{\gamma+\rho}\frac{\rho}{1-\rho} a_2(U^{-1}_{j+1,n})(1 + o_p(1))\right\}$$

$$= \ell_+ {U_{j+1,n}}^{-\gamma}\left\{\frac{1}{j}\sum_{i=1}^{j} \log \frac{U_{j+1,n}}{U_{i,n}} + \frac{A[\rho + \gamma(1-\rho)]}{(\gamma+\rho)(1-\rho)} a_2(U^{-1}_{j+1,n})\right\}(1 + o_p(1)).$$

Hence,

$$\log UH_{j,n} =_d \frac{P^{(1)}_n(j/n)}{\sqrt{k}} + \gamma \log(n/j) + \log \ell_+$$

$$+ \frac{P^{(2)}_{k,n}(j/k)}{\sqrt{k}(j/k)} + \frac{A[\rho + \gamma(1-\rho)]}{(\gamma+\rho)(1-\rho)} a_2\left(\frac{n}{j}\right)(1 + o_p(1)),$$

with

$$P^{(1)}_n(j/n) = \gamma\sqrt{k}\{\log U^{-1}_{j+1,n} - \log(n/j)\}, \qquad j = 1, \ldots, k,$$

and

$$P^{(2)}_{k,n}(j/k) = \sqrt{k}\left(\frac{j}{k}\right)\left\{\frac{1}{j}\sum_{i=1}^{j} \log\left(\frac{U_{j+1,n}}{U_{i,n}}\right) - 1\right\}, \qquad j = 1, \ldots, k.$$

Following Mason and Turova (1994), $P^{(2)}_{k,n}(t)$ is approximated by $\int_0^t (W(s)/s)\mathrm{d}s - W(t)$, while from Drees (1998) it follows that

$$\frac{1}{\gamma}(P^{(1)}_n(j/n) - P^{(1)}_n(k/n)) \approx \frac{W(j/k)}{j/k} - W(1).$$

For the case $\gamma > 0$, we therefore arrive at (12) with

$$Z_{0,k,n}\left(\frac{j}{k}\right) = P^{(2)}_{k,n}\left(\frac{j}{k}\right) + \frac{j}{k}\left[P^{(1)}_n\left(\frac{j}{n}\right) - P^{(1)}_n\left(\frac{k}{n}\right)\right].$$

We turn to the case $\gamma = 0$. In an analogous way and using Appendix A, we obtain

$$UH_{j,n} =_d a(U_{j+1,n}^{-1})\frac{1}{j}\sum_{i=1}^{j}\log\frac{U_{j+1,n}}{U_{i,n}} - \frac{1}{2}\frac{a^2(U_{j+1,n}^{-1})}{U(U_{j+1,n}^{-1})}\frac{1}{j}\sum_{i=1}^{j}\left(\log\frac{U_{j+1,n}}{U_{i,n}}\right)^2$$

$$= \ell_+\left\{1 + \left[\frac{1}{j}\sum_{i=1}^{j}\log\frac{U_{j+1,n}}{U_{i,n}} - 1\right] - \left(\log\frac{n}{j}\right)^{-1}(1 + o_p(1))\right\}.$$

Therefore,

$$\log UH_{j,n} = \log\ell_+ + \frac{P_{k,n}^{(2)}(j/k)}{\sqrt{k}j/k} - \left(\log\frac{n}{j}\right)^{-1}(1 + o_P(1)).$$

Finally, we deal with the case $\gamma < 0$. Again, we only give the proof for the case where $|\gamma| + \rho > 0$, the others being similar. We have, with the help of Appendix A,

$$UH_{j,n} =_d \left[U(\infty) - \ell_+ U_{j+1,n}{}^{-\gamma}\left\{-\frac{1}{\gamma} - \frac{A}{\gamma+\rho}a_2(U_{j+1,n}^{-1})\right\}\right].$$

$$\times\ \frac{1}{j}\sum_{i=1}^{j}\left\{[U(\infty)]^{-1}\ell_+ U_{j+1,n}^{-\gamma}\frac{(U_{j+1,n}/U_{i,n})^{\gamma} - 1}{\gamma}\right.$$

$$\left. + [U(\infty)]^{-1}\ell_+ A\frac{(U_{j+1,n}/U_{i,n})^{\gamma+\rho} - 1}{\gamma+\rho}U_{j+1,n}^{-\gamma}a_2(U_{j+1,n}^{-1})\right\}$$

$$= \frac{\ell_+}{\gamma}\left\{\frac{1}{j}\sum_{i=1}^{j}[U_{i,n}^{|\gamma|} - U_{j+1,n}^{|\gamma|}] + \frac{A\gamma}{1-\gamma-\rho}\left(\frac{j}{n}\right)^{|\gamma|}a_2\left(\frac{n}{j}\right)\right\}(1 + o_p(1))$$

$$= \frac{\ell_+}{1-\gamma}\left(\frac{j}{k}\right)^{|\gamma|}\left(\frac{k}{n}\right)^{|\gamma|}\left\{1 + (1-\gamma)\left(\frac{k}{j}\right)^{|\gamma|}\frac{Z_{\gamma,k,n}(j/k)}{\sqrt{k}j/k} + \frac{A(1-\gamma)}{1-\gamma-\rho}\left(\frac{j}{k}\right)^{|\rho|}a_2\left(\frac{n}{k}\right)\right\},$$

where

$$Z_{\gamma,k,n}(j/k) = \frac{1}{|\gamma|}\sqrt{k}\left(\frac{j}{k}\right)\left(\frac{k}{n}\right)^{-|\gamma|}\left\{\frac{1}{j}\sum_{i=1}^{j}\left(U_{j+1,n}^{|\gamma|} - U_{i,n}^{|\gamma|}\right) - \frac{|\gamma|}{1-\gamma}\left(\frac{j}{k}\right)^{|\gamma|}\left(\frac{k}{n}\right)^{|\gamma|}\right\},$$

$$j = 1, \ldots, k,$$

which now leads to the asymptotic representation (12) for $\log UH_{j,n}$ if $\gamma < 0$.

Appendix C: Asymptotic mean squared errors of the different estimators

We derive the asymptotic mean squared errors of the different estimators.

For the estimator $\hat{\gamma}_{k,n}^{\mathrm{H}}$,

$$\frac{1+\gamma^2}{k} + \left(\frac{1}{1-\tilde{\rho}} b\left(\frac{n}{k}\right)\right)^2, \qquad \text{if } \gamma \geqslant 0,$$
$$\frac{(1-\gamma)(1+\gamma+2\gamma^2)}{(1-2\gamma)k} + \left(\frac{1}{1-\tilde{\rho}} b\left(\frac{n}{k}\right)\right)^2, \qquad \text{if } \gamma < 0.$$

For the estimator $\hat{\gamma}^{\mathrm{Z}}_{k,n}$,

$$\frac{2[1+\gamma^2+\gamma]}{k} + \left(\frac{1}{(1-\tilde{\rho})^2} b\left(\frac{n}{k}\right)\right)^2, \qquad \text{if } \gamma \geqslant 0,$$
$$\frac{2(1-\gamma)[1+2\gamma+\gamma^2-2\gamma^3]}{(1-2\gamma)(1-\gamma)k} + \left(\frac{1}{(1-\tilde{\rho})^2} b\left(\frac{n}{k}\right)\right)^2, \qquad \text{if } \gamma < 0.$$

The AMSE of the moment estimator $\hat{\gamma}^{\mathrm{M}}_{k,n}$ can be found in Dekkers *et al*. (1989):

$$\begin{cases} \dfrac{1+\gamma^2}{k} + \left(\dfrac{1}{1-\tilde{\rho}} b\left(\dfrac{n}{k}\right)\right)^2, & \text{if } \gamma > 0, \\[2ex] \dfrac{1}{k} + b^2\left(\dfrac{n}{k}\right), & \text{if } \gamma = 0, \\[2ex] \dfrac{(1-\gamma)^2(1-2\gamma)(6\gamma^2-\gamma+1)}{(1-3\gamma)(1-4\gamma)k} + \left(\dfrac{1-2\gamma}{1-2\gamma-\tilde{\rho}} b\left(\dfrac{n}{k}\right)\right)^2, & \text{if } \gamma < \rho, \\[2ex] \dfrac{(1-\gamma)^2(1-2\gamma)(6\gamma^2-\gamma+1)}{(1-3\gamma)(1-4\gamma)k} + \left(\dfrac{1-2\gamma}{\tilde{\rho}(1-\tilde{\rho})} b\left(\dfrac{n}{k}\right)\right)^2, & \text{if } \rho \in \gamma < 0, \\[2ex] \dfrac{(1-\gamma)^2(1-2\gamma)(6\gamma^2-\gamma+1)}{(1-3\gamma)(1-4\gamma)k} + \left(\dfrac{(1-2\gamma)}{(1-\gamma)(1-3\gamma)} \dfrac{A(1-\gamma)^2 - 2\ell_+/U(\infty)}{A(1-\gamma)-\ell_+/U(\infty)} b\left(\dfrac{n}{k}\right)\right)^2, & \text{if } \gamma = \rho. \end{cases}$$

Drees *et al*. (2004) stated the following expressions for the AMSE of for the maximum likelihood estimator based on a generalized Pareto fit:

$$\mathrm{AMSE}\left(\hat{\gamma}^{\mathrm{ML}}_{k,n}\right) = \frac{(1+\gamma)^2}{k} + \left(\frac{\rho(\gamma+1)A}{(1-\rho)(1-\rho+\gamma)} a_2\left(\frac{n}{k}\right)\right)^2 \qquad \text{if } \gamma > -\frac{1}{2},\ \rho < 0.$$

Appendix D: The optimal values of k that minimize the different expressions for the AMSEs

We assume that the slowly varying parts of a_2 and L_2 are asymptotically equivalent to a constant. We give the optimal values of k that minimize the different expressions for the AMSEs.

For the estimator $\hat{\gamma}^{\mathrm{H}}_{k,n}$,

$$\begin{cases} \left(\dfrac{(1+\gamma^2)(1-\tilde{\rho})^2}{-2\tilde{\rho}}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, & \text{if } \gamma > 0, \\ \frac{1}{4}[b(n)]^{-5/2}(1+o(1)) & \text{if } \gamma = 0, \\ \left(\dfrac{(1+\gamma+2\gamma^2)(1-\tilde{\rho})^2(1-\gamma)}{(-2\tilde{\rho})(1-2\gamma)}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, & \text{if } \gamma < 0. \end{cases}$$

For the estimator $\hat{\gamma}^{Z}_{k,n}$,

$$\begin{cases} \left(\dfrac{2(1-\tilde{\rho})^4[1+\gamma^2+\gamma]}{(-2\tilde{\rho})}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, & \text{if } \gamma > 0, \\ \frac{1}{2}[b(n)]^{-5/2}(1+o(1)) & \text{if } \gamma = 0, \\ \left(\dfrac{2(1-\tilde{\rho})^4[1+2\gamma+\gamma^2-2\gamma^3]}{(-2\tilde{\rho})(1-2\gamma)}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, & \text{if } \gamma < 0. \end{cases}$$

For the estimator $\hat{\gamma}^{M}_{k,n}$,

$$\begin{cases} \left(\dfrac{(1+\gamma^2)(1-\tilde{\rho})^2}{-2\tilde{\rho}}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}. \\ \hfill \text{if } \gamma > 0, \\ \frac{1}{2}[b(n)]^{-3/2}(1+o(1)) \\ \hfill \text{if } \gamma = 0, \\ \left(\dfrac{(1-\gamma)^2(1-2\gamma-\tilde{\rho})^2(6\gamma^2-\gamma+1)}{(-2\tilde{\rho})(1-2\gamma)(1-3\gamma)(1-4\gamma)}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, \\ \hfill \text{if } \gamma < \rho, \\ \left(\dfrac{\tilde{\rho}^2(1-\gamma)^4(6\gamma^2-\gamma+1)}{(-2\tilde{\rho})(1-2\gamma)(1-3\gamma)(1-4\gamma)}\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, \\ \hfill \text{if } \rho < \gamma < 0, \\ \left(\dfrac{(1-3\gamma)(1-\gamma)^4(6\gamma^2-\gamma+1)}{(-2\tilde{\rho})(1-2\gamma)(1-4\gamma)}\left(\dfrac{A(1-\gamma)-\ell_+/U(\infty)}{A(1-\gamma)^2-2\ell_+/U(\infty)}\right)^2\right)^{1/(1-2\tilde{\rho})} [b(n)]^{-2/(1-2\tilde{\rho})}, \\ \hfill \text{if } \gamma = \rho. \end{cases}$$

For the estimator $\hat{\gamma}^{ML}_{k,n}$,

$$\left(\frac{(1-\rho)^2(\gamma-\rho+1)^2}{\rho^2(-2\rho)A^2}\right)^{1/(1-2\rho)} [a_2(n)]^{-2/(1-2\rho)} \qquad \text{if } \gamma > -\frac{1}{2},\ \rho < 0.$$

Appendix E: Minimal AMSE values

With the optimal values of k from Appendix D, we deduce the following minimal AMSE values. For the estimator $\hat{\gamma}^{H}_{k_{\text{opt}},n}$,

$$\begin{cases} \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}}{(1+\gamma^2)^{\tilde{\rho}}(1-\tilde{\rho})}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), & \text{if } \gamma > 0, \\ b^2(n), & \text{if } \gamma = 0, \\ \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}(1-2\gamma)^{\tilde{\rho}}}{(1-\gamma)^{\tilde{\rho}}(1-\tilde{\rho})(1+\gamma+2\gamma^2)^{\tilde{\rho}}}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), & \text{if } \gamma < 0. \end{cases}$$

For the estimator $\hat{\gamma}^{\mathrm{Z}}_{k_{\mathrm{opt}},n}$,

$$\begin{cases} \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}}{2^{\tilde{\rho}}(1-\tilde{\rho})^2[1+\gamma^2+\gamma]^{\tilde{\rho}}}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), & \text{if } \gamma > 0, \\ b^2(n), & \text{if } \gamma = 0, \\ \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho})}(1-2\gamma)^{\tilde{\rho}}}{2^{\tilde{\rho}}(1-\tilde{\rho})^2[1+2\gamma+\gamma^2-2\gamma^3]^{\tilde{\rho}}}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), & \text{if } \gamma < 0. \end{cases}$$

For the estimator $\hat{\gamma}^{\mathrm{M}}_{k_{\mathrm{opt}},n}$,

$$\begin{cases} \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}}{(1+\gamma^2)^{\tilde{\rho}}(1-\tilde{\rho})}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), \\ \qquad \text{if } \gamma > 0, \\ b^2(n), \\ \qquad \text{if } \gamma = 0, \\ \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}(1-\gamma)^{-2\tilde{\rho}}(1-2\gamma)^{1-\tilde{\rho}}(1-3\gamma)^{\tilde{\rho}}(1-4\gamma)^{\tilde{\rho}}}{(1-2\gamma-\tilde{\rho})(6\gamma^2-\gamma+1)^{\tilde{\rho}}}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), \\ \qquad \text{if } \gamma < \rho, \\ \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}(1-3\gamma)^{\tilde{\rho}}(1-4\gamma)^{\tilde{\rho}}}{\gamma(1-\gamma)^{1+2\tilde{\rho}}(1-2\gamma)^{\tilde{\rho}-1}(6\gamma^2-\gamma+1)^{\tilde{\rho}}}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), \\ \qquad \text{if } \rho < \gamma < 0, \\ \left(\dfrac{(-2\tilde{\rho})^{\tilde{\rho}}(1-3\gamma)^{\tilde{\rho}-1}(1-4\gamma)^{\tilde{\rho}}}{(1-\gamma)^{1+2\tilde{\rho})}(1-2\gamma)^{\tilde{\rho}-1}(6\gamma^2-\gamma+1)^{\tilde{\rho}}}\dfrac{A(1-\gamma)^2-2\ell_+/U(\infty)}{A(1-\gamma)-\ell_+/U(\infty)}\right)^{2/(1-2\tilde{\rho})}[b(n)]^{2/(1-2\tilde{\rho})}(1-2\tilde{\rho}), \\ \qquad \text{if } \gamma = \rho. \end{cases}$$

For the estimator $\hat{\gamma}^{\mathrm{ML}}_{k_{\mathrm{opt}},n}$,

$$\left(\frac{(1+\gamma)^{1-2\rho}(A\rho)(-2\rho)^{\rho}}{(1-\rho)(1-\rho+\gamma)}\right)^{2/(1-2\rho)}[a_2(n)]^{2/(1-2\rho)}(1-2\rho).$$

Acknowledgement

The authors are very grateful to the referee for a careful reading of the paper that led to significant improvements of the earlier draft.

References

Balkema, A. and de Haan, L. (1974) Residual life at great age. *Ann. Probab.*, **2**, 792–804.

Beirlant, J., Vynckier, P. and Teugels, J.L. (1996a) Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, **91**, 1659–1667.

Beirlant, J., Vynckier, P. and Teugels, J.L. (1996b) Excess functions and estimation of the extreme value index. *Bernoulli*, **2**, 293–318.

Beirlant, J., Dierckx, G., Goegebeur Y. and Matthys, G. (1999) Tail index estimation and an exponential regression model. *Extremes*, **2**, 177–200.

Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002) On exponential representations of log-spacings of extreme order statistics. *Extremes*, **5**, 157–180.

Castillo, E. and Hadi, A.S. (1997) Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc.*, **92**, 1609–1620.

Coles, S.G. and Powell, E.A. (1996) Bayesian methods in extreme value modelling: a review and new developments. *Internat. Statist. Rev.*, **64**, 119–136.

Csörgő, S. and Viharos, L. (1998) Estimating the tail index. In B. Szyszkowicz (ed.), *Asymptotic Methods in Probability and Statistics*, pp. 833–881. Amsterdam: North-Holland.

Danielsson, J.L., de Haan, L., Peng, L. and de Vries, C.G. (2001) Using a bootstrap based method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*, **76**, 226–248.

Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B*, **52**, 393–442.

de Haan, L. and Rootzén, H. (1993) On the estimation of high quantiles. *J. Statist. Plann. Inference*, **35**, 1–13.

de Haan, L. and Stadtmüller, U. (1996) Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, **61**, 381–395.

Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833–1855.

Draisma, G., de Haan, L., Peng, L. and Pereira, T.T. (1999) A bootstrap based method to achieve optimality in estimating the extreme value index. *Extremes*, **2**, 367–404.

Drees, H. (1998) On smooth statistical tail functionals. *Scand. J. Statist.*, **25**, 187–210.

Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, **75**, 149–172.

Drees, H., de Haan, L. and Resnick, S. (2000) How to make a Hill plot. *Ann. Statist.*, **28**, 254–274.

Drees, H., Ferreira, A. and de Haan, L. (2004) On maximum likelihood estimation of the extreme value index. *Ann. Appl.Probab.*, **14**, 1179–1201.

Ferreira, A. (2002) Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plann. Inference*, **53**, 83–102.

Ferreira, A., de Haan, L. and Peng, L. (2003) On optimizing the estimation of high quantiles of a probability distribution. *Statistics*, **37**, 401–434.

Feuerverger, A. and Hall, P. (1999) Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760–781.

Fraga Alves, M.I., de Haan, L. and Lin, T. (2003) Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods Statist.*, **12**, 155–176.

Groeneboom, P., Lopuhaä, H.P. and de Wolf, P.P. (2003) Kernel-type estimators for the extreme value index. *Ann. Statist.*, **31**, 1956–1995.

Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, **44**, 37–42.

Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251–261.

Kratz, M. and Resnick, S. (1996) The qq-estimator and heavy tails. *Comm. Statist. Stochastic Models*, **12**, 699–724.

Mason, D.M. and Turova, T.S. (1994) Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner and E. Simiu (eds), *Extreme Value Theory and Applications*. Dordrecht: Kluwer Academic.

Pickands, J. III (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.

Resnick, S. and Stărică, C. (1997) Smoothing the Hill estimator. *Adv. Appl. Probab.*, **29**, 271–293.

Schultze, J. and Steinebach, J. (1996) On least squares estimates of an exponential tail coefficient. *Statist. Decisions*, **14**, 353–372.

Smith, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.

Smith, R.L. (1989) Extreme value analysis of environmental times series: an application to trend detection in ground-level ozone. *Statist. Sci.*, **4**, 367–393.

Vanroelen, G. (2003) *Generalized regular variation, order statistics and real inversion formulas*. Doctoral thesis, Katholieke Universiteit Leuven, Leuven, Belgium.

Received april 2002 and revised January 2005