Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation

J. Danielsson
University of Iceland
L. de Haan, L. Peng, C.G. de Vries
Erasmus university Rotterdam

January 29, 1997

Abstract

We use a subsample bootstrap method to get a consistent estimate of the asymptotically optimal choice of the sample fraction, in the sense of minimal mean squared error, which is needed for tail index estimation. Unlike previous methods our procedure is fully self contained. In particular, the method is not conditional on an initial consistent estimate of the tail index; and the ratio of the first and second order tail indices is left unrestricted, but we require the ratio to be strictly positive. Hence the current method yields a complete solution to tail index estimation as it is not predicated on a more or less arbitrary choice of the number of highest order statistics.

Key Words and Phrases: Tail index, Bootstrap, Bias, Mean squared error.

1 Introduction

Let $X_1, X_2, \cdots$ be independent random variables with common distribution function F with a regularly varying tail

$$1 - F(x) = x^{-1/\gamma} L(x) \quad x \to \infty, \quad \gamma > 0$$

where L is a slowly varying function and $1/\gamma$ is the index of regular variation. Various estimators for estimating γ have been proposed (see Hill (1975), Pickands III (1975), de Haan and Resnick (1980), Hall (1982), Mason(1982), Davis and Resnick (1984), Csörgő, Deheuvels and Mason (1985), Hall and Welsh (1985)). These estimators are functions of the upper k order statistics of a sample size n and share a common problem. When k is small the variance of the estimator of γ is large and the use of large k introduces (asymptotic) bias in the estimation. Balancing the variance and bias components is important in the applications of extreme value theory, because this minimizes the asymptotic mean squared error (AMSE).

If the underlying probability distribution is known, the asymptotically optimal value of k can be determined in minimal mean squared error sense, assuming that the second order condition (see Dekkers and de Haan (1993) and de Haan and Peng (1994)) holds:

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \tag{1.1}$$

for all $x > 0$, where U is the right continuous inverse function of the function $1/(1-F)$(notation: $U = (\frac{1}{1-F})^{-}$), and where $\rho < 0$ is the second order parameter. We note that the limit function in (1.1) is necessarily of the stated form in (1.1), where $|A(t)|$ is a regularly varying function with index ρ (see Geluk and de Haan (1987)). We write $|A| \in RV_\rho$.

The asymptotically optimal value of k depends on the unknown parameter γ and the unknown second order scale function A. So if the only source of information is the sample, this optimal value is not available. Hall (1990) obtained the asymptically optimal value of k adaptively by using a bootstrap method. In Hall's method the initial value of k has to be specified, in the sense that it would yield a consistent estimate of γ. Moreover, a very strict second order condition, $A(t) = ct^{-1}$ is required. The problem of choosing the right k-value is therefore not really solved. But Hall's idea of using a subsample bootstrap procedure is applicable when it is applied to a statistic for which the asymptotic mean has a known value independent of the values of γ and ρ. This statistic can then be used to estimate the AMSE without having first to estimate its mean as in Hall(1990). For such a statistic one can take just the difference between two alternative estimators for γ. Hence the asymptotic mean of this statistic is trivially equal to zero, but the AMSE as a function of k still converges at the same rate.

In summary, this paper proposes a bootstrap method to obtain the asymptotically optimal value of k adaptively under a more general second order condition $A(t) \sim ct^\rho$, $t \to \infty$ ($\rho < 0$), and such that the initial value of k does not have to be specified. An altogether different approach to the problem is taken in a recent manuscript by Drees and Kaufmann (1996).

2 Main Results

Let $X_{n,1} \leq \cdots \leq X_{n,n}$ be the order statistics of $X_1, \cdots, X_n$. Hill's estimator is defined by

$$\gamma_n(k) := \frac{1}{k}\sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}.$$

Various authors have considered the asymptotic normality of γ_n (see Haeusler and Teugels (1985), Csörgő and Mason (1985)). We can minimize the mean squared error of γ_n to get the asymptotically optimal choice of k, but it depends on the unknown parameter γ and function $A(t)$ (see Dekkers and de Haan (1993)). Here we turn to the powerful bootstrap tool.

The mean squared error of γ_n is defined as

$$MSE(n,k) := E(\gamma_n(k) - \gamma)^2.$$

The idea is to estimate this MSE via the bootstrap and to extract out the MSE minimizing k value. But as is shown in Hall(1990), one can not use the usual approach to bootstrap $MSE(n,k)$ because it seriously underestimates bias. Therefore one draws resamples $\mathcal{X}_{n_1}^* = \{X_1^*, \cdots, X_{n_1}^*\}$ from $\mathcal{X}_n = \{X_1, \cdots, X_n\}$ with replacement. Let $n_1 < n$ and $X_{n_1,1}^* \leq \cdots \leq X_{n_1,n_1}^*$ denote the order statistics of $\mathcal{X}_{n_1}^*$ and define

$$\gamma_{n_1}^*(k_1) := \frac{1}{k_1}\sum_{i=1}^{k_1} \log X_{n_1,n_1-i+1}^* - \log X_{n_1,n_1-k_1}^*.$$

Hall (1990) uses the bootstrap estimate

$$M\hat{S}E(n_1,k_1) = E((\gamma_{n_1}^*(k_1) - \gamma_n(k))^2 | \mathcal{X}_n).$$

In this setup k has to be chosen such that $\gamma_n(k)$ is consistent. This then permits an estimate of k_1 for sample size n_1. The problem is, however, that

k itself is unknown. So we want to replace $\gamma_n(k)$ in the above expression by a more suitable statistic.

Define

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

Note that $M_n(k)/(2\gamma_n(k))$ is a consistent estimator of γ. We propose to use the following bootstrap estimate

$$Q(n_1, k_1) := E((M^*_{n_1}(k_1) - 2(\gamma^*_{n_1}(k_1))^2)^2|\mathcal{X}_n)$$

where $M^*_{n_1}(k_1) = \frac{1}{k_1}\sum_{i=1}^{k_1}(\log X^*_{n_1,n_1-i+1} - \log X^*_{n_1,n_1-k_1})^2$.

It can be shown that the statistic $M_n(k)/(2\gamma_n(k)) - \gamma_n(k)$ is asymptotically normal with asymptotic mean equal to 0. But as is shown in the following two theorems, the k-value that minimizes $MSE(n,k)$ and the k-value that minimizes $E(M_n(k) - 2(\gamma_n(k))^2)^2$ are of the same general order (with respect to n), under some conditions.

Theorem 1. *Suppose (1.1) holds and $k \to \infty$, $k/n \to 0$. Determine $k_0(n)$ such that $MSE(n,k)$ is minimal. Then*

$$k_0(n) = \frac{n}{s^{-}\left(\frac{\gamma^2(1-\rho)^2}{n}\right)}(1+o(1)) \in RV_{-2\rho/(1-2\rho)}, \qquad \textit{as } n \to \infty$$

where s^- is the inverse function of s, with s given by

$$A^2(t) = \int_t^\infty s(u)\,du(1+o(1)) \qquad \textit{as } t \to \infty.$$

For the existence of such a monotone function see Lemma 2.9 of Dekkers and de Haan (1993).

Theorem 2. *Suppose (1.1) holds and $k \to \infty$, $k/n \to 0$. Determine $\bar{k}_0(n)$ such that $E(M_n(k) - 2(\gamma_n(k))^2)^2$ is minimal. Then*

$$\bar{k}_0(n) = \frac{n}{s^{-}\left(\frac{\gamma^2(1-\rho)^4}{n\rho^2}\right)}(1+o(1)), \qquad \textit{as } n \to \infty.$$

Corollary 1.

$$\frac{\bar{k}_0(n)}{k_0(n)} \to (1-\frac{1}{\rho})^{\frac{2}{1-2\rho}} \quad (n \to \infty).$$

The next theorem shows that the optimal k_1 for a subsample of size n_1 can be estimated consistently. The idea behind the proof is as follows. Note that in the subsample procedure the bootstrapped values $X^*_{n_1,i}$ are still drawn from the full sample. Averaging over the subsample bootstrapped version of Q then amounts to calculating this statistic on the full sample with too many(i.e. too low) order statistics so that the bias dominates the variance. The next theorem shows that the argmin $Q(n_1, k_1)$ as a function of k_1 is asymptotic to a deterministic sequence. In contrast argmin $Q(n, k)$ as a function of k is only asymptotic to a random sequence. So, the motive for using the subsample bootstrap procedure is therefore different from that of Hall(1990).

Theorem 3. *Suppose (1.1) holds and* $k_1 \to \infty$, $k_1/n_1 \to 0$, $n_1 = O(n^{1-\epsilon})$ *for some* $0 < \epsilon < 1$. *Determine* $k^*_{1,0}(n_1)$ *such that*

$$E((M^*_{n_1}(k_1) - 2(\gamma^*_{n_1}(k_1))^2)^2|\mathcal{X}_n)$$

is minimal. Then

$$\frac{k^*_{1,0}(n_1)s^-(\frac{\gamma^2(1-\rho)^4}{n_1\rho^2})}{n_1} \xrightarrow{p} 1, \qquad \text{as } n \to \infty.$$

While Theorem 3 gives the optimal k_1 for sample size n_1, we would like to use the full sample. This can be achieved modulo a conversion factor.

Corollary 2. *Suppose (1.1) holds for* $A(t) \sim ct^\rho$, $t \to \infty$ *and* $k_1 \to \infty$, $k_1/n_1 \to 0$, $n_1 = O(n^{1-\epsilon})$ *for some* $0 < \epsilon < 1$. *Then*

$$\frac{k^*_{1,0}(n_1)}{\bar{k}_0(n)}/(\frac{n_1}{n})^{\frac{2\rho}{2\rho-1}} \xrightarrow{p} 1, \qquad \text{as } n \to \infty.$$

The conversion factor can be calculated consistently as follows.

Theorem 4. *Let* $n_1 = O(n^{1-\epsilon})$ *for some* $0 < \epsilon < 1/2$ *and* $n_2 = (n_1)^2/n$. *Suppose (1.1) holds for* $A(t) \sim ct^\rho$, $t \to \infty$ *and* $k_i \to \infty$, $k_i/n_i \to 0$ $(i = 1, 2)$. *Determine* $k^*_{i,0}$ *such that*

$$E((M^*_{n_i}(k_i) - 2(\gamma^*_{n_i}(k_i))^2)^2|\mathcal{X}_n)$$

is minimal $(i = 1, 2)$. *Then*

$$\frac{(k^*_{1,0}(n_1))^2}{k^*_{2,0}(n_2)}(\frac{(\log k^*_{1,0}(n_1))^2}{(2\log n_1 - \log k^*_{1,0}(n_1))^2})^{\frac{\log n_1 - \log k^*_{1,0}(n_1)}{\log n_1}}/k_0(n) \xrightarrow{p} 1 \tag{2.1}$$

as $n \to \infty$.

Remark 1. *From Theorem 4 we can achieve the optimal choice of k asymptotically, therefore we can get the asymptotically optimal Hill's estimator which is calculated using the asymptotically optimal choice of k.*

Corollary 3. *Suppose the conditions of Theorem 4 hold. Define*

$$\hat{k}_0(n) := \frac{(k^*_{1,0}(n_1))^2}{k^*_{2,0}(n_2)} \left(\frac{(\log k^*_{1,0}(n_1))^2}{(2\log n_1 - \log k^*_{1,0}(n_1))^2} \right)^{\frac{\log n_1 - \log k^*_{1,0}(n_1)}{\log n_1}}.$$

Then $\gamma_n(\hat{k}_0)$ has the same asymptotic efficiency as $\gamma_n(k_0)$.

3 Proofs

Let $Y_1, \cdots, Y_n$ be independent random variables with common distribution function $G(x) = 1 - x^{-1}, (x \geq 1)$. Let $Y_{n,1} \leq \cdots \leq Y_{n,n}$ be the order statistics of $Y_1, \cdots, Y_n$. Note that $\{X_{n,n-i}\}_{i=1}^n \stackrel{d}{=} \{U(Y_{n,n-i})\}_{i=1}^n$ with the function U defined in the Introduction.

Lemma 1. *Let $0 < k < n$ and $k \to \infty$. We have*

(1) for $n \to \infty$, $Y_{n,n-k}/(\frac{n}{k}) \to 1$ in probability.

(2) for $n \to \infty$, (P_n, Q_n) is asymptotically normal with means zero, variance 1 and 20 respectively and covariance 4, where

$$P_n := \sqrt{k}\{\frac{1}{k}\sum_{i=1}^k \log Y_{n,n-i+1} - \log Y_{n,n-k} - 1\}$$

and

$$Q_n := \sqrt{k}\{\frac{1}{k}\sum_{i=1}^k (\log Y_{n,n-i+1} - \log Y_{n,n-k})^2 - 2\}.$$

Proof. Similar to the proof of Lemma 3.1 of Dekkers and de Haan (1993). □

Proof of Theorem 1. We are going to use the method of Dekkers and de Haan (1993), which we shall outline, since a similar reasoning is used in the proofs of Theorem 2 and Theorem 3.

Relation (1.1) is equivalent to the regular variation of the function $|\log U(t) - \gamma \log t - c_0|$ with index ρ for some constant c_0 (see Geluk and de Haan (1987), II.1). Then (1.1) holds with $A(t) = \rho(\log U(t) - \gamma \log t - c_0)$. Applying extended Potter's inequalities to the function A we get that for each $0 < \epsilon < 1$ there exists $t_0 >$ such that for $t \geq t_0$ and $tx \geq t_0$

$$(1-\epsilon)x^\rho e^{-\epsilon|\log x|} - 1 \leq \frac{\log U(tx) - \log U(t) - \gamma \log t}{A(t)/\rho} \leq (1+\epsilon)x^\rho e^{\epsilon|\log x|} - 1. \tag{3.1}$$

Applying this relation with t replaced by $Y_{n,n-k}$ and x replaced by $Y_{n,n-i}/Y_{n,n-k}$, adding the inequalities for $i = 1, 2, \cdots, k$ and dividing by k we get

$$\gamma_n \approx \gamma + \gamma P_n/\sqrt{k} + \rho^{-1} A(Y_{n,n-k})(1 \pm \epsilon)\{\frac{1}{k}\sum_{i=1}^{k}(\frac{Y_{n,n-i+1}}{Y_{n,n-k}})^{\rho \pm \epsilon} - 1\}.$$

Now

$$\{\frac{Y_{n,n-i+1}}{Y_{n,n-k}}\}_{i=1}^{k} \stackrel{d}{=} \{Y_i\}_{i=1}^{k}$$

with $Y_1, \cdots, Y_k$ i.i.d. with common distribution function $1 - 1/x$. Hence by the weak law of large numbers

$$\gamma_n \approx \gamma + \gamma P_n/\sqrt{k} + \rho^{-1}(1 \pm \epsilon)(\frac{1}{1-\rho \mp \epsilon} - 1)A(Y_{n,n-k}),$$

i.e.

$$\gamma_n = \gamma + \gamma P_n/\sqrt{k} + (1-\rho)^{-1} A(\frac{n}{k}) + o_p(A(\frac{n}{k}))$$

(note that in the latter term we have replaced $Y_{n,n-k}$ by n/k which can be done since $|A|$ is regularly varying). Hence

$$E(\gamma_n - \gamma)^2 \approx \gamma^2/k + A^2(\frac{n}{k})/(1-\rho)^2.$$

We can assume (see Lemma 2.9 of Dekkers and de Haan(1993)) that A^2 has a monotone derivative s which is then regularly varying with index $2\rho - 1$. Consequently $s^-(1/t)$ (s^- denoting the inverse of s) is regularly varying with index $1/(1-2\rho)$. The result is then obtained by minimizing the right hand side of the equation above. $\square$

Proof of Theorem 2. From the proof of Theorem 1 we get

$$\gamma_n \stackrel{d}{=} \gamma + \gamma P_n/\sqrt{k} + d_1 A(Y_{n,n-k}) + o_p(A(n/k)) \tag{3.2}$$

with $d_1 = \frac{1}{1-\rho}$ and hence

$$\gamma_n^2 \stackrel{d}{=} \gamma^2 + 2\gamma^2 P_n/\sqrt{k} + 2\gamma d_1 A(Y_{n,n-k}) + o_p(A(n/k)). \tag{3.3}$$

Similarly

$$M_n \stackrel{d}{=} 2\gamma^2 + \gamma^2 Q_n/\sqrt{k} + d_2 A(Y_{n,n-k}) + o_p(A(n/k)) \tag{3.4}$$

where $d_2 = \frac{2\gamma(2-\rho)}{(1-\rho)^2}$. The rest of the proof is similar to that of Theorem 1. $\square$

Proof of Theorem 3. Let G_n denote the empirical distribution function of n independent, uniformly distributed random variables. As n is large enough and $n_1 = O(n^{1-\epsilon})$, we have

$$1/2 \le \sup_{0<t\le n_1(\log n_1)^2} |tG_n^-(\frac{1}{t})| \le 2 \quad \text{a.s.} \tag{3.5}$$

and

$$\sup_{t\ge 2} |\sqrt{t}(G_n(\frac{1}{t}) - \frac{1}{t})| \le \frac{\log n}{\sqrt{n}} \quad \text{a.s.}$$

(see equation (10) and (17) of Chapter 10.5 of Shorack and Wellner (1986)). Hence

$$\sup_{4\le t\le n_1(\log n_1)^2} |\sqrt{\frac{1}{G_n^-(\frac{1}{t})}}[G_n(G_n^-(\frac{1}{t})) - G_n^-(\frac{1}{t})]| \le \frac{\log n}{\sqrt{n}} \quad \text{a.s.}$$

Therefore for all $4 \le t \le n_1(\log n_1)^2$

$$|tG_n^-(\frac{1}{t}) - 1| \le \frac{2\sqrt{t}\log n}{\sqrt{n}} \quad \text{a.s.} \tag{3.6}$$

Let F_n denote the empirical distribution function of $\mathcal{X}_n$, $U_n = (\frac{1}{1-F_n})^-$. Now we use (3.1), (3.5), (3.6),

$$\begin{cases} |\log y| \le 2|y-1| \quad \text{for all } 1/2 \le y \le 2 \\ |y^{-\rho} - 1| \le (-\rho)(2^{-\rho-1} \vee 2^{1+\rho})|y-1| \quad \text{for all } 1/2 \le y \le 2 \end{cases}$$

and

$$\begin{aligned}
\log U_n(t) &= \log F_n^-(1-\tfrac{1}{t}) \\
&\stackrel{d}{=} \log F^-(G_n^-(1-\tfrac{1}{t})) \\
&= \log U(\tfrac{1}{1-G_n^-(1-\frac{1}{t})}) \\
&\stackrel{d}{=} \log U(\tfrac{t}{tG_n^-(\frac{1}{t})}).
\end{aligned}$$

From this we conclude that for any $0<\epsilon<1$ there exists $t_0>4$ such that for $t_0<t<n_1(\log n_1)^2$ and $t_0<tx<n_1(\log n_1)^2$

$$\begin{aligned}
& \tfrac{\log U_n(tx)-\log U_n(t)-\gamma\log x}{A(t)/\rho} \\
\stackrel{d}{=} \quad & \tfrac{\log U(\frac{tx}{txG_n^-(\frac{1}{tx})})-\log U(tx)-\gamma\log\frac{1}{txG_n^-(\frac{1}{tx})}}{A(t)/\rho} \\
& -\tfrac{\log U(\frac{t}{tG_n^-(\frac{1}{t})})-\log U(t)-\gamma\log\frac{1}{tG_n^-(\frac{1}{t})}}{A(t)/\rho} \\
& +\tfrac{\log U(tx)-\log U(t)-\gamma\log x}{A(t)/\rho} \\
& +\tfrac{\gamma\log\frac{1}{txG_n^-(\frac{1}{tx})}}{A(t)/\rho}+\tfrac{\gamma\log\frac{1}{tG_n^-(\frac{1}{t})}}{A(t)/\rho} \\
\leq \quad & (1+\epsilon)(txG_n^-(\tfrac{1}{tx}))^{-\rho}e^{\epsilon|\log(txG_n^-(\frac{1}{tx}))|}-1+\epsilon \\
& -(1-\epsilon)(tG_n^-(\tfrac{1}{t}))^{-\rho}e^{-\epsilon|\log(tG_n^-(\frac{1}{t}))|}+1+\epsilon \\
& +(1+\epsilon)x^\rho e^{\epsilon|\log x|}-1+\epsilon \\
& +|\tfrac{\gamma\rho}{A(t)}|2(|txG_n^-(\tfrac{1}{tx})-1|+|tG_n^-(\tfrac{1}{t})-1|) \quad \text{a.s.} \\
\leq \quad & (1+\epsilon)[(txG_n^-(\tfrac{1}{tx}))^{-\rho}-1]e^{\epsilon\log 2} \\
& +(1+\epsilon)e^{\epsilon\log 2}-1+\epsilon \\
& -(1-\epsilon)[(tG_n^-(\tfrac{1}{t}))^{-\rho}-1]e^{-\epsilon\log 2} \\
& -(1-\epsilon)e^{-\epsilon\log 2}+1+\epsilon \\
& +(1+\epsilon)x^\rho e^{\epsilon|\log x|}-1+\epsilon \\
& +|\tfrac{\gamma\rho}{A(t)}|\tfrac{4\sqrt{t}\log n}{\sqrt{n}}(\sqrt{x}+1) \quad \text{a.s.} \\
\leq \quad & (1+\epsilon)(-\rho)(2^{-\rho-1}\vee 2^{1+\rho})|txG_n^-(\tfrac{1}{tx})-1|e^{\epsilon\log 2} \\
& +(1+\epsilon)e^{\epsilon\log 2}-1+\epsilon \\
& +(1-\epsilon)(-\rho)(2^{-\rho-1}\vee 2^{1+\rho})|tG_n^-(\tfrac{1}{t})-1|e^{-\epsilon\log 2} \\
& -(1-\epsilon)e^{-\epsilon\log 2}+1+\epsilon \\
& +(1+\epsilon)x^\rho e^{\epsilon|\log x|}-1+\epsilon \\
& +|\tfrac{\gamma\rho}{A(t)}|\tfrac{4\sqrt{t}\log n}{\sqrt{n}}(\sqrt{x}+1) \quad \text{a.s.} \\
\leq \quad & [(-\rho)(2^{-\rho+1}\vee 2^{3+\rho})+2|\tfrac{\gamma\rho}{A(t)}|]\tfrac{2\sqrt{t}\log n}{\sqrt{n}}(\sqrt{x}+1) \\
& +(1+\epsilon)e^{\epsilon\log 2}-(1-\epsilon)e^{-\epsilon\log 2} \\
& +(1+\epsilon)x^\rho e^{\epsilon|\log x|}-1+3\epsilon \quad \text{a.s.}
\end{aligned} \tag{3.7}$$

Similarly

$$\begin{array}{ll} & \frac{\log U_n(tx)-\log U_n(t)-\gamma\log x}{A(t)/\rho} \\ \geq & -[(-\rho)(2^{-\rho+1}\vee 2^{3+\rho})+2|\frac{\gamma\rho}{A(t)}|]\frac{2\sqrt{t}\log n}{\sqrt{n}}(\sqrt{x}+1) \\ & -(1+\epsilon)e^{\epsilon\log 2}+(1-\epsilon)e^{-\epsilon\log 2} \\ & +(1-\epsilon)x^{\rho}e^{-\epsilon|\log x|}-1-3\epsilon \quad \text{a.s.} \end{array} \tag{3.8}$$

Inequalities (3.7) and (3.8) are valid in probability with t replaced by Y_{n_1,n_1-k_1} and tx replaced by $Y_{n_1,n_1-i}(i=0,\cdots,k_1-1)$ since

$$4\leq Y_{n_1,n_1-i}\leq Y_{n_1,n_1}(i=1,\cdots,k_1) \text{ in probability}$$

and

$$Y_{n_1,n_1}/(n_1(\log n_1)^2)\to 0 \text{ in probability}$$

for $n_1\to\infty$ and $k_1/n_1\to 0$.

We now minimize

$$E((M^*_{n_1}(k_1)-2(\gamma^*_{n_1}(k_1))^2)^2|\mathcal{X}_n).$$

Note that conditionally, given $\mathcal{X}_n$, P_{n_1} is once again a normalized sum of i.i.d. random variables from an exponential distribution. Hence, when n_1 increases, the distribution of P_{n_1} approaches a normal one. Similarly for Q_{n_1}.

We proceed as in the proof of Theorem 2 and use

$$\gamma^*_{n_1}(k_1)\stackrel{d}{=}\gamma+\gamma P_{n_1}/\sqrt{k_1}+d_1A(Y_{n_1,n_1-k_1})+o_p(A(n_1/k_1))+O(\frac{\log n\sqrt{n_1/k_1}}{\sqrt{n}}),$$

$$(\gamma^*_{n_1}(k_1))^2\stackrel{d}{=}\gamma^2+\frac{2\gamma^2P_{n_1}}{\sqrt{k_1}}+2\gamma d_1A(Y_{n_1,n_1-k_1})+o_p(A(\frac{n_1}{k_1}))+O(\frac{\log n\sqrt{n_1/k_1}}{\sqrt{n}})$$

and

$$M^*_{n_1}(k_1)\stackrel{d}{=}2\gamma^2+\frac{\gamma^2Q_{n_1}}{\sqrt{k_1}}+d_2A(Y_{n_1,n_1-k_1})+o_p(A(\frac{n_1}{k_1}))+O(\frac{\log n\sqrt{n_1/k_1}}{\sqrt{n}}).$$

Note that the term $\frac{\log n\sqrt{n_1/k_1}}{\sqrt{n}}=o(1/\sqrt{k_1})$, so that it can be neglected in the minimization process. The statement of Theorem 3 follows. $\square$

Proof of Corollary 2. Follows easily from Theorem 2 and Theorem 3 and the fact that

$$t^{\frac{1}{2\rho-1}} s^{-}(1/t) \to (-2\rho c^2)^{\frac{1}{1-2\rho}}.$$

□

Proof of Theorem 4. Since $k_{1,0}^* \in RV_{\frac{-2\rho}{1-2\rho}}$ in probability, we have

$$\frac{\log k_{1,0}^*}{\log n_1} \xrightarrow{p} \frac{-2\rho}{1-2\rho}$$

(see Proposition 1.7.1 of Geluk and de Haan (1987)), i.e.,

$$\frac{\log k_{1,0}^*}{-2\log n_1 + 2\log k_{1,0}^*} \xrightarrow{p} \rho. \tag{3.9}$$

Write the result of Corollary 2 for $k_{1,0}^*$ and $k_{2,0}^*$:

$$\begin{cases} \frac{k_{1,0}^*}{\bar{k}_0} / (\frac{n_1}{n})^{\frac{2\rho}{2\rho-1}} \xrightarrow{p} 1 \\ \frac{k_{2,0}^*}{\bar{k}_0} / (\frac{n_2}{n})^{\frac{2\rho}{2\rho-1}} \xrightarrow{p} 1. \end{cases}$$

Hence

$$\bar{k}_0 k_{2,0}^* / (k_{1,0}^*)^2 \xrightarrow{p} 1, \tag{3.10}$$

and by Corollary 1

$$\frac{(k_{1,0}^*(n_1))^2}{k_{2,0}^*(n_2)k_0(n)} \xrightarrow{p} (1 - \frac{1}{\rho})^{\frac{2}{1-2\rho}}.$$

An application of the estimate of ρ from (3.9) gives the result. □

Proof of Corollary 3. We now have a random sequence $\hat{k}_0(n)$ with the property

$$\lim_{n\to\infty} \frac{\hat{k}_0(n)}{k_0(n)} = 1 \quad \text{in probability.}$$

Theorem 4.1 of Hall and Welsh (1985) now guarantees that $\gamma_n(\hat{k}_0(n))$ achieves the optimal rate. □

4 Simulation Experiment

In order to study the performance of our method we set up a monte carlo experiment for some fat tail distributions. Results are reported for the completely skewed stable distribution with characteristic exponent 1/2, the symmetric Cauchy, the Student-t distribution with respectively 4 and 11 degrees of freedom, and the type II extreme value distribution with exponent 1, 5 and 11 respectively. Note that the Student distribution does not satisfy Hall's second order condition, but it can be handled with our method. The sample size was in each case $n = 20000$, and we experimented with three different subsample sizes n_1. Theorem 3 only gives a range within which n_1 has to be chosen. The number of simulations is 250. Per simulation 250 bootstrapped subsamples and 250 subsubsample bootstraps were created. The table gives the theoretical value of the inverse tail index γ and the optimal $k(n)$. The table reports the mean and empirical RMSE(root of mean squared error) of γ_n and the mean of $\hat{k}(n)$. The average run time per distribution was 15 minutes, and the total run time for the entire simulation study was 109 minutes on a Pentium Pro 200. Hence, for a single data set the procedure is quite fast.

Judging from the table this fully automatic procedure seems to perform reasonably well. The RMSE values are about the same for all distributions except for the type II extreme value distribution, which are lower. On the basis of the RMSE criterion it appears to be best (in finite samples) to choose n_1 on the low side. This corresponds with opting for a smaller bias component in the RMSE.

n_1	n_2	mean γ_n	rmse γ_n	mean $\hat{k}_0$
Distribution: stable (0.5) $\gamma = 2.00$ $k_0(n) = 4204$				
2000	200	2.028	0.048	5568.133
4000	800	2.040	0.064	6583.690
6666	2221	2.049	0.071	7050.276
Distribution: t(1) $\gamma = 1.00$ $k_0(n) = 1527$				
2000	200	1.012	0.037	1501.939
4000	800	1.018	0.040	1789.549
6666	2221	1.018	0.047	2057.432
Distribution: t(4) $\gamma = 0.25$ $k_0(n) = 110$				
2000	200	0.270	0.045	125.946
4000	800	0.278	0.055	165.617
6666	2221	0.283	0.053	207.490
Distribution: t(11) $\gamma = 0.0909$ $k_0(n) = 24$				
2000	200	0.135	0.058	36.171
4000	800	0.144	0.066	53.312
6666	2221	0.149	0.069	65.907
Distribution: extreme(1) $\gamma = 1.00$ $k_0(n) = 1474$				
2000	200	1.032	0.045	2403.471
4000	800	1.031	0.053	2806.829
6666	2221	1.035	0.053	3112.811
Distribution: extreme(5) $\gamma = 0.20$ $k_0(n) = 1474$				
2000	200	0.206	0.009	2403.471
4000	800	0.206	0.011	2806.829
6666	2221	0.207	0.011	3112.811
Distribution: extreme(11) $\gamma = 0.0909$ $k_0(n) = 1474$				
2000	200	0.094	0.004	2403.471
4000	800	0.094	0.005	2806.829
6666	2221	0.094	0.005	3112.811

References

[1] S. Csörgő, P. Deheuvels and D. M. Mason (1985). Kernel estimator of the tail index of a distribution. *Ann. Statist. 13, 1050-1077.*

[2] S. Csörgő and D.M. Mason (1985). Central limit theorems for sums of extreme values. *Math. Proc. Camb. Phil. Soc. 98, 547-558.*

[3] A.L.M. Dekkers and L. de Haan (1993). Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis 47(2), 173-195.*

[4] R. Davis and S.I. Resnick (1984). Tail estimates motivated by extreme value theory. *Ann. Statist. 12, 1467-1487.*

[5] H. Drees and E. Kaufmann (1996). Selection of the optimal sample fraction in univariate extreme value estimation. *Unpublished manuscript.*

[6] J. Geluk and L. de Haan (1987). Regular Variation, Extensions and Tauberian Theorems. *CWI Tract 40, Amsterdam.*

[7] L. de Haan and S.I. Resnick (1980). A simple asymptotic estimate for the index of a stable distribution. J. Roy. Statist. Soc. Ser. B 42, 83-87.

[8] P. Hall (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. B 42, 37-42.*

[9] P. Hall (1990). Using the bootstrap to estimate maen squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis 32, 177-203.*

[10] P. Hall and A.H. Welsh (1985). Adaptive estimate of parameters of regular variation. *Ann. Statist. 13, 331-341.*

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

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

[13] D.M. Mason (1982). Law of large numbers for sum of extreme values. *Ann. Probab. 10, 754-764.*

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

[15] G. Shorack and J. Wellner (1986). Empirical Processes with Applications to Statistics. *John Wiley & Sons.*