Extreme Value Index Estimators and Smoothing Alternatives: A Critical Review

Tsourti, Zoi and Panaretos, John

2003

EXTREME VALUE INDEX ESTIMATORS AND SMOOTHING ALTERNATIVES: A CRITICAL REVIEW

John Panaretos and Zoi Tsourti
Department of Statistics
Athens University of Economics and Business, Greece

1. Introduction

Extreme value theory is of major importance in many fields of applications where extreme values may appear and have detrimental effects. Such fields range from hydrology (Smith, 1989, Davison & Smith, 1990, Coles & Tawn, 1996, Barão & Tawn, 1999) to insurance (Beirlant et al., 1994, Mikosch, 1997, McNeil, 1997, Rootzen & Tajvidi, 1997) and finance (Danielsson & de Vries, 1997, McNeil, 1998; 1999, Embrechts et al., 1998; 1999, Embrechts, 1999). Actually, extreme value theory is a blend of a variety of applications and sophisticated mathematical results on point processes and regular varying functions.

The cornerstone of extreme value theory is Fisher-Tippet's theorem for limit laws for maxima (Fisher & Tippet, 1928). According to this theorem, if the maximum value of a distribution function (d.f.) tends (in distribution) to a non-degenerate d.f. then this limiting d.f. can ***only*** be the Generalized Extreme Value (GEV) distribution:

$$H_\theta(x) = H_{\gamma;\mu,\sigma}(x) = \exp\left\{-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\},$$

where $1+\gamma\frac{x-\mu}{\sigma} > 0$, and $\theta = (\gamma,\mu,\sigma) \in \Re\times\Re\times\Re_+$.

A comprehensive sketch of the proof can be found in Embrechts et al. (1997).

The random variable (r.v.) X (the d.f. F of X, or the distribution of X) is said to belong to the maximum domain of attraction of the extreme value distribution H_γ if there exist constants $c_n > 0$, $d_n \in \Re$ such that $c_n^{-1}(M_n - d_n) \xrightarrow{d} H_\gamma$ holds. We write $X \in MDA(H_\gamma)$ (or $F \in MDA(H_\gamma)$).

In this chapter, we deal with the estimation of the shape parameter γ known also as the *tail index or the extreme-value index*. In section 2 the general theoretical background is provided. In section 3, several existing estimators for *γ* are presented, while, in section 4, some smoothing methods on specific estimators are given and extended to other estimators, too. In section 5, techniques for dealing with the issue of choosing the threshold value of k, the number of upper order statistics required for the stimationj of γ, are discussed. Finally, concluding remarks are given in section 6.

2. Modelling Approaches

A starting point for modelling the extremes of a process is based on distributional models derived from asymptotic theory. The *parametric approach* to modelling extremes is based on the assumption that the data in hand (X_1, X_2, ..., X_n) form an i.i.d. sample from an exact GEV d.f. In this case, standard statistical methodology from parametric estimation theory can be utilized in order to derive estimates of the parameters θ. In practice, this approach is adopted whenever the dataset consists of maxima of independent samples (e.g., in hydrology where we have disjoint time periods). This method is often called method of *block maxima* (initiated by Gumbel, 1958). Such techniques are discussed in DuMouchel (1983), Hosking (1985), Hosking et al. (1985), Smith (1985), Scarf (1992), Embrechts et al. (1997) and Coles and Dixon (1999). However, this approach may seem restrictive and not very realistic since the grouping of data into epochs is sometimes rather arbitrary, while by using only the block maxima, we may lose important information (some blocks may contain several among the largest observations, while other blocks may contain none). Moreover, in the case that we have a few data, block maxima cannot be actually implemented.

In this chapter, we examine another widely used approach, the so-called 'Maximum Domain of Attraction or Non-Parametric Approach' (Embrechts et al., 1997). In the present context we prefer the term '*semi-parametric*' since this term reflects the fact that we make only partial assumptions about the unknown d.f. F.

So, essentially, we are interested in the distribution of the maximum (or minimum) value. Here is the point where extreme-value theory gets involved. According to the Fisher-Tippet theorem, the limiting d.f. of the (normalized) maximum value (if it exists) is the GEV d.f. $H_\theta = H_{\gamma;\mu,\sigma}$. So, without making any assumptions about the unknown d.f. F (apart from some continuity conditions which ensure the existence of the limiting d.f.), extreme-value theory provides us with a fairly sufficient tool for describing the behavior of extremes of the distribution that the data in hand stem from. The only issue that remains to be resolved is the estimation of the parameters of the GEV d.f. $\theta = (\gamma, \mu, \sigma)$. Of these parameters, the *shape parameter* γ is the one that attracts most of the attention, since it is the parameter that determines, in general terms, the behavior of extremes.

According to extreme-value theory, these are the parameters of the GEV d.f. that the maximum value follows asymptotically. Of course, in reality, we only have a finite sample and, in any case, we cannot use only the largest observation for inference. So, the procedure followed in practice is that we assume that the asymptotic approximation is achieved for the largest k observations (where k is large but not as large as the sample size n), which we subsequently use for the estimation of the parameters. However, the choice of k is not an easy task. On the contrary, it is a very controversial issue. Many authors have suggested alternative methods for choosing k, but no method has been universally accepted.

3. Semi-Parametric Extreme-Value Index Estimators

In this section, we give the most prominent answers to the issue of parameter estimation. We mainly concentrate on the estimation of the shape parameter γ due to its (already stressed) importance. The setting on which we are working is :

Suppose that we have a sample of i.i.d r.v.'s $X_1, X_2, ..., X_n$ (where $X_{1:n} \geq X_{2:n} \geq ... \geq X_{n:n}$ are the corresponding descending order statistics) from an unknown *continuous* d.f. F. According to extreme-value theory, the normalized maximum of such a sample follows asymptotically a GEV d.f. $H_{\gamma;\mu,\sigma}$, i.e., $F \in MDA\left(H_{\gamma;\mu,\sigma}\right)$.

In the remainder of this section, we describe the most known suggestions to the above question of estimation of extreme-value index γ, ranging from the first contributions, of 1975, in the area to very recent modifications and new developments.

3.1 The Pickands Estimator

The Pickands estimator (Pickands, 1975), is the first suggested estimator for the parameter $\gamma \in \Re$ of GEV d.f and is given by the formula

$$\hat{\gamma}_P = \frac{1}{\ln 2} \ln\left(\frac{X_{(k/4):n} - X_{(k/2):n}}{X_{(k/2):n} - X_{k:n}}\right).$$

The original justification of Pickands's estimator was based on adopting a percentile estimation method for the differences among the upper-order statistics. A more formal justification is provided by Embrechts et al. (1997).

The properties of Pickands's estimator were mainly explored by Dekkers and de Haan (1989). They proved, under certain conditions, weak and strong consistency, as well as asymptotic normality. Consistency depends only on the behavior of k, while asymptotic normality requires more delicate conditions (2^{nd} order conditions) on the underlying d.f. F, which are difficult to verify in practice. Still, Dekkers and de Haan (1989) have shown that these conditions hold for various known and widely-used d.f.'s (normal, gamma, GEV, exponential, uniform, Cauchy).

A particular characteristic of Pickands's estimator is the fact that the largest observation is not explicitly used in the estimation. One can argue that this makes sense since the largest observation may add too much uncertainty.

Generalizations of Pickands's estimator have been introduced sharing its virtues and rendering its problems. Innovations are related to both alternative values of the multiplicative spacing parameter '2' as well as convex combinations over different k values (Yun, 2000; Segers, 2001a).

3.2 The Hill Estimator

The most popular tail index estimator is the Hill estimator, (Hill, 1975) which, however, is restricted to the Fréchet case $\gamma > 0$. The Hill estimator is provided by the formula

$$\hat{\gamma}_{H} = \frac{1}{k}\sum_{i=1}^{k} \ln X_{i:n} - \ln X_{k+1:n} .$$

The *original derivation* of the Hill estimator relied on the notion of conditional maximum likelihood estimation method.

The statistical behavior and properties of the Hill estimator have been studied by many authors separately, and under diverse conditions. Weak and strong consistency as well as asymptotic normality of the Hill estimator hold under the assumption of i.i.d. data (Embrechts et al., 1997). Similar (or slightly modified) results have been derived for data with several types of dependence or some other specific structures (see for example Hsing, 1991, as well as Resnick and Stărică, 1995, 1996, and 1998).

Note that the conditions on k and d.f. F that ensure the consistency and asymptotic normality of the Hill estimator are the same as those imposed for the Pickands estimator. Such conditions have been discussed by many authors, such as Davis and Resnick (1984), Haeusler and Teugels (1985), de Haan and Resnick (1998).

Though the Hill estimator has the apparent disadvantage that is restricted to the case $\gamma>0$, it has been widely used in practice and extensively studied by statisticians. Its popularity is partly due to its simplicity and partly due to the fact that in most of the cases where extreme-value analysis is called for, we have long-tailed d.f.'s (i.e., $\gamma>0$). However, its popularity generated a tempting problem, namely to try to extend the Hill estimator to the general case $\gamma \in \Re$. Such an attempt, led Beirlant et al. (1996) to the so-called adapted Hill estimator, which is applicable for $\gamma \in \Re$. Recent generalizations of the Hill estimator for $\gamma \in \Re$ are presented by Gomes and Martins (2001).

3.3 The Moment Estimator

Another estimator that can be considered as an adaptation of the Hill estimator, in order to obtain consistency for all $\gamma \in \Re$, has been proposed by Dekkers et al. (1989). This is the moment estimator, given by

$$\hat{\gamma}_{M} = M_1 + 1 - \frac{1}{2}\left(1 - \frac{(M_1)^2}{M_2}\right)^{-1} \quad , \quad \text{where} \quad M_j \equiv \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{i:n} - \ln X_{(k+1):n}\right)^j ,$$

j=1, 2.

Weak and strong consistency, as well as asymptotic normality of the moment estimator have been proven by its creators Dekkers et al. (1989).

3.4 The Moment-Ratio Estimator

Concentrating on cases where $\gamma > 0$, the main disadvantage of the Hill estimator is that it can be severely biased, depending on the 2nd order behavior of the underlying d.f. F. Based on an asymptotic 2nd order expansion of the d.f. F, from which one gets the bias of the Hill estimator, Danielsson et al. (1996) proposed the moment-ratio estimator defined by

$$\hat{\gamma}_{MR} = \frac{1}{2} \cdot \frac{M_2}{M_1} \ .$$

They proved that $\hat{\gamma}_{MR}$ has a lower asymptotic square bias than the Hill estimator (when evaluated at the same threshold, i.e., for the same k), though the convergence rates are the same.

3.5 Peng's and W estimators

An estimator related to the moment estimator $\hat{\gamma}_M$ is Peng's estimator, suggested by Deheuvels et al. (1997):

$$\hat{\gamma}_L = \frac{M_2}{2M_1} + 1 - \frac{1}{2}\left(1 - \frac{(M_1)^2}{M_2}\right)^{-1} \ .$$

This estimator has been designed to somewhat reduce the bias of the moment estimator.

Another related estimator suggested by the same authors is the W estimator

$$\hat{\gamma}_W = 1 - \frac{1}{2}\left(1 - \frac{(L_1)^2}{L_2}\right)^{-1}, \text{ where } L_j \equiv \frac{1}{k}\sum_{i=1}^{k}\left(X_{i:n} - X_{(k+1):n}\right)^j, \ j=1, 2.$$

As Deheuvels et al. (1997) mentioned, $\hat{\gamma}_L$ is consistent for any $\gamma \in \Re$ (under the usual conditions), while $\hat{\gamma}_W$ is consistent only for $\gamma < 1/2$. Moreover, under appropriate conditions on F and $k(n)$, $\hat{\gamma}_L$ is asymptotically normal. Normality holds for $\hat{\gamma}_W$ only for $\gamma < 1/4$.

3.6 Estimators based on QQ plots

One of the approaches concerning Hill's derivation is the 'QQ-plot' approach (Beirlant et al., 1996). According to this, the Hill estimator is approximately the slope of the line fitted to the upper tail of Pareto QQ plot. A more precise estimator, under this approach, has been suggested by Kratz and Resnick (1996), who derived the following estimator of γ:

$$\hat{\gamma}_{qq} = \frac{\sum_{i=1}^{k} \ln \frac{i}{k+1} \left\{ \sum_{j=1}^{k} \ln X_{j:n} - k \ln X_{i:n} \right\}}{k \sum_{i=1}^{k} \left(\ln \frac{i}{k+1} \right)^2 - \left(\sum_{i=1}^{k} \ln \frac{i}{k+1} \right)^2} .$$

The authors proved weak consistency and asymptotic normality of the qq-estimator (under conditions similar to the ones imposed for the Hill estimator). However, the asymptotic variance of the qq-estimator is twice the asymptotic variance of the Hill estimator, while similar conclusions are drawn from simulations of small samples. On the other hand, one of the advantages of the qq-estimator over the Hill estimator is that the residuals (of the Pareto plot) contain information which potentially can be utilized to confront the bias in the estimates when the approximation is not exactly valid.

A further enhancement of this approach (estimation of γ based on Pareto QQ plot) is presented by Beirlant et al. (1999). They suggest the incorporation, in the estimation, of the covariance structure of the order statistics involved. This leads to a regression model formulation, from which a new estimator of γ can be constructed. This estimator is proved to be particularly useful in the case of bias of the standard estimators.

3.7 Estimators based on Mean Excess Plots

A graphical tool for assessing the behavior of a d.f. F is the mean excess function (MEF). The limit behavior of MEF of a distribution gives important information on the tail of that distribution function (Beirlant et al., 1995). MEF's and the corresponding mean excess plots (MEP's), are widely used in the first exploratory step of extreme-value analysis, while they also play an important role in the more systematic steps of tail index and large quantiles estimation. MEF is essentially the expected value of excesses over a threshold value u. The formal definition of MEF (Beirlant et al., 1996) is as follows:

Let X be a positive r.v. X with d.f. F and with finite first moment. Then MEF of X is

$$e(u) = E\left(X - u \middle| X > u\right) = \frac{1}{\overline{F}(u)} \int_{u}^{x_F} \overline{F}(y) dy , \quad \text{for all } u>0.$$

The corresponding MEP is the plot of points $\{u, e(u), \text{for all } u > 0\}$.

The empirical counterpart of MEF based on sample $\left(X_1, X_2, ..., X_n\right)$, is

$$\hat{e}(u)=\frac{\sum_{i=1}^{n}(X_i-u)\mathbf{1}_{(u,\infty)}(X_i)}{\sum_{i=1}^{n}\mathbf{1}_{(u,\infty)}(X_i)}, \text{ where } \mathbf{1}_{(u,\infty)}(x)=1 \text{ if } x>u, 0 \text{ otherwise.}$$

Usually, the MEP is evaluated at the points. In that case, MEF takes the form

$$E_k=\hat{e}\left(X_{(k+1):n}\right)=\frac{1}{k}\sum_{i=1}^{k}X_{i:n}-X_{(k+1):n}, \quad k=1,\ldots, n.$$

If $X\in MDA(H_\gamma)$, γ>0, then it's easy to show that

$$e_{\ln X}(\ln u)=E\left(\ln X-\ln u \middle| X>u\right)\to\gamma, \text{ as } u\to\infty.$$

Intuitively, this suggests that if the MEF of the logarithmic-transformed data is ultimately constant, then $X\in MDA(H_\gamma)$, γ>0 and the values of MEF converge to the true value γ.

Replacing u, in the above relation, by a high quantile $Q\left(1-\frac{k}{n}\right)$, or empirically by $X_{(k+1):n}$, we find that the estimator $e_{\ln X}(X_{(k+1):n})$ will be a consistent estimator of γ in case $X\in MDA(H_\gamma)$, γ>0. This holds when $k/n\to 0$ as $n\to\infty$. Notice that the empirical counterpart of $e_{\ln X}(X_{(k+1):n})$ is the well-known Hill estimator.

3.8 Kernel Estimators

Extreme-value theory dictates that if $F\in MDA\left(H_{\gamma;\mu,\sigma}\right)$, $\gamma>0$, then it holds that $F^{\leftarrow}(1-x)\in RV_{-\gamma}$, where $F^{\leftarrow}(.)$ is the generalized inverse (quantile) function corresponding to d.f. F. Csörgő et al. (1985) showed that for 'suitable' kernel functions K, it holds that

$$\int_0^{1/\lambda}\left\{\ln F^{\leftarrow}(1-u\lambda)d\{uK(u)\}\right\}\to\gamma, \text{ as } \lambda\to 0.$$

Substituting $F^{\leftarrow}$ by its empirical counterpart $F_n^{\leftarrow}$ (which is a consistent estimator of $F^{\leftarrow}$), they propose

$$\hat{\gamma}_{Kernel}=\left(\int_0^{1/\lambda}K(u)du\right)^{-1}\left(\sum_{j=1}^{n}\frac{j}{n\lambda}K\left(\frac{j}{n\lambda}\right)\left\{\ln X_{j:n}-\ln X_{(j+1):n}\right\}\right)$$

as an estimator of γ, where $\lambda=\lambda(n)$ is a bandwidth parameter, and K is a kernel function satisfying specific conditions. Under these conditions the authors prove asymptotic consistency and normality of the derived estimator. A more general class of kernel-type estimators for $\gamma\in\Re$ is given by Groeneboom et al. (2001).

3.9 The ‘k-records’ Estimator

A statistical notion that is closely related to extreme-value analysis is that of records, or, more generally, k-records. The k-record values (for definition see Nagaraja, 1988) are themselves revealing the extremal behavior of the d.f. F, so they can also be used to assess the extreme-value parameter $\gamma \in \Re$. Berred (1995) constructed the estimator :

$$\hat{\gamma}_{\text{rec}} = \ln \frac{X^{(k)}(n) - X^{(k)}(n-k)}{X^{(k)}(n-k) - X^{(k)}(n-2k)} .$$

Under the usual conditions on k(n) (though notice that now the meaning of k(n) is different than before), he proves weak and strong consistency of $\hat{\gamma}_{\text{rec}}$ while, by imposing 2nd order conditions on F (similar to the previous cases), he also shows asymptotic normality of $\hat{\gamma}_{\text{rec}}$.

3.10 Other Semi-Parametric Estimation Methods

Up to now, we have described analytically the best-known semi-parametric methods of estimation of parameter γ (extreme-value index) of $H_{\gamma;\mu,\sigma}$. Still, there is a vast literature on alternatives estimation methods. The applicability of extreme-value analysis on a variety of different fields led scientists with different background to work on this subject and consequently derive many and different estimators. The Pickands, Hill and, recently, the moment estimators, continue to be the basis. If nothing else, most of the other proposed estimators constitute efforts to render some of the disadvantages of these three basic estimators, while others aim to generalize the framework of these. In the sequel, we present a number of such methods.

As one may notice, apart from estimators applicable for any $\gamma \in \Re$, estimation techniques have been developed valid only for a specific range of values of γ. This is due to the fact that H_γ, for γ in a specific range, may lead to families of d.f.’s of special interest. The most typical types are estimation methods for $\gamma > 0$ which correspond to d.f.’s with regularly varying tails (here the Hill estimator is included). Moreover, estimators for $\gamma \in (0,1/2)$ are of particular interst since H_γ, $\gamma \in (0,1/2)$ represents α-stable distributions ($\gamma=1/\alpha$).

Estimators for the index $\gamma > 0$, have also been proposed by Hall (1982), Feuerverger and Hall (1999), and Segers (2001b). More restricted estimation techniques for α-stable d.f.’s are described in Mittnik et al. (1998) as well as in Kogon and Williams (1998). Sometimes, the interest of authors is focused merely on the estimation of large quantiles, which in any case is what really matters in practical situations. Such estimators have been proposed by Davis and Resnick (1984) (for $\gamma > 0$) and Boos (1984) (for $\gamma = 0$).

Under certain conditions on the 2nd order behavior of the underlying distribution the error of the Hill estimator consists of two components: the systematic bias and a stochastic error. These quantities are functions of unknown parameters, prohibiting their determination and, thus, the correction of the Hill estimator. Hall (1990), suggested the use of bootstrap resamples of small size for computing a series of values of γ to estimate its bias. This approach has been further explored and extended by Pictet et al. (1998). Furthermore, they developed a jackknife algorithm for the assessment of the stochastic error component of the Hill estimator. The bootstrap (jackknife) methodology in estimation of the extreme value index has also been discussed by Gomes et al. (2000), where generalized jackknife estimators are presented as affined combinations of Hill estimators. As the authors mention, this methodology could also be applied to other classical extreme value index estimators.

3.11 Theoretical Comparison of Estimators

So far, we have mentioned several alternative estimators for the extreme-value index γ. All of these estimators share some common desirable properties, such as weak consistency and asymptotic normality (though these properties may hold under slightly different, unverifiable in any case, conditions on F and for different ranges of the parameter γ). On the other hand, simulation studies or applications on real data can end up in large differences among these estimators. In any case, there is no 'uniformly better' estimator (i.e., an estimator that is best under all circumstances). Of course, Hill, Pickands and moment estimators are the most popular ones. This could be partly due to the fact that they are the oldest ones. The rest of the existing estimators will be introduced later. Actually, most of these have been introduced as alternatives to Hill, Pickands or moment estimators and some of them have been proven to be superior in some cases. In the literature, there are some comparison studies of extreme-value index estimators (either theoretically or via Monte-Carlo methods), such as those by Rosen and Weissman (1996), Deheuvels et al. (1997), Pictet et al. (1998), and Groeneboom et al. (2001). Still, these studies are confined to a small number of estimators. Moreover, most of the authors that introduce a new estimator compare it with some of the standard estimators (Hill, Pickands, Moment).

3.12 An Alternative Approach: The Peaks Over Threshold Estimation Method

All the previously discussed semi-parametric estimation methods were based on the notion of maximum domains of attraction of the generalized extreme-value d.f. Still, further results in extreme-value theory describe the behavior of large observations that exceed high thresholds, and these are the results which lend themselves to the so-called 'Peaks Over Threshold' (POT, in short) models. The distribution which comes to the fore in this case is the generalized Pareto distribution (GPD). Thus, the estimation of the extreme-value parameter γ or the large quantiles of the underlying d.f.'s can be alternatively estimated via the GPD instead of the generalized extreme-value distribution.

The GPD can be fitted to data consisting of excesses of high thresholds by a variety of methods including the maximum likelihood method (ML) and the method of probability weighted moments (PWM). MLEs must be derived numerically because the minimal sufficient statistics for the GPD are the order statistics and there is no obvious simplification of the non-linear likelihood equation. Grimshaw (1993) provides an algorithm for estimating the MLEs for GPD. ML and PWM methods have been compared for data of GPD both theoretically and in simulation studies by Hosking and Wallis (1987) and Rootzén and Tajvidi (1997). A graphical method of estimation (Davison & Smith, 1990) is also suggested

Here, an important practical problem is the choice of the level u of the excesses. This is analogous to the problem of choosing k (number of upper order statistics) in the previous estimators. There are theoretical suggestions on how to do this, based on a compromise between bias and variance – a higher level can be expected to give less bias, but instead gives fewer excesses, and hence a higher variance. However, these suggestions don't quite solve the problem in practice. Practical aid can be provided by QQ plots, mean excess plots and experiments with different levels u. If the model produces very different results for different choices of u, the results obviously should be viewed with more caution (Rootzén & Tajvidi, 1997).

4. Smoothing and Robustifying Procedures for Semi-Parametric Extreme-Value Index Estimators

In the previous section, we presented a series of (semi-parametric) estimators for the extreme value index γ. Still, one of the most serious objections one could raise against these methods is their sensitivity towards the choice of k (number of upper order statistics used in the estimation). The well-known phenomenon of bias-variance trade-off turns out to be unresolved, and choosing k seems to be more of an art than a science.

Some refinements of these estimators have been proposed, in an effort to produce unbiased estimators even when a large number of upper order statistics is used in the estimation (see, for example, Peng, 1998, or Drees, 1996). In the next section we present a different approach on this issue. We go back to elementary notions of extreme-value theory, and statistical analysis in general, and try to explore methods to render (at least partially) this problem. The procedures we use are (i) smoothing techniques and (ii) robustifying techniques.

4.1 Smoothing Extreme-Value Index Estimators

The essence of semi-parametric estimators of extreme-value index γ, is that we use information of only the most extreme observations in order to make inference about the behavior of the maximum of a d.f. An exploratory way to subjectively choose the number k is based on the plot of the estimator $\hat{\gamma}(k)$ versus k. A stable region of the plot indicates a valid value for the estimator. The need for a stable region results from adapting theoretical limit theorems which are proved subject to the conditions $k(n) \to \infty$ and $k(n)/n \to 0$. However, the search for a stable region in the plot is a standard but problematic and ill-defined practice. Since extreme events by definition are rare, there is only little data (few observations) that can be utilized and this inevitably involves an added statistical uncertainty. Thus, sparseness of large observations and the unexpectedly large differences between them, lead to a high volatility of the part of the plot that we are interested in and makes the choice of k very difficult. That is, the practical use of the estimator on real data is hampered by the high volatility of the plot and bias problems and it is often the case that volatility of the plot prevents a clear choice of k. A possible solution would be to smooth 'somehow' the estimates with respect to the choice of k (i.e., make it more insensitive to the choice of k), leading to a more stable plot and a more reliable estimate of γ. Such a method was proposed by Resnick and Stărică (1997, 1999) for smoothing Hill and moment estimators, respectively.

4.1.1 Smoothing the Hill Estimator

Resnick and Stărică (1997) proposed a simple averaging technique that reduces the volatility of the Hill-plot. The smoothing procedure consists of averaging the Hill estimator values corresponding to different values of the order statistics p. The formula of the proposed averaged-Hill estimator is :

$$av\hat{\gamma}_H(k) = \frac{1}{k-[ku]} \sum_{p=[ku]+1}^{k} \hat{\gamma}_H(p) \quad ,$$

where u<1, and [x] denotes the smallest integer greater than or equal to x.

The authors proved that through averaging (using the above formula), the variance of the Hill estimator can be considerably reduced and the volatility of the plot tamed. The smoothed graph has a narrower range over its stable regime, with less sensitivity to the value of k. This fact diminishes the importance of selecting the optimal k. The smoothing techniques make no (additional) unrealistic or uncheckable assumptions and are always available to complement the Hill plot. Obviously, when considerable bias is present, the averaging technique offers no improvement.

Resnick and Stărică (1997) derived the adequacy (consistency and asymptotic normality) of the averaged-Hill estimator, as well as its improvement over the Hill estimator (smaller asymptotic variance). Since the asymptotic variance is an increasing function of u, one would like to choose u as small as possible to ensure a maximum decrease in the variance. However, the choice of u is limited by the sample size. Due to the averaging, the smaller the u, the fewer the points one gets on the plot of averaged Hill. Therefore, an equilibrium should be reached between variance reduction and a comfortable number of points on the plot. This is a problem similar to the variance-bias trade-off encountered in the simple extreme-value index estimators.

4.1.2 Smoothing the Moment Estimator

Resnick and Stărică (1999) also applied their idea of smoothing to the more general moment estimator $\hat{\gamma}_M$, essentially generalizing their reasoning of smoothing the Hill estimator.

The proposed smoothing technique consists of averaging the moment estimator values corresponding to different numbers of order statistics p. The formula of the proposed averaged-moment estimator, for given $0 < u < 1$, is :

$$\mathrm{av}\hat{\gamma}_M(k) = \frac{1}{k-[ku]} \sum_{p=[ku]+1}^{k} \hat{\gamma}_M(p) \ .$$

In practice, the authors suggest to take u=0.3 or u=0.5 depending on the sample size (the smaller the sample size the larger u should be).

In this case, the consequent reduction in asymptotic variance is not so profound. The authors actually showed that through averaging (using the above formula), the variance of the moment estimator is considerably reduced only in the case $\gamma < 0$. In the case $\gamma > 0$ the simple moment estimator turns out to be superior than the averaged-moment estimator. For $\gamma \approx 0$ the two moment estimators (simple and averaged) are almost equivalent. These conclusions hold asymptotically, and have been verified via a graphical comparison, since the analytic formulas of variances are rather complicated to be compared directly. A full treatment of this issue and proofs of the propositions can be found in Resnick and Stărică (1999).

4.2 Robust Estimators Based on Excess Plots

As we have previously mentioned the MEP constitutes a starting point for the estimation of extreme-value index. In practice, strong random fluctuations of the empirical MEF and of the corresponding MEP are observed, especially in the right part of the plot (i.e., for large values of u), since there we have fewer data. But this is exactly the part of plot that mostly concerns us; that is the part that theoretically informs us about the tail behavior of the underlying d.f. Consequently, the calculation of the 'ultimate' value of MEF can be largely influenced by only a few extreme outliers, which may not even be representative of the general 'trend.' The result of Drees and Reiss (1996) that the empirical MEF is an inaccurate estimate of the Pareto MEF, and that the shape of the empirical curve heavily depends on the maximum of the sample, is striking.

In an attempt to make the procedure more robust, that is less sensitive to the strong random fluctuations of the empirical MEF at the end of the range, the following adaptations of MEF have been considered (Beirlant et al., 1996):

- Generalized Median Excess Function $M^{(p)}(k) = X_{([pk]+1):n} - X_{(k+1):n}$

(for p=0.5 we get the simple median excess function).

- Trimmed Mean Excess Function $T^{(p)}(k) = \frac{1}{k-[pk]} \sum_{j=[pk]+1}^{k} X_{j:n} - X_{(k+1):n}$.

The general motivations and procedures explained for the MEF and its contribution to the estimation of γ hold here as well. Thus, alternative estimators for $\gamma>0$ are :

- $\hat{\gamma}_{gen.med} = \frac{1}{\ln(1/p)}\left(\ln X_{([pk]+1):n} - \ln X_{(k+1):n}\right)$

which for p=0.5 gives $\hat{\gamma}_{med} = \frac{1}{\ln(2)}\left(\ln X_{([k/2]+1):n} - \ln X_{(k+1):n}\right)$

(the consistency of this estimator is proven by Beirlant et al., 1996).

- $\hat{\gamma}_{trim} = \frac{1}{k-[pk]} \sum_{j=[pk]+1}^{k} \ln X_{j:n} - \ln X_{(k+1):n}$

It is worth noting that robust estimation of the tail index of a two-parameter Pareto distribution is presented by Brazauskas and Serfling (2000). The corresponding estimators are of generalized quantile type. The authors distinguish the trimmed mean and generalized median type as the most competitive trade-offs between efficiency and robustness.

5. More Formal Methods for Selecting k

In the previous sections we have presented some attempts to derive extreme-value index estimators, smooth enough, so that the plot $\{k, \hat{\gamma}(k)\}$ is an adequate tool for choosing k and consequently deciding on the estimate $\hat{\gamma}(k)$. However, such a technique will always be a subjective one and there are cases where we need a more objective solution. Actually, there are cases where we need a quick, automatic, clear-cut choice of k. So, for reasons of completeness, we present some methods for choosing k in extreme-value index estimation. Such a choice of k is, essentially, an ‘optimal choice,’ in the sense that we are looking for the optimal sequence $k(n)$ that balances the variance and bias of the estimators. This optimal sequence $k_{opt}(n)$ can be determined when the underlying distribution F is known, provided that the d.f. has a second order expansion involving an extra unknown parameter. Adaptive methods for choosing k were proposed for special classes of distributions (see Beirlant et al., 1996 and references in Resnick and Stărică, 1997). However, such second order conditions are unverifiable in practice. Still Dekkers and de Haan (1993) prove that such conditions hold for some well-known distributions (such as the Cauchy, the uniform, the exponential, and the generalized extreme-value distributions). Of course, in practice we do not know the exact analytic form of the underlying d.f. So, several approximate methods, which may additionally estimate (if needed) the 2^{nd} order parameters, have been developed. Notice, that the methods existing in the literature are not generally applicable to any extreme-value index estimator but are designed for particular estimators in each case.

Drees and Kaufmann (1998), proposed a sequential approach to construct a consistent estimator of k that works asymptotically without any prior knowledge about the underlying d.f. Recently, a simple diagnostic method for selecting k has been suggested by Guillou and Hall (2001). They performed a sort of hypothesis testing on log-spacings by appropriately weighting them. Both of these approaches have been originally introduced for the Hill estimator, but can be extended to other extreme-value index estimators, too.

5.1 The Regression Approach

Recall that according to the graphical justification of the Hill estimator, this estimator can be derived as the estimation of the slope of a line fitted to the k upper-order statistics of our dataset. In this sense, the choice of k can be reduced to the problem of choosing an anchor point to make the linear fit optimal. In statistical practice, the most common measure of optimality is the mean square error.

In the context of the Hill estimator (for $\gamma > 0$) and the adapted Hill estimator (for $\gamma \in \Re$), Beirlant et al. (1996) propose the minimization of the asymptotic mean square error of the estimator as an appropriate optimality criterion. They have suggested using

$$\mathrm{MSE}_{\mathrm{opt}}(k) = \frac{1}{k}\sum_{j=1}^{k} w_{j,k}^{\mathrm{opt}}\left(\ln X_{j:n} - \left[\ln X_{(k+1):n} + \gamma \ln\frac{k+1}{j}\right]\right)$$

as a consistent estimate (as $n \to \infty$, $k \to \infty$, $k/n \to 0$) of asymptotic mean square error of Hill estimator ($w_{j,k}^{\mathrm{opt}}$ is a sequence of weights).

Theoretically, it would suffice to compute $\mathrm{MSE}_{\mathrm{opt}}$ for every relevant value of k and look for the minimal MSE value with respect to k. Note that in the above expression neither γ (true value of extreme-value index) nor the weights $w_{j,k}^{\mathrm{opt}}$, which depend on a parameter ρ of the 2^{nd} order behavior of F, are known. So, Beirlant et al. (1996), propose an iterative algorithm for the search of the optimum k.

5.2 The Bootstrap Approach

Draisma et al. (1999) developed a purely sample-based method for obtaining the optimal sequence $k_{opt}(n)$. They, too, assume a second order expansion of the underlying d.f., but the second (or even the first) order parameter is not required to be known. In particular, their procedure is based on a double bootstrap. They are concerned with the more general case $\gamma \in \Re$, and their results refer to the Pickands and the moment estimators.

As before, they want to determine the value of k, $k_{opt}(n)$, minimizing the asymptotic mean square error $E_F(\hat{\gamma}(k) - \gamma)$, where $\hat{\gamma}$ refers either to the Pickands estimator $\hat{\gamma}_P$ or to the moment estimator $\hat{\gamma}_M$. However, in the above expression there are two unknown factors: the parameter γ and the d.f. F. Their idea is to replace γ by a second estimator $\hat{\gamma}_+$ (its form depending on whether we use the Pickands or the moment estimator) and F by the empirical d.f. F_n. This is determined by bootstrapping. The authors prove that minimizing the resulting expression, which can be calculated purely on the basis of the sample, still leads to the optimal sequence $k_{opt}(n)$ again via a bootstrap procedure.

The authors test their proposed bootstrap approach on various d.f.'s (such as those of the Cauchy, the generalized Pareto, and the generalized extreme-value distributions) via simulation. The general conclusion is that the bootstrap procedure gives reasonable estimates (in terms of mean square error of the extreme-value index estimator) for the sample fraction to be used. So, such a procedure takes out the subjective element of choosing k. However, even in such a procedure an element of subjectivity remains, since one has to choose the number of bootstrap replications (r) and the size of the bootstrap samples (n_1).

Similar bootstrap-based methods for selecting k have been presented by Danielsson and de Vries (1997) and Danielsson et al. (2000), confined to $\gamma > 0$, with results concerning only the Hill estimator $\hat{\gamma}_H$. Moreover, Geluk and Peng (2000) apply a 2-stage non-overlapping subsampling procedure, in order to derive the optimal sequence $k_{opt}(n)$ for an alternative tail index estimator (for $\gamma > 0$) for finite moving average time series.

6. Discussion and Open Problems

The wide collection of estimators of the extreme value index which characterizes the tails of most distributions, has been the central issue of this chapter. We presented the main approaches for the estimation of γ, with special emphasis to the semi-parametric one. In sections 3 and 4 several such estimators are provided (Hill, Moment, Pickands, among others). Some modifications of these proposed in the literature based on smoothing and robustifying procedures have also been considered since the dependence of these estimators on the very extreme observations which can display very large deviations, is one of their drawbacks.

Summing up, there is not a uniformly best estimator of the extreme-value index. On the contrary, the performance of estimators seems to depend on the distribution of data in hand. From another point of view, one could say that the performance of estimators of the extreme-value index depends on the value of the index itself. So, before proceeding to the use of any estimation formula it would be useful if we could get an idea about the range of values where the true γ lies in. This can be achieved graphically via QQ and mean excess plots. Alternatively, there exist statistical tests that tests such a hypothesis. (See, for example, Hosking, 1984, Hasofer & Wang, 1992, Alves & Gomes, 1996, and Marohn, 1998; 2000, Segers & Teugels, 2000).

However, the 'Achilles heel' of semi-parametric estimators of the extreme-value index is its dependence and sensitivity on the number k of upper order statistics used in the estimation. No hard and fast rule exists for confronting this problem. Usually, the scientist subjectively decides on the number k to use, by looking at appropriate graphs. More objective ways for doing this are based on regression or bootstrap. The bootstrap approach is a newly suggested and promising method in the field of extreme-value analysis. Another area of extreme-value index estimation where bootstrap methodology could turn out to be very useful is in the estimation (and, consequently, elimination) of the bias of extreme-value index estimators. The bias is inherent in all of these estimators, but it is not easy to be assessed theoretically because it depends on second order conditions on the underlying distribution of data, which are usually unverifiable. Bootstrap procedures could approximate the bias without making any such assumptions.

Finally, we should mention that a new promising branch of extreme-value analysis is that of multivariate extreme-value methods. One of the problems in extreme-value analysis is that, usually, one deals with few data which leads to great uncertainty. This drawback can be aleviated somehow, by the simultaneous use of more than one source of information (variables), i.e., by applying multivariate extreme-value analysis. Such an approach is attempted by Embrechts, de Haan and Huang (1999) and Caperaa and Fougeres (2001). This technique has already been applied to the field of hydrology. See, for example, de Haan and de Ronde (1998), de Haan and Sinha (1999) and Barão and Tawn (1999).

7. References

Alves, M.I.F. and Gomes, M.I. (1996). Statistical Choice of Extreme Value Domains of Attraction – A Comparative Analysis. *Communication in Statistics (A) – Theory and Methods,* 25 (4), 789-811.

Barão, M.I. and Tawn, J.A. (1999). Extremal Analysis of Short Series with Outliers: Sea-Levels and Athletic Records. *Applied Statistics,* 48 (4), 469-487.

Beirlant, J., Broniatowski, M., Teugels, J.L. and Vynckier, P. (1995). The Mean Residual Life Function at Great Age: Applications to Tail Estimation. *Journal of Statistical Planning and Inference,* 45, 21-48.

Beirlant, J., Dierckx, G., Goegebeur, Y. and Matths, G. (1999). Tail Index Estimation and an Exponential Regression Model. *Extremes,* 2(2), 177-200.
Beirlant, J., Teugels, J.L. and Vynckier, P. (1994). Extremes in Non-Life Insurance. *Extremal Value Theory and Applications*, Galambos, J., Lenchner, J. and Simiu, E. (eds.), Dordrecht, Kluwer, 489-510.
Beirlant, J., Teugels, J.L. and Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press, Leuven.
Berred, M. (1995). K-record Values and the Extreme-Value Index. *Journal of Statistical Planning and Inference,* 45, 49-63.
Boos, D.D. (1984). Using Extreme Value Theory to Estimate Large Percentiles. *Technometrics,* 26 (1), 33-39.
Brazauskas, V. and Serfling., R. (2000). Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via generalized Quantile Statistics. *Extremes,* 3(3), 231-249.
Caperaa, P. and Fougeres, A.L. (2001). Estimation of a Bivariate Extreme Value Distribution. *Extremes,* 3(4), 311-329.
Coles, S.G. and Dixon, M.J. (1999). Likelihood-Based Inference for Extreme Value Models. *Extremes,* 2 (1), 5-23.
Coles, S.G. and Tawn, J.A. (1996). Modelling Extremes of the Areal Rainfall Process. *Journal of the Royal Statistical Society, B* 58 (2), 329-347.
Csörgő, S., Deheuvels, P. and Mason, D. (1985). Kernel Estimates of the Tail Index of a Distribution. *The Annals of Statistics,* 13 (3), 1055-1077.
Danielsson, J. and de Vries, C.G. (1997). Beyond the Sample: Extreme Quantile and Probability Estimation. *Preprint*, Erasmus University, Rotterdam.
Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (2000). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis,* 76, 226-248.
Danielsson, J., Jansen, D.W. and deVries, C.G. (1996). The Method of Moment Ratio Estimator for the Tail Shape Distribution. *Communication in Statistics (A) – Theory and Methods,* 25 (4), 711-720.
Davis, R. and Resnick, S. (1984). Tail Estimates Motivated by Extreme Value Theory. *The Annals of Statistics,* 12 (4), 1467-1487.
Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society, B* 52 (3), 393-442.
De Haan, L. and de Ronde, J. (1998). Sea and Wind: Multivariate Extremes at Work. *Extremes,* 1 (1), 7-45.
De Haan, L. and Resnick, S. (1998). On Asymptotic Normality of the Hill Estimator. *Stochastic Models,* 14, 849-867.
De Haan, L. and Sinha, A.K. (1999). Estimating the Probability of a Rare Event. *The Annals of Statistics,* 27 (2), 732-759.
Deheuvels, P., de Haan, L., Peng, L. and Pereira, T.T. (1997). Comparison of Extreme Value Index Estimators. NEPTUNE T400:EUR-09.
Dekkers, A.L.M. and de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation. *The Annals of Statistics,* 17 (4), 1795 - 1832.
Dekkers, A.L.M. and de Haan, L. (1993). Optimal Choice of Sample Fraction in Extreme-Value Estimation. *Journal of Multivariate Analysis,* 47(2), 173-195.
Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A Moment Estimator for the Index of an Extreme-Value Distribution. *The Annals of Statistics,* 17 (4), 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 (4), 367-404.
Drees, H. (1996). Refined Pickands Estimators with Bias Correction. *Communication in Statistics (A) – Theory and Methods,* 25 (4), 837-851.
Drees, H. and Kaufmann, E. (1998). Selecting the Optimal Sample Fraction in Univariate Extreme Value Estimation. *Stochastic Processes and their Applications,* 75, 149-172.
Drees, H. and Reiss, R.D. (1996). Residual Life Functionals at Great Age. *Communication in Statistics (A) – Theory and Methods,* 25 (4), 823-835.

DuMouchel, W.H. (1983). Estimating the Stable Index α in Order to Measure Tail Thickness: a Critique. *The Annals of Statistics,* 11, 1019-1031.
Embrechts, P. (1999). Extreme Value Theory in Finance and Insurance. *Preprint*, ETH.
Embrechts, P., de Haan, L. and Huang, X. (1999). Modelling Multivariate Extremes. *ETH Preprint*.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
Embrechts, P., Resnick, S. and Samorodnitsky, G. (1998). Living on the Edge. *RISK Magazine,* 11 (1), 96-100.
Embrechts, P., Resnick, S. and Samorodnitsky, G. (1999). Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal,* 26, 30-41.
Feuerverger, A. and Hall, P. (1999). Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution. *The Annals of Statistics,* 27(2), 760-781.
Fisher, R.A. and Tippet, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proc. Cambridge Phil. Soc.,* 24 (2), 163-190. (in Embrechts et al. 1997).
Geluk, J.L. and Peng, L. (2000). An Adaptive Optimal Estimate of the Tail Index for MA(1) Time Series. *Statistics and Probability Letters,* 46, 217-227.
Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill Estimator – Asymptotic versus Finite Sample Properties. *Journal of Statistical Planning and Inference,* 93, 161-180.
Gomes, M.I., Martins, M.J., Neves, M. (2000). Alternatives to a Semi-Parametric estimator of parameters of Rare Events-The Jackknife Methodology, *Extremes,* 3(3), 207-229.
Grimshaw, A. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics,* 35, 185-191.
Groeneboom, P., Lopuhaa, H.P. and de Wolf, P.P. (2001). Kernel-Type estimators for the Extreme Value Index. *Annals of Statistics*, (to appear).
Guillou, A. and Hall, P. (2001). A Diagnostic for Selecting the Threshold in Extreme Value Analysis. *Journal of the Royal Statistical Society Series B*, 63 (2), 293-350.
Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York. (in Kinnison, 1985).
Haeusler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation. *The Annals of Statistics,* 13 (2), 743-756.
Hall, P. (1982). On Some Simple Estimates of an Exponent of Regular Variation. *Journal of the Royal Statistical Society,B* 44 (1), 37-42.
Hall, P. (1990). Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameter in NonParametric Problems. *Journal of Multivariate Analysis*, 32, 177-203.
Hasofer, A.M. and Wang, Z. (1992). A Test for Extreme Value Domain of Attraction. *Journal of the American Statistical Association,* 87, 171-177.
Hill, B.M. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *The Annals of Statistics,* 3 (5), 1163-1174.
Hosking, J.R.M. (1984). Testing Whether the Shape Parameter Is Zero in the Generalized Extreme-Value Distribution. *Biometrika,* 71(2), 367-374.
Hosking, J.R.M. (1985). Maximum-Likelihood Estimation of the Parameter of the Generalized Extreme-Value Distribution. *Applied Statistics,* 34, 301-310.
Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics,* 29 (3), 333-349.
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 (3), 251-261.
Hsing, T. (1991). On Tail Index Estimation Using Dependent Data. *Annals of Statistics*, 19 (3), 1547-1569.
Kogon, S.M. and Williams, D.B. (1998). Characteristic Function Based Estimation of Stable Distribution Parameters. In R. Adler et al, (eds.), "A Practical Guide to Heavy Tails: Statistical Techniques and Applications". Boston: Birkhauser.
Kratz, M. and Resnick, S.I. (1996). The QQ Estimator and Heavy Tails. *Communication in Statistics – Stochastic Models,* 12 (4), 699-724.
Marohn, F. (1998). Testing the Gumbel Hypothesis via the Pot-Method. *Extremes,* 1(2), 191-213.

Marohn, F. (2000). Testing Extreme Value Models. *Extremes,* 3(4), 363-384.
McNeil, A.J. (1997). Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin,* 27, 117-137.
McNeil, A.J. (1998). On Extremes and Crashes. A short-non-technical article. *RISK,* January 1998, 99.
McNeil, A.J. (1999). Extreme Value Theory for Risk Managers. *Internal Modelling and CAD II* published by RISK Books, 93-113.
Mikosch, T. (1997). Heavy-Tailed Modelling in Insurance. *Communication in Statistics – Stochastic Models,* 13 (4), 799-815.
Mittnik, S., Paolella, M.S. and Rachev, S.T. (1998). A Tail Estimator for the Index of the Stable Paretian Distribution. *Communication in Statistics – Theory and Method,* 27 (5), 1239-1262.
Nagaraja, H.N. (1988). Record Values and Related Statistics – A Review. *Communication in Statistics – Theory and Methods,* 17 (7), 2223-2238.
Peng, L. (1998). Asymptotically Unbiased Estimators for the Extreme-Value Index. *Statistics and Probability Letters,* 38, 107-115.
Pickands, J. (1975). Statistical Inference Using Extreme order statistics. *The Annals of Statistics,* 3 (1), 119-131.
Pictet, O.V., Dacorogna, M.M. and Muller, U.A., (1998). Hill, Bootstrap and Jackknife Estimators for Heavy Tails. In R. Adler et al. (Eds.), "A Practical Guide to Heavy Tails: Statistical Techniques and Applications". Boston: Birkhauser.
Resnick, S. and Stărică, C. (1995). Consistency of Hill's Estimator for Dependent Data. *Journal of Applied Probability,* 32, 139-167.
Resnick, S. and Stărică, C. (1996). Asymptotic Behaviour of Hill's Estimator for Autoregresive Data. Preprint, Cornell University. (Available as TR1165.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html.)
Resnick, S. and Stărică, C. (1997). Smoothing the Hill Estimator. *Advances in Applied Probability,* 29, 271-293.
Resnick, S. and Stărică, C. (1998). Tail Index Estimation for Dependent Data. *Annals of Applied Probability,* 8 (4), 1156-1183.
Resnick, S. and Stărică, C. (1999). Smoothing the Moment Estimator of the Extreme Value Parameter. *Extremes,* 1(3), 263-293.
Rootzén, H. and Tajvidi, N. (1997). Extreme Value Statistics and Wind Storm Losses: A Case Study. *Scandinavian Actuarial Journal*, 70-94.
Rosen, O. and Weissman, I. (1996). Comparison of Estimation Methods in Extreme Value Theory. *Communication in Statistics – Theory and Methods,* 25 (4), 759-773.
Scarf, P.A. (1992). Estimation for a Four Parameter Generalized Extreme Value Distribution. *Communication in Statistics – Theory and Method,* 21 (8), 2185-2201.
Segers, J. (2001a). On a Family of Generalized Pickands Estimators. *Preprint*, Katholieke University Leuven.
Segers, J. (2001b). Residual Estimators, *Journal of Statistical Planning and Inference,* 98, 15-27.
Segers, J. and Teugels, J. (2000). Testing the Gumbel Hypothesis by Galton's Ratio. *Extremes,* 3(3), 291-303.
Smith, R.L. (1985). Maximum-Likelihood Estimation in a Class of Non-Regular Cases. *Biometrika,* 72, 67-90.
Smith, R.L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. *Statistical Science*, 4 (4), 367-393.
Yun, S. (2000). A Class of Pickands-Type Estimators for the Extreme Value Index. *Journal of Statistical Planning and Inference,* 83, 113-124.

Received: January 2002, Revised: June 2002