OPTIMAL ASYMPTOTIC ESTIMATION OF SMALL EXCEEDANCE PROBABILITIES

Ana Ferreira*
ISA, Universidade Técnica de Lisboa, Portugal
on leave at EURANDOM, Technical University Eindhoven, The Netherlands

Abstract. For the estimation of the probability of a tail set beyond the range of the observations an estimator based on Pareto tails can be used. We calculate the optimum number of upper order statistics used for this estimator, in the mean square error sense. Moreover an adaptive procedure is given to find this optimum in a practical situation.

AMS 1991 subject classifications. Primary 62E20, 62G05, 62G30; Secondary 62G09.

Key words and phrases. Generalized Pareto distribution, tail probability estimation, extreme values.

1 INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a sample of n i.i.d. random variables, with common (but unknown) distribution function F. The aim is to estimate an extreme exceedance probability that is, given a 'high' value x one wants to estimate $1 - F(x)$.

On the one hand, if x is well into the sample range then it is known that $1 - F(x)$ can be estimated via the empirical distribution function ([7] Einmahl, 1990). On the other hand, if x is at the boundary or outside the range of the observations (and then we shall call it a 'high' value) then alternative approaches have to be considered. Empirically this means $P(X > x) \leq 1/n$, and hence we will denote x by x_n and define $p_n = P(X > x_n)$. Therefore in this paper we consider the cases $np_n \to c \geq 0$ where c is a finite real constant, as $n \to \infty$. Note that 'well into the sample range' means $np_n \to \infty$ and in this case the use of the empirical distribution function to estimate p_n is preferred.

For the main conditions we assume that F belongs to the domain of attraction of the Generalized Extreme Value distribution for some real extreme value index γ ([10] Gnedenko, 1943), shortly $F \in \mathcal{D}_\gamma(GEV)$, $\gamma \in \mathbb{R}$.

*Email for correspondence: Ferreira@eurandom.tue.nl

Equivalently, one may write it in terms of tail probabilities

$$\lim_{t\to\infty} t\left\{1 - F\left(b(t) + xa(t)\right)\right\} = 1 - H_\gamma(x) \tag{1.1}$$

for all x, for which $0 < H_\gamma(x) < 1$, and $a(t) > 0$, $b(t)$ are suitable normalising functions; $H_\gamma(x)$ is the Generalized Pareto (GP) distribution function given by

$$H_\gamma(x) = 1 - (1+\gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, \gamma \in \mathbb{R}\,. \tag{1.2}$$

For the exceedance probability estimator we use (as in [4] Dekkers, Einmahl and de Haan, 1989; [5] Dijk and de Haan, 1992),

$$\hat{p}_n(k) = \frac{k}{n} \max\left\{0, \left(1 + \hat{\gamma}_n(k)\frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)\right\}^{-1/\hat{\gamma}_n(k)} \tag{1.3}$$

where $\hat{a}(t)$, $\hat{b}(t)$ and $\hat{\gamma}_n(k)$ are estimators of $a(t)$, $b(t)$ and γ, respectively, n is the sample size and k is an intermediate sequence i.e., $k = k(n)$ such that $k(n)/n \to 0$ and $k(n) \to \infty$, as $n \to \infty$. A motivation for (1.3) may be given as follows. Relations (1.1) and (1.2) suggest, for large x,

$$1 - F(x) \approx \frac{1}{t}\left\{1 + \gamma\frac{x - b(t)}{a(t)}\right\}^{-1/\gamma}, \quad \text{as} \quad t \to \infty\,. \tag{1.4}$$

Then take for t the quantity n/k. Since the condition $\{1 + \hat{\gamma}_n(k)(x_n - \hat{b}(\frac{n}{k}))\,\hat{a}(\frac{n}{k})\} \geq 0$ may not be fulfilled (although asymptotically it will be zero), in practice the maximum value between this quantity and zero is taken.

Our main result concerns the characterization of an optimal rate for the number of upper order statistics, k, to use in (1.3), given a sample of size n. The reasoning is in the same line as in [2] Danielsson, de Haan, Peng and de Vries (1997) and [6] Draisma, de Haan, Peng and Pereira (1998) on extreme value index estimation, and [9] Ferreira, de Haan and Peng (1999) on endpoint and high quantiles estimation. We obtain an optimal rate (in the sense of (2.5)) of order n to some negative power (cf. Theorem 2.1). However this asymptotic result might not be adequate for practical applications. Firstly, it contains parameters ($\tilde{c}$ and ρ') that can not be estimated with sufficient accuracy. Secondly, even in cases when $\tilde{c}$ and ρ' are known, the asymptotic optimal result may be far from the real optimum. Still, our optimal rate may be applied in the adapted bootstrap procedure, as suggested in the aforementioned papers, to optimize the performance of $\hat{p}_n(k)$.

An alternative approach to our problem is given in the work by [13] Hall and Weissman (1997). However they concentrate only on models with positive γ. Moreover, our result allows smaller exceedance probabilities, therefore covering more interesting situations. Other related work on tail estimation includes [3] Davis and Resnick (1984) and [14] Smith (1987).

The paper is organized as follows. Section 2 deals with our main result, namely Theorem 2.1. Following this, a simulation study is carried out on section 3, including results from the adaptive bootstrap. In section 4 we present auxiliary Lemmas and the proof of Theorem 2.1.

Finally, some notation. Let a_n and b_n be two sequences, then $a_n \sim b_n$ means $\lim_{n\to\infty} a_n/b_n = 1$. Let $\gamma_- = \min(\gamma, 0)$ and $\gamma_+ = \max(\gamma, 0)$.

2 ASYMPTOTIC OPTIMAL RATE

Let us assume the first order regular variation condition that is, condition (1.1) in terms of the function $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ (the arrow denoting the generalized inverse function),

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \tag{2.1}$$

for all positive x, where $a(t)$ is a suitable positive function (in particular it might be the same as in (1.1)). Also assume the second order regular variation condition for U. Specifically, suppose there exists a function $A(t) \to 0$, as $t \to \infty$, with constant sign near infinity such that, for all positive x

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \frac{1}{\rho}\left[\frac{x^{\gamma+\rho} - 1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right] \tag{2.2}$$

where $\rho \le 0$ is a real constant. For a detailed analysis on these conditions we refer to [12] de Haan and Stadtmüller (1996).

In order to achieve our main result, a second order condition for the function $\log U$ is needed. We use Theorem A in [6] Draisma, de Haan, Peng and Pereira (1998) (here Lemma 4.2 in section 4). In particular,

$$\lim_{t\to\infty} \frac{\frac{\log U(tx)-\log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-} - 1}{\gamma_-}}{\tilde{A}(t)} = \frac{1}{\rho'}\left[\frac{x^{\gamma_- + \rho'} - 1}{\gamma_- + \rho'} - \frac{x^{\gamma_-} - 1}{\gamma_-}\right] \tag{2.3}$$

where $\tilde{A}(t) \in RV_{\rho'}$ and $\rho' < 0$. Note that the limit function is as in (2.2) but where ρ and γ were replaced by ρ' and γ_-, respectively. Basically ρ' is determined from ρ and γ (see (4.2)).

Denote the right endpoint of a distribution function F by x_0 i.e., $x_0 = \sup\{x : F(x) < 1\}$. Under the assumption $F \in \mathcal{D}_\gamma(GEV)$, x_0 is finite if $\gamma < 0$ and infinite if $\gamma > 0$ (both situations may happen if $\gamma = 0$).

By optimal asymptotic estimation it is meant to minimize, with respect to k, the asymptotic mean square error function, as. $E(\hat{p}_n(k) - p_n)^2$. From the proof of Theorem 2.1 we have that the random variable

$$\sqrt{k}\, r_\gamma(a_n) \left(\frac{\hat{p}_n(k)}{p_n} - 1 \right) \tag{2.4}$$

where $r_\gamma(x) = -\gamma/\log x$ if $\gamma > 0$ and $r_\gamma(x) = -\gamma x^\gamma$ if $\gamma < 0$, and $a_n = k/(np_n)$, equals, in distribution, to

$$\left(N + bias_{\gamma,\rho'} \tilde{A}(\frac{n}{k}) \sqrt{k} \right) (1 + o_p(1))$$

where N is a normally distributed random variable with zero mean and variance var_γ (see (2.12) and (2.13) for var_γ and $bias_{\gamma,\rho'}$, respectively). So, we have that the asymptotic bias component of (2.4) is $bias_{\gamma,\rho'} \tilde{A}(\frac{n}{k})\sqrt{k}$. Let $\tilde{A}(\frac{n}{k})\sqrt{k} \to \lambda$, as $n \to \infty$. Then we have three possibilities: *1)* $\lambda = 0$, *2)* $0 < \lambda < \infty$ and *3)* $\lambda = \infty$. If the intermediate sequence $k = k(n)$ is such that *1)* holds, then (2.4) is asymptotically normally distributed with mean zero. Our approach assume *2)* holds i.e., we deal with the intermediate sequences where (2.4) has, asymptotically, a non zero bias component.

Thus we seek for the asymptotic optimal rate

$$k_0(n) = \arg \inf_k \text{ as. } E(\hat{p}_n(k) - p_n)^2 \,, \tag{2.5}$$

where p_n is the true exceedance probability one wants to estimate and as. $E()$ means the asymptotic expectation according to the limit distribution of $(\hat{p}_n(k) - p_n)$ as discussed earlier. In order to express the restriction to intermediate sequences for k but include the optimal, one must restrain k to some bounds, for instance that it ranges from $\log n$ to $n/(\log n)$. Nonetheless any particular bounds do not play a role in the general result.

The estimators used in (1.3) are (as in [4] Dekkers, Einmahl and de Haan, 1989; [5] Dijk and de Haan, 1992)

$$\hat{\gamma}_n(k) \quad = \quad M_n^{(1)} + 1 - \frac{1}{2}(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}})^{-1}, \tag{2.6}$$

and

$$\hat{a}(\frac{n}{k}) \quad = \quad X_{n-k,n} M_n^{(1)} (1 - \hat{\gamma}_n^-(k)) \tag{2.7}$$

where

$$M_n^{(j)} \quad = \quad \frac{1}{k} \sum_{i=0}^{k-1} (\log \ X_{n-i,n} \ - \log \ X_{n-k,n})^j , \qquad j = 1, 2 \tag{2.8}$$

and

$$\hat{\gamma}_n^-(k) \quad = \quad 1 - \frac{1}{2}(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}})^{-1} = \hat{\gamma}_n(k) - M_n^{(1)}. \tag{2.9}$$

Moreover take $\hat{b}(\frac{n}{k}) = X_{n-k,n}$.

Theorem 2.1. *Suppose $U(t)$ satisfies (2.1) and (2.2). Assume $\rho < 0$, $\gamma > -1/2$, $\gamma \neq 0$, $\gamma \neq \rho$ and $np_n \to$ constant (finite, ≥ 0) as $n \to \infty$. Also let, as $n \to \infty$,*

$$\begin{aligned} &\log x_n = o\left(n^{\frac{-\rho'}{1-2\rho'}}\right) && \text{if } \gamma > 0 \ ; \\ &(x_0 - x_n)^{-1} = o\left(n^{-\frac{\gamma+\rho'}{1-2\rho'}}\right) && \text{if } \gamma < 0 \ , \end{aligned} \tag{2.10}$$

and $\tilde{A}(t) \sim \tilde{c}\, t^{\rho'}$, with $\tilde{c}$ a non zero real constant, as $t \to \infty$.

Then, there exists an optimal $k_0 = k_0(n)$ minimizing the asymptotic mean square error,

$$k_0(n) = \arg\inf_k \ as.\ E(\hat{p}_n(k) - p_n)^2 \ ,$$

such that

$$k_0(n) \sim \begin{cases} \left(\frac{var_\gamma}{-2\rho' \tilde{c}^2 \ bias^2_{\gamma,\rho'}}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} & ,\gamma > 0 \\ \\ \left(\frac{var_\gamma}{\tilde{c}^2 \ bias^2_{\gamma,\rho'}} \ \frac{1+2\gamma}{-2\rho'-2\gamma}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} & ,\gamma < 0 \end{cases} \tag{2.11}$$

where

$$var_\gamma = \begin{cases} 1 + \gamma^2 & ,\gamma > 0 \\ \frac{(1-\gamma)^2(1-3\gamma+4\gamma^2)}{\gamma^2(1-2\gamma)(1-3\gamma)(1-4\gamma)} & ,\gamma < 0 \end{cases} \tag{2.12}$$

and

$$bias_{\gamma,\rho'} = \begin{cases} \frac{\gamma\rho'-\gamma-\rho'}{\rho'(1-\rho')^2} & ,(\lim_{t\to\infty} U(t) - a(t)/\gamma = 0 \ and \ 0 < \gamma < -\rho) \\ & \quad or\ \gamma \geq -\rho \\ \frac{\gamma\rho'+\rho'^2-\gamma-2\rho'}{(1-\rho')^2} & ,(\lim_{t\to\infty} U(t) - a(t)/\gamma \neq 0 \ and \ 0 < \gamma < -\rho) \\ \frac{-3\gamma^2+\gamma+2\gamma^3-2\rho'+2\gamma\rho'+\gamma^2\rho'+\rho'^2}{\gamma(1-\gamma)(1-\gamma-\rho')(1-2\gamma-\rho')} & ,\rho < \gamma < 0 \\ \frac{(\gamma-1)\rho'}{\gamma(1-\gamma-\rho')(\gamma+\rho')(1-2\gamma-\rho')} & ,\gamma < \rho\ . \end{cases} \tag{2.13}$$

Remark 2.2. Note that $k_0(n)$ in the Theorem does not depend on x_n.

Remark 2.3. We restrict attention to $\gamma > -1/2$, since otherwise the extrapolation should be based on extreme rather than intermediate order statistics (cf. [1] Aarssen and de Haan, 1994).

Remark 2.4. Concerning the conditions in (2.10) in Theorem 2.1:

1. Since ρ' and γ are unknown (2.10) may alternatively be written as

$$\begin{aligned} \log x_n &= o\left(n^{\epsilon}\right), \epsilon > 0 \qquad \text{if } \gamma > 0 \\ \left(x_0 - x_n\right)^{-1} &= o\left(n^{\epsilon}\right), \epsilon > 0 \qquad \text{if } \gamma < 0\,. \end{aligned} \tag{2.14}$$

2. In fact (2.10) is only required when $np_n \to 0$ as $n \to \infty$.

3. Conditions (2.10) are equivalent to

$$\begin{aligned} \lim_{n\to\infty} \frac{\log p_n}{n^{-\rho'/(1-2\rho')}} &= 0 \qquad \text{if } \gamma > 0 \\ \lim_{n\to\infty} \frac{p_n^{\gamma}}{n^{-(\gamma+\rho')/(1-2\rho')}} &= 0 \qquad \text{if } \gamma < 0 \end{aligned}$$

   or, equivalently,

$$\lim_{n\to\infty} \frac{\log a_{0n}}{\sqrt{k_0(n)}} = 0 \qquad \text{if } \gamma > 0 \tag{2.15}$$

$$\lim_{n\to\infty} \frac{a_{0n}^{-\gamma}}{\sqrt{k_0(n)}} = 0 \qquad \text{if } \gamma < 0 \tag{2.16}$$

   where $a_{0n} = k_0(n)/(np_n)$. Note that if $\gamma > 0$, (2.15) implies (2.16); conversely if $\gamma < 0$, (2.16) implies (2.15). Therefore our optimal sequence satisfy condition (2.10) in [5] Dijk and de Haan (1992).

Remark 2.5. Suggested by Bahadur-Kiefer representation we can see a correspondence between our result and optimal quantile estimation. Let the quantile estimator for a given exceedance probability p_n be ([4] Dekkers, Einmahl and de Haan, 1989)

$$\hat{x}_{p_n}(k) = X_{n-k,n} + \hat{a}(\frac{n}{k}) \frac{(\frac{k}{np_n})^{\hat{\gamma}_n(k)} - 1}{\hat{\gamma}_n(k)}\,. \tag{2.17}$$

Assume the usual conditions, namely $np_n \to c(\geq 0)$ and $k = k(n)$ such that $k(n)/n \to 0$ and $k(n) \to \infty$, as $n \to \infty$. As referred in [5] Dijk and de Haan (1992) and subsequently analysed in [8] Einmahl (1995), under similar conditions

$$\frac{\sqrt{k}}{a(\frac{n}{k})\; q_\gamma(a_n)}\;(\hat{x}_{p_n}(k) - x_n) \quad \text{and} \quad -\frac{n\; a_n^{\gamma+1}}{q_\gamma(a_n)\sqrt{k}}\;((1-\hat{p}_n(k)) - (1-p_n)) \tag{2.18}$$

have exactly the same asymptotic distribution, normal with mean value zero and variance (2.12) where, as $x \to \infty$,

$$q_\gamma(x) \sim \begin{cases} x^\gamma(\log x)/\gamma & \gamma > 0 \\ (\log x)^2/2 & \gamma = 0 \\ 1/\gamma^2 & \gamma < 0 \end{cases}$$

and $a_n = k/(np_n)$. Later, in [9] Ferreira, de Haan and Peng (1999) the results on quantile estimation in [11] de Haan and Rootzén (1993) were extended and the $k_0(n)$ minimizing the as. $E(\hat{x}_{p_n}(k) - x_n)^2$ was obtained. In the same way, Theorem 2.1 extends the ones in [5] Dijk and de Haan (1992) on exceedance probability estimation. Moreover, the later equality in limit distribution (2.18) still holds if we take in each case the k equal to the respective k_0, the optimal one that minimizes the asymptotic mean square error given in Theorem 2.1 on exceedance probability estimation and the one given in Theorem 2.3 on quantile estimation ([9] Ferreira, de Haan and Peng (1999)). Since

$$\frac{\sqrt{k}}{a(\frac{n}{k})q_\gamma(a_n)}\,\frac{q_\gamma(a_n)\sqrt{k}}{na_n^{\gamma+1}} = \frac{p_n^{\gamma+1}}{c}(1+o(1)) \qquad (n \to \infty)$$

(where $a(t) \sim c_1 t^\gamma$, as $t \to \infty$, with c_1 a real constant, from the second order regular variation condition) we have that

$$as.E(\hat{p}_n(k_0) - p_n)^2 \sim \left(\frac{p_n^{\gamma+1}}{c_1}\right)^2 as.E(\hat{x}_{p_n}(k_0) - x_n)^2 \quad (n \to \infty)$$

and so the minimization of the asymptotic mean square error with respect to k is exactly the same in both cases.

Example 2.6. Generalized Extreme Value distribution. Let $GEV_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}$, $1 + \gamma x > 0$, $\gamma \in \mathbb{R}$. Then $U(t) = ((-\log(1-1/t))^{-\gamma} - 1)/\gamma$, $t > 1$, where $\lim_{t\to\infty} U(t) = x_0 = -1/\gamma$ if $\gamma < 0$ and $x_0 = \infty$ if $\gamma > 0$. The second order parameter, ρ, equals -1 if $\gamma \neq 1$ and -2 otherwise. Possible choices for the auxiliary functions in the regular variation conditions are $a(t) \sim t^\gamma$ and $A(t) \sim (\gamma-1)t^{-1}/2$ if $\gamma \neq 1$; $t^{-2}/6$ if $\gamma = 1$, as $t \to \infty$. Note that $\lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0$. Hence, from (2.11) the asymptotic optimal rate may be calculated.

Example 2.7. Reversed Burr distribution. A random variable Y has Burr distribution function with parameters β, λ and τ if $F_Y(y) = 1 - \beta^\lambda/(\beta + y^\tau)^\lambda$, $y > 0$, $\beta, \lambda, \tau > 0$. We shall denote the distribution function of $X = -Y^{-1}$ by Reversed Burr distribution, say $RB_{\beta,\lambda,\tau}$, which is given by $F_X(x) = 1 - \beta^\lambda/(\beta + (-x)^{-\tau})^\lambda$, $x < 0 = x_0$, $\beta, \lambda, \tau > 0$. This random variable has been referred to in financial applications. The extreme value parameters are $\gamma = -1/(\lambda\tau)$ and $\rho = -1/\lambda$. Note that in order to properly use this model with the suggested methods it must be shifted by a positive constant, say a, so that $x_0 = a > 0$. Therefore in this case $U(t) = a - \beta^{-1/\tau}(t^{1/\lambda} - 1)^{-1/\tau}$, $t > 1$, and $\lim_{t\to\infty} U(t) = x_0 = a$. Possible choices for the auxiliary functions, in the regular variation conditions are $a(t) = \beta^{-1/\tau}t^{-1/\lambda\tau}/\lambda\tau$ and $A(t) = (1 + \tau)t^{-1/\lambda}/\lambda\tau$, as $t \to \infty$. Hence, from (2.11) the asymptotic optimal rate may be calculated.

3 SIMULATION RESULTS

For the simulations we use the distribution families presented in the examples in the previous section, Generalized Extreme Value distribution and Reversed Burr distribution. Specifically $RB_{4,4,2}$, $GEV_{-.1}$, $GEV_{.5}$ and GEV_1. We opted by fixing two exceedance probabilities to be estimated: $1/(n \log n) = .000010875$ and $1/n = .0001$. Then to each distribution function they correspond to a different quantile given by,

$$x_n = \frac{(-\log(1 - p_n))^{-\gamma} - 1}{\gamma} + a \quad \text{and} \quad x_n = -\left(\frac{\beta}{p_n^{1/\lambda}} - \beta\right)^{-1/\tau} + a\ ,$$

for GEV_γ and $RB_{\beta,\lambda,\tau}$, respectively. The parameter a stands for a positive shift in the data set, in order the sample be constituted of positive values. Three methods to estimate the exceedance probability were considered: *1)* (1.3) with k determined by the adaptive bootstrap method, resumed in Appendix A; *2)* (1.3) with k equal to the intermediate sequence $\sqrt{n}$; *3)* empirical distribution function, by calculating the number of values in the sample greater than the respective quantile. Of course the last approach is only used in the case $p_n = 1/n$.

The simulation results are presented in tables 1 and 2 (see also figures 1, 2 and 3). They are based on 100 independent simulations of samples of size $n = 10000$ of each distribution and are resumed in terms of mean, root mean square error and $n*$ (or $n**$). In the first two cases, *1)* and *2)*, $n*$ equals the number of simulations with valid results in 100 (and so in this cases the previous descriptive statistics and graphics use exactly these $n*$ values). When using (1.3) we distinguish the $\hat{p}_n(k) = 0$ case to be non-valid. In the bootstrap (for the technical details,

	1 - bootstrap			2 - $k(n)=\sqrt{n}$		
	mean ($\times 10^4$)	rootmse ($\times 10^4$)	$n*$	mean ($\times 10^4$)	rootmse ($\times 10^4$)	$n*$
$GEV_{-.1}$ (a=4)	.188	.255	86	.180	.227	86
$RB_{4,4,2}$ (a=649)	.223	.244	48	.224	.250	74
$GEV_{.5}$(a=2)	.116	.042	96	.102	.093	100
GEV_1(a=1)	.130	.060	82	.144	.120	100

Table 1: Estimation of $p_n = 1/(n \log n)$, based on 100 independent repetitions of samples of size $n = 10000$; $n*$ is the number of valid simulations in 100.

	1 - bootstrap			2 - $k(n)=\sqrt{n}$			3 - empirical d.f.		
	mean ($\times 10^3$)	rootmse ($\times 10^3$)	$n*$	mean ($\times 10^3$)	rootmse ($\times 10^3$)	$n*$	mean ($\times 10^3$)	rootmse ($\times 10^3$)	$n**$
$GEV_{-.1}$ (a=4)	.105	.069	89	.093	.070	100	.087	.082	64
$RB_{4,4,2}$ (a=649)	.102	.064	67	.105	.072	97	.102	.087	33
$GEV_{.5}$(a=2)	.106	.037	81	.089	.051	100	.101	.093	67
GEV_1(a=1)	.113	.038	82	.113	.061	100	.124	.119	71

Table 2: Estimation of $p_n = 1/n$, based on 100 independent repetitions of samples of size $n = 10000$; $n*$ is the number of valid simulations in 100.

namely the explanation of the quantities $k_0^*(n_1)$ and $k_0^*(n_2)$, we refer to Appendix A) sometimes happens to obtain $k_0^*(n_1)$ less or equal to $k_0^*(n_2)$, or the intermediate consistent estimate of p_n involved in the algorithm is equal to zero, or $\hat{k}_0(n)$ is equal to 0,1 or it is greater than the sample size; all of these also considered non-valid simulations. In *3)*, $n**$ denotes the number of non zero estimates in 100 or, in other words, the number of samples with at least one observation greater than the given x_n.

In general we claim that the bootstrap procedure gave the most accurate results. Nonetheless it involves more effort than the other two approaches.

4 PROOFS

We start giving some auxiliary Lemmas. The following one is part of Lemma 4.3 in [9] Ferreira, de Haan and Peng (1999).

Lemma 4.1. *Suppose condition (2.2) holds with $\rho < 0$. Then*

$$\lim_{\substack{t\to\infty \\ x\to\infty}} \frac{\frac{U(tx)-U(t)}{a(t)} \frac{\gamma}{x^{\gamma}-1} - 1}{A(t)} = \frac{-1}{\rho+\gamma_-}.$$

Proof. See [9] Ferreira, de Haan and Peng (1999). □

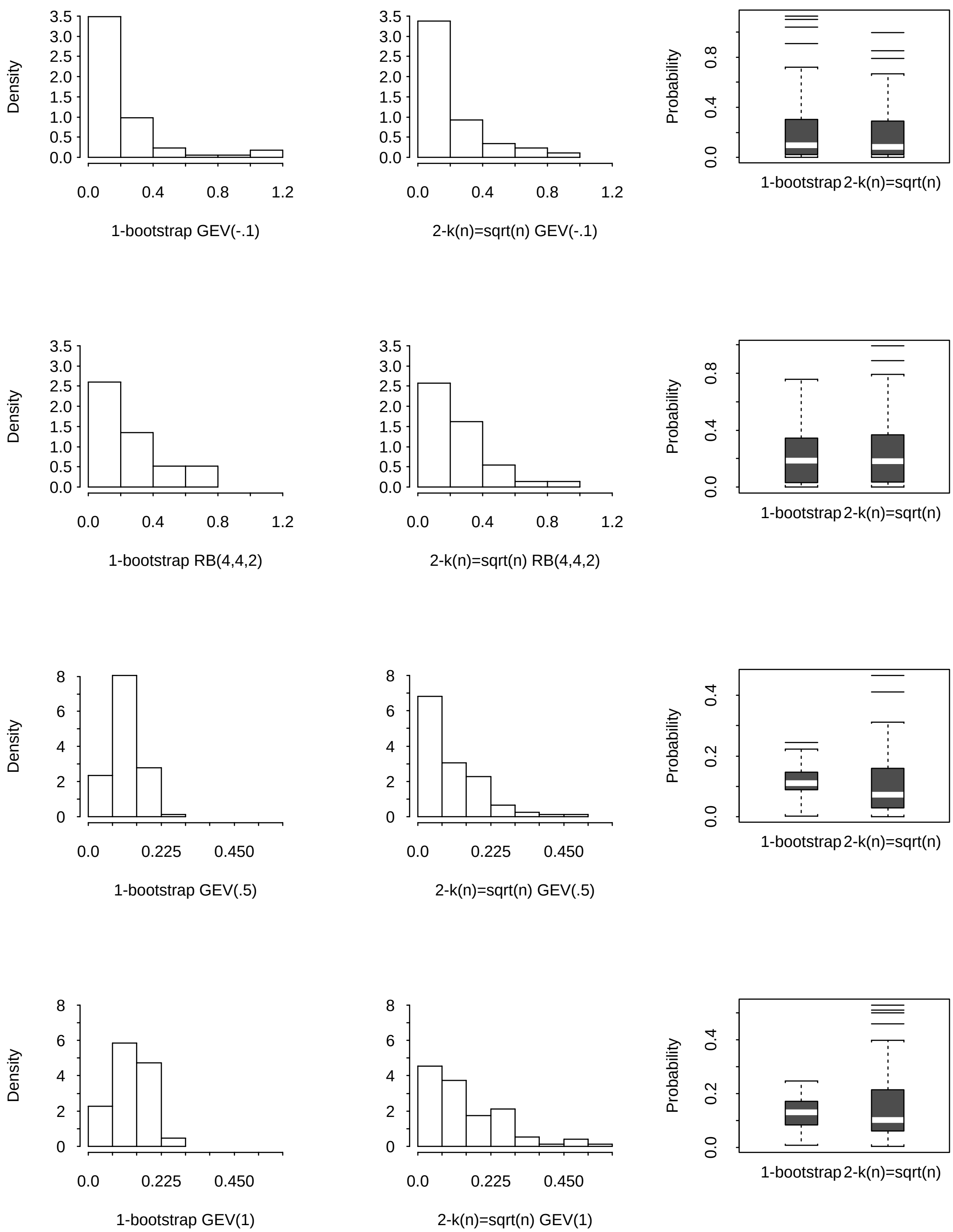

Figure 1: Histograms, with area equal to 1, and boxplots of the estimates of $p_n = 1/(n \log n)$.

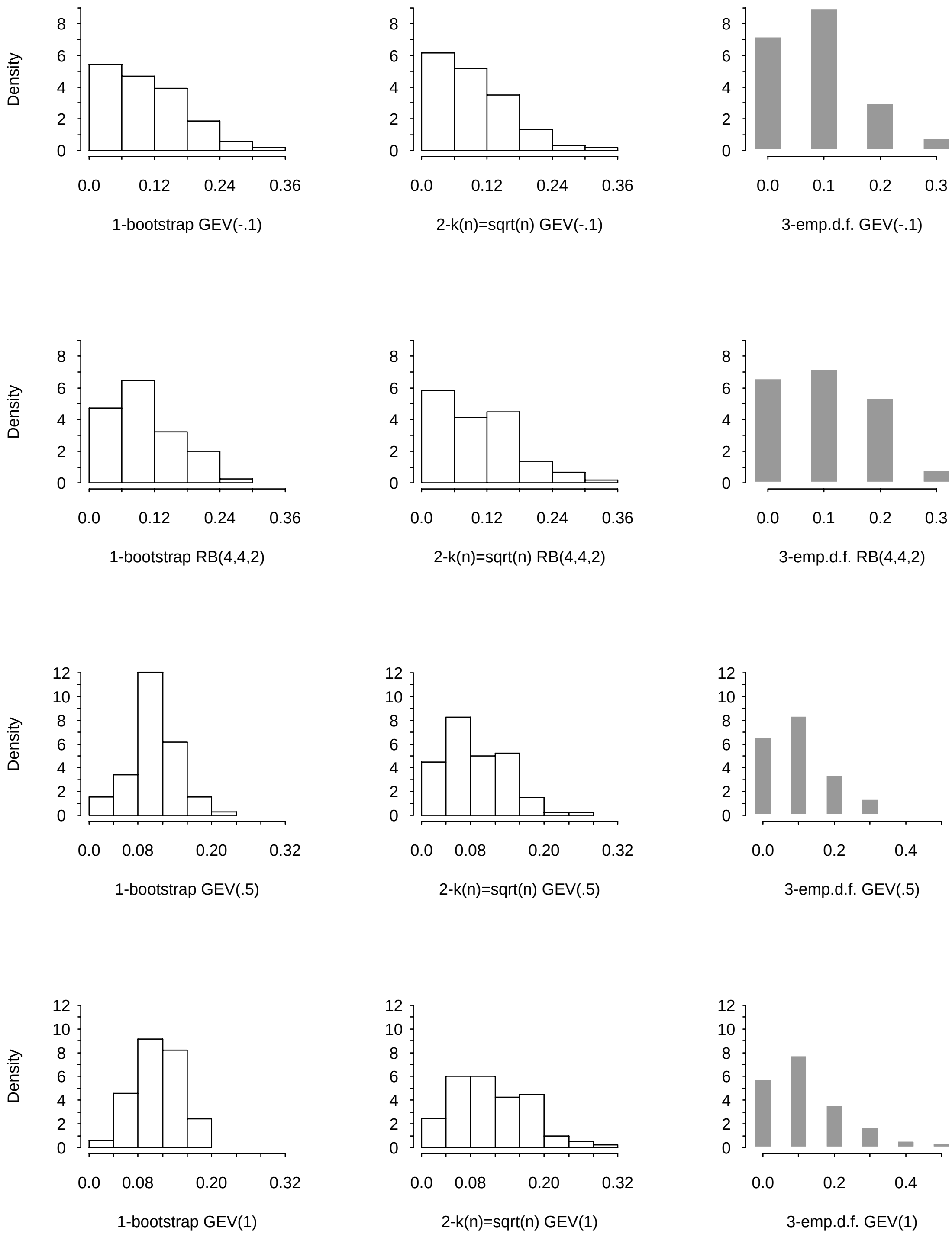

Figure 2: Histograms, with area equal to 1, of the estimates of $p_n = 1/n$.

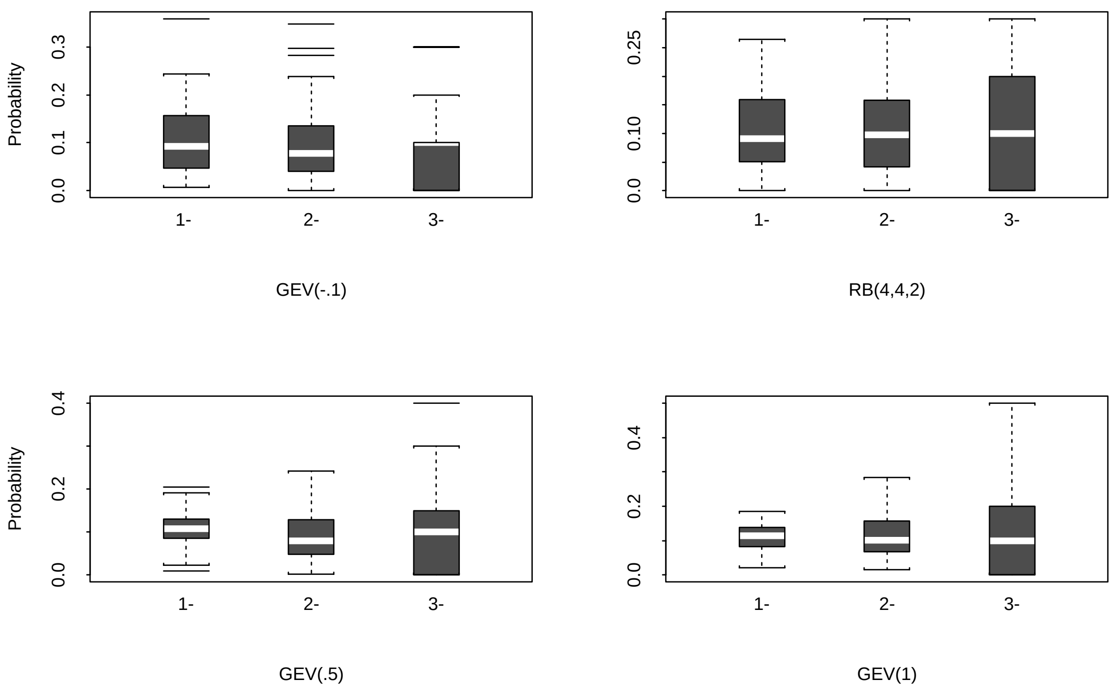

Figure 3: Boxplots of the estimates of $p_n = 1/n$; 1-bootstrap; 2-$k(n) = \sqrt{n}$; 3-empirical d.f.

Lemma 4.2. *Assume condition (2.2) holds with $\rho < 0$ and $U(\infty) > 0$. Suppose that $\gamma \neq \rho$. Then*

$$\lim_{t\to\infty} \frac{\frac{a(t)}{U(t)} - \gamma_+}{A(t)} = c \in [-\infty, \infty]$$

where

$$c = \begin{cases} 0 & \text{if } \gamma < \rho \\ \frac{\gamma}{\gamma+\rho} & \text{if } \gamma > -\rho \\ \frac{\gamma}{\gamma+\rho} & \text{if } 0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0 \\ \pm\infty & \text{if } \rho < \gamma \leq 0 \\ \pm\infty & \text{if } 0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0 \\ \pm\infty & \text{if } \gamma = -\rho \; . \end{cases}$$

Furthermore

$$\lim_{t\to\infty} \frac{\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-} - 1}{\gamma_-}}{\tilde{A}(t)} = \frac{1}{\rho'} \left[\frac{x^{\gamma_- + \rho'} - 1}{\gamma_- + \rho'} - \frac{x^{\gamma_-} - 1}{\gamma_-} \right] \tag{4.1}$$

where

$$\tilde{A}(t) = \begin{cases} A(t) & \text{if } c = 0 \\ \gamma_+ - \frac{a(t)}{U(t)} & \text{if } c = \pm\infty \\ \rho A(t)/(\gamma+\rho) & \text{if } c = \gamma/(\gamma+\rho), \end{cases}$$

$\tilde{A}(t) \in RV_{\rho'}$ *and*

$$\rho' = \begin{cases} -\gamma & \text{if } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0) \\ \gamma & \text{if } \rho < \gamma \leq 0 \\ \rho & \text{if } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0) \\ & \quad \text{or } \gamma < \rho \text{ or } \gamma \geq -\rho\ . \end{cases} \tag{4.2}$$

Proof. See [6] Draisma, de Haan, Peng and Pereira (1998). □

Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with distribution function $1 - 1/y,\ y > 1$. Then $U(Y_1)$, $U(Y_2)$, ... are i.i.d. F.

Lemma 4.3. *Let*

$$M_j = \frac{M_n^{(j)} U^j(Y_{n-k,n})}{a^j(Y_{n-k,n})} - l_j\ , \qquad j = 1, 2$$

with

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1}\{\log U(Y_{n-i,n}) - \log U(Y_{n-k,n})\}^j,$$
$$1/l_1 = 1 - \gamma_-\ ,$$
$$1/l_2 = (1-\gamma_-)(1-2\gamma_-)/2\ .$$

Then under the conditions of Lemma 4.2, for $k = k(n) \to \infty$ and $k(n)/n \to 0$ $(n \to \infty)$

$$M_j = \frac{P_j}{\sqrt{k}} + d_j \tilde{A}(\frac{n}{k}) + o_p(\frac{1}{\sqrt{k}}) + o_p(\tilde{A}(\frac{n}{k})), \quad j = 1, 2$$

where (P_1, P_2) is normally distributed with mean vector zero and covariance matrix

$$\begin{cases} EP_1^2 = \frac{1}{(1-\gamma_-)^2(1-2\gamma_-)} \\ EP_2^2 = \frac{4(5-11\gamma_-)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)(1-4\gamma_-)} \\ E(P_1P_2) = \frac{4}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)} \end{cases}$$

and

$$\begin{cases} d_1 = \frac{1}{(1-\gamma_-)(1-\rho'-\gamma_-)} \\ d_2 = \frac{2(3-2\rho'-4\gamma_-)}{(1-\gamma_-)(1-2\gamma_-)(1-\rho'-\gamma_-)(1-\rho'-2\gamma_-)}\ . \end{cases}$$

Proof. See [9] Ferreira, de Haan and Peng (1999). □

The following is a compilation of Lemmas 4.6-4.11 in [9] Ferreira, de Haan and Peng (1999). In particular they state the consistency of the estimators $\hat{\gamma}_n(k)$ and $\hat{a}(\frac{n}{k})$.

Lemma 4.4. *Under the conditions of Lemma 4.2 with $\rho < 0$, for $k = k(n) \to \infty$ and $k(n)/n \to 0$ $(n \to \infty)$*

$$\frac{\hat{\gamma}_n(k)}{\gamma} = 1 + \left(\frac{\gamma_+}{\gamma} - \frac{4}{\gamma l_1 l_2}\right) M_1 + \frac{2}{\gamma l_2^2} M_2 + \frac{q_{\gamma,\rho} l_1}{\gamma} \tilde{A}(\frac{n}{k})$$

with

$$q_{\gamma,\rho} = \lim_{t\to\infty} \frac{a(t)/U(t) - \gamma_+}{\tilde{A}(t)} = \begin{cases} 0 & \text{if } \gamma < \rho \\ \gamma/\rho & \text{if } (\lim_{t\to\infty} U(t) - a(t)/\gamma_+ = 0 \\ & \quad \text{and } 0 < \gamma < -\rho) \text{ or } \gamma > -\rho \\ -1 & \text{if } (\lim_{t\to\infty} U(t) - a(t)/\gamma_+ \neq 0 \text{ and } 0 < \gamma < -\rho) \\ & \quad \text{or } \rho < \gamma \leq 0 \text{ or } \gamma = -\rho \end{cases},$$

$$\frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} = 1 + \frac{l_2 + 4l_1}{l_1 l_2} M_1 - \frac{2l_1}{l_2^2} M_2 + \gamma \frac{B}{\sqrt{k}} + o_p(A(\frac{n}{k}))$$

with B a standard normal random variable, independent of (P_1, P_2), and

$$\frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})} = \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})) \ .$$

Proof. See [9] Ferreira, de Haan and Peng (1999). □

Proof of Theorem 2.1. Let $a_n = k/(np_n)$. Then

$$\hat{p}_n(k) = \frac{k}{n} \max\left\{0, \left(1 + \hat{\gamma}_n(k) \frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)\right\}^{-1/\hat{\gamma}_n(k)}$$

$$= \frac{k}{n} \max\left\{0, \left(1 + \gamma \frac{\hat{\gamma}_n(k)}{\gamma} \frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})} \left[\frac{U(\frac{n}{k}a_n) - U(\frac{n}{k})}{a(\frac{n}{k})} + \frac{U(\frac{n}{k}) - U(Y_{n-k,n})}{a(\frac{n}{k})}\right]\right)\right\}^{-1/\hat{\gamma}_n(k)}$$

which, by the previous Lemmas, for large values of n, has the same limit behaviour as

$$\begin{aligned} \frac{k}{n} \Bigg\{ 1 + \gamma \frac{\hat{\gamma}_n(k)}{\gamma} \frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})} \Bigg[\frac{a_n^\gamma - 1}{\gamma} + \frac{a_n^\gamma - 1}{\gamma} \frac{-1}{\rho + \gamma_-} A(\frac{n}{k}) + \frac{a_n^\gamma - 1}{\gamma}\, o(A(\frac{n}{k})) \\ - \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})) \Bigg] \Bigg\}^{-1/\hat{\gamma}_n(k)} \quad . \end{aligned} \tag{4.3}$$

The reasoning is divided in the two cases, γ positive and γ negative. First suppose $\gamma > 0$. Then $a_n^{-\gamma} \to 0$ as $n \to \infty$ and so (4.3) becomes

$$\frac{k}{n} \left\{ 1 + \gamma \frac{\hat{\gamma}_n(k)}{\gamma} \frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})} \right.$$

$$\left[\frac{a_n^\gamma - 1}{\gamma} + a_n^\gamma\left(-\frac{1}{\gamma(\rho+\gamma_-)}A(\frac{n}{k}) + o(A(\frac{n}{k}))\right) - \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k}))\right]\Bigg\}^{-1/\hat\gamma_n(k)} .$$

Using the expansions given in Lemma 4.4 this becomes

$$\frac{k}{n}\left\{1+\gamma\left[1+\left(-\frac{l_2+4l_1}{l_1l_2}+\frac{\gamma_+}{\gamma}-\frac{4}{\gamma l_1 l_2}\right)M_1+\left(\frac{2l_1}{l_2^2}+\frac{2}{\gamma l_2^2}\right)M_2-\frac{\gamma B}{\sqrt{k}}+\frac{q_{\gamma,\rho}l_1}{\gamma}\tilde A(\frac{n}{k})\right]\right.$$

$$\left[\frac{a_n^\gamma - 1}{\gamma} + a_n^\gamma\left(-\frac{1}{\gamma(\rho+\gamma_-)}A(\frac{n}{k}) + o(A(\frac{n}{k}))\right) - \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k}))\right]\Bigg\}^{-1/\hat\gamma_n(k)} .$$

Since $A(\frac{n}{k}) = 1_{\{\gamma<\rho\}}\tilde A(\frac{n}{k})$, after working out the multiplication we get

$$\begin{aligned} & p_n\left(\frac{k}{np_n}\right)\left\{a_n^\gamma + a_n^\gamma\left[g_1M_1+g_2M_2-\frac{\gamma B}{\sqrt{k}}+g_3\tilde A(\frac{n}{k})\right]\right\}^{-1/\hat\gamma_n(k)} \\ = \;& p_n\left\{a_n^{\gamma-\hat\gamma_n(k)}\left(1+g_1M_1+g_2M_2-\frac{\gamma B}{\sqrt{k}}+g_3\tilde A(\frac{n}{k})\right)\right\}^{-1/\hat\gamma_n(k)} \end{aligned} \qquad (4.4)$$

where g_1, g_2 and g_3 are non zero real constants depending on γ and ρ'.

In the optimal case $a_n^{\gamma-\hat\gamma_n(k)}$ must converge to one in probability. Note that the second factor in the main brackets converges to one. In fact, we know that there exists a sequence $k = k(n)$ such that $a_n^{\gamma-\hat\gamma_n(k)} \to 1$ $(n\to\infty)$ in probability: take for $k_0(n)$, for example, the optimal one in tail index estimation (for this sequence we have $\hat\gamma_n(k_0)-\gamma = O((k_0(n))^{-1/2})$ and $\log a_{0n}/\sqrt{k_0}= \log(k_0/(np_n))/\sqrt{k_0} \to 0$ - for the later see Remark 2.4.3.). Note that the power $-1/\hat\gamma_n(k)$ has no influence since $-1/\hat\gamma_n(k) \to -1/\gamma$ $(n\to\infty)$ in probability. Therefore, as $n\to\infty$, if $a_n^{\gamma-\hat\gamma_n(k)} \to 1$, then $(a_n^{\gamma-\hat\gamma_n(k)}-1)/((\gamma-\hat\gamma_n(k))\log a_n) \to 1$ in probability and so, at least in the optimal case, the second factor in (4.4) does not contribute asymptotically. Hence the simplified expansion

$$p_n\left\{1+(\gamma-\hat\gamma_n(k))\log a_n + o((\gamma-\hat\gamma_n(k))\log a_n)\right\}^{-1/\hat\gamma_n(k)} .$$

Disregarding terms of smaller order, we get

$$p_n\left\{1-\frac{1}{\gamma}\,\frac{\gamma}{\hat\gamma_n(k)}(\gamma-\hat\gamma_n(k))\log a_n\right\}$$

or,

$$p_n - \frac{p_n}{\gamma}(\gamma-\hat\gamma_n(k))\log a_n .$$

Hence, as $n\to\infty$,

$$\begin{aligned} \gamma^2 as.E(\hat p_n(k)-p_n)^2 \;&\sim\; p_n^2(\log a_n)^2E(\gamma-\hat\gamma_n(k))^2 \\ &\sim\; p_n^2(\log\frac{k}{np_n})^2\left(\frac{var_\gamma}{k}+bias^2_{\gamma,\rho'}\;\tilde c^2\;(\frac{n}{k})^{2\rho'}\right) \end{aligned}$$

where var_γ and $bias_{\gamma,\rho'}$ are from the representation in Lemma 4.4; they are known constants calculated from the covariance matrix given in Lemma 4.3. Thus taking the derivative with respect to k in the last expression and equating it to zero, one gets the result. For more details see [9] Ferreira, de Haan and Peng (1999), e.g. proof of Proposition 4.12.

Next suppose $\gamma < 0$. Then $a_n^\gamma \to 0$ as $n \to \infty$ and so, by Lemma 4.4 relation (4.3) leads to

$$\frac{k}{n}\left\{1+\gamma\left[1-\frac{4}{\gamma l_1 l_2}M_1+\frac{2}{\gamma l_2^2}M_2+\frac{q_{\gamma,\rho}l_1}{\gamma}\tilde{A}(\frac{n}{k})\right]\right.$$

$$\left[1-\frac{l_2+4l_1}{l_1 l_2}M_1+\frac{2l_1}{l_2^2}M_2-\gamma\frac{B}{\sqrt{k}}+o_p(A(\frac{n}{k}))\right]$$

$$\left.\left[\frac{a_n^\gamma-1}{\gamma}+\frac{1}{\gamma(\rho+\gamma)}A(\frac{n}{k})-\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))\right]\right\}^{-1/\hat{\gamma}_n(k)}$$

or

$$\frac{k}{n}\left\{a_n^\gamma+\left(\frac{l_2+4l_1}{l_1 l_2}+\frac{4}{\gamma l_1 l_2}\right)M_1-\left(\frac{2l_1}{l_2^2}+\frac{2}{\gamma l_2^2}\right)M_2+\frac{1}{\rho+\gamma}A(\frac{n}{k})-\frac{q_{\gamma,\rho}l_1}{\gamma}\tilde{A}(\frac{n}{k})\right\}^{-1/\hat{\gamma}_n(k)} .$$

Since $A(\frac{n}{k}) = 1_{\{\gamma<\rho\}}\tilde{A}(\frac{n}{k})$ one may simply write

$$\begin{aligned} & p_n\left(\frac{k}{np_n}\right)\left\{a_n^\gamma+g_4M_1+g_5M_2+g_6\tilde{A}(\frac{n}{k})\right\}^{-1/\hat{\gamma}_n(k)} \\ = \quad & p_n\left\{a_n^{\gamma-\hat{\gamma}_n(k)}+a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6\tilde{A}(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)} \\ = \quad & p_n\left\{1+\left(a_n^{\gamma-\hat{\gamma}_n(k)}-1\right)+a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6\tilde{A}(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)} \end{aligned}$$

where g_4, g_5 and g_6 are non zero real constants depending on γ and ρ'.

Next we prove that

$$\frac{a_n^{\gamma-\hat{\gamma}_n(k)}-1}{a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6\tilde{A}(\frac{n}{k})\right)} \to 0 \qquad (n\to\infty) \tag{4.5}$$

in probability. Note that $\gamma-\hat{\gamma}_n(k) = O_p(g_4M_1+g_5M_2+g_6\tilde{A}(\frac{n}{k}))$. Therefore (4.5) is of the same order as

$$\begin{aligned} & \left|\frac{a_n^{\gamma-\hat{\gamma}_n(k)}-1}{a_n^{-\hat{\gamma}_n(k)}(\gamma-\hat{\gamma}_n(k))}\right| = \left|\frac{a_n^\gamma-a_n^{\hat{\gamma}_n(k)}}{\gamma-\hat{\gamma}_n(k)}\right| = \\ = \quad & \left|\frac{\log a_n}{\gamma-\hat{\gamma}_n(k)}\int_{\hat{\gamma}_n(k)}^{\gamma}a_n^s ds\right| \le (\log\ a_n)a_n^{\max(\hat{\gamma}_n(k),\gamma)} \to 0 \ \text{ as } \ n\to\infty\ . \end{aligned}$$

Hence the numerator in (4.5) is of smaller order than the denominator, and so (4.3) simplifies to

$$p_n \left\{ 1 + a_n^{-\hat{\gamma}_n(k)} \left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right) \right\}^{-1/\hat{\gamma}_n(k)} \quad . \tag{4.6}$$

In the optimal case $a_n^{-\hat{\gamma}_n(k)} \left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right)$ must converge to zero in probability and in fact there exists a sequence such that this holds (for example take the optimal one in tail index estimation together with condition (2.10)). So, expanding (4.6) again we get, neglecting terms of lower order,

$$p_n \left\{ 1 - \frac{a_n^{-\gamma}}{\hat{\gamma}_n(k)} \left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right) \right\}$$

or,

$$p_n - \frac{p_n}{\gamma} a_n^{-\gamma} \left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right) \quad .$$

Hence, as $n \to \infty$,

$$\begin{aligned} as.E(\hat{p}_n(k) - p_n)^2 &\sim E\left[\frac{p_n}{\gamma} a_n^{-\gamma} \left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right) \right]^2 \\ &\sim \left(\frac{p_n^{\gamma+1}}{\gamma} \right)^2 \left(\frac{n}{k} \right)^{2\gamma} E\left(g_4 M_1 + g_5 M_2 + g_6 \tilde{A}(\frac{n}{k}) \right)^2 \\ &\sim \left(\frac{p_n^{\gamma+1}}{\gamma} \right)^2 \left(\frac{n}{k} \right)^{2\gamma} \left(\frac{var_\gamma}{k} + bias^2_{\gamma,\rho'} \ \tilde{c}^2 \ (\frac{n}{k})^{2\rho'} \right) \end{aligned}$$

where var_γ and $bias_{\gamma,\rho'}$ come from the representation in Lemma 4.4; they are known constants calculated from the covariance matrix given in Lemma 4.3. Thus taking the derivative with respect to k in the last expression and equating it to zero, one gets the result. For more details see [9] Ferreira, de Haan and Peng (1999), e.g. proof of Proposition 4.12. Note that in order to assure that a minimum is in fact attained one must assume the extra condition $\gamma > -1/2$. □

A ADAPTIVE BOOTSTRAP ON EXCEEDANCE PROBABILITY ESTIMATION

Without going into details, we remark that theoretically the adaptive bootstrap method to estimate the optimal rate is still valid on exceedance probability estimation. The proof follows the same line as in, e.g., [6] Draisma, de Haan, Peng and Pereira (1998) on extreme value index estimation and [9] Ferreira, de Haan and Peng (1999) on endpoint and high quantiles estimation.

The main step is Theorem 2.1 in section 2. In the following we shall just give the main steps necessary to implement it.

In order to use the bootstrap method an alternative estimator to the exceedance probability must be taken. We still use (1.3) but with

$$\hat{\bar{\gamma}}_n(k) = \sqrt{M_n^{(2)}/2} + 1 - \frac{2}{3}(1 - \frac{M_n^{(1)} M_n^{(2)}}{M_n^{(3)}})^{-1}, \tag{A.1}$$

in place of (2.6). Note that a similar substitution must be done in (2.7). For $M_n^{(3)}$ just take $j = 3$ in (2.8). Denote this alternative estimator of the exceedance probability by $\hat{\bar{p}}_n(k)$. Then we need the asymptotic variance and bias of the random variable $\sqrt{k}\,(r_\gamma(a_n)/p_n)(\hat{p}_n(k) - \hat{\bar{p}}_n(k))$ where $r_\gamma(a_n)$ is the same as in (2.4). Following a similar reasoning as in the proof of Theorem 2.1 one gets

$$\overline{var}_\gamma = \begin{cases} \frac{1+\gamma^2}{4} & ,\gamma > 0 \\ \frac{(1-\gamma)^2(1-6\gamma+35\gamma^2-78\gamma^3+72\gamma^4)}{4\gamma^2(1-2\gamma)(1-3\gamma)(1-4\gamma)(1-5\gamma)(1-6\gamma)} & ,\gamma < 0 \end{cases} \tag{A.2}$$

and

$$\overline{bias}_{\gamma,\rho'} = \begin{cases} -\frac{\rho'+\gamma-\gamma\rho'}{2(1-\rho')^3} & ,\gamma > 0 \\ \frac{2-12\gamma+22\gamma^2-12\gamma^3-5\rho'+22\gamma\rho'-21\gamma^2\rho'+6\rho'^2-12\gamma\rho'^2-2\rho'^3}{2\gamma(1-\gamma)(1-\gamma-\rho')(1-2\gamma-\rho')(1-3\gamma-\rho')} & \\ +\frac{-2+14\gamma-34\gamma^2+34\gamma^3-12\gamma^4+6\rho'-30\gamma\rho'+46\gamma^2\rho'-22\gamma^3\rho'}{2\gamma(1-\gamma)(1-\gamma-\rho')(1-2\gamma-\rho')(1-3\gamma-\rho')\sqrt{(1-\gamma)(1-2\gamma)}} & \\ +\frac{-6\rho'^2+18\gamma\rho'^2-12\gamma^2\rho'^2+2\rho'^3-2\gamma\rho'^3}{2\gamma(1-\gamma)(1-\gamma-\rho')(1-2\gamma-\rho')(1-3\gamma-\rho')\sqrt{(1-\gamma)(1-2\gamma)}} & ,\rho < \gamma < 0 \\ \frac{(1-\gamma)\rho'}{2\gamma(1-\gamma-\rho')(1-2\gamma-\rho')(1-3\gamma-\rho')} & ,\gamma < \rho\,. \end{cases} \tag{A.3}$$

The bootstrap procedure follows: *Step 1)* Select randomly and independently n_1 times ($n_1 = O(n)$) a member from the sample $\{X_1, X_2, \ldots, X_n\}$. Indicate the result by $X_1^*, X_2^*, \ldots X_{n_1}^*$. Form the order statistics $X_{1,n_1}^* \le X_{2,n_1}^* \le \cdots \le X_{n_1,n_1}^*$ and compute the quantities $\hat{p}_n(k)$ and $\hat{\bar{p}}_n(k)$. We denote the resulting quantities by $\hat{p}_n^*(k)$ and $\hat{\bar{p}}_n^*(k)$ for $k = 1, 2, \ldots, n_1 - 1$. Form $q_{n_1,k}^* = (\hat{p}_n^*(k) - \hat{\bar{p}}_n^*(k))^2$ on the basis of these bootstrap estimators; *Step 2)* Repeat step 1 r times independently. This results in a sequence $q_{n_1,k,s}^*$, $k = 1, 2, \ldots, n_1 - 1$ and $s = 1, 2, \ldots, r$. Calculate $\frac{1}{r}\sum_{s=1}^r q_{n_1,k,s}^*$; *Step 3)* Minimize $\frac{1}{r}\sum_{s=1}^{r} q_{n_1,k,s}^*$ with respect to k but reject values which are very small or very near to n_1. Denote the value of k where the minimum is obtained by $k_0^*(n_1)$; *Step 4)* Repeat step 1 up to 3 independently with the number n_1 replaced by $n_2 = (n_1)^2/n$. So n_2 is smaller than n_1. This results in $k_0^*(n_2)$; *Step 5)* Calculate

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)} \; \frac{var_{\hat{\gamma}}\; \overline{bias}^2_{\hat{\gamma},\hat{\rho}'}}{\overline{var}_{\hat{\gamma}}\; bias^2_{\hat{\gamma},\hat{\rho}'}}$$

with $\hat{\gamma}$ any consistent estimator of γ (we have used (1.3) with $k = \sqrt{n}$) and, $\hat{\rho}' = \hat{\rho}'_{n_1}(k_0^*) = \log k_0^*(n_1)/(-2\log n_1 + 2\log k_0^*(n_1))$ a consistent estimator of ρ'. This $\hat{k}_0(n)$, which is obtained adaptively, is asymptotically as good as the optimal number of order statistics in (2.11).

Acknowledgments. The author gratefully acknowledges Prof. Laurens de Haan for his support throughout the work presented in this paper.

References

[1] Aarssen, K. and de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4, 259-281.

[2] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (1997). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Analysis, to appear.*

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

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

[5] Dijk, V. and de Haan, L. (1992). On the estimation of the exceedance probability of a high level. In: P.K. Sen and I.A. Salama, Eds., *Order Statist. and Nonparametrics: Theory and Applications*, North-Holland, Amsterdam, 79-92.

[6] Draisma, G., de Haan, L., Peng, L. and Pereira, T.T. (1998). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes, to appear.*

[7] Einmahl, J.H.J. (1990). The empirical distribution function as a tail estimator. *Statistica Neerlandica* 44, 79-82.

[8] Einmahl, J.H.J. (1995). A Bahadur-Kiefer Theorem beyond the largest observation. *J. Multivariate Analysis* 55, 29-38.

[9] Ferreira, A. de Haan, L. and Peng, L. (1999). Adaptive estimators for the endpoint and high quantiles of a probability distribution. *Technical Report, EURANDOM, Technical University Eindhoven.*

[10] Gnedenko, B.V. (1943). Sur la distribution limite du terme du maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.

[11] de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles. *J. Statistical Planning and Inference* 35, 1-13.

[12] de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Australian Math. Soc. (Series A)* 61, 381-395.

[13] Hall, P. and Weissman, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* 25, 1311-1326.

[14] Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* 15, 1174-1207.