Regression Estimator for the Tail Index

László Németh¹ · András Zempléni¹

Published online: 24 June 2020
© The Author(s) 2020

Abstract

Estimating the tail index parameter is one of the primal objectives in extreme value theory. For heavy-tailed distributions the Hill estimator is the most popular way to estimate this parameter. Several recent publications' aim was to improve the Hill estimator, using different methods, for example the bootstrap, or the Kolmogorov–Smirnov metric. These methods are asymptotically consistent, but for tail index $\xi > 0.5$ the estimations fail to approach the theoretical value for realistic sample sizes. In this paper, we introduce new empirical methods, which combine the advantages of the Kolmogorov–Smirnov approach and the bootstrap. We demonstrate that our estimators are able to estimate large tail index parameters well and might also be useful for relatively small sample sizes. As an application, we consider the classic Danish fire data set and the most destructive natural disasters in Europe.

Keywords Tail index · Bootstrap · Hill estimation · Kolmogorov–Smirnov distance

Mathematics Subject Classification 62G32 · 62F40 · 60G70

1 Introduction

In many applications of probability theory and statistics the most important problem of research is to estimate the high quantiles of a distribution, for example solvency margin calculations for risky investments or estimating the loss caused by a natural disaster, which may be observed once within a given time period (e.g. 100 years). These type of problems could be solved using tools provided by extreme value theory, summarized by [10, 16].

In the 1920's [17] described the limit behaviour of the maximum of i.i.d. samples, which is the basis of many applications of extreme value theory. Later another approach emerged, where the extremal model is based on the values over a high threshold. Its

✉ László Németh
lnemeth@caesar.elte.hu

¹ Department of Probability Theory and Statistics, Institute of Mathematics, Eötvös Loránd University, Budapest, Hungary

theoretical background was developed by [1, 28], the statistical applications by [9], among others—summarized by [25]. Both approaches depend on the tail behaviour of the underlying distribution, which can be measured by the tail index.

Theoretical results for estimating the tail index were proved by [2, 8, 13] and others. Hill [24] constructed an estimator for the tail index parameter, using the k largest values of the ordered sample. Although finding the optimal number of the used sample elements is still a challenge, Hill's estimator remained popular as [11, 14] introduced theoretical methods to calculate the optimal sample fraction. The work of [3, 23] opened the way for numerous methods to find the optimal threshold using bootstrap simulations (introduced by [15]) such as the double bootstrap method by [6], improved by [29], and other resampling-based algorithms by [18–20, 23]. The Kolmogorov–Smirnov method was developed by [7], this does not need simulations.

Hill's estimator is asymptotically consistent for both threshold selection methods, however simulations show that for some sample distributions we need more than 10,000 observations for a reasonably accurate estimation. In contrast to the previous methods we present a new algorithm that skips the direct threshold selection for the sample and calculates the tail index estimation via the tail indices of simulated bootstrap subsamples. In this way it results in acceptable estimators for moderate size samples, like $n \in (200, 2000)$ too.

In this paper, after a short mathematical overview of extreme value theory in Sect. 1.1, we present the Hill estimator (Sect. 2) with possible ways finding the optimal k . Additionally, we analyse the theoretical background of Kolmogorov–Smirnov method and present an approximating algorithm to make double bootstrap method faster. In Sect. 3 we present a new algorithm for estimating the tail index parameter. The steps of calculation can be seen in Sect. 3.1. In a simulation study, we compare the methods (Sects. 3.2, 3.3), present a model selection scheme (Sect. 3.4) and use them on real life datasets (Sect. 4).

1.1 Mathematical Overview

Let $X_1, X_2, \dots, X_n, \dots$ be independent and identically distributed (i.i.d) random variables with distribution function F , and let $M_n = \max(X_1, X_2, \dots, X_n)$. If there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} < x\right) \rightarrow G^*(x)$$

as $n \rightarrow \infty$ and G^* is a non-degenerate distribution function, then it is said that F is in the maximum domain of attraction of G^* . The Fisher–Tippett theorem [17] claims that the limit distribution G^* is the standardized extreme value distribution:

$$G_\xi^*(x) = \begin{cases} \exp\left\{-(1 + \xi x)^{-\frac{1}{\xi}}\right\}, & \text{if } \xi \neq 0, \\ \exp\{-e^{-x}\}, & \text{if } \xi = 0, \end{cases} \quad (1)$$

where $1 + \xi x > 0$ holds. The parameter ξ is called the tail index of the distribution.

In real life applications, a three-parametric family (including location and scale parameters) is used, called generalized extreme value distribution (GEV). For every GEV distribution G there exist real numbers $a > 0$ and b , such that $G^*(x) = G(ax + b)$ for every x . The tail index parameter is invariant under such linear transformations.

Another approach to the investigation of the tail behaviour is the peaks over threshold (POT) model of [1, 28], where the extremal model is based on the values over a threshold u . Let x_F be the right endpoint of the distribution F (finite or infinite). If the distribution of the standardized excesses over the threshold has a limit, that must be the generalized Pareto distribution:

$$\mathbb{P}\left(\frac{X-u}{\sigma_u} < x | X > u\right) \rightarrow F_\xi(x), \quad \text{if } u \rightarrow x_F,$$

where

$$F_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0, \\ 1 - e^{-x}, & \text{if } \xi = 0. \end{cases} \quad (2)$$

The two models result in the same parameter ξ in Eqs. (1) and (2) for any given initial distribution F , for which the limit exists.

An $\ell(x)$ function is called slowly varying if $\lim_{t \rightarrow \infty} \ell(tx)/\ell(t) = 1$ for every $x > 0$. A sufficient condition for the limit theorems above, in case of positive tail index parameter (this implies $x_F = \infty$) is that

$$1 - F(x) \sim x^{-\frac{1}{\xi}} \ell(x), \quad \text{if } x \rightarrow \infty.$$

Finding a suitable function $\ell(x)$ and calculating the parameters of the limiting distribution of extremes is only feasible for special known distributions. In case of real life problems finding $\ell(x)$ is unrealistic, therefore estimating parameter ξ is in the main focus.

2 Methods for Defining the Threshold in Hill Estimator

For a tail index $\xi > 0$ Hill [24] proposed an estimator: let $X_1, X_2, \dots, X_n$ be a sample from a distribution function F and $X_1^* \leq X_2^* \leq \dots \leq X_n^*$ the order statistic. The Hill estimator for the tail index is

$$\hat{\xi} = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1}^* - \log X_{n-k}^*.$$

Similarly to the POT model, the Hill estimator also uses the largest values of the sample. The threshold is defined as the $(k+1)$ th largest observation.

The Hill estimator strongly depends on the choice of k . It is important to mention that by the work of [26], $\hat{\xi}$ is a consistent estimator for the tail index only if

$k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. Using a too small k leads to large variance; however, for a too large k , the estimator is likely to be biased. Therefore, proposing a method for choosing the optimal k for the Hill estimator has been in the focus of research since its publication, see, e.g. [6, 7, 19, 20, 22] and others.

2.1 Double Bootstrap

One of the most accurate estimators, the double bootstrap method was introduced by [6] and improved by [29]. One can find a proposed k by minimizing the asymptotic mean squared error of Hill estimator. Define

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

For $i = 1$ one can observe, that the $\hat{\xi} = M^{(1)}(n, k)$ is the Hill estimation. Let us define $M(n, k) = M^{(2)}(n, k) - 2(M^{(1)}(n, k))^2$ and k_2 as the optimal threshold for $M(n, k)$. Since

$$\frac{k_2}{k_1} \sim \left(1 - \frac{1}{\rho} \right)^{1/(1-2\rho)} \quad (3)$$

as proved by [6], where ρ is a regularity parameter (which can be estimated in a consistent way), it is possible to estimate k_2 and ρ instead of k_1 . The double bootstrap method consists of the following steps:

- Choose $\varepsilon \in (0, 1/2)$, and set $m_1 = \lfloor n^{1-\varepsilon} \rfloor$ to ensure consistency as $n \rightarrow \infty$. Estimate $E(M(m_1, r)^2 | X_1, X_2, \dots, X_n)$ by drawing m_1 size bootstrap samples from the empirical distribution function F_n and minimize it in r . Denote the minimum by r_1 .
- Set $m_2 = \lfloor m_1^2/n \rfloor$ and minimize $E(M(m_2, r)^2 | X_1, X_2, \dots, X_n)$ the same way as in the first step, denote the minimum by r_2 .
- Estimate the regularity parameter ρ , which is important for the further calculations, by $\hat{\rho} = \log(r_1)/(-2 \log(m_1) + 2 \log(r_1))$.
- Now one can estimate the optimal k using the approximation (3) by

$$\hat{k} = \frac{r_1^2}{r_2} \left(1 - \frac{1}{\hat{\rho}} \right)^{1/(2\hat{\rho}-1)}.$$

The Hill estimator based on the double bootstrap method provides appropriate tail index estimate for $\xi > 0$, but usually results in a long computation time because of the needed high number of bootstrap simulations. For smaller sample sizes the acceptable range is limited to $\xi > 0.5$, as one can see the poor simulation results of [7] for lower tail indices.

2.2 Hall's Methodology for Optimal k -Selection

In recent years many methods have arisen from the influence of [23] which use bootstrap samples to estimate the tail index parameter or improve the results of previous estimators. One of these methods appeared first in [18] and was concretized by [19], which paper also corrected the k selection, as follows.

Let $X_1, X_2, \dots, X_n$ be the sample. First, one must fix an initial k_{aux} threshold (usually $k_{\text{aux}} = 2\sqrt{n}$), an estimator $\gamma^{(1)}(k)$ for $1/\xi$ (i.e. Hill with k as the threshold), and $n_1 = n^\varepsilon$, $n_2 = n_1^2/n$ subsample sizes, where n is the sample size, $\varepsilon \in (0.5, 1)$ is arbitrary (usually $\varepsilon = 0.85$ is used). The steps of the algorithm are the following:

1. Compute $\gamma^{(1)}(k)$ for every $k = 1, 2, \dots, n-1$.
2. Simulate bootstrap samples of size n_1 and n_2 , B times. Calculate

$$t_{n_i,l}(k) = \gamma_{n_i,l}^{(1)}(k) - \gamma^{(1)}(k_{\text{aux}}),$$

for every k , where $\gamma_{n_i,l}^{(1)}(k)$ is the estimate by using n_i -element ($i = 1, 2$) bootstrap samples, $l = 1, 2, \dots, B$ and k as threshold.

3. Let $k_0(n_i) = \arg \min_{n_1, n_2 \in 1, 2, \dots, n-1} \frac{1}{B} \sum_{l=1}^B (t_{n_i,l}(k))^2$.
4. As a corrected threshold one may use $k_0 = \frac{k_0(n_1)}{k_0(n_2)}$ if $k_0 \in (1, n-1)$.
5. That way the corrected estimate for the tail index is $1/\gamma(k_0)$.

This algorithm can be really effective, however it has the same disadvantage as the double bootstrap method, namely finding $k_0(n_i)$ requires the calculations to be repeated for every k which takes approximately as long time as calculating the optimal r_1, r_2 in the double bootstrap method.

2.2.1 Acceleration Algorithm

For both the double bootstrap and Hall's method, the computational time could be a crucial problem. Minimizing among a large interval of possible k or t_{n_i} values needs a lot of calculations and especially the bootstrap simulations increase the needed time. These problems could lead us to use possibly less simulations, or execute the minimization over a smaller set.

We experienced that the absolute error of estimate is high by using small and large k values, while it is smaller in the neighbourhood of the optimum. The error function in k was usually smooth and showed a "U shaped" line (similar property was already mentioned by [19]). This motivates an approximation algorithm to find the minimal error value, instead of scanning all of the numbers $k = 1, \dots, n$, which takes unnecessary long simulation time. We propose the following step by step approximation algorithm:

1. Calculate the mean squared error only for $k = c \cdot \sqrt{n}$ ($c \in \mathbb{Z}^+$) and execute the minimization among them. Denote the minimum by m_a .

2. Focus on the minimum m_a and its $2 \cdot \sqrt{n}$ wide surroundings and fix an integer b . Now minimize among the numbers $m_a \pm c \cdot b$ ($c \in \mathbb{Z}^+$) in the examined region. Denote the minimum by m_b .
3. Continue the calculation in the $2 \cdot b$ wide neighbourhood of m_b and minimize among these values.

This way, one can find the global optimum for k_0, r_1, r_2 in double bootstrap and Hall's method in most of the cases. However, sometimes the algorithm stops in a local optimum. Therefore, using the acceleration algorithm we trade-off some accuracy of estimation for faster computation time. In fact, as $n \rightarrow \infty$, the error rate decreases, since finding the optimal k is not crucial for the estimator—an adequate k results in similar estimate. In addition, the steps can be modified, to focus better on the potential optimal k values, making the calculation even faster.

In Table 1 one can see the comparison of Hall and accelerated Hall method. The mean of absolute error is similar for 200–1000 size samples, while the computation time is strongly in favour of accelerated method.

The acceleration algorithm can provide even better results in some cases (e.g. Table 1, $n = 200$), however it is caused by the uncertainty of bootstrap simulations—original method may stop in a wrong value, while accelerated finds the right one. Using more bootstrap simulations this paradox diminishes.

2.3 Kolmogorov–Smirnov Distance Metric

Another approach for tail index estimation was suggested by [7] which minimizes the distance between the tail of the empirical distribution function and the fitted Pareto distribution with the estimated tail index parameter. For the minimization the Kolmogorov–Smirnov distance of the quantiles can be used. Assume that the distribution of the sample is in the maximum domain of attraction of a GEV distribution which implies that the distribution function can be written in the following form:

$$F(x) = 1 - Cx^{-\frac{1}{\xi}} + o\left(x^{-\frac{1}{\xi}}\right).$$

Rearranging this equation the value of x can be approximated by

Table 1 Comparison of accelerated and classical Hall method for Fréchet distribution with tail index 1

Method	$n = 200$		$n = 500$		$n = 1000$	
	Error	Time	Error	Time	Error	Time
Hall	0.177	1.56	0.134	4.401	0.104	8.579
Accelerated hall	0.172	1.046	0.135	1.634	0.104	2.051

The sample size varies between 200 and 1000. The error is the mean absolute error, average time is measured in seconds. The values were estimated by 500 Monte Carlo simulations

$$x = \left(\frac{P(X > x)}{C} \right)^{-\xi}.$$

One can construct a quantile function by using this approximation. The probability $P(X > x)$ can be replaced by the observed relative frequency j/n , and ξ may be estimated using the Hill estimator for a k , moreover C can be approximated by $\frac{k}{n}(X_{n-k+1}^*)^{1/\xi}$ as the highest observations follow a Pareto distribution in the limit. These substitutions result in an estimator for the quantiles as a function of j and k .

$$q(j, k) = \left(\frac{k}{j} (X_{n-k+1}^*)^{1/\xi} \right)^{\frac{j}{k}}.$$

This quantile estimator was first proposed by [31] as a maximum likelihood estimator using only the k upper ordered statistics.

In this procedure, the optimal k for the Hill estimator is chosen as the k , which minimizes the Kolmogorov–Smirnov distance between the empirical and the calculated quantiles

$$\hat{k} = \arg \min_k \left[\sup_{j \in 1, \dots, T} |X_{n-j}^* - q(j, k)| \right],$$

where T sets the fitting threshold (we call it KS threshold). Choosing T is almost arbitrary because for heavy tailed distributions the largest differences always appear for the highest quantiles. We applied the rule of choosing $T = \max(50, \sqrt{n})$. The advantages of the Kolmogorov–Smirnov method are that it is easy to programme and its computation time is short. As [7] mentioned, it is the best performing known method if $0 < \xi < 0.5$ and by our experience, it also works well for small sample sizes. However, for distributions with tail index $\xi > 0.5$ this technique results in highly biased estimates.

2.3.1 Asymptotics of Kolmogorov–Smirnov Method

Although the Kolmogorov–Smirnov method has a straightforward algorithm, its theoretical background and asymptotic behaviour is not presented yet. In Theorem 1 we state that under some conditions the Kolmogorov–Smirnov technique results in underestimates for the tail index. Moreover, the simulations show that the estimation is approximating the real tail index parameter if the number of observation increases. To the proof of the properties of the estimator we need the following lemma:

Lemma 1 *If $k \rightarrow \infty$, then $\frac{1}{k} \log \left(\prod_{j=1}^k \frac{k}{j} \right) \rightarrow 1$.*

Proof For $k \rightarrow \infty$ by using the Stirling formula we have

$$\begin{aligned}
\frac{1}{k} \log \left(\prod_{j=1}^k \frac{k}{j} \right) &= \frac{1}{k} \log \left(\frac{k^k}{k!} \right) \sim \frac{1}{k} \log \left(\frac{k^k}{\sqrt{2\pi k} \frac{k^k}{e^k}} \right) \\
&= \frac{1}{k} \log \left(\frac{e^k}{\sqrt{2\pi k}} \right) = \frac{1}{k} \left(\log(e^k) - \log(\sqrt{2\pi k}) \right) \\
&= 1 - \frac{1}{k} \log(\sqrt{2\pi k}) \rightarrow 1.
\end{aligned}$$

□

Theorem 1 Assume that $F(x) = 1 - Cx^{-\frac{1}{\xi}} + o\left(x^{-\frac{1}{\xi}}\right)$, (i.e. the distribution is in the maximum domain of attraction of an extreme value distribution). The Hill estimator using the Kolmogorov–Smirnov k -selection technique has a negative bias as $n \rightarrow \infty$, $k \rightarrow \infty$, while $k/n \rightarrow 0$.

Proof Let $X_1, X_2, \dots, X_n$ be our sample, and let $X_1^*, X_2^*, \dots, X_n^*$ be the ordered sample. Let

$$q_n(j, k) = \left(\frac{k}{j} (X_{n-k+1}^*)^{1/\xi} \right)^\xi \quad (4)$$

be a quantile estimator for $\mathbb{E}(X_{n-j+1}^*)$, similarly to [7]. It is asymptotically unbiased as it is a maximum likelihood estimator for a given quantile by [31].

Under the given conditions $X_n^*, X_{n-1}^*, \dots, X_{n-k}^*$ can be considered as a $k+1$ size ordered sample from a Pareto distribution with unknown parameters (see [25]). Therefore, a Glivenko–Cantelli type theorem for quantiles provides that $q_n(j, k)$ is a consistent estimator for the $(k+1-j)/(k+1)$ quantile of the unknown Pareto distribution which can also be approximated by X_{n-j+1}^* . Namely, for every $\varepsilon > 0$ there is an M such that if $k > M$ (and $n \gg M$) then the Kolmogorov–Smirnov distance between $q_n(j, k)$ and X_{n-j+1}^* will be less than ε .

Choose k according to the Kolmogorov–Smirnov method (2.3) to ensure the approximation of quantile estimation and let $\hat{\xi}$ be the Hill estimate using k if $n > k > M$.

$$\begin{aligned}
\mathbb{E}(\hat{\xi}) &= \mathbb{E}\left(\frac{1}{k} \sum_{j=1}^k \left(\log(X_{n-j+1}^*) - \log(X_{n-k}^*)\right)\right) \\
&= \frac{1}{k} \sum_{j=1}^k \mathbb{E}\left(\log\left(\frac{X_{n-j+1}^*}{X_{n-k}^*}\right)\right) \\
&\leq \frac{1}{k} \sum_{j=1}^k \log \mathbb{E}\left(\frac{X_{n-j+1}^*}{X_{n-k}^*}\right) \quad \text{by using Jensen's inequality} \\
&\sim \frac{1}{k} \sum_{j=1}^k \log\left(\frac{q_n(j, k)}{q_n(k+1, k)}\right) \quad \text{approximation by using (4)} \\
&= \frac{1}{k} \sum_{j=1}^k \log\left(\left(\frac{\frac{k}{j}(X_{n-k+1}^*)^{1/\xi}}{\frac{k}{k+1}(X_{n-k+1}^*)^{1/\xi}}\right)^\xi\right) \\
&= \frac{1}{k} \sum_{j=1}^k \log\left(\left(\frac{k/j}{k/(k+1)}\right)^\xi\right) \\
&= \frac{\xi}{k} \sum_{j=1}^k \log\left(\frac{k/j}{k/(k+1)}\right) \\
&= \xi \cdot \left(\frac{1}{k} \log\left(\prod_{j=1}^k \frac{k}{j}\right)\right) - \frac{\xi}{k} \left(\log\left(\prod_{j=1}^k \frac{k}{k+1}\right)\right) \\
&= \xi \cdot \left(\frac{1}{k} \log\left(\prod_{j=1}^k \frac{k}{j}\right)\right) - \xi \left(\log\left(\frac{k}{k+1}\right)\right) \rightarrow \xi + 0
\end{aligned}$$

□

Conjecture 1 *Using the conditions of Theorem 1 the Hill estimator using the Kolmogorov–Smirnov k -selection technique shows asymptotically unbiased behaviour.*

The consistency of the Hill estimator is well-known as $n \rightarrow \infty$, $k \rightarrow \infty$ and $k/n \rightarrow 0$. (see, e.g. [16]). The estimator is expected to behave regularly, implying the L_1 convergence. As an empirical observation by Monte Carlo simulations from standard generalized Pareto distribution (with $\xi = 1$), we could observe, that for an increasing sequence k the ratio between the two sides of Jensen's inequality in the proof of Theorem 1 tends to 1 (see Table 2). Similarly, in the approximation by quantile estimation the ratio of the two sides tends to 1 when $k \rightarrow \infty$ and $n \rightarrow \infty$ holds (see Table 3). These observations suggest that the Kolmogorov–Smirnov method results in asymptotically unbiased estimation for the tail index, under the conditions $k \rightarrow \infty$, $k/n \rightarrow 0$ and $n \rightarrow \infty$.

However $k \rightarrow \infty$ is not an evident condition and even if it is realized we cannot say anything about the speed of convergence. This could lead to biased tail index

Table 2 Ratio of the two sides of the Jensen's inequality, for different sample sizes and different k indices

	$k = 5$	$k = 25$	$k = 100$	$k = 500$	$k = 5000$
$n = 10^3$	1.326	1.104	1.032	1.008	
$n = 10^4$	1.303	1.095	1.03	1.007	1.0012
$n = 10^5$	1.307	1.087	1.029	1.007	1.0011
$n = 10^6$	1.295	1.084	1.029	1.007	1.001

Simulations based on standard GPD samples with $\xi = 1$ **Table 3** Ratio of the two sides of quantile estimate approximation, for different sample sizes and different k indices

	$k = 5$	$k = 25$	$k = 100$	$k = 500$	$k = 5000$
$n = 10^3$	0.495	0.714	0.899	1.001	
$n = 10^4$	0.633	0.841	0.934	0.984	0.998
$n = 10^5$	0.645	0.862	0.954	0.989	0.9981
$n = 10^6$	0.654	0.864	0.952	0.992	0.9981

Simulations based on standard GPD samples with $\xi = 1$

estimate using the Kolmogorov–Smirnov method, especially if ξ is large. For moderate size samples, this bias is estimated in Table 9, and it is always negative in accordance to Theorem 1.

As [7] presented, the Kolmogorov–Smirnov method provides acceptable estimators for moderate size samples if $\xi < 0.5$, but for tail index $\xi \geq 0.5$ more than 10,000 size samples might be needed for proper estimates. In most real-life applications such a large sample size is not available, therefore in cases of distributions with infinite variance the estimator may have a significant bias.

In the previous paragraphs we have dealt with two important properties of Kolmogorov–Smirnov method: consistency (originated from Hill estimators consistency) and negative bias for finite samples. These properties motivates us to perform a simulation study, in order to estimate the bias of Hill estimator based on the Kolmogorov–Smirnov method using numerous initial distributions, sample sizes and theoretical tail indices. This will lead us to our new tail index estimator.

2.3.2 Finite Sample Properties of Kolmogorov–Smirnov Method

In this section, we show that Kolmogorov–Smirnov method has the following properties:

- The distribution of $\hat{\xi}$ is similar for each parent distribution
- Independence of sample size in the interval $n \in (200, 10,000)$
- The bias is linear in ξ

First, we simulated 500 size i.i.d. samples from Fréchet, GPD, stable and Student distributions with fixed tail index parameter ξ . The estimates were calculated 30,000 times for each ξ . We experienced, that in accordance to the empirical

distribution of $\hat{\xi}$ is similar, GEV-like for each parent distribution as one can see in Fig. 1. It would be beneficial to characterize the empirical distribution more precisely, since the best fitting GEV distribution is still rejected by the goodness of fit tests (chi square test after discretization, $p < 10^{-16}$, which may be caused by the large sample size). Despite, the empirical distribution and the best fitted GEV distribution are in relation, thus the parameters of the GEV distribution (estimated by maximum likelihood) might characterize the empirical distribution well.

Second, we simulated from 200 to 10,000 sized samples. We fitted GEV distribution to tail index estimates of 1000 Monte Carlo simulations. We experienced, that the GEV parameters and expectation of the tail index were similar, regardless of the sample size – see Table 4. We do not have results for larger samples, since double bootstrap and Hall method are much more effective in this case, due to their asymptotic properties. We can state that, for $\xi \in (0.5, 4)$ the distribution of the Kolmogorov–Smirnov estimators does not depend on the sample size if $200 < n < 10,000$.

Finally, we wanted to determine the bias of Kolmogorov–Smirnov method for $\xi \in (0.5, 4)$. Therefore, we simulated 500 size samples from Fréchet distribution with various ξ parameters. We calculated the average of estimated ξ and the location parameter of the best fitted GEV distribution based on 30,000 simulations. As one can see in Fig. 2, the average and GEV parameters are in linear relation with the true ξ parameter. Therefore, correcting the estimator by a linear transformation we can improve the Kolmogorov–Smirnov method. When $\xi > 4$, we can not receive reasonable results, so the correction could not help (but these distributions do not have any practical relevance).

Fig. 1 Empirical density functions of Kolmogorov–Smirnov estimates for tail index $\xi = 1$ based on 1000 size samples using 5000 simulations. The samples are originated from different heavy tailed distributions. The best fitted GEV distribution is similar to the empirical distributions of the estimates (parameters: shape = -0.04 , location = 0.704 , scale = 0.273)

Table 4 The mean of the Kolmogorov–Smirnov estimates and the parameters of the fitted GEV distribution for KS estimates

	Mean	Location	Scale	Shape
$\xi = 0.5$				
$n = 200$	0.449	0.377	0.117	0.037
$n = 500$	0.448	0.378	0.113	0.043
$n = 2000$	0.446	0.378	0.112	0.032
$n = 10,000$	0.448	0.382	0.116	− 0.001
$\xi = 1$				
$n = 200$	0.852	0.700	0.278	− 0.028
$n = 500$	0.85	0.701	0.274	− 0.03
$n = 2000$	0.855	0.709	0.274	− 0.039
$n = 10,000$	0.856	0.713	0.276	− 0.062
$\xi = 2$				
$n = 200$	1.626	1.318	0.612	− 0.065
$n = 500$	1.628	1.334	0.605	− 0.094
$n = 2000$	1.634	1.356	0.601	− 0.115
$n = 10,000$	1.59	1.316	0.594	− 0.118
$\xi = 3$				
$n = 200$	2.397	1.960	0.953	− 0.121
$n = 500$	2.392	1.978	0.935	− 0.141
$n = 2000$	2.36	1.963	0.903	− 0.137
$n = 10,000$	2.441	2.09	0.903	− 0.217
$\xi = 4$				
$n = 200$	3.161	2.561	1.290	− 0.111
$n = 500$	2.962	2.478	1.159	− 0.177
$n = 2000$	3.137	2.59	1.232	− 0.128
$n = 10,000$	3.09	2.573	1.2	− 0.141

The test samples are from Fréchet distributions

3 The Proposed New Estimator

Based on the finite sample properties of the Kolmogorov–Smirnov method, we constructed a new, empirical estimator, that corrects the potential bias. As we experienced in Sect. 2.3.2, the Kolmogorov–Smirnov method neither depends, on the type of the sampling distribution, nor on the size of the sample for the investigated cases, therefore a linear transformation can correct the bias. Based on the Monte Carlo simulation we fitted a linear regression model to the average of estimated ξ , GEV parameters and sample size against the theoretical ξ . We experienced that:

- Both the mean of the estimates and the location parameter of the fitted GEV were in strong correlation with ξ , thus two regression models are feasible for bias correction:

One of the potential models is $\hat{\xi} = -0.119 + 1.603 \cdot \mu_f$, where μ_f is the location parameter of the best fitted GEV model to the KS estimates. The other

Fig. 2 Linear regression for the true tail index using mean and EVD parameters. Average of Kolmogorov–Smirnov estimates were calculated using 500 size Fréchet samples and 30,000 simulations

potential model is $\hat{\xi}^* = -0.1181 + 1.3301 \cdot \xi^*$, where ξ^* is the mean of simulated KS estimates.

- Despite the assumed independence, the sample size was a significant factor. However, the coefficient was too small ($p = 0.017$, $\text{coef} = -0.0000046$) to have important addition for small sample sizes so we omitted it.
- The covariance matrix of the fitted GEV parameters was analysed to choose the important factors. The eigenvalues (0.666, 0.0005, 0.00006) and the eigenvector of largest eigenvalue (0.89, 0.45, -0.09) suggests, that only the first two variables, i.e. the location and the scale parameter might have effect on $\hat{\xi}$. However, using both parameters in a linear regression model, only the location parameter shows significant effect ($p = 0.0003$).

However, for applying the correction models, we need to know the distribution of the Hill estimates on the sample. As [5] pointed out, m out of n bootstrap is a feasible way to estimate the distribution of a point estimator in the extreme-value setup. Therefore, the new bias corrected algorithm consists of a bootstrap simulation, and a transformation based on linear regression.

It is important to mention, that the method works best for $\xi \in (0.5, 4)$ and for sample sizes between 200 and 10,000. Outside this region the statements above might not hold, thus the regression coefficients may be different.

Now, we present a step-by-step algorithm for the bias correcting method, we call this model regression estimator:

3.1 Algorithm

Let $\underline{X} = X_1, X_2, \dots, X_n$ be independent and identically distributed observations.

1. Fix $\varepsilon \in (0.5, 1)$ (we use $\varepsilon = 0.85$) and $m = n^\varepsilon$ to ensure the asymptotic properties of m out of n bootstrap [4, 5].
2. Simulate $\underline{X}_i$ bootstrap sample of size m from $\underline{X}$. Repeat it M times. Calculate the Hill estimate of each $\underline{X}_i$ using Kolmogorov–Smirnov method, denote the tail index estimate by ξ_i . The $\xi_1, \xi_2, \dots, \xi_M$ tail index estimates form an empirical distribution for the real ξ .
3. Denote $\xi^* = (\sum_{j=1}^M \xi_j)/M$ the average of Kolmogorov–Smirnov estimates—as a value, that is closely related to the tail index of $\underline{X}$. Let $\hat{\xi}^* = -0.1181 + 1.3301 \cdot \xi^*$. We call this method mean based regression estimator (MRE).
4. Fit a generalized extreme value distribution to the $\xi_1, \xi_2, \dots, \xi_M$ estimates and denote the parameters as (μ_f, σ_f, ξ_f) .
Let $\hat{\xi} = -0.119 + 1.603 \cdot \mu_f$. We call this method fitted regression estimator (FRE).

3.2 Simulations

The properties of the regression estimators were tested for simulated data from Pareto, Fréchet, Student and symmetric stable parent distributions. We also tested the estimator on a mixed distribution—the lower 80% consist of an exponential core, while the upper 20% is a Pareto tail. We set the tail index parameter to the same $\xi \in (0.2, 4)$ value in each distribution for the samples of size 200, 1000 and 4000. For the simulation study we used the built-in `rgpd` (package: `fExtremes`), `rfrchet` (package: `VGAM`), `rt` and `rstable` (package: `stabledist`) functions of the R programming language. We compared the average of the estimated tail index values by FRE. Subsequently we calculated the average absolute error from the theoretical ξ . Using $m = n^{0.85}$, the subsample sizes were 90, 355 and 1153. We set a suitable KS threshold (maximal number of ordered statistic for KS method) to $\min(50, \sqrt{m})$. The results of the simulation can be seen in Table 8. The most important observations:

- For $\xi < 0.5$ the error rate is higher than the optimal, while for $\xi > 2$ one can see slight underestimation of the real tail index (except for the mixed distribution).
- For $\xi \in (0.5, 2)$ the estimator seems to work properly.
- We can say generally that higher sample size results in more accurate estimate for every distribution and tail index.
- Symmetric stable distribution with $\xi = 0.5$ is a special case: it is not in the domain of attraction of the Fréchet distribution, so extreme value model can not work on it. However it was an interesting question if the regression estimator works outside the domain of attraction or not—the answer is no.

Fig. 3 Average absolute error of methods for 200 and 500 size samples from Student ($1/\xi, 0$) distribution with varying tail index. Averages are based on 1000 simulations in each case. Subsample size, value of ε were set as it was presented in the description of each method

Fig. 4 Average absolute error of methods for Fréchet, GPD, Student and symmetric stable distributions under 0.7 tail index and varying sample sizes. Location and scale parameters were standardized for every distribution. Averages are based on 1000 simulations. Subsample size, value of ε were set as it was presented in the description of each method

Generally, we can say, that FRE works well for $\xi \in (0.2, 4)$, but its best region is $\xi \in (0.5, 2)$ for every type of parent distribution with proper extremal behaviour. We experienced similar properties for MRE too (Figs. 2, 3 and 4; Table 9).

3.3 Comparison of the Methods

In this section, we compare the effectiveness of Kolmogorov–Smirnov (2.3), double bootstrap (2.1) and Hall’s method (2.2) with accelerated algorithm to the two types of regression estimators. We try to decide which one is the most efficient on the average, absolute error and the computational time of estimation. We include different parent distributions (namely, Fréchet, symmetric stable, generalized Pareto, and Student) with varying tail indices and sample sizes.

First, we present the average of absolute error of different methods using 200 size samples from the Student distribution with $\xi \in (0.1, 1.5)$. The results in Fig. 3 imply that Kolmogorov–Smirnov method has the smallest error for $\xi < 0.5$, while in the $\xi \in (0.5, 1.5)$ region FRE and MRE seems to be the best. For larger ξ double bootstrap starts to outperform the regression estimators. For 500 size samples the optimal range for MRE and FRE lies around $\xi = 0.5$, while smaller tail indices can be estimated well with Kolmogorov–Smirnov method. For larger ξ values Hall’s and double bootstrap method are better options for the estimation.

Second, we analysed the effect of sample size on the estimates. We fixed $\xi = 0.7$, as a tail index value, where regression estimation works well. We present the average absolute error using different size samples for the methods in Fig. 4. The four parent distributions were analysed separately. We may conclude, that regression estimators seem to be the best for $n < 200$. In case of Student distribution FRE has the smallest error for $n < 500$, while for GPD the regression estimators have no competitors if $n < 800$. Double bootstrap and Hall method start to behave better for larger sizes. Concluding the figure, for high tail indices the best method depends on the sample size and distribution, FRE and MRE are the best for smaller samples.

Finally, for $\xi \in (0.333, 2)$ the mean of estimates, absolute error and computational time are presented in Table 9 for 200, 500 and 1000 sizes samples using different methods. The parent distribution is Fréchet, and the averages were calculated after 500 simulations. We can say, that for smaller sample sizes usually FRE or MRE has the best estimate and smallest error. For $n = 1000$ double bootstrap starts to be better and Hall’s method also outperforms the regression estimator. Kolmogorov–Smirnov method is the fastest in every cases, but works properly only for $\xi < 0.5$. Computational times of FRE and MRE methods are also competitive with the other accelerated algorithms.

We can conclude, that our new estimation method provides the best estimates among the examined options if the tail index is between 0.5 and 1.5. For small (less than 200) sized samples it always seems to be the best, but for moderate (200–500) it is preferred in case of Student or GPD—like parent distribution. However, for larger ξ and different type of parent distribution the optimal region may be wider.

Some remarks:

- The effectiveness of FRE and MRE comes from the property that the methods are not sensitive with respect to the sample size. Therefore, the advantage can disappear if $n > 1000$ due to consistency of other methods.

- Unlike for other examined methods, the distribution of FRE and MRE can be approximated by a normal distribution. One can see a comparison with double bootstrap in Fig. 5. Therefore, confidence interval calculations are easier.
- If $\xi \in (0.5, 4)$, the errors of the FRE and MRE estimators are not sensitive to the underlying distribution, unlike the double bootstrap or Hall method. This property results in different optimality regions, thus for different distributions the double bootstrap and Hall method needs different sample sizes to outperform the regression estimators.

3.4 Model Selecting Scheme

The results of Sect. 3.3 suggest, that different methods are optimal for different types of samples. Therefore, a model selecting scheme is beneficial for real life applications. For getting preliminary information about the sample we need to fit different distributions and calculate ξ in a fast, but maybe not reliable way. This is formulated in the model selection algorithm as follows:

1. Estimate the tail index using the Kolmogorov–Smirnov method (2.3). If the estimated $\hat{\xi}$ is smaller than 0.45, then the true tail index is likely smaller than 0.5, therefore usually the Kolmogorov–Smirnov is the best working method (according to [7] and simulations).

Fig. 5 QQ plots for case $\xi = 2$ with FRE and double bootstrap to sample sizes of 500 and 1000. We have got p -values of 0.32 and 0.05 by the Anderson–Darling test for FRE, but for double bootstrap both p -values were less than 0.0001

2. If $1.22 < \hat{\xi}$, then it is likely that for the true tail index $1.5 < \xi$ holds (see Fig. 2 in Sect. 3), thus the double bootstrap method or Hall's method gives the best estimates (Fig. 3).
3. If $0.45 < \hat{\xi} < 1.22$ and the sample size is less than 200, then it is likely that for the true tail index $0.5 < \xi < 1.5$ holds, thus the small sample implies that the FRE or MRE method gives the best estimate.
4. If $0.45 < \hat{\xi} < 1.22$, and the sample size is between 200 and 500, then the best fitting distribution determine the model. If GPD or the Student distribution fits the best, then usually FRE or MRE method gives the best estimate. On the other hand, if Fréchet or stable distribution fits the best, then usually double bootstrap or Hall method gives the best estimate.
5. For larger samples it is not clear which method is the best. As $n \rightarrow \infty$ the above rules loose their values, all the methods start to work well, and the best working method depends on the sample distribution.

4 Applications to Real Data

4.1 Danish Fire Losses

The Danish fire losses is a well known open dataset, which is available in an R package [27]. Previous discussions were published, e.g. by [12, 30]. The dataset contains 2167 fire losses, which occurred between 1980 and 1990. The true tail index parameter is between 0.5 and 1 by multiple estimators, therefore large sample size implies point no. 5 of the model selection scheme. It is not clear which method results in the best estimate, however the large sample size strongly supports the asymptotic double bootstrap and Hall method.

We estimated the tail index parameter by using the presented models, namely: double bootstrap, Kolmogorov–Smirnov and Hall's method. Additionally, we fitted Pareto distribution for the data. As one can see in Table 5, all of the estimators results in $0.5 < \hat{\xi} < 0.75$ values, therefore the model selection might not have a big significance. It is important to mention, that MRE and FRE results in the smallest estimates. This might be caused by the fact, that we used $n^{0.85} = 685$ subsample size, which is too large for such a big sample. However, we used the same method for regression estimator to subsample selection as the other estimators, it might not be optimal for m out of n bootstrap [4].

To check if we can improve regression estimates we considered smaller bootstrap subsample sizes, namely ones from the interval (50, 300). We calculated the tail index of the dataset with 10,000 bootstrap samples for different subsample sizes. The results can be seen in Table 6. By using 50–100 subsample size, we receive the values of most accurate methods. It might be a future task, to optimize ϵ based on the sample size, thus making the regression estimator more accurate for larger samples.

Table 5 Tail index estimation on Danish fire losses data using different methods

DB	Hall	KS	MRE	FRE	GPD
0.713	0.667	0.610	0.587	0.543	0.604

Bold values indicate the best performing methods

Table 6 MRE and FRE estimates for Danish fire losses data using different subsample sizes

Subsample size	FRE	MRE
50	0.689	0.702
100	0.687	0.68
150	0.644	0.646
200	0.604	0.621
300	0.555	0.598

4.2 European Natural Disaster Damages

As a second example, we analysed the damages of the most destructive storms and temperature related natural disasters of the past 50 years in Europe. The data originated from [21] database. First, we need to prepare the data for the tail index analysis. We summed up the damages occurring in different countries and days for each disaster. With these modifications the dataset contains 403 observations.

In the analysis we applied the presented tail index estimators. One can see in Table 7, that the estimates differ in a wide range, which can be explained by the small sample size and strange distribution—extremes are quite large, but more of them occurred in the observation period. The KS estimate is larger than 0.45, which suggests that the true tail index value is likely to be larger than 0.5. Since the sample size is also in the optimal range of regression estimator MRE or FRE can result in the best estimate for the data. One can see the histogram of data and the best fitted extreme value distribution using the estimated tail index in Fig. 6. The tail behaviour is best captured by the MRE, therefore regression estimates could result in the best high quantile estimates.

For further analysis, one might have a look at the time dependent structure of the data. Climate change could have an effect on the severity of storms and heat waves, which can turn up as a positive trend among the observations. Therefore, using time dependent models might improve the effectiveness of the other estimators.

5 Conclusion

Our new regression method (Sect. 3) provides an alternative to estimate the tail index for heavy-tailed distributions. We have shown its merits by the parameter estimation of known distributions, and presented that our method is also useful for real life data. The computation time is similar to the double bootstrap method. However as our algorithm also applies bootstrap techniques, one can use less bootstrap samples to lower the computation time further if needed—at the expense of the estimations' accuracy, or conversely.

Table 7 Tail index estimation on European extreme natural disasters data using different methods

DB	Hall	KS	MRE	FRE	GPD
0.066	1.153	0.461	0.78	0.822	1.804

Bold values indicate the best performing methods

Fig. 6 Histogram of European disasters and best fitted GEV model using tail estimates of different methods. Location and scale parameters were estimated by maximum likelihood

Our simulations showed in agreement to [7] that the best estimation is the Hill estimation based on the Kolmogorov–Smirnov method (Sect. 2.3), if the ξ parameter is less than 0.5. If the sample size is small ($n < 200$), FRE and MRE have the best properties in most cases. For moderate ($200 < n < 500$) sample sizes and $0.5 < \xi < 1.5$, the best working algorithm depends on the distribution of the sample. If it is GPD or Student-like, then usually the FRE and MRE can result in the best estimates. Otherwise using double bootstrap (2.1) or Hall method is suggested. If the size is more than 500, then in general the double bootstrap or Hall method could result in more accurate estimation. This model selection algorithm could be extended by comparing more methods or by more types of initial distributions.

Our regression estimation has two types (FRE and MRE), using the mean of the bootstrap samples or using the location parameter of the best fitted GEV distribution. Our experiments did not indicate which one is the more precise, therefore we suggest using both in a real life analysis. Further research could find the answer for this question.

Acknowledgements Open access funding provided by Eötvös Loránd University (ELTE). The project was supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 8 and 9.

Table 8 Average FRE tail index estimations for 200, 1000 and 4000 size samples, by using 500 Monte Carlo simulations, while the bootstrap sample size is $m = n^{0.85}$

ξ	0.2	0.33	0.5	1	2	3	4
$n = 200$							
GPD	0.45	0.522	0.617	1.012	1.868	2.808	3.647
Error	0.25	0.19	0.128	0.189	0.371	0.578	0.834
Fréchet	0.152	0.311	0.504	0.994	1.896	2.921	3.744
Error	0.054	0.056	0.088	0.193	0.408	0.635	0.851
Student	0.471	0.521	0.606	0.971	1.834	2.661	3.656
Error	0.271	0.188	0.113	0.191	0.42	0.678	0.869
Stable			0.401	0.965	1.89	2.854	3.735
Error			0.099	0.198	0.433	0.661	0.987
Exp + Pareto	0.456	0.525	0.638	1.014	2.084	3.172	4.493
Error	0.256	0.195	0.144	0.184	0.415	0.699	0.939
$n = 1000$							
GPD	0.325	0.411	0.535	0.999	1.961	2.948	3.853
Error	0.125	0.083	0.082	0.172	0.353	0.512	0.696
Fréchet	0.14	0.297	0.481	0.975	1.946	2.919	3.853
Error	0.062	0.06	0.079	0.162	0.358	0.537	0.7433
Student	0.262	0.354	0.501	0.981	1.889	2.94	3.866
Error	0.063	0.05	0.082	0.183	0.365	0.541	0.73
Stable			0.156	0.956	1.906	2.812	3.89
Error			0.344	0.175	0.377	0.551	0.833
Exp + Pareto	0.324	0.410	0.542	0.979	2.052	3.338	4.680
Error	0.124	0.084	0.086	0.167	0.372	0.612	0.877
$n = 4000$							
GPD	0.27	0.369	0.513	0.999	1.968	2.96	4.008
Error	0.07	0.056	0.075	0.157	0.327	0.472	0.647
Fréchet	0.138	0.29	0.486	0.981	1.966	2.968	3.979
Error	0.063	0.056	0.075	0.153	0.311	0.481	0.649
Student	0.203	0.319	0.48	0.982	1.978	2.909	3.976
Error	0.03	0.047	0.074	0.153	0.321	0.486	0.628
Stable			0.071	1.003	1.971	2.903	3.896
Error			0.429	0.163	0.338	0.509	0.731
Exp + Pareto	0.267	0.368	0.510	0.987	2.008	3.321	4.825
Error	0.068	0.053	0.070	0.156	0.319	0.531	0.919

The KS threshold is set to 50. We also give the mean absolute error from the theoretical parameter ξ

Table 9 Multiple tail index estimator for Fréchet distribution tested for different tail indices

ξ	Point of view	KS	Double bootstrap	MRE	FRE	Hall
$n = 200$						
0.333	Mean	0.305	0.353	0.303	0.311	0.368
	Error	0.076	0.065	0.065	0.056	0.058
	Time	0.006	0.277	0.684	0.707	0.978
0.5	Mean	0.456	0.542	0.501	0.504	0.555
	Error	0.126	0.097	0.095	0.088	0.092
	Time	0.006	0.251	0.598	0.626	0.981
1	Mean	0.878	1.07	1.017	0.994	1.108
	Error	0.288	0.189	0.196	0.193	0.17
	Time	0.006	0.273	0.628	0.65	1.041
2	Mean	1.614	2.077	1.953	1.896	2.233
	Error	0.639	0.354	0.389	0.408	0.371
	Time	0.006	0.251	0.581	0.602	1.008
$n = 500$						
0.333	Mean	0.302	0.352	0.305	0.315	0.357
	Error	0.079	0.046	0.043	0.034	0.045
	Time	0.012	0.266	0.346	0.353	1.506
0.5	Mean	0.443	0.526	0.492	0.495	0.537
	Error	0.134	0.078	0.06	0.054	0.066
	Time	0.013	0.224	0.28	0.302	1.428
1	Mean	0.826	1.049	1.017	0.993	1.065
	Error	0.293	0.132	0.122	0.116	0.135
	Time	0.01	0.211	0.267	0.275	1.637
2	Mean	1.669	2.14	2.047	1.993	2.108
	Error	0.577	0.271	0.25	0.253	0.247
	Time	0.011	0.244	0.301	0.301	1.422
$n = 1000$						
0.333	Mean	0.3	0.353	0.288	0.297	0.356
	Error	0.073	0.036	0.066	0.06	0.034
	Time	0.036	0.36	1.825	1.825	2.538
0.5	Mean	0.446	0.523	0.475	0.481	0.527
	Error	0.13	0.054	0.083	0.079	0.061
	Time	0.038	0.386	1.957	1.963	2.414
1	Mean	0.819	1.05	0.981	0.975	1.061
	Error	0.282	0.102	0.162	0.162	0.121
	Time	0.05	0.508	2.537	2.553	2.365
2	Mean	1.634	2.107	1.959	1.946	2.068
	Error	0.596	0.213	0.329	0.358	0.25
	Time	0.071	0.694	3.413	3.423	2.027

We used 250 bootstrap simulations and $n^{0.85}$ subsample size, where needed. The results are based on 500 Monte Carlo simulations in each case. The initial sample size is 200, 500 and 1000

Bold values indicate the best performing methods for each point of view

References

- Balkema A, de Haan L (1974) Residual life time at great age. *Ann Probab* 2:792–804
- Beirlant J, Vynckier P, Teugels JL (1996) Excess functions and estimation of the extreme-value index. *Bernoulli* 2:293–318
- Bickel PJ, Freedman DA (1981) Some asymptotic theory for the bootstrap. *Ann Stat* 9:1196–1217
- Bickel PJ, Sakov A (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Stat Sin* 18:967–985
- Bickel PJ, Götze F, van Zwet WR (1997) Resampling fewer than n observations: gains, losses and remedies on losses. *Stat Sin* 7:1–31
- Danielsson J, de Haan L, Peng L, de Vries CG (2001) Using a bootstrap method to choose the sample fraction in tail index estimation. *J Multivar Anal* 76:226–248
- Danielsson J, Ergun LM, De Haan L, de Vries CG, (2019) Tail index estimation: quantile driven threshold selection. <https://www.bankofcanada.ca/wp-content/uploads/2019/08/swp2019-28.pdf>
- Davis RA, Resnick S (1984) Tail estimates motivated by extreme value theory. *Ann Stat* 12:1467–1487
- Davison AC, Smith RL (1990) Models for exceedances over high thresholds. *J Roy Stat Soc: Ser B (Methodol)* 52:393–442
- de Haan L, Ferreira A (2006) *Extreme value theory: an introduction*. Springer, Berlin
- Dekkers ALM, de Haan L (1993) Optimal choice of sample fraction in extreme-value estimation. *J Multivar Anal* 47:173–195
- Del Castillo J, Padilla M (2016) Modeling extreme values by the residual coefficient of variation. *SORT-Stat Oper Res Trans* 40:303–320
- Drees H (1998) A general class of estimators of the extreme value index. *J Stat Plan Inference* 98:95–112
- Drees H, Kaufmann E (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch Process Appl* 75:149–172
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7:1–26
- Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling extremal events for insurance and finance*. Springer, Berlin
- Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math Proc Camb Philos Soc* 24:180–190
- Gomes MI (1994) Metodologias Jackknife e Bootstrap em Estatística de Extremos. *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística* 31–46
- Gomes MI, Oliveira O (2001) The bootstrap methodology in statistics of extremes: choice of the optimal sample fraction. *Extremes* 4:331–358
- Gomes MI, Pestana D, Caeiro F (2009) A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Stat Probab Lett* 79:295–303
- Guha-Sapir D, EM-DAT: The emergency events database. In: Université catholique de Louvain (UCL)—CRED. Brussels, Belgium <https://www.emdat.be>
- Hall P (1982) On some simple estimates of an exponent of regular variation. *J Roy Stat Soc Ser B (Methodol)* 44:37–42
- Hall P (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J Multivar Anal* 32:177–203
- Hill BM (1975) A simple general approach to inference about the tail index. *Ann Stat* 3:1163–1174
- Leadbetter MR (1991) On a basis for 'Peaks over Threshold' modeling. *Stat Probab Lett* 12:357–362
- Mason DM (1982) Law of large number for sums of extreme values. *Ann Probab* 10:754–764
- Pfaff B, McNeil A (2018) evir: Extreme Values in R. In: R package version 1.7-4. <https://CRAN.R-project.org/package=evir>
- Pickands J (1975) Statistical inference using extreme order statistics. *Ann Stat* 3:119–131
- Qi Y (2008) Bootstrap and empirical likelihood methods in extremes. *Extremes* 11:81–97
- Resnick SI (1997) Discussion of the Danish data on large fire insurance losses. *Astin Bull* 27:139–151
- Weissman I (1978) Estimation of parameters and larger quantiles based on the k largest observations. *J Am Stat Assoc* 73:812–815