CHAPTER 5

SMOOTHING AND ROBUSTIFYING PROCEDURES FOR SEMI-PARAMETRIC EXTREME-VALUE INDEX ESTIMATORS

5.1 Introduction

In the previous chapter, we presented a series of (semi-parametric) estimators for the extreme value index γ. Apart from the inherent attractiveness that they have due to the fact that they rely on the well-established theoretical background of extreme-value theory, each one of them has been studied extensively and its advantages and disadvantages have been revealed according to the type of d.f. that they are applied to. For almost all of these estimators (at least for the basic ones) consistency and asymptotic normality have been proven. 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.

Previously, we have mentioned some refinements of these estimators, 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). From another standpoint, adaptive methods for choosing k were proposed for special classes of distributions (see Beirlant et al., 1996 and references in Resnick and Stărică, 1997). The criterion for the choice of k is minimization of the asymptotic mean square error (AMSE). Csörgő et al. (1985) introduced a family of kernel estimators (described in the previous chapter) and for the same special class of distributions, they determined the optimal choice of λ (corresponding to our k) and of the kernel function K to be used in the estimation. Their optimality criterion is also based on minimizing AMSE. Indeed, all of these approaches provide us with a single choice of k that is optimal in 'some sense'. Still, none of these has been widely accepted. No hard and fast rule seems to exist. It is interesting to note that Pickands, when introducing Pickands estimator (Pickands, 1975), also suggested an ad-hoc method for choosing k, but that method was never actually widely adopted, contrary to the estimator itself. Another common unfortunate feature of semi-parametric estimators of γ is that the optimal choice of k depends mainly on 2nd order assumptions of the underlying d.f. F. As we have repeatedly mentioned, these conditions are not verifiable in practice. Remember that unbiased versions of these estimators also depend on 2nd order conditions on F. Additionally, the formulas for minimization of AMSE are usually only asymptotic formulas and the asymptotic equivalence is not helpful for finite samples. So, it seems that again we are led to a dead-end. Accordingly, it is not an exaggeration to say that choosing k (number of upper order statistics used in the estimation procedure) is the Achilles heel of semi-parametric estimation methods for extreme-value index γ. In the sections to follow we attempt a different approach towards 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:

- Smoothing techniques
- Robustyfing Techniques

Moreover, new developments, based on bootstrap, for choosing the number of upper order statistics used in the estimation of the index are also presented.

5.2 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 behaviour 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 search for a stable region in the plot is a standard but problematic and ill-defined practice. The need for a stable region results from adapting theoretical limit theorems which are proved subject to the condition that the number of upper order statistics k used in the estimation tends to infinity ($k(n) \to \infty$) but k is a diminishing proportion of the sample size ($k(n)/n \to 0$).

But, since extreme events by definition are rare, there is only little data (few observations) that can be utilized and this inevitably involves an added large 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 $\hat{\gamma}$. 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 the Hill and moment estimator respectively.

5.2.1 Smoothing Hill Estimator

Resnick and Stărică (1997) propose 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 order statistics p. The formula of the proposed averaged-Hill estimator is :

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

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

Notice that the order of number of terms involved in the averaging is k. Therefore when $n, k \to \infty$ we will be averaging larger and larger numbers of Hill estimator's values with a consequent reduction in asymptotic variance. Indeed, the authors prove 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 estimate of γ suggested by averaged-Hill plot can be used as a basis for further studies which would, for example, use bootstrap techniques to correct for bias. We should also note that though the technique is simple and obvious, it turns out to be quite useful in practice, while asymptotic normality can be proved for averaged-Hill estimator under the same conditions required for the normality of the original Hill estimator. 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. Still, the procedure, while unable to solve the bias problem, is usually more informative than the standard practice.

Theoretical (Asymptotic) Investigation of Averaged-Hill Estimator

In order to study the (asymptotic) behaviour of averaged-Hill estimator, Resnick and Stărică (1997) proceed to the following definitions:

Definitions (Resnick and Stărică, 1997)

On $[0,\infty)$ we have the empirical processes:

Tail Empirical Process $E_{k,n}(y) \equiv \frac{1}{k}\sum_{i=1}^{n} \varepsilon_{X_i/b(n/k)}\left[y^{-\gamma},\infty\right]$,

Normalized Tail Empirical Process $\mathsf{E}_{k,n}(y) \equiv \sqrt{k}\left(E_{k,n}(y) - y\right)$,

where $b(t) = F^{\leftarrow}(1-t^{-1})$ and $\varepsilon_x(\cdot)$ is an indicator function,

while on $(0,\infty)$ we have the processes:

Hill Process $H_{k,n}(y) \equiv \hat{\gamma}_H([ky])$, and

Normalized Hill Process $\mathsf{H}_{k,n}(y) \equiv \sqrt{k}\left(H_{k,n}(y) - \gamma\right)$.

By associating the tail empirical process to the sequence (X_1, X_2, ..., X_n), the authors show the weak convergence of the normalized tail empirical process to a Brownian motion and deduce from this the convergence of the normalized Hill process. By a process of integration the asymptotic behaviour of the smoothed estimator is obtained. In the lines to follow we provide a brief outline of the reasoning that Resnick and Stărică (1997) adopt in order to derive the adequacy (consistency and asymptotic normality) of the averaged-Hill estimator, as well as its improvement over Hill estimator (smaller asymptotic variance). We omit the technical details, while the proofs of the propositions and theorem sited below can be found in Resnick and Stărică (1997).

Necessary Conditions (these are essentially the conditions needed to prove asymptotic normality of the Hill estimator, as well)

Condition 1:

There exist constants $\gamma > 0$, $\rho \le 0$, $K \in \Re$ such that

$$\lim_{t\to\infty} \frac{x^{1/\gamma} F(tx) - F(t)}{F(t)g(t)} = K\frac{x^{\rho}-1}{\rho} \qquad (1),$$

where $g \in RV_{\rho}$, the convergence in (1) is uniform on [1,∞), and

$$\int_1^{\infty}\left(x^{1/\gamma}\frac{\overline{F}(xu)}{\overline{F}(x)}-1\right)u^{-(1+\gamma)/\gamma}du \to \frac{K\gamma^2}{1-\gamma\rho}g(x) \text{, as } x\to\infty .$$

Condition 2:

The sequence k(n), such that $k \to \infty$, and $k/n \to 0$, satisfies

$$\sqrt{k}\,g\left(b\left(\frac{n}{k}\right)\right) \to 0 \text{, as } n \to \infty .$$

The derivation of the asymptotic behaviour of the tail empirical process is the first step in the endeavor of smoothing the Hill estimator

Proposition: Asymptotic Behaviour of the Normalized Tail Empirical Process
(Resnick and Stărică, 1997)

Assume that Conditions 1, 2 hold, as $n \to \infty$. Then

$$E_{k,n}(y) \equiv \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n}\varepsilon_{X_i/b(n/k)}\left[y^{-\gamma},\infty\right]-y\right) \to W(y)$$

in $D[0,\infty)$, where $\{W(t), t \ge 0\}$ is a standard Brownian motion.

(detailed description of Brownian motion and its properties can be found in Ross, 1996)

Based on the above proposition and on the remark that the Hill process can 'almost' be expressed as a functional of the tail empirical process and its inverse process, the asymptotic behaviour of Hill process is derived.

Proposition: Asymptotic Behaviour of the Normalized Hill Process (Resnick and Stărică, 1997)

Assume that Conditions 1, 2 hold. Then as $n \to \infty$,

$$H_{k,n}(y) = \sqrt{k}\left(\hat{\gamma}_H([ky]) - \gamma\right) \to \frac{\gamma}{y} W(y)$$

in $D(0,\infty)$, where $\{W(t), t \geq 0\}$ is a standard Brownian motion.

These results, concerning stochastic process, lead to the derivation of the asymptotic distribution of the averaged-Hill estimator.

Proposition: Asymptotic Distribution of the Averaged-Hill Estimator (Resnick and Stărică, 1997)

Assume that Conditions 1, 2 hold, $n \to \infty$, and let $s \geq 1$, $v \geq 2$ be fixed. Then

$$\sqrt{k}\left(\frac{1}{ks(v-1)} \sum_{p=[ks]}^{[ksv]} \hat{\gamma}_H(p) - \gamma\right) \to N(0,c),$$

where $c = \gamma^2 \frac{2}{s(v-1)}\left(1 - \frac{\ln v}{v-1}\right)$ (v corresponds to 1/u in the previous defining formula of the averaged-Hill estimator).

The above formula of variance consists a formal justification of the averaging procedure. It is apparent that the variance is decreasing with respect to both s and u, while for their minimum values s=1 and v=2 the variance of the averaged-Hill estimator is $0.614\gamma^2$, i.e. there is a 38.6% reduction compared to the variance of the original Hill estimator. In essence, the bigger v (smaller u) and s the better. As long as s is concerned it is reasonable to take s=1. But then, since the asymptotic variance is a decreasing function of u, one would like to choose v as big as possible to ensure the maximum decrease of the variance. However, the choice of v is limited by the sample size. Due to the averaging, the larger the v, 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. Notice, that this is a problem similar to the variance-bias trade-off encountered in the simple extreme-value index estimators. The figure below graphically illustrates the reduction in the variance of the original Hill estimator when we use smoothing.

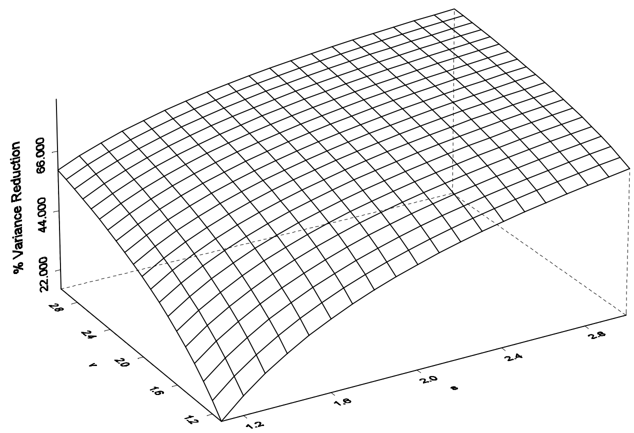
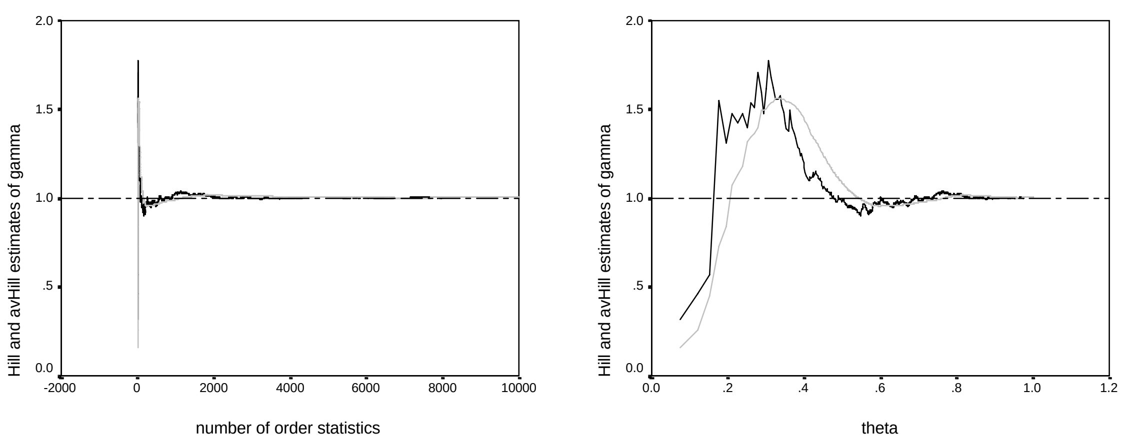


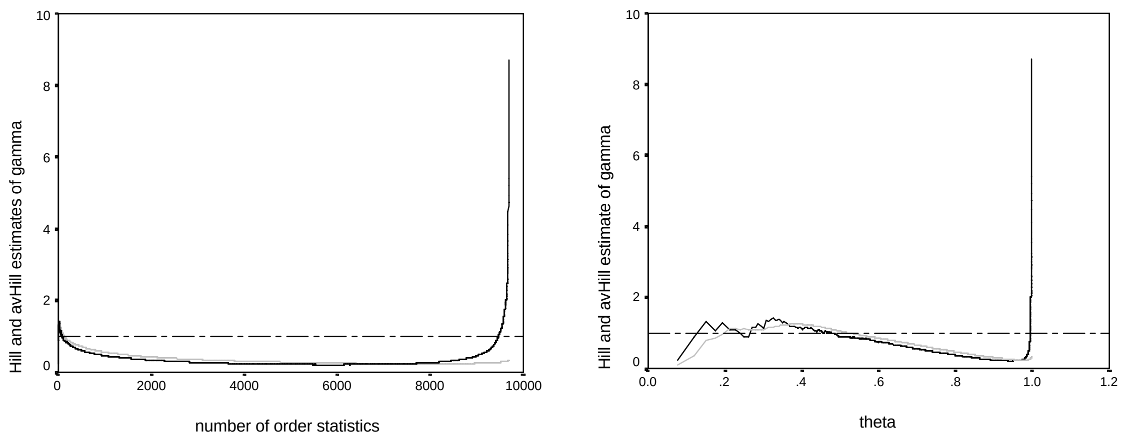
Figure 5.1. Percentage of variance reduction due to smoothing of Hill estimator, versus smoothing parameters s, v

In the figures that follow, we can see the effect of averaging Hill estimator to data generated from some well-known and widely applicable distributions.



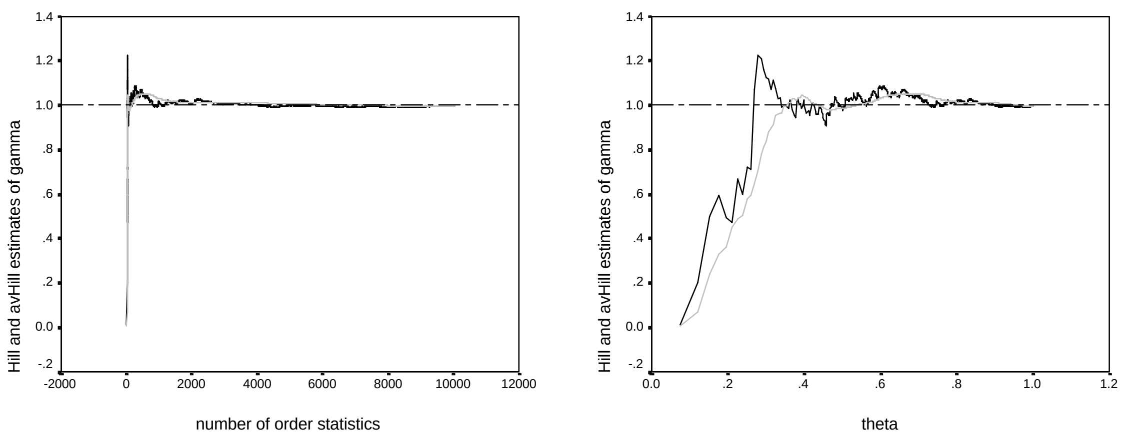
(a) (b)

Figure 5.2. Hill (black line) and avHill (gray line) plot (a) and alternative plot (b) for 10,000 Pareto observations (γ=1, smoothing parameter u=0.3)



(a) (b)

Figure 5.3. Hill (black line) and avHill (gray line) plot (a) and alternative plot (b) for 10,000 Cauchy (10,1) observations (γ=1, smoothing parameter u=0.3)



(a) (b)

Figure 5.4. Hill (black line) and avHill (gray line) plot (a) and alternative plot (b) for 10,000 log-gamma(1,1) observations (γ=1, smoothing parameter u=0.3)

5.2.2 Smoothing Moment Estimator

The previous method referred to the smoothing of Hill estimator, i.e. it was restricted to cases where $\gamma > 0$. Hence, it would be desirable to proceed to the construction of a more general smoothing procedure that would correspond to an estimator covering the whole range $\gamma \in \Re$. The moment estimator $\hat{\gamma}_M$, described in the previous chapter, fulfills the requirement of being applicable to all $\gamma \in \Re$. So, 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 Hill estimator. As was the case for the Hill estimator, they propose a modification of this estimator obtained through an averaging technique.

In practice, one makes a plot of $\{(k, \hat{\gamma}_M(k)), 1 \le k \le n\}$ and attempts to infer a value of γ from a stable region of the plot. As was the case for the Hill estimator, this is quite a subjective procedure, which is marred by the volatility of the plot and considerations of the percentage of the display space that the plot occupies in a neighborhood of the true value. Methods adaptively estimating the 'optimal' k, which minimizes the asymptotic mean square error, exist here as well and are welcome but they have unproven practical value (see comments in previous section). The smoothing procedure that Resnick and Stărică (1999) propose reduces the volatility of the plot and makes the selection of an estimate based on the plot more secure.

So, 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 :

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

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).

Notice that, again, the order of number of terms involved in the averaging is k. Therefore when $n, k \to \infty$ we will be averaging larger and larger numbers of moment estimator values.

Still, in this case the consequent reduction in asymptotic variance is not so profound. The authors actually show that through averaging (using the above formula), the variance of the moment estimator can be considerably reduced only in the case $\gamma < 0$.

In this case all the remarks that were made for averaged-Hill (smoothed graph, diminishing importance of selecting k and so on) are valid, too. Still, in the case $\gamma > 0$ the simple moment estimator turns out to be superior that the averaged-moment estimator. Of course in that case, as the authors argue, if we could know that indeed

$\gamma > 0$, we could simple use the averaged-Hill estimator. For $\gamma \approx 0$ the two moment estimators (simple and averaged) are almost equivalent. These conclusions hold asymptotically, and have been shown via a graphical comparison, since the analytic formulas of variances are rather complicated to be compared directly.

Theoretical (Asymptotic) Investigation of Averaged-Moment Estimator

In order to study the (asymptotic) properties of the smoothed moment estimator, Resnick and Stărică (1999) prove, first of all the weak convergence of a normalized tail empirical process to a Brownian motion which implies the convergence of a process closely related to the moment estimator to a functional of a Brownian motion (note that, here, the definition of tail empirical process is similar but not identical to the tail empirical process defined in the context of averaged-Hill estimator). By an integrating procedure, they derive the asymptotic behaviour of the averaged-moment estimator. Still, the situation here is somewhat more complicated than was for the averaged-Hill estimator. So, the authors do not prove asymptotic normality of the averaged-moment estimator, but instead show that it converges to a more complicated functional. Even the variance of the new estimator is complicated enough, show that only numerically (graphically) can we see that the averaged estimator has larger variance when $\gamma > 0$ and smaller variance when $\gamma < 0$, compared to the original moment estimator. So, in the sequel we display only the main results which reveal the asymptotic behaviour of averaged moment estimator, as well as the required conditions (these conditions are not more restrictive than the conditions imposed to F, in order to prove asymptotic normality of the simple moment estimator). Full treatment of this issue and proofs of the propositions can be found in Resnick and Stărică (1999).

Condition 1:

Let $\tilde{\gamma} = \min(\gamma, 0)$ and assume there exists constant $\tilde{\rho} \leq 0$, and function $\tilde{A}$ (ultimately of constant sign and $\tilde{A} \to 0$) such that

$$\lim_{t\to\infty} \frac{\dfrac{b_Y(tx) - b_Y(t)}{a_Y(t)/b_Y(t)} - \dfrac{x^{\tilde{\gamma}} - 1}{\tilde{\gamma}}}{\tilde{A}(t)} = \tilde{H}(x) \quad (1),$$

for a function $\widetilde{H}$ not a multiple of $\left(x^{\widetilde{\gamma}}-1\right)/\widetilde{\gamma}$,

where $Y=\ln\left(X_{+}\right)$, $b_X(t)=F^{\leftarrow}(1-t^{-1})$, and

$$a_X(t)=\begin{cases}(-\gamma)\left(b_X(\infty)-b_X(t)\right) & \text{if } \gamma<0,\\ b_X(te)-b_X(t) & \text{if } \gamma=0,\\ \gamma b_X(t) & \text{if } \gamma>0.\end{cases}$$

<u>Condition 2</u>

The sequence k(n), such that $k\to\infty$, and $k/n\to 0$, satisfies

$$\sqrt{k}\,\widetilde{A}\left(\frac{n}{k}\right)\to 0,\text{ as } n\to\infty.$$

Proposition 1 (Resnick and Stărică, 1999)**:**

Assume that $F\in D\left(H_{\gamma}\right)$, and Conditions 1, 2 hold. Then, the following convergence properties hold in $D(0,\infty)$ as $n\to\infty$:

for $\gamma\geq 0$,

$$\sqrt{k}\left(1-\frac{1}{2}\left(1-\frac{\left(M_1([ky])\right)^2}{M_2([ky])}\right)^{-1}\right)\to\frac{1}{y}\left\{B(y)+\int_0^y B(u)u^{-1}\ln(u^{-1})du+(\ln y-2)\int_0^y B(u)u^{-1}du\right\}$$

and

$$\sqrt{k}\left(\hat{\gamma}_M([ky])-\gamma\right)\to\frac{1}{y}\left\{\int_0^y B(u)u^{-1}\ln(u^{-1})du+(\gamma+\ln y-2)\int_0^y B(u)u^{-1}du+(1-\gamma)B(y)\right\},$$

while for $\gamma<0$,

$$\sqrt{k}\left(1-\frac{1}{2}\left(1-\frac{\left(M_1([ky])\right)^2}{M_2([ky])}\right)^{-1}-\gamma\right)\to-\frac{(1-2\gamma)(1-\gamma)^2}{\gamma}y^{-1}\times$$

$$\left\{y^{\gamma}\int_0^y B(u)u^{-\gamma-1}du-(1-2\gamma)y^{2\gamma}\int_0^y B(u)u^{-2\gamma-1}du-\frac{\gamma}{1-\gamma}B(y)\right\}$$

and

$$\sqrt{k}\left(\hat{\gamma}_M([ky])-\gamma\right) \to -\frac{(1-2\gamma)(1-\gamma)^2}{\gamma} y^{-1} \times$$
$$\left\{ y^{\gamma}\int_0^y B(u)u^{-\gamma-1}du-(1-2\gamma)y^{2\gamma}\int_0^y B(u)u^{-2\gamma-1}du-\frac{\gamma}{1-\gamma}B(y)\right\}.$$

Proposition 2 (Resnick and Stărică, 1999):

Assume the hypotheses of proposition 1 hold and that $0<u<1$. Then in $D(0,\infty)$ and for $n\to\infty$, it holds that

For $\gamma \geq 0$

$$\sqrt{k}\left(\frac{1}{k(1-u)}\sum_{p=[ku]+1}^{k}\hat{\gamma}_M(p)-\gamma\right)\to$$
$$\frac{1}{1-u}\times\int_u^1\frac{1}{y}\left\{-\int_0^y B(u)u^{-1}\ln u du+(\gamma+\ln y-2)\int_0^y B(u)u^{-1}du+(1-\gamma)B(y)\right\}dy$$

and for $\gamma<0$

$$\sqrt{k}\left(\frac{1}{k(1-u)}\sum_{p=[ku]+1}^{k}\hat{\gamma}_M(p)-\gamma\right)\to-\frac{(1-2\gamma)(1-\gamma)^2}{\gamma(1-u)}\times$$
$$\int_u^1\frac{1}{y}\left\{y^{\gamma}\int_0^y B(u)u^{-\gamma-1}du-(1-2\gamma)y^{2\gamma}\int_0^y B(u)u^{-2\gamma-1}du-\frac{\gamma}{1-\gamma}B(y)\right\}.$$

5.3 Robust Estimators Based on Excess Plots

In the derivation of Hill estimator, we have mentioned that one of Hill estimator's motivating ideas was based on mean excess plot. In the sequel we explore this approach, underline its deficiencies and try to find some 'better' (actually, more robust) alternatives in the context of excess plots.

A graphical tool for assessing the behaviour of a d.f. F is the mean excess function (MEF), also known as mean residual life. More precisely, the limit behaviour 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 also of interest in actuarial studies (especially in reinsurance), survival analysis and reliability; it is related to other well-known functions such as the Lorenz curve and the hazard function. MEF is essentially the expected value of excesses over a threshold value u. The formal definition is as follows

Definition: *Mean Excess Function, Mean Excess Plot* (Beirlant et al., 1996)

Let X be a positive r.v. X with d.f. F and with finite first moment. Then the mean excess function (MEF) of X is

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

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

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

$$\hat{e}(u) = \frac{\sum_{i=1}^{n} \left(X_i - u\right) 1_{(u,\infty)}\left(X_i\right)}{\sum_{i=1}^{n} 1_{(u,\infty)}\left(X_i\right)}.$$

Usually, the MEP is evaluated at the points $u = X_{i:n}$, where $\left(X_{1:n} \geq X_{2:n} \geq ... \geq X_{n:n}\right)$ are the (descending) order statistics of the sample. In that case, MEF takes the form

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

<u>Properties of MEF</u>

A detailed investigation of MEF and its properties, as well as long lists of the formulae of MEF for several known d.f.'s, can be found in Embrechts, et al. (1997) and in Beirlant et al. (1996). In the sequel we present the most important properties that are useful for the present scope.

- Any continuous d.f. F is uniquely determined by its MEF. The "1-1" relationship can be described by the formula

$$1-F(x)=\frac{e(0)}{e(x)}\exp\left\{-\int_0^x \frac{1}{e(u)}du\right\},\ x\geq 0.$$

- If X~Exp(λ), then $e(u)=\lambda^{-1}$, for all u>0.

Based on this property we can discriminate the d.f.'s as follows. If the MEF of a d.f. ultimately (i.e. as $u\to x_F$) increases, then the d.f. is heavier-tailed than the exponential d.f. and is called sub-exponential, while if MEF ultimately decreases, the corresponding d.f. has weaker than exponential tails, and is called super-exponential. Ultimately constant behaviour of MEF implies that the d.f. is medium tailed, i.e. it has tails equivalent to exponential tails. Still, as Drees and Reiss (1996) mention, when the MEF is increasing, yet ultimately there is a concave tendency, then the underlying d.f. will be close (in the upper-tail) to a Weibull d.f. with parameter c<1, i.e. the d.f. is medium tailed. A more thorough discussion on sub-exponentiality is cited in the appendix of Embrechts, et al. (1997).

- If X is of Pareto type with index 1/γ (i.e. $X\in MDA(H_\gamma)$, γ>0), then

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

Intuitively, this last property 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 the MEF converge to the true value of 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 1-F is regularly varying with index –1/γ (whenever $X_{(k+1):n}$ diverges, with arbitrarily large probability, to ∞). 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.

However, though theoretically the values of MEF consistently estimate the parameter γ, in practice strong random fluctuations of the empirical MEF and 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 exactly is the part of plot that mostly concerns us, that is the part that theoretically informs us about the tail behaviour 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'. It is striking 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.

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

where [x] is the largest integer not larger that x
(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 γ>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).

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

The performance of both of these families of alternative estimators is going to be explored (via simulation) in the next section.

5.4 Simulation Comparison of Extreme-Value Index Estimators

5.4.1 Details of Simulation Study

In this section, we try to investigate and compare, via Monte Carlo methods, the performance of several of the extreme-value index estimators introduced in the fourth chapter, as well as the performance of the modifications suggested previously in the current chapter. Apart from the standard form of estimators, we apply to all these the averaging procedure presented in section 5.2. Resnick and Stărică (1997, 1999) suggest (and prove the adequacy and good properties of) this procedure only in the context of Hill and moment estimator. Anyhow, we apply it to other extreme-value index estimators, so as to empirically evaluate its performance to these estimators. In addition, apart from these mean-averaged estimators, we apply analogously a median-averaging procedure to our estimators. Moreover, for $\gamma > 0$, we also examine estimators based on median excess plot as well as on trimmed mean excess plot (see section 5.3.). The following table 5.1 contains all the estimators that are included in the present simulation study. These estimators are compared with respect to the distributions of table 5.2.

Table 5.1. Estimators included in the simulation study

Estimators	Formula
Standard Estimators	
Pickands estimator (for $\gamma \in \Re$)	$\hat{\gamma}_P = \frac{1}{\ln 2} \ln\left(\frac{X_{M:n} - X_{2M:n}}{X_{2M:n} - X_{4M:n}} \right)$
Hill estimator (for $\gamma > 0$)	$\hat{\gamma}_H = \frac{1}{k} \sum_{i=1}^{k} \ln X_{i:n} - \ln X_{k+1:n}$
Adapted Hill estimator (for $\gamma \in \Re$)	$\hat{\gamma}_{adH} = \frac{1}{k} \sum_{i=1}^{k} \ln(UH_i) - \ln(UH_{k+1})$
Moment estimator (for $\gamma \in \Re$)	$\hat{\gamma}_M = M_1 + 1 - \frac{1}{2}\left(1 - \frac{(M_1)^2}{M_2} \right)^{-1}$, $M_j \equiv \frac{1}{k} \sum_{i=1}^{k} \left(\ln X_{i:n} - \ln X_{(k+1):n} \right)^j$

Moment-Ratio estimator (for γ>0)	$\hat{\gamma}_{MR} = \frac{1}{2} \cdot \frac{M_2}{M_1}$
QQ estimator (for γ>0)	$\hat{\gamma}_{qq} = \frac{\sum_{i=1}^{k} \ln \frac{i}{k+1} \left\{ \sum_{j=1}^{k} \ln X_{j:n} - k \ln X_{i:n} \right\}}{k \sum_{i=1}^{k} \left(\ln \frac{i}{k+1} \right)^2 - \left(\sum_{i=1}^{k} \ln \frac{i}{k+1} \right)^2}$
Peng's estimator (for $\gamma \in \Re$)	$\hat{\gamma}_L = \frac{M_2}{2M_1} + 1 - \frac{1}{2} \left(1 - \frac{(M_1)^2}{M_2} \right)^{-1}$
W estimator (for γ<1/2)	$\hat{\gamma}_W = 1 - \frac{1}{2} \left(1 - \frac{(L_1)^2}{L_2} \right)^{-1}$
Mean-Averaged Estimators	(s=0.5)
Averaged Pickands estimator (for $\gamma \in \Re$)	$av\hat{\gamma}_P(k) = \frac{1}{(1-s)k} \sum_{p=[ks]+1}^{k} \hat{\gamma}_P(p)$
Averaged Hill estimator (for γ>0)	$av\hat{\gamma}_H(k) = \frac{1}{(1-s)k} \sum_{p=[ks]+1}^{k} \hat{\gamma}_H(p)$
Averaged adapted Hill estimator (for $\gamma \in \Re$)	$av\hat{\gamma}_{adH}(k) = \frac{1}{k(1-s)} \sum_{p=[ks]+1}^{k} \hat{\gamma}_{adH}(p)$
Averaged moment estimator (for $\gamma \in \Re$)	$av\hat{\gamma}_M(k) = \frac{1}{k(1-s)} \sum_{p=[ks]+1}^{k} \hat{\gamma}_M(p)$
Averaged moment-ratio estimator (for γ>0)	$av\hat{\gamma}_{MR}(k) = \frac{1}{k(1-s)} \sum_{p=[ks]+1}^{k} \hat{\gamma}_{MR}(p)$
Averaged QQ estimator (for γ>0)	$av\hat{\gamma}_{qq}(k) = \frac{1}{(1-s)k} \sum_{p=[ks]+1}^{k} \hat{\gamma}_{qq}(p)$
Averaged Peng's estimator (for $\gamma \in \Re$)	$av\hat{\gamma}_L(k) = \frac{1}{k(1-s)} \sum_{p=[ks]+1}^{k} \hat{\gamma}_L(p)$
Averaged W estimator (for γ<1/2)	$av\hat{\gamma}_W(k) = \frac{1}{k(1-s)} \sum_{p=[ks]+1}^{k} \hat{\gamma}_W(p)$
Median-Averaged Estimators	(s=0.5)
Median-Averaged Pickands estimator (for $\gamma \in \Re$)	$med.av\hat{\gamma}_P(k) = med\{\hat{\gamma}_P(p), p = [ks]+1,\ldots,k\}$
Median-Averaged Hill estimator (for γ>0)	$med.av\hat{\gamma}_H(k) = med\{\hat{\gamma}_H(p), p = [ks]+1,\ldots,k\}$
Median-Averaged adapted Hill estimator (for $\gamma \in \Re$)	$med.av\hat{\gamma}_{adH}(k) = med\{\hat{\gamma}_{adH}(p), p = [ks]+1,\ldots,k\}$
Median-Averaged moment estimator (for $\gamma \in \Re$)	$med.av\hat{\gamma}_M(k) = med\{\hat{\gamma}_M(p), p = [ks]+1,\ldots,k\}$

Median-Averaged moment-ratio estimator (for $\gamma>0$)	$med.av\hat{\gamma}_{MR}(k) = med\{\hat{\gamma}_{MR}(p), p=[ks]+1,\dots,k\}$
Median-Averaged QQ estimator (for $\gamma>0$)	$med.av\hat{\gamma}_{qq}(k) = med\{\hat{\gamma}_{qq}(p), p=[ks]+1,\dots,k\}$
Median-Averaged Peng's estimator (for $\gamma\in\Re$)	$med.av\hat{\gamma}_{L}(k) = med\{\hat{\gamma}_{L}(p), p=[ks]+1,\dots,k\}$
Median-Averaged W estimator (for $\gamma<1/2$)	$med.av\hat{\gamma}_{W}(k) = med\{\hat{\gamma}_{W}(p), p=[ks]+1,\dots,k\}$
Estimators based on Excess Plot	
Estimator based on Median Excess Plot (for $\gamma>0$)	$\hat{\gamma}_{med} = \frac{1}{\ln(2)}\left(\ln X_{([k/2]+1):n} - \ln X_{(k+1):n}\right)$
Estimators based on Trimmed Mean Excess Plot (for $\gamma>0$) p=0.01, 0.05, 0.10	$\hat{\gamma}_{trim} = \frac{1}{k-[pk]}\sum_{j=[pk]+1}^{k}\ln X_{j:n} - \ln X_{(k+1):n}$

Table 5.2. Distributions used in the simulation study

Name	Other parameters	Extreme-value index γ
Burr	(τ, λ) = (0.25, 1), (0.55, 1), (0.5, 2), (1, 0.5)	0.25, 0.55, 1, 2
Fréchet	-	0.25, 0.55, 1, 2
Log-gamma	α=1, λ = 4, 1/0.55, 1, 0.5	0.25, 0.55, 1, 2
Log-logistic	-	0.25, 0.55, 1, 2
Pareto	-	0.25, 0.55, 1, 2
Weibull	λ=1, τ=0.5, 1.5	0
Exponential	-	0
Log-normal	μ=100, σ=1	0
Normal	μ=10, σ=1	0
Gamma	α=1, β=0.5, 1.5	0
Beta	(α, β) = (0.5, 3), (2, 3), (0.5, 0.5), (2, 0.5)	-1/3, -2
Uniform	(α, β) = (0,1), (5, 10)	-1

In particular, from each of these distributions, 1000 samples were generated of moderate size (n=100) and 500 samples of large size (n=1000), based on which the performance of the estimators is examined. In our study, the performance of any estimator of γ, is evaluated in terms of the bias, standard error and root mean square error of the estimator based on k upper order statistics (where k ranges from 1 up to sample size). The root mean square error (RMSE), being a combination of standard deviation and bias, is essentially the basis for comparisons of estimators. In the next section we summarize the main results of the simulation study. More details can be found in the tables in the appendix.

5.4.2 Discussion of Simulation Results

Before proceeding to the discussion of the results, it should be noted that the performance of the estimators did not seem to remain stable for data stemming from different distributions. For this reason, in the sequel we provide the main findings of the simulation study distinguishing for each different class of distributions. More general remarks are provided at the end of this section.

- **Burr Distribution**

In the case of Burr distribution, the estimation of extreme-value index, which ranges in the interval $(0,+\infty)$, seems to depend on the value of extreme-value index itself. So, for γ=0.25 (i.e., more generally speaking, for $\gamma < \frac{1}{2}$) the best estimator is Hill estimator (it has the smallest rmse and in most cases the smallest bias). Moreover, in small samples (n=100) Hill, also, has the smallest std, while for n=1000 Moment-Ratio displays the smallest std. The mean averaging procedure improves the performance of Pickands estimator. The same general result holds for γ=0.55 but only for small samples. For γ=0.55 (large samples) and for γ=1 Moment-Ratio outperforms the Hill estimator. Finally, for γ>1 (i.e. for γ=2) Moment estimator is preferable for large sample sizes (n=1000), while W is best when it comes to small sample sizes. As long as trimming effect is concerned, no clear effect exists.

- **Fréchet Distribution**

For samples of small size (n=100), Hill estimator, in general, displays the best performance. For Fréchet data with γ up to 1, Hill estimator has the smallest rmse and, most of the times, the smallest bias and std. However, when γ exceeds 1 (in our case for γ=2) the variability of Hill is much increased and W stands out as the best estimator. When we are dealing with large simulated data-sets (n=1000), though, the performance of

Hill remains in the same level, Moment-Ratio estimator improves greatly (its bias and, consequently, its rmse is reduced significantly) and turns out to be the best estimator. Again this holds for values of γ smaller than or equal to 1, since for $\gamma=2$, Moment estimator has the smallest rmse. As long as the averaging procedures are concerned, the only improvement is observed in the case of mean-averaged Pickands estimator.

- **Log-Gamma Distribution**

For large sample sizes (n=1000) and γ up to 1, the best estimator is, undeniably, the Moment-Ratio estimator. It has the smallest bias, std, as well as rmse. However, when γ exceeds 1, the behavior of Moment-Ratio deteriorates (mainly due to the large increase of std) and though, still, Moment-Ratio has the smallest bias, it is Moment the estimator with the smallest rmse. A point that is worthy to be pointed out here is that while the behavior of most estimators deteriorates as γ exceeds 1, the performance of W estimator is improved. Actually, now, W has the smallest std. The situation is quite different when we are dealing with small samples (n=100). In those cases, it is the Hill the best estimator (smallest bias, std, rmse). Of course the behavior of Moment-Ratio is not disappointing since it is the second best estimator (not differing much from the Hill). The only case where Hill is surpassed by another estimator is for $\gamma=2$ and for small k (k=12, 20), when W estimator displays the smallest std and rmse. No satisfactory improvement can be attributed to the averaging procedures.

- **Log-Logistic Distribution**

When we are dealing with large samples the situation is quite clear. The best estimator is the Moment-Ratio estimator, since it always has the smallest rmse (in many cases it also has minimum bias or std). However, the evaluation of estimators gets more complicated for small sample sizes (n=100). Generally speaking, we could say that among all estimators the estimator based on 10% trimmed mean excess plot is the best estimator (smallest rmse). If we confine our comparison among the standard estimators, then Hill and Moment-Ratio estimators are the most preferable. For $\gamma<1$ and for small and moderate choice of k (k=12, 20) it is the Hill which has the smallest rmse (while the second smallest rmse is achieved by the Moment-Ratio), while when a larger portion of data is used in the estimation (k=40) the Moment-Ratio stands out as the best estimator (followed by the Hill). Irrespectively of the choice of k, in the case of $\gamma>1$, the best estimator is W.

- **Pareto Distribution**

As was the case for distributions previously discussed, the Moment-Ratio estimator tends to be, in general, preferable. For large samples and γ up to 1, Moment-Ratio has the smallest rmse and the smallest bias (for γ smaller than 1 it also has the smallest std). However, when γ exceeds 1, Moment estimator "outguns" Moment-Ratio. Even then, Moment-Ratio has the smallest bias, while W has the smallest std. Again, for small samples the situation is somewhat different. For γ up to 1, Hill is the best estimator (smallest rmse, bias and, usually, smallest std). But for $\gamma=2$ and for k small (k=12, 20) its variability is much increased while the behavior of W is improved leading to a minimum rmse attributed to W. It is useful to add that Moment-Ratio is the second best estimator here.

- **Weibull Distribution**

Weibull d.f. belongs to the maximum domain of attraction of Gumbel d.f., i.e. the extreme value index equals zero. That means that not all of the estimators under examination should actually be applied to that type of data. Indeed, according to theoretical results, Hill, Moment-Ratio and QQ estimators are not consistent for $\gamma \leq 0$. However, in the context of our simulation study we applied all estimators even in cases that they are not applicable (for comparison reasons). The simulation results here do not suggest that any single estimator is uniformly best. So, for large samples and $\tau=0.5$ Peng is the best estimator, while for $\tau=1.5$ the result depends on the number of upper order statistics used (k). For large k, again Peng is the best estimator. However for smaller k (k=12, 20) QQ and Moment estimators display better performances (though even then the performance of Peng is not bad). The situation is somehow similar for small samples (n=100). That is, for $\tau=0.5$ Peng is preferable but for $\tau=1.5$, it is outperformed by Moment and Moment-Ratio estimators.

- **Exponential Distribution**

For large sample sizes (n=1000) Moment estimator has the best performance (for all choices of k). Apart from having the smallest rmse, it also exhibits minimum bias. As long as standard error is concerned, Moment-Ratio has the smallest values (though Moment-Ratio is theoretically applicable only for $\gamma > 0$). The mean averaging procedure improves the performance of Pickands estimator (though it still remains inferior to other estimators). Estimators based on trimmed mean excess plots have a better performance than simple Hill estimator (the larger the proportion of trimming the better the performance of the estimator, while the median excess estimator is worse than Hill.

For small samples (n=100) Moment estimators perform better only when a large portion of the sample is used in the estimation (k=40). Even then when we are dealing with averaged estimators, Pickands outperforms the Moment. When a smaller k is used in the estimation of extreme-value index (k=12 or 20) Moment-Ratio performs better (in terms of rmse) and has the smallest std, while Pickands has the smallest bias.

Generally speaking, for estimating the extreme-value index in the case of exponential d.f.s Moment estimator is most preferable for large samples, while for small samples we should opt for mean-averaged Pickands estimators.

- **Log-Normal Distribution**

For large samples the situation is quite clear, since Peng is the best estimator for all choices of k. However, things are not so simple for n=100. In that case, the choice of best estimator seems to depend on the choice of k. So, for large k, Peng is the best estimator, while for small k Moment and Moment-Ratio outperform it.

- **Normal Distribution**

In the case of normally distributed data the zero extreme-value index seems to be better estimated by the Moment-Ratio estimators. This holds for all choices of k, for small and for large sample sizes. The only problem seems to be the fact that Moment-Ratio estimator is not theoretically applicable when extreme-value index equals zero. Other estimators with satisfactory behaviour are Hill and Q-Q estimators (both of which are, again, not applicable for zero extreme-value index). As long as averaging procedures are concerned, we should note that mean averaging greatly improves the performance of the standard Pickands estimator. However, mean averaged Pickands estimator is not one of the best estimators.

- **Gamma Distribution**

Here, for the first time the mean averaging procedure provides rather promising results. Though there is not a uniformly best estimator (for all choices of k and sample size n), the mean averaged Pickands estimator displays the most satisfactory behaviour (in most cases it has the smallest rmse). It is really impressive its improvement over the simple, standard Pickands estimator. Other estimators with acceptable performance are Peng's, Moment, Moment-Ratio and Hill estimator.

- **Beta Distribution**

Beta distribution belongs to the maximum domain of attraction of Weibull d.f, since it has finite upper end-point. That means that the extreme-value index that we are trying to estimate is negative and, as previously mentioned, not all estimators are applicable here.
When we are dealing with large samples the situation is quite clear. Moment estimator is uniformly the best estimator (with the smallest rmse). Also, Peng's performance is almost as good as Moment's. Moreover, here it is evident the beneficial effect of the averaging procedures (the mean as well as the median averaging). Mean averaging procedure substantially improves the performance of Pickands estimator, while both averaging procedures improve the performance of Moment and Peng's estimator. So, in most cases it is the mean averaged Moment estimator that has the smallest rmse among all tested estimators. These results hold for different choices of k and parameters α,β of the beta distribution (i.e. for different values of the extreme-value index). However, this is not so when it comes to small sample sizes. Actually, for small sample sizes (n=100 in our case) the situation is much more complicated. The performance of the estimators depends not only on the value of k but also on the value of the extreme-value index γ itself. So, for values of γ close to 0 (-1/3 in particular) mean averaged Pickands estimator seems to be the best choice, while among the standard estimators Moment-Ratio and Moment are more preferable. As the value of γ draws away from zero (γ=-2) the performance of all

estimators deteriorates (the values of rmse's are evidently larger). The behaviour of Moment, Moment-Ratio and Peng's estimator is very disappointing. Pickands is the best estimator among standard estimators. Moreover, the effect of averaging procedures is much less significant. Anyhow, it is the median averaged Pickands and Moment estimators that exhibit the smallest rmse.

- **Uniform Distribution**

Uniform distribution is also bounded to the right and its tails are characterized by extreme-value index equal to -1. The situation here is somewhat similar to the case of Beta distribution. More specifically, for large sample sizes (n=1000) Moment estimator is the most preferable estimator, followed by Peng's. Averaging procedures (mean and median) improve their performances, while mean averaging also improves the performance of Pickands estimator. Here it is the median averaging procedure that provides better results than the mean averaging. These results hold for small samples of uniformly distributed data. The only difference in this case is that for small k (k=12) standard estimators display very large rmse's and the 'good' estimators previously mentioned are outperformed by estimators such as Hill and Moment-Ratio.

<u>General Comments</u>

- *For $\gamma>0$, MDA(Fréchet)*

For large sample sizes Moment-Ratio seems to be the most preferable estimator. It is usually the best estimator, in terms of minimum rmse. Even in cases that other estimators outperform it, it is one of the bests, while in no case does it display very unsatisfactory performance. It is interesting to note that W estimator tends to be appropriate for distributions with extreme-value index γ larger than 1, though for smaller values of γ its performance can be very unsatisfactory. So it may be risky to use that estimator, since in real-life applications the value of γ is unknown. For small samples (in our case n=100) Hill estimator turns out to be the best choice, while Moment-Ratio and Moment estimators can also be regarded as safe options. Among averaging procedures, only the mean averaging of Pickands estimator is effective. However, the improvement is not large enough to out-beat the other standard estimators. On the other hand, the trimming procedure concerning the Hill estimator slightly improves the performance of Hill. This result combined with the fact that standard Hill estimator is, in some cases, the best estimator lead to even better results. Still further exploration on this issue is required.

- *For γ=0, MDA(Gumbel)*

The maximum domain of attraction of Gumbel distribution contains a wide range of distributions differing a lot. Consequently there is not a uniformly superior estimator. However, by looking more carefully the above results one could deduce the conclusion that Peng's is the most preferable estimator of the extreme-value index. Moment and (surprisingly) Moment-Ratio also display an adequate behaviour. The usefulness of averaging procedures in these cases should also be stressed out. These procedures have an obvious profitable impact on Pickands estimator, so that mean-averaged Pickands estimator can also be regarded as an adequate estimator of zero γ.

- *For γ<0, MDA(Weibull)*

This class contains upper-bounded distributions. Though the shape of distributions differs a lot from the distributions with γ=0, the behaviour of extreme-value index estimators in these two classes of d.f.'s displays great analogues. Here, Moment and Peng's estimators are undeniably the most preferable estimators and the beneficial effect of both mean and median averaging procedures is even more evident. Moreover, as we deviate from zero (and positive) values of γ, the inadequacy of estimators such as Hill, Moment-Ratio and so on is more clear.

5.5 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, cut and clear 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 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. We have already mentioned that 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 Cauchy, uniform, exponential, generalized extreme-value). 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. In the sequel, we describe two such approaches that stem from totally different reasoning (though in all cases the ultimate objective is to find a sequence k(n) that minimizes the asymptotic mean square error of the estimator).

5.5.1 Regression Approach

Remember that according to the graphical justification of 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 data-set (i.e. to the right of the anchor point $\left(-\ln\left(\frac{k+1}{n+1}\right), \ln X_{k+1:n}\right)$). 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 Hill estimator (for $\gamma > 0$) and 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

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

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

Theoretically, it would suffice to compute MSE_{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 $w_{j,k}^{opt}$, which depend on a parameter ρ of the 2nd order behaviour of F, are known. So, Beirlant et al. (1996) propose an iterative algorithm for the search of the optimum k. The idea of this algorithm is as follows:

- An initial value k_0 is found by minimizing a function of data
- Original estimators of γ, ρ (γ_0, ρ_0) are estimated based on k_0.
- Based on the estimates of γ, ρ (γ_i, ρ_i) a new k_i is found that minimizes the MSE of the above expression
- Again revised estimates of γ, ρ are obtained
- And so on, until the improvement in the estimated MSE is smaller than some tolerance level.

5.5.2 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 Pickands and moment estimator.

As before, they want to determine the value of k, $k_{opt}(n)$, minimizing the asymptotic mean square error $E_F\left(\hat{\gamma}(k)-\gamma\right)$, $\hat{\gamma}$ refers either to Pickands $\hat{\gamma}_P$ or 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 depends on whether we use Pickands or moment estimator) and F by the empirical d.f. F_n. This amounts to 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 $k_{opt}(n)$ with the help of a second bootstrap.

The proposed algorithm can be summarized in the following steps

Step 1:

Select with replication a random sample of size n_1 (n_1<<n) from the original sample of size n.

Denote this (bootstrapped) sample as $\left(X_1^*,\ldots,X_{n_1}^*\right)$ and its order statistics as $\left(X_{1:n1}^* \geq \ldots \geq X_{n1:n1}^*\right)$.

Based on this sample compute the extreme-value index estimators $\hat{\gamma}_M^*(k)$ (or $\hat{\gamma}_P^*(k)$) and $\hat{\gamma}_+^*(k)$ for k=1,2,...,n_1.

Compute the quantity $q_{n_1}^*(k) = \left(\hat{\gamma}_s^*(k) - \hat{\gamma}_+^*\right)$, for k=1,2,...,$n_1$ (s = 'P' or 'M').

Step 2:

Repeat this procedure r times independently (r can be taken as large as necessary), i.e. obtain the sequences $q_{n_1,i}^*(k)$ for i=1,...,r.

Compute $S_{n_1}(k) = \frac{1}{r}\sum_{i=1}^{r} q_{n_1,i}^*(k)$

Step 3:

Minimize $S_{n_1}(k)$ with respect to k. Let $\overline{k}_{opt}^*(n_1) = \arg\min_k S_{n_1}(k)$.

Step 4:

Repeat step 1 to 3 independently with the number n_1 replaced by $n_2 = \left(n_1\right)^2 / n \quad < n_1$.

Obtain, thus, $\overline{k}_{opt}^*(n_2) = \arg\min_k S_{n_2}(k)$.

Step 5:

Calculate $k_{opt}(n)$ on the basis of $\overline{k}_{opt}^*(n_1)$ and $\overline{k}_{opt}^*(n_2)$. The exact formula for estimation can be found in Draisma et al. (1999).

The authors test their proposed bootstrap procedure on various d.f.'s (such as Cauchy, generalized Pareto, generalized extreme-value) 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 we have 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.