Excess functions and estimation of the extreme-value index

JAN BEIRLANT*, PETRA VYNCKIER and JOZEF L. TEUGELS

Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

A general class of estimators of the extreme-value index is generated using estimates of mean, medium and trimmed excess functions. Special cases yield earlier proposals in the literature, such as Pickands' (1975) estimator. A particular restatement of the mean excess function yields an estimator which can be derived from the slope at the right upper tail from a generalized quantile plot. From this viewpoint algorithms can be constructed to search for the number of extremes needed to minimize the mean square error of the estimator. Basic asymptotic properties of this estimator are derived. The method is applied in case studies of size distributions for alluvial diamonds and of wind speeds.

Keywords: extreme value theory; parameter estimationweighted least squares

1. Introduction

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of positive independent and identically distributed (i.i.d.) observations from some distribution function F. Denote the order statistics based on the first n observations by

$$X_1^* \leqslant X_2^* \leqslant \cdots \leqslant X_n^*.$$

We assume that for some constants $a_n > 0$ and b_n and some $\gamma \in \mathbb{R}$

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

where $G_\gamma(x)$ denotes the so-called *extreme-value distributions*, given by

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

The parameter γ is called the *extreme-value index*.

We consider the problem of estimating γ under the model (1) from a sample of size n. Applications are numerous and can, for example, be found in econometrics, insurance, reliability theory, environmetrics, geology and climatology. An application to geology can be found in Caers *et al.* (1996). In that paper the data, which are sizes of precious stones, are modelled by a log-hyperbolic distribution, which can be shown to satisfy (1) with $\gamma > 0$. In the present paper we will demonstrate the results on two examples of alluvial diamond

*To whom correspondence should be addressed

deposits in Guinea, West Africa. The Bougban data set covers a part of the Bougban River more than 20 kilometres in length and consists of 683 stone sizes. The Bimboko deposit contains 356 data points. A second illustration is an application to wind-speed modelling. Data sets of daily fastest-mile speeds from three cities in the USA with different wind regimes will be analysed: Des Moines (Iowa), Grand Rapids (Michigan) and Albuquerque (New Mexico). The differences between the cities are reflected in the γ values: positive values of γ entail a heavy polynomially decreasing tail (referred to as a Pareto tail), a zero value can be regarded as a moderate tail behaviour, while negative values of γ allow for a distributional model with finite right end-point. Moreover, one can use formulae, as in Dekkers *et al.* (1989), for estimating large quantiles as well as the end-point of the underlying distribution.

We first refer to two estimators of γ that have been proposed under the general model as described above. Pickands (1975) estimates γ by

$$\hat{\gamma}_{k,n}^{\mathrm{P}} = \frac{1}{\log 2} \log \left(\frac{X_{n-\lfloor k/4 \rfloor}^* - X_{n-\lfloor k/2 \rfloor}^*}{X_{n-\lfloor k/2 \rfloor}^* - X_{n-k}^*} \right),$$

where k denotes the number of extremes used in the estimation procedure, and where $\lfloor x \rfloor$ denotes the integer part of x ($\lceil x \rceil = \lfloor x \rfloor + 1$). Pickands' estimator was extensively studied in Dekkers and de Haan (1989).

A second estimator, called *the moment estimator*, was proposed in Dekkers *et al.* (1989):

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

where

$$\beta_{k,n}^{(l)} = k^{-1} \sum_{j=1}^{k} (\log X_{n-j+1}^* - \log X_{n-k}^*)^l \qquad (l = 1, 2).$$

This estimator can be regarded as an adaptation of *Hill's estimator* (Hill 1975), $\beta_{k,n}^{(1)}$, which is only a consistent estimator for γ where one knows that $\gamma > 0$. Hill's estimator can be understood as follows. First, $\gamma > 0$ implies that X is of *Pareto type*, that is,

$$\frac{1 - F(x\lambda)}{1 - F(x)} \to \lambda^{1/\gamma} \qquad (2)$$

as $x \to \infty$ for all $\lambda > 1$. Introducing the so-called *mean residual life function*, e, of any positive random variable Y, given by

$$e(x) = E(Y - x | Y > x) \qquad (x > 0),$$

it is known that the mean residual life function of $Y = \log X$, with X of Pareto type, tends to γ as $x \to \infty$ (see, for example, Beirlant *et al.* 1996). Taking $x = \log X_{n-k}^*$ we find $\beta_{k,n}^{(1)}$ as a natural empirical estimator of the corresponding value of $e(x)$. The remaining terms in the moment estimator were then added to obtain consistency when the parameter γ is real-valued.

In the present context of estimating the real-valued parameter γ no really practical algorithms have yet been proposed in order to make an optimal choice of the number of extremes k in the estimation procedure. Theoretical work — see, for example, Haeusler and Teugels (1985) and Csörgő *et al.* (1985) in the Pareto case — has clearly shown that under a minimum mean square error criterion the optimal value of k typically depends on the particular tail at hand, and more specifically on the rate in (2) (see Bingham *et al.* 1987, Section 3.12.1). In practice, practitioners try to make inferences on γ from a stable part of the sequence of estimators of γ as a function of k. This, however, presents serious problems in non-simulated applications due to the high volatility of estimators appearing in extreme-value theory. This can be seen from the figures in earlier papers on the subject (for example, Dekkers *et al.* 1989) and in Section 4 of the current paper. Sometimes there is no stable part in the picture, and even if a stable part is present, the substantial variance makes it difficult to choose within it.

Apart from providing insight into the link between excess functions and novel estimators for the extreme-value index, the present paper seriously tries to remedy this last problem in the choice of k, crucial in applications. More specifically, we will show that generalized quantile plots can be constructed which, under (1), show an ultimately linear behaviour with slope approaching γ at the points corresponding to the extreme data. Estimating this slope by a regression fit then provides an estimator for γ. This is the subject of Section 2. In Section 3 we will then show how the optimal k can be tracked down by looking for the best linear regression fit through the quantile plot based on a weighted least-squares error criterion. In Section 4 we comment on the behaviour of our algorithm on some simulated data sets. We further illustrate the algorithm on the diamond data examples already mentioned, as well as on the US wind-speed data. Proofs and computational details are deferred to Appendices A and B.

2. Excess functions and estimators of the extreme value index

2.1. PARETO DISTRIBUTIONS

Before we introduce our new estimator(s), we recall some basic notions on the Pareto case ($\gamma > 0$) leading to the Hill estimator among others. In this particular case (1) can be restated in terms of the tail function U, defined as

$$U(x) = Q\left(1 - \frac{1}{x}\right) \qquad (x > 1)$$

where

$$Q(p) = F^i(p) = \inf\{y : F(y) \geqslant p\} \qquad (p \in (0,1)).$$

It was shown by de Haan (1970) that where $\gamma > 0$, (1) is equivalent to

$$U(x) = x^{\gamma}\ell(x) \tag{3}$$

where ℓ denotes a *slowing varying function*, that is, one which satisfies

$$\lim_{x \to \infty} \frac{\ell(\lambda x)}{\ell(x)} = 1 \qquad (\lambda > 0).$$

Relation (3) states that U is *regularly varying* with index γ (notation: $U \in \Re_\gamma$).

Based on a sample of size n, γ can be estimated consistently using relation (3) by inspecting the scatterplot with coordinates

$$\left(-\log\left(\frac{j}{n+1} \right), \log X^*_{n-j+1} \right) \qquad (j = 1, \ldots, n)$$

which can be seen as an approximation of the scatterplot

$$\left(-\log\left(\frac{j}{n+1} \right), \log U\left(\frac{n+1}{j} \right) \right) \qquad (j = 1, \ldots, n).$$

Indeed, under (3) with $\gamma > 0$, the theoretical quantiles $\log U((n+1)/j)$ ultimately (for $j/n \to 0$) stand in linear relation, with slope γ, to the corresponding quantiles $-\log(j/(n+1))$ of the standard exponential distribution. The basic property here is that for any slowly varying function l we have that $\log l(x)/\log x \to 0$ as $x \to \infty$.

The slope to the right of the point with coordinates $(-\log((k+1)/(n+1)), \log X^*_{n-k})$ is then typically estimated by the quotient of average increase

$$\frac{k^{-1} \sum_{j=1}^{k} \log X^*_{n-j+1} - \log X^*_{n-k}}{k^{-1} \sum_{j=1}^{k} \log\left(\frac{n+1}{j} \right) - \log\left(\frac{n+1}{k+1} \right)}$$

which leads to the Hill estimator

$$H_{k,n} = k^{-1} \sum_{j=1}^{k} \log X^*_{n-j+1} - \log X^*_{n-k}. \tag{4}$$

Indeed, the denominator in the above quotient can be safely approximated by 1 using Stirling's formula.

Beirlant *et al.* (1996) proposed to estimate the optimal k from the minimum value of the sequence of weighted mean square error expressions

$$\mathrm{MSE}(k) = \frac{1}{k} \sum_{j=1}^{k} w^{\mathrm{opt}}_{j,k} \left(\log\left(\frac{X^*_{n-j+1}}{X^*_{n-k}} \right) - \gamma \log\left(\frac{k+1}{j} \right) \right)^2,$$

where $w^{\mathrm{opt}}_{j,k}$ denotes the weights to be optimized in such a way that $\mathrm{MSE}(k)$ consistently estimates the asymptotic mean square error of the Hill estimator $H_{k,n}$.

2.2. GENERALIZED QUANTILE PLOTS

Next we show how the Pareto case of Section 2.1 can be extended to the general model (1) without prior knowledge of the sign of γ.

Our first theorem proposes functions that exhibit the regularly varying behaviour with index γ under (1) as does U in (3) in the Pareto model. The first part of the theorem states that when $\gamma < 1$, the *mean excess function* $e_U := e \circ U$ is regularly varying with index γ. The restriction to $\gamma < 1$ appears from the fact that otherwise e may not be defined. This restriction can be lifted in two ways. One way is to replace the empirical mean excess values obtained at $x = n/k$,

$$E_{k,n} = k^{-1} \sum_{j=1}^{k} X_{n-j+1}^* - X_{n-k}^*$$

by approximations

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

obtained by a one-term Taylor expansion of the logarithmic function around X_{n-k}^* (to be justified later). Remark then that $UH_{k,n}$ can be seen as an estimator of the product function $U(x)H(x)$ where $H(x) = e_{\log} \circ \log U(x)$ with $e_{\log}$ the mean residual life function of the log-transformed data. We henceforth write

$$UH(x) = U(x)H(x) = U(x) \int_1^{\infty} \{\log U(wx) - \log U(x)\} \frac{\mathrm{d}w}{w^2}.$$

Another way to lift the restriction is to replace $E_{k,n}$ with more robust versions, such as a generalized median excess functional

$$M_{k,n}^{(p)} = X_{n-\lfloor pk \rfloor}^* - X_{n-k}^* \qquad (0 < p < 1),$$

or a trimmed mean excess functional

$$E_{k,n}^{(p)} = \frac{1}{k - \lfloor pk \rfloor} \sum_{i=\lfloor pk \rfloor + 1}^{k} X_{n-i+1}^* - X_{n-k}^* \qquad (0 < p < 1).$$

These functions estimate

$$m_U^{(p)}(x) = U\left(\frac{x}{p}\right) - U(x)$$

and

$$e_U^{(p)}(x) = \frac{x}{1-p} \int_x^{x/p} \frac{U(v)}{v^2} \, \mathrm{d}v - U(x)$$

at $x = n/k$, respectively.

We can now state the analytic results on which our estimation procedures will be based. Proofs are deferred to Appendix A.

Theorem 1.

(*i*) *Restricting to the case* $\gamma < 1$, *F satisfies* (1) *if and only if* $e_U \in \Re\gamma$.

(*ii*) *If F satisfies* (1) *for some* $\gamma \in \mathbb{R}$ *then* $UH \in \Re_\gamma$.

(*iii*) *F satisfies* (1) *for some* $\gamma \in \mathbb{R}$ *if and only if for all* $0 < p < 1$, $m_U^{(p)} \in \Re_\gamma$ *and* $e_U^{(p)} \in \Re_\gamma$.

In the following we concentrate on the function UH and consequently derive estimators for γ. Note that the use of estimators based on $M_{k,n}^{(p)}$ or $E_{k,n}^{(p)}$ requires the estimation of the nuisance parameter p. Theorem 1(ii) then entails that under the general model (1) $\log UH(x) \sim \gamma \log x$ as $x \to \infty$. From this it follows that a generalized quantile plot

$$\left(-\log\left(\frac{j}{n}\right), \log UH_{j,n} \right) \qquad (j = 1, \ldots, n - 1) \tag{5}$$

eventually will be linear for the smaller *j* values.

In Figure 1 this is illustrated for the wind-speed data from three US cities. They are the daily fastest-mile speeds measured by anemometers 10 m above the ground. The line structures in the generalized quantile plots are the result of an inherent grouping of the data due to loss of accuracy during the data collecting process. For the Des Moines daily wind-speed maxima ($n = 5478$) the generalized quantile plot (5) clearly shows an increasing behaviour, which reflects a heavy tail for the underlying distribution. The flattening trend in the Grand Rapids data set ($n = 5478$) suggests a weaker tail with $\gamma = 0$, while for Albuquerque ($n = 6939$) even a negative γ value, resulting in a distribution with a finite right end-point, can be expected.

2.3. ESTIMATION OF THE EXTREME-VALUE INDEX

An estimate of the extreme-value index γ can thus be obtained by estimating the slope of the generalized quantile plot (5) from a point $(-\log(k/n), \log UH_{k,n})$ to the right of which linearity starts to appear. An estimation algorithm for the slope of the right tail of the quantile plot can be found by applying a right-sided weighted least-squares method. The problem of deciding on the number *k* of extremes to be used can hence be regarded as a diagnostic regression problem, that is, deciding on the point from which an 'optimal' linear fit is obtained through the quantile plot.

The equation of the line through $(-\log((k + 1)/n), \log UH_{k+1,n})$ with slope γ is given by

$$y = \log UH_{k+1,n} + \gamma\left(x + \log\left(\frac{k+1}{n}\right) \right).$$

The fitting of a straight line through the points with coordinates $(\log(j/n), \log UH_{j,n})$ ($j = 1, \ldots, k$) with a least-squares algorithm leads to the minimization of

$$\sum_{j=1}^{k} w_{j,k}\left(\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right) - \gamma \log\left(\frac{k+1}{j}\right) \right)^2$$
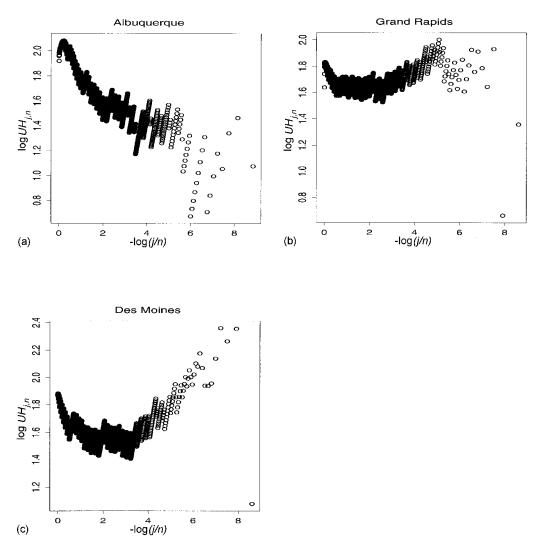
Figure 1.Generalised quantile plot (5). (a) Albuquerque; (b) Grand Rapids; (c) Des Moines

with respect to γ. Equating the first-order derivative with respect to γ to zero yields

$$\hat{\gamma}_{k,n} = \frac{k^{-1} \sum_{j=1}^{k} w_{j,k} \log\left(\frac{k+1}{j}\right) \{\log UH_{j,n} - \log UH_{k+1,n}\}}{k^{-1} \sum_{j=1}^{k} w_{j,k} \left(\log\left(\frac{k+1}{j}\right)\right)^2}. \tag{6}$$

Introduce the weight function $\tilde{w}(j, k) := w_{j,k} \log((k + 1)/j)$. The numerator in (6) can be rewritten as

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

If we take $w_{j,k}$ of the form $w(j/k)$ and introduce $K(t) = -t^{-1} \int_0^t w(x) \log x \, \mathrm{d}x$, then $j^{-1} \Sigma_{i=1}^{j} \tilde{w}(i, k)$ can be seen as an approximation of $K(j/k)$. The denominator in (6) is then a Riemann approximation of the integral $\int_0^1 K(t) \, \mathrm{d}t$.

We conclude that a weighted least-squares estimation method for the slope of the generalized quantile plot leads to a broad class of estimators for γ which can be compared with the class of kernel estimators introduced by Csörgő *et al.* (1985) for the subcase $\gamma > 0$:

$$\hat{\gamma}_{k,n} = \frac{\displaystyle\sum_{j=1}^{k} \frac{j}{k} K\left(\frac{j}{k}\right) \{\log UH_{j+1,n} - \log UH_{j,n}\}}{\displaystyle k^{-1} \sum_{j=1}^{k} K\left(\frac{j}{k}\right)} \tag{7}$$

where

$$K(t) = -\frac{1}{t} \int_0^t w(x) \log x \, \mathrm{d}x \qquad (t \in (0, 1))$$

is a smoothing kernel. Formal replacement of the UH values in (7) by $M^{(p)}$ and $E^{(p)}$ values yields other sets of estimators.

The choice of the uniform $(0,1)$ kernel in (7) yields the direct adaptation of Hill's estimator

$$H_{k,n}^{*2} := k^{-1} \sum_{i=1}^{k} \log UH_{j,n} - \log UH_{k+1,n}, \tag{8}$$

which can be rewritten as

$$H_{k,n}^{*2} = \tilde{H}_{k,n} + HH_{k,n}$$

where

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

is the result of a Hill-type operation on the Hill-type statistics, and

$$\tilde{H}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j}^* - \log X_{n-k-1}^*$$

is the Hill estimator with the largest observation deleted. This estimator can be compared with the moment estimator. Both estimators depart from the Hill estimator and add a

term to the Hill estimator in order to allow for an extreme-value index estimator in the full sense.

Pickands' (1975) estimator is a kernel estimator based on the median excess statistics $M_{k,n}^{(0.5)}$ with $K(u) = u^{-1}$ for $u \in (0.5, 1)$ and 0 elsewhere.

2.4. ASYMPTOTIC PROPERTIES OF $H_{k,n}^{*2}$

From now on we will concentrate on the above generalization $H_{k,n}^{*2}$ of the Hill estimator given in (8). We end this section with a statement of the basic asymptotic properties of this estimator of the extreme-value index.

In our next theorem we give some natural and fairly general conditions under which asymptotic normality holds for $H_{k,n}^{*2}$. Other results on the asymptotic normality of estimators based on trimmed mean excess or median excess estimators have been derived too and can be found in Vynckier (1996).

Generalizing the methods of proof developed in Dekkers *et al.* (1989), we derive approximations of an extreme-sum process which constitutes a somewhat different version of the process discussed in Mason and Turova (1994).

Let $\{\zeta_{k,n}(t); 1/k_n \leqslant t \leqslant 1\}$ denote the *mean excess process*, defined by

$$\zeta_{k,n}(t) = \sqrt{k_n} \left\{ \frac{X_{n-\lceil tk_n \rceil}^* k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} (\log X_{n-i+1}^* - \log X_{n-\lceil tk_n \rceil}^*)}{UH\left(\frac{n}{tk_n}\right)} - t \right\} \qquad (1/k_n \leqslant t \leqslant 1).$$

We will show that

$$\sqrt{k}(H_{k,n}^{*2} - \gamma) = \sqrt{k} \left(\int_0^1 \{\log UH_{\lceil tk_n \rceil, n} - \log UH_{k,n}\} dt - \gamma \right)$$

$$= \sqrt{k} \left(\int_0^1 \log \left(\frac{UH(n/tk)}{UH(n/k)} \right) dt - \gamma \right) + \int_{1/k}^1 t^{-1} \zeta_{k,n}(t) \, d\{tI(t)\} + o_P(1) \quad (9)$$

where I denotes the indicator of the unit interval and $\lim_{k/n \to 0} o_P(1) = 0$ in probability. The next theorem states an approximation theorem for the mean excess process which will allow us to obtain the asymptotic distribution of the estimator $H_{k,n}^{*2}$. The limit process will be distributed as

$$Z_\gamma(t) = \begin{cases} (1-\gamma)\left(t^\gamma \int_0^t W(s)s^{-1-\gamma} \, ds - W(t) \right) + \gamma t(\tilde{W}(1) - W(1)) & \gamma < 0 \\ (1+\gamma)W(t) + \gamma t(\tilde{W}(1) - W(1)) & \gamma \geqslant 0, \end{cases}$$

where $\{W(t); t \geqslant 0\}$ and $\{\tilde{W}(t); t \geqslant 0\}$ are independent standard Wiener processes.

We now state our key theorem upon which rely all proofs of the asymptotic normality of the estimators. We set, for any $\rho \leqslant 0$ and any $t > 0$,

$$k_\rho(t) = \begin{cases} \dfrac{t^\rho - 1}{\rho} & \text{if } \rho < 0 \\[2ex] \log t & \text{if } \rho = 0. \end{cases}$$

Theorem 2. *Suppose F satisfies* (1) *with $\gamma \in \mathbb{R}$; moreover,*

(*i*) *for $\gamma > 0$: there exist $\gamma_1 \leqslant 0$ and a positive function $b_1(b_1(x) \to 0$ as $x \to \infty)$ such that*

$$\frac{(tx)^{-\gamma} U(tx)}{x^{-\gamma} U(x)} - 1 \sim b_1(x) k_{\gamma_1}(t) \qquad (t \to \infty);$$

(*ii*) *for $\gamma = 0$: there exist positive functions $a, b_2(b_2(x) \to 0$ as $x \to \infty)$ such that*

$$\lim_{z \to \infty} \frac{U(x)(\log U(tx) - \log U(x)) - a(x) \log t}{b_2(x)} = \pm \frac{(\log t)^2}{2} \qquad (t \geqslant 1);$$

(*iii*) *for $\gamma < 0$; there exist $\gamma_3 \leqslant 0$ and a positive function, $b_3(b_3(x) \to 0$ as $x \to \infty)$ such that*

$$\frac{(tx)^{-\gamma}(U(\infty) - U(tx))}{x^{-\gamma}(Y(\infty) - U(x))} - 1 \sim b_3(x) k_{\gamma_3}(t) \qquad (x \to \infty).$$

Suppose also $\lim_{n \to \infty} k_n = \infty$ and:

(*iv*) *for $\gamma > 0 : \sup_{1/k_n \leqslant t \leqslant 1} \sqrt{k_n} b_1(n/t k_n) \to 0$* $(n \to \infty)$
(*v*) *for $\gamma = 0 : \sup_{1/k_n \leqslant t \leqslant 1} \sqrt{k_n} b_2(n/t k_n)/a(n/t k_n) \to 0$* $(n \to \infty)$
(*vi*) *for $\gamma < 0 : \sup_{1/k_n \leqslant t \leqslant 1} \sqrt{k_n} b_3(n/t k_n) \to 0$* $(n \to \infty)$.

Then for any $\delta \in (0, \frac{1}{2})$, on a suitable probability space there exist sequences $\{\tilde{\zeta}_{k,n}\}$ of probabilistically equivalent versions of $\{\zeta_{k,n}\}$ and a Gaussian process Z_γ as described above such that, as $n \to \infty$,

$$\sup_{1/k_n \leqslant t \leqslant 1} |t^{-1/2 + \delta}(\tilde{\zeta}_{k,n}(t) - Z_\gamma(t))| \to_P 0.$$

The assumptions of Theorem 2 are similar to those of Theorem 3.1 in Dekkers *et al.* (1989) and are found to be quite general. Some similar alternative conditions are discussed in the same paper.

By Karamata's theorem (see, for example, Bingham *et al.* 1987) statement (*ii*) of Theorem 1 is equivalent to stating that $UH(x)$ $(x > 0)$ can be written in the form

$$x^\gamma c(x) \exp\left(\int_1^x \frac{\varepsilon(u)}{u}\, \mathrm{d}u\right)$$

where $\varepsilon(x) \to 0$ $(x \to \infty)$ and where $c(x) \to c$ $(x \to \infty)$. In what follows we restrict ourselves to the case where $c(x)$ is a constant function, equal to 1, say, in which case we

speak about *normalized* regularly and slowly varying functions. It now follows that the expression for the bias of $H_{k,n}^{*2}$ as found in the first term on the right-hand side of (9) can be rewritten as

$$\int_0^1 \varepsilon\left(\frac{n}{kt}\right) dt. \tag{10}$$

If we use this analysis of the bias terms, we can now formulate the basic asymptotic result for our extreme-value index estimator.

Theorem 3. *When UH is a normalized regularly varying function with index γ such that $\sqrt{k_n} \int_0^1 \varepsilon(n/kt)\, dt \to 0$ as $n \to \infty$, we have under conditions (i)–(vi) of Theorem 2 that*

$$\sqrt{k_n}\{H_{k_n,n}^{*2} - \gamma\} \to_d \int_0^1 t^{-1} Z_\gamma(t)\, dt - Z_\gamma(1).$$

Calculation of the covariance structure of the Z_γ process leads to the asymptotic variance for the estimator:

$$\int_0^1 t^{-1} Z_\gamma(t)\, dt - Z_\gamma(1) =_d \begin{cases} N(0, (1+\gamma)^2) & \gamma \geqslant 0 \\ N\left(0, \dfrac{(1-\gamma)(1+\gamma+2\gamma^2)}{(1-2\gamma)}\right) & \gamma < 0. \end{cases}$$

3. Optimal choice of k

In this section we provide an algorithm for choosing k optimally. Through regression diagnostics such as mean square error goodness-of-fit plots, we will estimate the asymptotic mean square error (AMSE) of the adapted Hill estimator of the extreme value index given by

$$\mathrm{AMSE}(H_{k,n}^{*2}) = \frac{1}{k}\mathrm{var}\left(\int_0^1 t^{-1} Z_\gamma(t)\, dt - Z_\gamma(1)\right) + \left(\int_0^1 \varepsilon(n/kt)\, dt\right)^2.$$

3.1. CONDITIONS ON $\tilde{l}$

We will carry out the above-mentioned study under the following assumption on the slowly varying function $\tilde{l}$ defined by

$$\tilde{l}(x) = x^{-\gamma} UH(x).$$

Condition. *There exist a real constant $\rho \leqslant 0$ and a rate function g satisfying $g(x) \to 0$ as $x \to \infty$, such that, for all $\lambda > 1$,*

$$\frac{\tilde{\ell}(\lambda x)}{\tilde{\ell}(x)} - 1 \sim g(x) k_\rho(\lambda) \qquad (x \to \infty). \tag{11}$$

This condition is based on the concept of slow variation with remainder (see, for example, Section 3.12.1 of Bingham et al. 1987) and can be checked to follow from the conditions (i)–(iii) in Theorem 2.

Assumption (11) entails that as $x \to \infty$

$$\frac{x}{g(x)}\int_x^\infty \frac{\log \tilde{\ell}(t) - \log \tilde{\ell}(x)}{t^2}\, dt = \int_1^\infty \frac{\log\left(\frac{\tilde{l}(ux)}{\tilde{\ell}(x)}\right)}{g(x)}\frac{du}{u^2} \to \int_1^\infty \frac{u^\rho - 1}{\rho}\frac{du}{u^2}$$

$$= \frac{1}{\rho}\left\{\frac{1}{1-\rho} - 1\right\} = \frac{1}{1-\rho},$$

so that g can be represented as

$$g(x) \sim (1-\rho)x\int_x^\infty \frac{\log \tilde{\ell}(t) - \log \tilde{\ell}(x)}{t^2}\, dt$$

$$= (1-\rho)\left\{\left(x\int_x^\infty \frac{\log UH(t)}{t^2}\, dt - \log UH(x)\right) - \gamma\right\}.$$

Where the slowly varying function $\tilde{\ell}$ (see (2)) is normalized one can show (see, for example, 1.3.4 in Bingham et al. 1987) that

$$\varepsilon(x) = \frac{x\tilde{\ell}'(x)}{\tilde{\ell}(x)} \qquad \text{almost everywhere,}$$

from which it easily follows with (10) that

$$g\left(\frac{n}{k}\right) = (1-\rho)\mathrm{ABias}(H_{k,n}^{*2})$$

where $\mathrm{ABias}(H_{k,n}^{*2})$ denotes the asymptotic bias of $H_{k,n}^{*2}$ as $k/n \to 0$.

3.2. ESTIMATION OF THE NUISANCE PARAMETER ρ

The non-positive index ρ can easily be seen to be the index of regular variation of the function $(x\int_x^\infty t^{-2}\log UH(t)\, dt - \log UH(x)) - \gamma$.

To estimate the auxiliary index ρ, we will use the estimator

$$HR_{m,n} = \frac{\log\left|\dfrac{H_{\lfloor(m+k)/4\rfloor}^{*2} - H_{\lfloor(m+k)/2\rfloor}^{*2}}{H_{\lfloor m/2\rfloor}^{*2} - H_m^{*2}}\right|}{\log\left(\dfrac{2m}{m+k}\right)} \qquad (m = k+1,\ldots,n-2), \qquad (12)$$

where k denotes the position of an initial estimate $\hat{\gamma}_{k,n}$ of γ.

The estimator $HR_{m,n}$ can be seen as an estimator for the slope of the straight-line pattern in the plot

$$\left(-\log\left(\frac{m}{n}\right); \log|H^{*2}_{\lfloor m/2\rfloor} - H^{*2}_m|\right) \qquad (m = k,\ldots,n-2). \qquad (13)$$

In this, by using a Pickands' operation one avoids the estimation of the unknown parameter γ. Only the position k of an initial estimate is of importance as the estimates are only relevant for $m > k$. The position of m is determined by an unweighted least-squares principle.

This is illustrated for a simulation of size $n = 1500$ from the standard log-normal distribution on where $\rho = 0$. In this case the unweighted mean square error goodness-of-fit plot for (13) (Figure 2a), reaches its minimum at $m = 909$ (Figure 2b), which corresponds with the estimate $HR_{909,1500} = -0.311$ for ρ in (12) (Figure 2c).
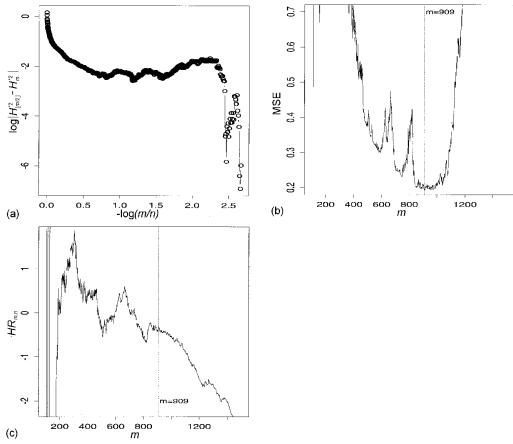

Figure 2. Simulation of 1500 standard lognormal data. (a) Plot (13) with $k = 104$; (b) unweighted mean square error: $\dfrac{1}{m}\sum_{j=1}^{m}\left(\log\left|\dfrac{H^{*2}_{j/2,n} - H^{*2}_{j,n}}{H^{*2}_{(m+1)/2,n} - H^{*2}_{m+1,n}}\right| - HR_{m,n}\log\dfrac{m+1}{j}\right)^2$; (c) Plot (12) of $HR_{m,n}$ against m with $k = 104$.

3.3. THE OPTIMAL MSE PLOTS

Our aim now is to estimate nonparametrically the AMSE of the adapted Hill estimator using a weighted mean square error expression

$$\text{MSE}_{\text{opt}}(k) = \frac{1}{k}\sum_{j=1}^{k} w_{j,k}^{\text{opt}}\left(\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right) - \gamma\log\left(\frac{k+1}{j}\right)\right)^2$$

for some sequence of weights $w_{j,k}^{\text{opt}}$, to be defined next, which depend on γ and ρ. The value of k_{opt} minimizing $\text{MSE}_{\text{opt}}(k)$ $(k = 1, 2, \ldots)$ is then used in estimating γ by $H_{k_{\text{opt}},n}^{*2}$. By this method we will avoid the direct estimation of $g(n/k)$, which would give rather unstable graphics. We also remark that the above-mentioned method in a sense corresponds to using an adaptive kernel estimator in order to decide on the optimal k.

The factor γ in $\text{MSE}_{\text{opt}}(k)$ denotes the real value of the extreme-value index which needs to be estimated. Also the weights $w_{j,k}^{\text{opt}}$ will be seen to depend on the unknown parameters γ and ρ. This problem can be circumvented by replacing them by a value $H_{k,n}^{*2}$ (or $HR_{m,n}$) of the corresponding Hill statistics for some value of k (or m). This will be done through an iterative procedure, described in Appendix B.

The optimal weights can be calculated as follows. For any sequence of weights $w_{j,k}$ we observe that

$$\text{MSE} = \frac{1}{k}\sum_{j=1}^{k} w_{j,k}\text{E}\left(\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right) - \gamma\log\left(\frac{k+1}{j}\right)\right)^2$$

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

$$\left. + \log\left(\frac{UH\left(\frac{n}{j}\right)}{UH\left(\frac{n}{k+1}\right)}\right) - \gamma\log\left(\frac{k+1}{j}\right)\right)^2$$

which can be approximated asymptotically using Theorem 2 and assumption (11) by

$$\text{MSE} \sim \frac{1}{k}\sum_{j=1}^{k} w_{j,k}\text{E}\left(\frac{1}{\sqrt{k+1}}\left\{\frac{Z_\gamma\left(\frac{j}{k+1}\right)}{\frac{j}{k+1}} - Z_\gamma(1)\right\}\right.$$

$$\left. + \log\left(\frac{\left(\frac{n}{j}\right)^\gamma \tilde{\ell}\left(\frac{n}{j}\right)}{\left(\frac{n}{k+1}\right)^\gamma \tilde{\ell}\left(\frac{n}{k+1}\right)}\right) - \gamma\log\left(\frac{k+1}{j}\right)\right)^2$$

$$= \frac{1}{k}\sum_{k=1}^{k} w_{j,k} \mathrm{E}\left(\frac{1}{\sqrt{k+1}}\left\{ \frac{Z_{\gamma}\left(\frac{j}{k+1}\right)}{\frac{j}{k+1}} - Z_{\gamma}(1)\right\} + g\left(\frac{n}{k+1}\right)\left\{ \frac{\left(\frac{j}{k+1}\right)^{-\rho}-1}{\rho}\right\}\right)^{2}$$

$$= \frac{1}{k}\sum_{j=1}^{k} w_{j,k}\frac{1}{k+1}\mathrm{E}\left(\frac{Z_{\gamma}\left(\frac{j}{k+1}\right)}{\frac{j}{k+1}} - Z_{\gamma}(1)\right)^{2}$$

$$+ (1-\rho)^{2}\frac{1}{k}\sum_{j=1}^{k} w_{j,k}\left(\frac{\left(\frac{j}{k+1}\right)^{-\rho}-1}{\rho}\right)^{2}\mathrm{ABias}^{2}(H_{k+1,n}^{*2})$$

$$= a_{k}\mathrm{Avar}(H_{k+1,n}^{*2}) + b_{k}\mathrm{ABias}^{2}(H_{k+1,n}^{*2})$$

where

$$a_{k} = \begin{cases} \dfrac{1}{k}\sum_{j=1}^{k} w_{j,k}\left(\left(\dfrac{j}{k+1}\right)^{-1}-1\right) & (\gamma \geqslant 0) \\[4mm] \dfrac{1}{k}\sum_{j=1}^{k} w_{j,k}\left(\left(\dfrac{j}{k+1}\right)^{-1} - \dfrac{4}{1+\gamma+2\gamma^{2}}\left(\dfrac{j}{k+1}\right)^{-\gamma} + \dfrac{3-\gamma-2\gamma^{2}}{1+\gamma+2\gamma^{2}}\right) & (\gamma < 0) \end{cases}$$

$$b_{k} = (1-\rho)^{2}\frac{1}{k}\sum_{j=1}^{k} w_{j,k}\left(\frac{\left(\frac{j}{k+1}\right)^{-\rho}-1}{\rho}\right)^{2}.$$

Hence, for two linear independent sequences of weights, $w_{j,k}^{(1)}$ and $w_{j,k}^{(2)}$, with corresponding constants $a_{k}^{(i)}$ and $b_{k}^{(i)}$ $(i=1,2)$, the optimal weights are obtained by taking linear combinations:

$$w_{j,k}^{\mathrm{opt}} = \delta_{1,k}w_{j,k}^{(1)} + \delta_{2,k}w_{j,k}^{(2)}$$

where

$$\delta_{1,k} = \frac{b_{k}^{(2)}-a_{k}^{(2)}}{a_{k}^{(1)}b_{k}^{(2)}-b_{k}^{(1)}a_{k}^{(2)}}, \qquad \delta_{2,k} = \frac{a_{k}^{(1)}-b_{k}^{(1)}}{a_{k}^{(1)}b_{k}^{(2)}-b_{k}^{(1)}a_{k}^{(2)}}.$$

The authors found the choice of the weight functions $w_{j,k}^{(1)} = j/(k+1)$ and $w_{j,k}^{(2)} = -\log(j/(k+1))$ to be quite appropriate.
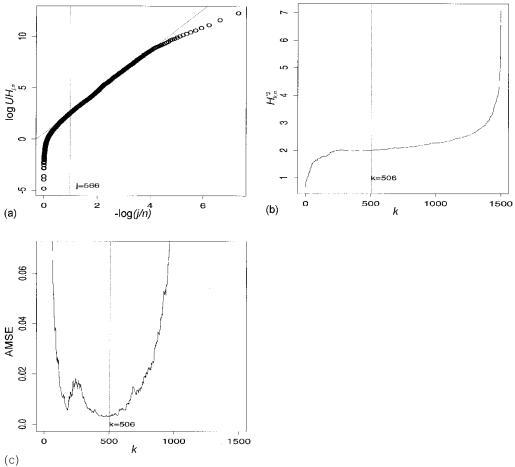
4. Some simulations and a practical example

In this section the iteration procedure will be demonstrated for three simulated data sets and for the Bimboko example, using the weights $(j/(k+1))$, $-\log(j/(k+1))$.
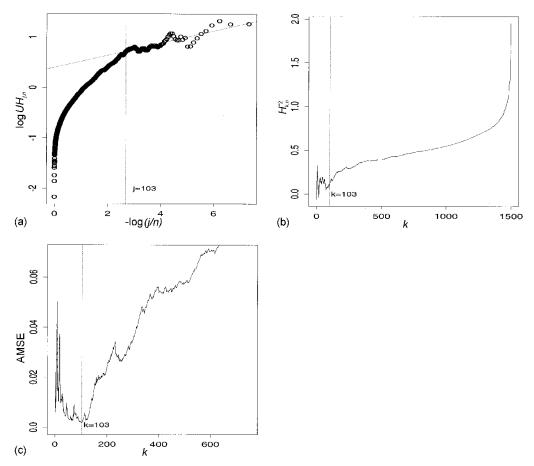
4.1. A SIMULATION FOR POSITIVE γ

The first simulation of size $n = 1500$ is drawn from a Burr distribution with $\gamma = 2$:

$$1 - F(x) = \left(\frac{1}{1+x}\right)^2 \qquad (x > 0).$$

Figure 3. Simulated Burr dataset. (a) Generalised quantile plot (5) with fitted line at position $j = 566$; (b) $H_{k,n}^{*2}$ versus k; (c) estimated $AMSE(H_{k,n}^{*2})$ versus k in the final iteration step.

Figure 4. Simulated standard lognormal dataset. (a) Generalised quantile plot with fitted line at position $j = 103$; (b) $H_{k,n}^{*2}$ versus k; (c) estimated $AMSE(H_{k,n}^{*2})$ versus k in the final iteration step.

The fitted line on the generalized quantile plot (Figure 3a) has been fixed at the initial position $k_0 = 566$ with corresponding value $H_{566,1500}^{*2} = 2.010$. In Figure 3b the estimator $H_{k,n}^{*2}$ is plotted against k. The second and final iteration step, shown in Figure 3c, reaches its minimum at $\hat{k}_{opt} = 506$ with corresponding $H_{506,1500}^{*2} = 1.993$.

As seen in Figure 3c some local minima in the estimated AMSE can show up. A simulation study confirmed that such secondary minima are dominated by the 'right' global minimum during the iteration.

4.2. A SIMULATION FOR $\gamma = 0$

For the log-normal data set of Figure 2 with $\gamma = 0$, the iteration also stopped after two

steps. In Figure 4b the estimator $H_{k,n}^{*2}$ is plotted against k, while Figure 4c shows the final estimator for the asymptotic MSE. Here $\hat{k}_{opt}$ is 103, leading to the estimate 0.119 for γ. This final estimate has been used for the fitted line on the generalized quantile plot (Figure 4a).

We notice that the log-normal distribution is one of the most difficult cases in extreme-value statistics in the class $\gamma = 0$, as the slowly varying function $\tilde{l}$ in (11) behaves logarithmically or, equivalently, $\rho = 0$. This entails that the bias disappears logarithmically slow. Hence, a relatively small value of k_{opt} has to be expected, which is reflected in the estimated AMSE plot. Of course, in practice one will have at one's disposal more specific techniques, such as log-normal, exponential and Weibull quantile plots.
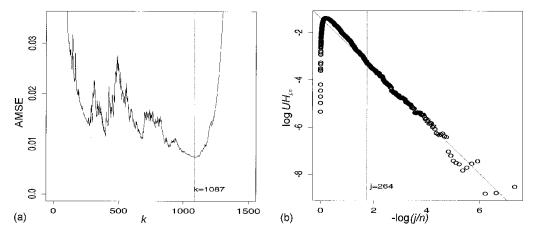
4.3. A SIMULATION FOR NEGATIVE γ

A third simulation is a sample of 1500 data points from a reversed Burr distribution where $\gamma = -1$, defined by

$$1 - F(x) = \left(\frac{2}{1 + (1-x)^{-0.5}} \right)^2 \qquad (0 < x < 1).$$

For this example five iteration steps were needed in order to estimate the optimal k value. Starting with an initial estimate $\gamma_0 = -1.212$ at $k_0 = 1087$, the position of the minimum of the unweighted mean square error (Figure 5a), the iterative procedure resulted in a final estimate at position $\hat{k}_{opt} = 264$, yielding an estimate $H_{264,1500}^{*2}$ of -1.099 for γ. The corresponding fitted line on the generalized quantile plot (5) is shown in Figure 5b.

4.4. THE BIMBOKO DEPOSIT

For our final example we consider the Bimboko dataset introduced in Section 1. The clearly increasing trend in the generalized quantile plot (5) (Figure 6) convinces us of the Pareto-type behaviour of the underlying distribution. This result corresponds with the geological finding of the adequacy of the log-hyperbolic distribution for describing sizes of diamonds in situations such as the alluvial diamond deposit at the Bimboko River. In geological literature the log-normal distribution for homogeneous deposits and mixtures of this distribution for non-homogeneous deposits have been put forward. Due to the sorting process during transportation by the river, it is plausible to assume a linear relation between the logarithmic mean and variance. This constraint, together with the choice of the generalized inverse Gaussian distribution as mixing distribution for the variance, results in the log-hyperbolic distribution (Barndorff-Nielsen 1977). For the Bimboko data set, the iterative procedure for finding the optimal k value stopped after two steps at position $\hat{k}_{opt} = 209$, yielding an estimate of 0.716 for γ, which is the extreme-value index of the log-hyperbolic distribution. The corresponding fitted line on the generalized quantile plot has been overlaid in Figure 6.

Figure 5. Simulated reversed Burr dataset. (a) Unweighted mean squares error on generalised quantile plot; (b) generalised quantile plot (5) with fitted line at position $j = 264$.

4.5. CONCLUDING REMARKS

1. In order to estimate ρ adequately, the position k of the initial value for γ in (12) must not be too large. For this reason k can be taken at a local minimum in the mean square error, rather than at the global minimum. This was the case for the reversed Burr data set (see Figure 5a) where in the first iteration step k has been set at 399, which is the local minimum, rather than at 1087. For the initial value γ_0 itself, $H^{*2}_{1087,1500} = -1.212$.
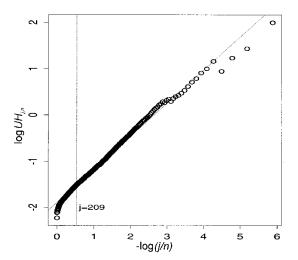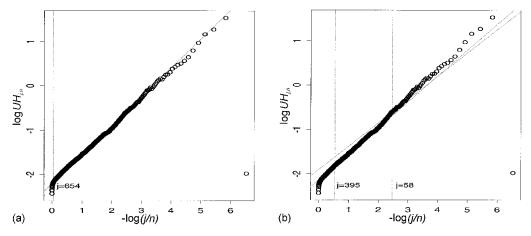

Figure 6. Bimboko deposit. Generalised quantile plot (5) with fitted line at position $j = 209$.

Figure 7. Bougban deposit. Generalised quantile plot. (a) Fitted line at position $j = 654$; (b) Fitted line at position $j = 58$ and $j = 395$.

2. The estimation of the ρ parameter in the iteration is of secondary importance. The procedure used above was found to be quite robust to variations in this ρ parameter.

3. Finally, we comment on the choice of the weights $w^{(1)}$, $w^{(2)}$. By simulations it turned out that the weights $(j/(k + 1), -\log(j/(k + 1))$ gave the most stable numerical results. However, in the presence of outliers in the generalized quantile plot (5), the weights $(1, -\log(j/(k + 1))$ gave better results. This can be seen in Figure 7, where the results of the iteration procedure are shown for the Bougban data set. In Figure 7a the 'best fitted' line through the generalized quantile plots for the weights $(1, -\log(j/(k + 1))$ was reached after three iteration steps and gives an estimate of 0.651 for γ at $k_{\mathrm{opt}} = 654$. The procedure with weights $(j/(k + 1), -\log(j/(k + 1))$ ended after two iteration steps in a loop, jumping from $k = 58$ ($H^{*2}_{58,683} = 0.526$) to $k = 395$ ($H^{*2}_{395,683} = 0.584$). The corresponding estimates of γ give unsatisfactory fits, as can be seen in Figure 7b.

Appendix A: Proofs

Proof of Theorem 1

We confine ourselves to the proof of (ii); demonstrations of the other parts can be found in Vynckier (1996).

First suppose that F satisfies (1). Then,

$$UH(x) = U(x) \int_1^\infty \{\log U(wx) - \log U(x)\} \frac{\mathrm{d}w}{w^2}.$$

Hence Lemma 2.5 in Dekkers *et al.* (1989), together with the dominated convergence

theorem, yields that, for some positive function a,

$$\frac{UH(x)}{a(x)} = U(x) \int_1^\infty \left[\frac{\log U(tx) - \log U(x)}{a(x)} \right] t^{-2}\,\mathrm{d}t \to \begin{cases} \int_1^\infty t^{-2} \log t\,\mathrm{d}t = 1 & (\text{if } \gamma \geqslant 0) \\[2mm] \int_1^\infty \left[\frac{t^\gamma - 1}{\gamma} \right] t^{-2}\,\mathrm{d}t = \frac{1}{1-\gamma} & (\text{if } \gamma < 0). \end{cases}$$

It also follows from Lemma 2.5 in Dekkers *et al.* (1989) that under (1) a is regularly varying with index γ which implies that UH is regularly varying with index γ. $\qquad\square$

Proof of Theorem 2

We confine ourselves to the proof of Theorem 2 where $\gamma < 0$. The other results can be proved in analogous fashion. Recall the following result: let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables with common distribution function $1 - 1/x$ $(x > 1)$; then

$$(X_1, X_2, \ldots) =_\mathrm{d} (U(Y_1), U(Y_2), \ldots)$$

and for all n,

$$(X_1^*, X_2^*, \ldots, X_n^*) =_\mathrm{d} (U(Y_1^*), U(Y_2^*), \ldots, U(Y_n^*)).$$

Furthermore, the random variables $Y_i^\gamma (i = 1, \ldots, n)$ have as common distribution function $F_\gamma(x) = x^{-1/\gamma}$ $(0 < x < 1)$. Moreover, the vector

$$\left(\left(\frac{Y_{n-k_n+1}^*}{Y_{n-k_n}^*} \right)^\gamma, \left(\frac{Y_{n-k_n+2}^*}{Y_{n-k_n}^*} \right)^\gamma, \ldots, \left(\frac{Y_n^*}{Y_{n-k_n}^*} \right)^\gamma \right)$$

has the same distribution as the vector of order statistics $(Z_{k_n}^*, \ldots, Z_1^*)$ of a sample of size k_n from F_γ, independent of $Y_{n-k_n}^*$.

We denote the corresponding quantile functions by Q_γ and the empirical quantile function $\hat{Q}_{k_n}$, that is,

$$\hat{Q}_{k_n}(t) = Z_j^* \text{ if } \frac{j-1}{k_n} < t \leqslant \frac{j}{k_n}, \qquad (j = 1, \ldots, k_n).$$

The empirical quantile process q_{k_n} based on $\hat{Q}_{k_n}$ will now be of basic importance:

$$q_{k_n}(t) = -\frac{t^{\gamma+1}}{\gamma} \sqrt{k_n} \{ \hat{Q}_{k_n}(t) - Q_\gamma(t) \}, \qquad (0 \leqslant t < 1).$$

Lemma 1. *If* $\{ B(t); 0 \leqslant t \leqslant 1 \}$ *is a Brownian bridge, then for any* $\delta_1 \in (0, \frac{1}{2})$

$$\sup_{0 \leqslant t \leqslant 1} \left| \frac{B(t)}{t^{-\delta_1+1/2}} \right| = O_\mathrm{P}(1).$$

For each n one can define a Brownian bridge $\{B_n(t); 0 \leqslant t \leqslant 1\}$ such that for any $\delta_2 \in (0, \frac{1}{2})$

$$\sup_{0 \leqslant c \leqslant 1} k_n^{1/2-\delta_2} |q_{k_n}(t) - B_n(t)| \to_P 0.$$

Moreover, for each n one can define a Wiener process $\{\tilde{W}_n(t); 0 \leqslant t \leqslant 1\}$, independent of B_n, such that

$$\left| \sqrt{k_n} \left\{ \left(\frac{k_n}{n} Y^*_{n-k_n} \right)^\gamma - 1 \right\} - \gamma \tilde{W}_n(1) \right| \to_P 0.$$

Proof. The first statement can be derived for example from (2.6) in Mason and Turova (1994). The second statement follows immediately from Theorem 4.5.7 in M. Csörgő and Révész (1981). Finally, the method of proof of statement (2.9) in Mason and Turova (1994) can be applied to the specific distribution at hand, but can also be generalized in this case to all values of γ, in order to derive the final statement in the Lemma.   □

First, we decompose $\zeta_{k,n}$ as

$$\zeta_{k,n}(t) = \sqrt{k_n} \left(\frac{a(Y^*_{n-\lfloor tk_n \rfloor})}{a\left(\frac{n}{tk_n}\right)} - 1 \right) (1 - \gamma)$$

$$\times \left(k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} \frac{U(Y^*_{n-\lceil tk_n \rceil})(\log U(Y^*_{n-i+1}) - \log U(Y^*_{n-\lceil tk_n \rceil}))}{a(Y^*_{n-\lfloor tk_n \rfloor})} \right)$$

$$+ \sqrt{k_n} \left((1 - \gamma) k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} \frac{U(Y^*_{n-\lceil tk_n \rceil})(\log U(Y^*_{n-i+1}) - \log U(Y^*_{n-\lceil tk_n \rceil}))}{a(Y^*_{n-\lfloor tk_n \rfloor})} - t \right)$$

$$+ \sqrt{k_n} \left(\frac{a\left(\frac{n}{tk_n}\right)}{UH\left(\frac{n}{tk_n}\right)} - (1 - \gamma) \right) \left(k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} \frac{X^*_{n-\lceil tk_n \rceil}(\log X^*_{n-i+1} - \log X^*_{n-\lceil tk_n \rceil})}{a\left(\frac{n}{tk_n}\right)} \right)$$

$$=: \Delta_n^{(1)}(t) + \Delta_n^{(2)}(t) + R_n^{(1)}(t).$$

We now show that under the given conditions for any $\delta \in (0, \frac{1}{2})$ as $n \to \infty$

$$\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta} |\Delta_n^{(1)}(t) - (\gamma B_n(t) - \gamma \tilde{W}(1))| \to_P 0, \tag{A.1}$$

$$\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta} |\Delta_n^{(2)}(t) - ((1-\gamma)t^\gamma \int_0^t B_n(s)s^{-1-\gamma} \, ds - B_n(t))| \to_P 0. \tag{A.2}$$

$$\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta} |R_n^{(1)}(t)| \to_P 0. \tag{A.3}$$

To prove (A2), take $\delta \in (0, \frac{1}{2})$. Observe that

$$
\Delta_n^{(2)}(t) = \sqrt{k_n} \left((1-\gamma)k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} \frac{U(Y_{n-\lceil tk_n \rceil}^*) \left(\log U\left(\left\{ \frac{Y_{n-i+1}^*}{Y_{n-\lceil tk_n \rceil}^*} \right\} Y_{n-\lceil tk_n \rceil}^* \right) - \log U(Y_{n-\lceil tk_n \rceil}^*) \right)}{a(Y_{n-\lfloor tk_n \rfloor}^*)} - t \right)
$$

$$
= \sqrt{k_n} \left(\frac{1-\gamma}{\gamma} k_n^{-1} \sum_{i=1}^{\lceil tk_n \rceil} \left[\left(\frac{Y_{n-i+1}^*}{Y_{n-\lceil tk_n \rceil}^*} \right)^\gamma - 1 \right] - t \right) + R_n^{(2)}(t)
$$

$$
=_\mathrm{d} \sqrt{k_n} \left(\frac{1-\gamma}{\gamma} k_n^{-1} \left(\sum_{i=1}^{\lceil tk_n \rceil} Z_i^* \right) \Big/ Z_{\lceil tk_n \rceil+1}^* - \gamma^{-1} t \right) + R_n^{(2)}(t) + R_n^{(3)}(t)
$$

where $\sup_{t \in (0,1)} |R_n^{(3)}(t)| = O(k_n^{-1/2})$ as $n \to \infty$ and $k_n \to \infty$.

Since $Z_{\lceil tk_n \rceil+1}^* = \hat{Q}_{k_n}(t)$ and $k_n^{-1} \Sigma_{i=1}^{\lceil tk_n \rceil} Z_i^* = \int_{k^{-1}}^t \hat{Q}_k(v)\,\mathrm{d}v + O_\mathrm{P}(k_n^{-1})$ uniformly in $t \in [k_n^{-1}, 1]$, we obtain

$$
\Delta_n^{(2)}(t) - \left((1-\gamma)t^\gamma \int_0^t B_n(v)v^{-1-\gamma}\,\mathrm{d}v - B_n(t) \right)
$$

$$
= (1-\gamma)t^\gamma \int_{k^{-1}}^t (q_k(v) - B_n(v))v^{-1-\gamma}\,\mathrm{d}v - (q_k(t) - B_n(t))
$$

$$
+ \left((1-\gamma)t^\gamma \int_{k^{-1}}^t q_k(v)v^{-1-\gamma}\,\mathrm{d}v - q_k(t) \right) \left(\frac{t^{-\gamma}}{\hat{Q}_k(t)} - 1 \right)
$$

$$
+ \sqrt{k} \left(\frac{1-\gamma}{\gamma} \int_0^t Q_\gamma(v)\,\mathrm{d}v - \gamma^{-1}tQ_\gamma(t) \right) \frac{1}{\hat{Q}_k(t)} + R_n^{(4)}(t)
$$

$$
=: S_1(t) + S_2(t) + S_3(t) + R_n^{(4)}(t).
$$

where $\sup_{t \in (0,1)} |R_n^{(4)}(t)| = O(k_n^{-1/2})$ as $n \to \infty$ and $k_n \to \infty$.

From $Q_\gamma(v) = v^{-\gamma}$ one checks that $S_3(t) = 0$.

Lemma 1 entails that, for every $t \in [1/k_n, 1]$,

$$
t^{-1/2+\delta} |S_1(t)| \leqslant C_\gamma \sup_{1/k_n \leqslant t \leqslant 1} \frac{|q_{k_n}(t) - B_n(t)|}{t^{1/2-\delta}}
$$

where C_γ is a constant depending on γ and δ. The right-hand side of this expression tends to zero in probability as $n \to \infty$ and $k_n \to \infty$.

Application of Lemma 1 with $\delta_2 = \delta/4$ shows that uniformly in $t \in [1/k_n, 1]$

$$
q_{k_n}(t) = B_n(t) + O_\mathrm{P}(k_n^{\delta/4-1/2}),
$$

or

$$\frac{t^{-\gamma}}{\hat{Q}_{k_n}(t)} - 1 = -\gamma k_n^{-1/2} t^{-1} B_n(t) + O_{\mathrm{P}}(k_n^{-1+\delta/4})|\gamma|t^{-1} \qquad (k_n \to \infty).$$

To show that $\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta}|S_2(t)|$ tends to zero in probability as $n \to \infty$ and $k_n \to \infty$, we write

$$S_2(t) = S_1(t)\left(\frac{t^{-\gamma}}{\hat{Q}_k(t)} - 1\right) + S_4(t)$$

with

$$S_4(t) = \left((1-\gamma)t^\gamma \int_0^t B(v)v^{-1-\gamma}\,\mathrm{d}v - B(t)\right)\left(\frac{t^{-\gamma}}{\hat{Q}_k(t)} - 1\right).$$

First,

$$t^{-1/2+\delta}|S_1(t)|\left|\frac{t^{-\gamma}}{\hat{Q}_k(t)} - 1\right|$$

$$\leqslant |\gamma t^{-1/2+\delta/2} S_1(t)|\left(k^{-1/2}t^{-1/2} \sup_{1/k_n \leqslant t \leqslant 1}\left|\frac{B(t)}{t^{1/2-\delta/2}}\right| + t^{-1+\delta/2}R_n^{(5)}(t)\right)$$

where $\sup_{t \in [1/k_n, 1]}|R_n^{(5)}(t)| = O(k_n^{-1+\delta/4})$ as $k_n \to \infty$. Using the first statement in Lemma 1 with $\delta_1 = \delta/2$ and the above observations about S_1, one obtains that for $n \to \infty$, $k_n \to \infty$, uniformly in $t \in [1/k, 1]$ the first term in the expansion of S_2 tends to zero in probability. Finally, using the same methods as before, one shows that

$$\sup_{1/k_n \leqslant t \leqslant 1} t^{\delta/2-1/2}\left|(1-\gamma)t^\gamma \int_0^t B(v)v^{-1-\gamma}\,\mathrm{d}v - B(t)\right| = O_{\mathrm{P}}(1),$$

so that with the help of Lemma 1 one finds that

$$\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta}|S_4(t)| = O_{\mathrm{P}}(k_n^{-\delta/4}) \qquad (k_n \to \infty).$$

To derive $\sup_{1/k_n \leqslant t \leqslant 1} t^{-1/2+\delta}|R_n^{(2)}(t)| \to_{\mathrm{P}} 0$, one basically relies on Lemma 3.5 in Dekkers *et al.* (1989). Where $\gamma < 0$ part (iii) of that lemma can be invoked assuming (iii) in Theorem 2. Hence we have to show that, uniformly in $t \in [1/k_n, 1]$ as $n \to \infty$, $k_n \to \infty$,

$$\sqrt{k_n}b_3(Y_{n-\lceil tk_n \rceil}^*)t^{-1/2+\delta}A_n(t) \to_{\mathrm{P}} 0,$$

where $A_n(t)$ is a linear combination of terms of the form

$$\frac{1}{k_n}\sum_{i=1}^{\lceil tk_n \rceil}\left(\frac{Y_{n-i+1}^*}{Y_{n-\lceil tk_n \rceil}^*}\right)^\nu$$

where ν is a positive constant. Again with the preceding argument concerning (A.2) we derive that $t^{-1/2+\delta}A_n(t)$ is bounded in probability, uniformly in $t \in [1/k_n, 1]$ as $n \to \infty$, $k_n \to \infty$.

Finally, it follows from assumption (vi) and the last statement in Lemma 1 that $\sqrt{k_n}b_3(Y^*_{n-\lceil tk_n\rceil}) \to_P 0$, $(n \to \infty, k_n \to \infty)$. This concludes the proof of (A.2).

The proof of (A.1) and (A.3) can now be given using the same techniques as in the preceeding derivations. □

Appendix B: Algorithm

In the following, the iterative procedure to find the AMSE is described. Here $w^{\mathrm{opt}}_{j,k}(\gamma_\eta, \rho_\eta)$ denotes the value of $w^{\mathrm{opt}}_{j,k}$ with γ and ρ replaced by estimates γ_η, ρ_η.

1. Determine k_0 minimizing

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

2. Let $\gamma_0 = H^{*2}_{k_0,n}$.
3. Plot $(-\log(m/n); \log|H^{*2}_{\lfloor m/2\rfloor} - H^{*2}_m|)$ $(m = k, \ldots, n-2)$ and determine m_0 by minimizing

$$\frac{1}{m-k_0}\sum_{i=k_0}^{m-1}\left(\log\left|\frac{H^{*2}_{\lfloor i/2\rfloor} - H^{*2}_i}{H^{*2}_{\lfloor m/2\rfloor} - H^{*2}_m}\right| - HR_{m,n}\log\left(\frac{m}{i}\right)\right)^2.$$

4. For $\eta = 1, 2, \ldots$:
   (a) determine k_η minimizing

$$\frac{1}{k}\sum_{j=1}^{k}w^{\mathrm{opt}}_{j,k}(\gamma_{\eta-1}, \rho_{\eta-1})\left(\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right) - \gamma_{\eta-1}\log\left(\frac{k+1}{j}\right)\right)^2$$

$$\text{for } k = \left\lfloor\frac{\eta}{100}\right\rfloor, \ldots, n \tag{14}$$

   (b) let $\gamma_n = H^{*2}_{k_\eta,n}$;
   (c) let $\rho_\eta = HR_{m_\eta,n}$ with m_η determined by minimizing

$$\frac{1}{m-k_\eta}\sum_{i=k_\eta}^{m-1}\left(\log\left|\frac{H^{*2}_{\lfloor i/2\rfloor} - H^{*2}_i}{H^{*2}_{\lfloor m/2\rfloor} - H^{*2}_m}\right| - HR_{m,n}\log\left(\frac{m}{i}\right)\right)^2.$$

5. Let $k_{\mathrm{opt}} = k_\eta$ when $|k_\eta - k_{\eta-1}|$ is smaller than some tolerance level.

Acknowledgements

References

Barndorff-Nielsen, O. (1977) Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London Ser. A*, **353**, 401–419.

Beirlant, J., Vynckier, P. and Teugels, J.L. (1996) Tail index estimation, Pareto quantile plots and regression diagnostics. *J. Amer. Statist. Assoc.* To appear.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.

Caers, J., Vynckier, P., Beirlant, J. and Rombouts, L. (1996) Extreme value analysis of diamond size distributions. *Math. Geol.*, **28**, 25–43.

Csörgő, S., Deheuvels, P. and Mason, D. (1985) Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050–1077.

Csörgő, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics*. Budapest: Akadémiai Kiado.

de Haan, L. (1970) *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. Math. Centre Tract 32. Amsterdam: Centrum voor Wiskunde en Informatica.

Dekkers, A.L.M. and de Haan, L. (1989) On the estimation of the extreme-value index and large quantile estimation, *Ann. Statist.*, **17**, 1795–1833.

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.

Haeusler, E. and Teugels, J.L. (1985) On asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.*, **13**, 743–756.

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

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

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

Vynckier, P. (1996) Tail estimation, quantile plots and regression diagnostics. PhD thesis, KU Leuven.