



Threshold selection in univariate extreme value analysis

Laura Fee Schneider¹ · Andrea Krajina¹ · Tatyana Krivobokova²

Received: 28 June 2019 / Revised: 29 December 2020 / Accepted: 3 January 2021 /
Published online: 16 April 2021
© The Author(s) 2021

Abstract

Threshold selection plays a key role in various aspects of statistical inference of rare events. In this work, two new threshold selection methods are introduced. The first approach measures the fit of the exponential approximation above a threshold and achieves good performance in small samples. The second method smoothly estimates the asymptotic mean squared error of the Hill estimator and performs consistently well over a wide range of processes. Both methods are analyzed theoretically, compared to existing procedures in an extensive simulation study and applied to a dataset of financial losses, where the underlying extreme value index is assumed to vary over time.

Keywords Peaks-over-threshold approach · Power laws · Hill estimator · Tuning parameter selection · Bias estimation

AMS 2000 Subject Classifications 62G32 · 62G05 · 62F12 · 97M30

1 Introduction

Extreme value analysis of heavy-tailed distributions is important for various applications. In seismology and climatology, for example, statistics of extremes are used to

✉ Laura Fee Schneider
lfee.schneider@gmail.com

Andrea Krajina
akrajina@gmail.com

Tatyana Krivobokova
tatyana.krivobokova@univie.ac.at

¹ Institute for Mathematical Stochastics, Georg-August University Göttingen, Goldschmidtstrasse 7, 37077 Göttingen, Germany

² Institute for Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria

study earthquakes (Beirlant et al. 2018) or heavy precipitation (Carreau et al. 2017). Another important field of research is analysing high financial losses, which becomes particularly interesting if the losses depend on covariates (Chavez-Demoulin et al. 2016; Hambuckers et al. 2018).

To investigate the behaviour of heavy tails, we consider random variables from the max-domain of attraction (DoA) of a Fréchet distribution. Let X_1, \dots, X_n be independent and identically distributed (i.i.d.) random variables with a distribution function F , where F is in the DoA of an extreme value distribution (evd) G_γ with extreme value index $\gamma > 0$. This means there exist sequences $a_n > 0$ and b_n real, s.t.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) := \exp\left(-x^{-1/\gamma}\right).$$

In this situation the following first order condition holds,

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad x > 0, \tag{1}$$

i.e. the survival function $1 - F$ is regularly varying with index $-1/\gamma$. Distributions fulfilling this condition are called Pareto-type distributions, because they only differ from the Pareto distribution by a slowly varying function $\ell_F(x)$, i.e. $1 - F(x) = x^{-1/\gamma} \ell_F(x)$.

We can interpret the quotient in Eq. 1 as a conditional probability, from which it follows directly that

$$\begin{aligned} \frac{X_1}{t} \Big| X_1 > t &\xrightarrow{\mathcal{D}} P, \text{ as } t \rightarrow \infty \text{ and } P \sim \text{Pareto}\left(1, \frac{1}{\gamma}\right), \\ \log\left(\frac{X_1}{t}\right) \Big| X_1 > t &\xrightarrow{\mathcal{D}} E, \text{ as } t \rightarrow \infty \text{ and } E \sim \text{Exp}\left(\frac{1}{\gamma}\right), \end{aligned} \tag{2}$$

where Pareto $(1, 1/\gamma)$ denotes the Pareto distribution with the scale parameter 1 and the shape parameter $1/\gamma$. Thus, for a sufficiently large threshold t the data above this threshold can be modelled by a Pareto or an exponential distribution. In this article we concentrate on the exponential approximation and utilize it for inference on the extreme value index. It is common to consider the threshold $t = X_{(n-k,n)}$ and choose the sample fraction k instead of t , where $X_{(1,n)} \leq \dots \leq X_{(n,n)}$ denote the order statistics of a sample of size n . In this case, a natural estimator for γ under the exponential approximation of the log-spacings $Y_{(i,k)} := \log(X_{(n-i+1,n)}) - \log(X_{(n-k,n)})$ is their mean, the Hill estimator (Hill 1975),

$$\hat{\gamma}_k := \frac{1}{k} \sum_{i=1}^k \log\left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}}\right) = \frac{1}{k} \sum_{i=1}^k Y_{(i,k)}. \tag{3}$$

The Hill estimator is still among the most popular and well-known estimators for the extreme value index, although its sample path as a function in k can be highly unstable and estimation, therefore, crucially depends on the choice of the sample fraction k . This dependence highlights the difficulties in estimating γ in the peaks-over-threshold approach. To select a threshold above which the data can be used for statistical inference in the tail is one of the most fundamental problems in the field of extreme value analysis.

Due to the importance of this task, an appropriate choice of the threshold has been discussed extensively in extreme value research over the last decades and suggested solutions cover a variety of methodologies. We give a short summary of different types of approaches and stress the specific difficulties that arise. We mainly concentrate on the methods considered in the simulation study in Section 4. More comprehensive reviews about threshold selection can be found in Scarrott and MacDonald (2012) and Dey and Yan (2016).

One basic concept in threshold selection is data visualisation, which is also discussed more deeply in Kratz and Resnick (1996) and Drees et al. (2000). Popular graphical diagnostics used in this context are the Zipf plot, Hill plot, QQ-plot or the mean-excess plot, to name a few. A major drawback of these methods is their subjectivity due to the necessarily personal interpretation of the plot. Further, it is a burden to choose each threshold manually, especially in high dimensional settings or when analysing many samples. Easier ways to select the sample fraction are rule-of-thumb approaches such as using the upper 10% of the data (DuMouchel 1983) or $k = \sqrt{n}$ (Ferreira et al. 2003). Reiss and Thomas (2007) present a procedure that tries to find a region of stability among the estimates of the extreme value index. Their method depends on a tuning parameter, whose choice is further analysed in Neves and FragaAlves (2004). To our knowledge, no theoretical analysis exists for this approach.

Besides these, and similar, heuristic approaches, there is a class of theoretically motivated procedures that target the optimal sample fraction for specific estimation tasks, such as quantile estimates (Ferreira et al. 2003), estimation of high probabilities (Hall and Weissman 1997) or the Hill estimator, see below. Further, some suggestions compare the empirical distribution to the fitted generalized Pareto distribution (GPD) via goodness-of-fit tests (Northrop and Coleman 2014; Wadsworth 2016; Bader et al. 2018) or by minimizing the distance between them (Pickands 1975; Gonzalo and Olmo 2004; Clauset et al. 2009), where the latter approach is theoretically analysed in (Drees et al. 2018). Alternatively, Bayesian procedures can be considered for threshold selection, e.g., the cross-validation approach suggested in Northrop et al. (2017) in the context of ocean storm severity.

Of particular interest to us are methods that aim to estimate the sample fraction $k_{\text{opt}} = \arg \min_{k>0} \mathbb{E}(\hat{\gamma}_k - \gamma)^2$ which minimizes the asymptotic mean squared error (AMSE) of the Hill estimator. To construct an estimator for k_{opt} , Drees and Kaufmann (1998) utilize the Lepskii method and an upper bound on the maximum random fluctuation of $\hat{\gamma}_k$ around γ . To apply their approach it is necessary to choose several tuning parameters and to obtain consistent initial estimates for γ and a second-order parameter ρ . In Guillou and Hall (2001), a test statistic is constructed based on an accumulation of log-spacings, which is tested against a critical value that has to be chosen. Danielsson et al. (2001) introduce a double bootstrap approach to estimate the optimal sample fraction. They need to choose the number of bootstrap samples and a parameter n_1 . For n_1 , a data-driven, but computationally expensive selection method is provided, where the whole bootstrap procedure is repeated for various possible values of n_1 . Another estimator for k_{opt} is given in Beirlant et al. (2002), which employs least squares estimates from an exponential regression approach. The method depends on an estimate for ρ and a sample fraction k_0 . A different approach is

taken in Goegebeur et al. (2008), where the properties of kernel test statistics regarding the estimation of bias are used in order to construct an estimator for the AMSE/ γ and minimize it with respect to k . Here the choice or estimation of the second-order parameter ρ is needed.

In this paper, we contribute to the problem of threshold selection by introducing two new and easily applicable methods, which do not require the user to manually choose any tuning parameters. The first method presented in Section 2 is inspired by the idea of testing the exponential approximation. We estimate the integrated square error (ISE) of the exponential density under the assumption that the log-spacings are indeed exponentially distributed. The error functional we obtain, called the inverse Hill statistic (IHS), is very easy to compute and does not depend on any tuning parameters. Since this criterion is variable for small k , it can be additionally smoothed to improve the performance. The minimizing oracle sample fraction of IHS is asymptotically smaller than k_{opt} , as it is stricter against deviation from an exponential approximation. However, this estimator performs remarkably well in small samples, as illustrated in our simulation study.

In the second approach, we suggest a smooth estimator for the AMSE of the Hill estimator, called SAMSEE (smooth AMSE estimator). This estimator is based on a preliminary estimate of γ and an estimate of the bias of the Hill estimator. For the preliminary gamma estimate we use a generalized Jackknife approach in Gomes et al. (2000). By minimizing SAMSEE we estimate the optimal sample fraction k_{opt} . For estimation, the choice of a large sample fraction K is necessary, for which we present a data-driven selection procedure in Section 3. SAMSEE utilizes the idea of fixing $\rho = -1$, which is justified by good performance in simulations and leads to a simpler and more robust estimator than employing an estimate $\hat{\rho}$.

After introducing the two novel threshold selection methods in Sections 2 and 3 we compare these methods to various other approaches in a numerical analysis in Section 4. In Section 5 the importance of automated threshold selection procedures is illustrated in an application, where we non-parametrically estimate an extreme value index that varies over time. This example illustrates how the new methods enable the selection of a threshold that depends on covariates.

The proof of Theorem 3, which describes the asymptotic behaviour of the bias estimator, as well as the auxiliary results, can be found in Appendix A. Appendix B provides additional results from our simulation study.

2 Inverse Hill statistic

In this section, we introduce the first threshold selection procedure by analysing the integrated square error (ISE) between the exponential density h_γ and its parametric estimator $h_{\hat{\gamma}_k}$ employing the Hill estimator,

$$\text{ISE}(k) := \int \{h_\gamma(x) - h_{\hat{\gamma}_k}(x)\}^2 dx = \frac{1}{2\gamma} - \frac{2}{\gamma + \hat{\gamma}_k} + \frac{1}{2\hat{\gamma}_k},$$

where $h_\gamma(x) := \frac{1}{\gamma}e^{-x/\gamma}$ and $\hat{\gamma}_k$ is the Hill estimator. The first term of ISE is constant and thus plays no role for selecting k . The last term of ISE is known, but the

second term is not. Therefore, we cannot minimize ISE directly. Instead, we want to estimate and minimize its expectation under the exponential approximation. This is based on the idea of considering the hypothesis H_0 that the log-spacings $Y_{(i,k)}$ are indeed exponentially distributed. Under H_0 the Hill estimator is gamma distributed as the sum of independent exponential random variables and the mean of ISE (MISE) can be calculated explicitly. We observe that MISE is a decreasing function in k under the exponential approximation,

$$\text{MISE}(k) - \frac{1}{2\gamma} := \mathbb{E}_{H_0}[\text{ISE}(k)] - \frac{1}{2\gamma} = -\frac{1}{\gamma}C(k) + \frac{k}{2(k-1)\gamma}, \tag{4}$$

where $C(k) := 2 \exp(k)k^k \Gamma(1 - k, k)$ and $\Gamma(a, b) := \int_b^\infty t^{a-1} e^{-t} dt$ denotes the upper incomplete gamma function. One can show that $C(k) \approx 1 + 1/(4k)$, i.e. it converges to 1 very fast, s.t. we obtain

$$\lim_{k \rightarrow \infty} \mathbb{E}_{H_0} \left[\frac{2}{\gamma + \hat{\gamma}_k} \right] = \frac{1}{\gamma} = \mathbb{E}_{H_0} \left[\frac{k-1}{k\hat{\gamma}_k} \right].$$

This provides an estimator for the first term in Eq. 4 under H_0 . However, owing to its high variability for small k , we instead want to find an estimator of the form $w/\hat{\gamma}_k$ for some w depending on k that minimizes the MSE of that estimator under the exponential approximation. To do so, we approximate its MSE in the following way,

$$\mathbb{E}_{H_0} \left[\left(\frac{w}{\hat{\gamma}_k} - \frac{2}{\hat{\gamma}_k + \gamma} \right)^2 \right] \approx \frac{w^2 k^2}{\gamma^2 (k-1)(k-2)} - \frac{2wk}{\gamma^2 (k-1)} + \frac{1}{\gamma^2}. \tag{5}$$

The approximation depends on similar functions as $C(k)$, which quickly become constant. The MSE in Eq. 5 is minimized for $w = (k - 2)/k$. Thus, we suggest the inverse Hill statistic

$$\text{IHS}(k) := \frac{1}{2\hat{\gamma}_k} - \frac{k-2}{\hat{\gamma}_k k} = \frac{4-k}{2\hat{\gamma}_k k}$$

to estimate $\text{MISE}(k) - (2\gamma)^{-1}$ and the sample fraction selected via minimizing IHS, $\hat{k}_{\text{IHS}} := \arg \min_{1 < k < n} \{\text{IHS}(k)\}$.

By minimizing IHS we select a sample fraction where IHS starts increasing and thereby contradicts the behaviour of MISE under H_0 . This criterion can be compared to hypothesis testing with a high significance level α , which implies seeking high confidence when deciding to not reject H_0 . Further properties of \hat{k}_{IHS} are analysed theoretically in Section 2.1 and for finite samples in a numerical study in Section 4.

Figure 1 illustrates that IHS exhibits high fluctuations for small k , which might make the automated selection of the minimum highly variable. To control this problematic behaviour we smooth IHS. More specifically, we want to estimate $\mathbb{E}[\text{IHS}]$ by considering the regression problem

$$\text{IHS}(k) = \mathbb{E}[\text{IHS}(k)] + \sigma \epsilon_k, \quad k = 1, \dots, n,$$

where $\sigma > 0$ and $\mathbb{E}[\epsilon_k] = 0$. Due to the structure of the Hill estimator, the random variables ϵ_k are highly dependent, which needs to be taken into account in estimation. In our simulations, we apply a Bayesian non-parametric procedure introduced by Serra et al. (2018) to obtain a smooth estimator for the expectation of IHS – denoted

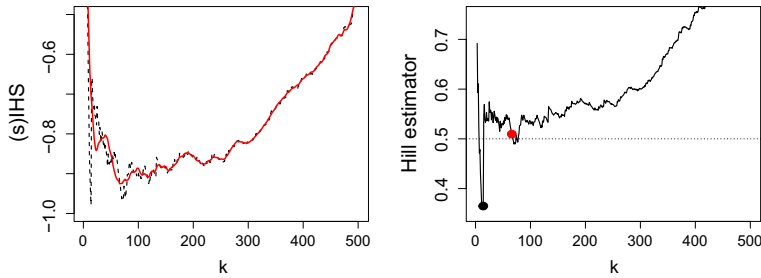


Fig. 1 On the left, the IHS (dashed) and sIHS (red) are plotted for a Fréchet(2) sample of size $n = 500$. On the right the Hill plot for the same sample with the minimizing k of IHS (black) and sIHS (red) is shown, where the dotted line marks the true value of $\gamma = 0.5$

as sIHS – comprising less variation for small k , as illustrated in Fig. 1, and thereby improving automatic threshold selection. The approach implemented in the `eBSC` R-package is completely data-driven and non-parametric: both mean and covariance structure of the data are estimated non-parametrically with the smoothing parameters chosen from the data. Alternatively, one can employ any other smoothing technique, which accounts for dependence in the error (Opsomer et al. 2001; Krivobokova and Kauermann 2007; Lee et al. 2010). A very good choice is the well-established function `gamm` from the `mgcv` package. However, for this procedure some parameters need to be fixed in advance. We experimented with several settings and found that choosing the number of knots to be at least 40 (or more) and specifying the correlation structure to `corAR1()` delivers results very much comparable with the `eBSC` function.

We finally want to remark on the relation between IHS and ISE, which is given by

$$\text{IHS} + \frac{1}{2\gamma} = \text{ISE} + \frac{2}{k\hat{\gamma}_k} + \frac{\hat{\gamma}_k - \gamma}{\hat{\gamma}_k(\hat{\gamma}_k + \gamma)}. \tag{6}$$

This equation points out that minimizing IHS does not minimize ISE, as IHS takes an additional bias term into account, and that the sign of the bias of $\hat{\gamma}_k$ plays a role. If the bias is negative, we suggest using

$$\text{IHS}^-(k) := \frac{4+k}{2\hat{\gamma}_k k} \quad \text{and} \quad \hat{k}_{\text{IHS}^-} := \arg \min_{1 < k < n} \text{IHS}^-(k)$$

instead of IHS, where the two cases can easily be distinguished by analysis of the Hill estimator for large k . If the bias of the Hill estimator is positive, which is the standard case, we can see from Eq. 6 that IHS selects smaller k (larger thresholds) than ISE. This is not surprising, because we estimate the expectation of the ISE under the hypothesis that the exponential approximation holds. This is a much more conservative error functional, meaning it is more strict against deviation from the exponential distribution.

In conclusion, with IHS we do not aim to estimate k_{opt} but to find a sample fraction where we can be very certain that the exponential approximation still holds. The performance of sIHS in small samples is illustrated in simulations and an application in Sections 4 and 5.

2.1 Theoretical analysis of IHS

In order to understand the IHS asymptotically, we consider the second-order condition,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \tag{7}$$

for $x > 0$ and with second-order parameter $\rho < 0$. Here, $A(t)$ denotes a function converging to zero as t goes to infinity and $|A|$ is regularly varying with index ρ . Further, U is defined by $U(x) := F^{\leftarrow}\left(1 - \frac{1}{x}\right)$, where F^{\leftarrow} denotes the left inverse of the distribution function F . In this setting the following asymptotic normality statements for the Hill estimator $\hat{\gamma}_k$ hold.

Theorem 1 (Theorem 3.2.5 in de Haan and Ferreira, 2006) *Let X_1, \dots, X_n be i.i.d. random variables with distribution function $F \in \text{DoA}(G_\gamma)$ for $\gamma > 0$. If Eq. 7 holds and k is an intermediate sequence, i.e. $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$, then*

$$\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{(1-\rho)}, \gamma^2\right),$$

as $n \rightarrow \infty$ and with $\lambda := \lim_{k \rightarrow \infty} \sqrt{k}A(n/k)$.

Theorem 2 *Under the conditions of Theorem 1, it holds as $n \rightarrow \infty$ that*

$$\sqrt{k}\left(\frac{1}{\hat{\gamma}_k} - \frac{1}{\gamma}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{-\lambda}{(1-\rho)\gamma^2}, \frac{1}{\gamma^2}\right).$$

Proof Apply the delta method to Theorem 1. □

Following the reasoning in de Haan and Ferreira (2006), page 78, the minimizing point of the AMSE of the Hill estimator can be found explicitly if considering $A(t) = ct^\rho$ with $c \neq 0$. In this special case, the minimizing sample fraction can be expressed as

$$k_{\text{opt}} = \left(\frac{\gamma^2(1-\rho)^2}{-2\rho c^2}\right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}. \tag{8}$$

Under the same assumption, we can calculate the minimizing point k_{IHS} of the asymptotic expectations of IHS and IHS^- . Let $\mathbb{A}\mathbb{E}$ denote the asymptotic expectation referring to the expectation of the limiting distribution in Theorem 2. Then

$$\begin{aligned} k_{\text{IHS}} &:= \arg \min_k \mathbb{A}\mathbb{E}[\text{IHS}] = \arg \min_k \left\{ \frac{2}{\gamma k} + \frac{A(n/k)}{2\gamma^2(1-\rho)} \cdot \frac{k-4}{k} \right\} \\ &\approx \arg \min_k \left\{ \frac{2}{\gamma k} + \frac{A(n/k)}{2\gamma^2(1-\rho)} \right\} = \left(\frac{4\gamma(1-\rho)}{-\rho c}\right)^{1/(1-\rho)} n^{-\rho/(1-\rho)}. \end{aligned}$$

It is easy to check that the same formula holds for IHS^- if c is replaced by its absolute value. Comparing k_{opt} and k_{IHS} for a fixed $\rho > -\infty$ we obtain that

$$\frac{k_{\text{IHS}}}{k_{\text{opt}}} \approx \left(\frac{-\rho}{32} \cdot k_{\text{IHS}}\right)^{-1/(1-2\rho)} \approx d \cdot n^{\frac{\rho}{(1-2\rho)(1-\rho)}} \rightarrow 0, \tag{9}$$

as $n \rightarrow \infty$ and for a constant d depending on ρ , γ and c . This supports what Eq. 6 already suggested: minimizing IHS gives asymptotically a smaller k than k_{opt} . Thus, k_{IHS} asymptotically performs suboptimally for the Hill estimator but still is an intermediate sequence leading to consistent estimates asymptotically. For finite samples the ratio crucially depends on ρ , and k_{IHS} can also be larger than k_{opt} , as illustrated in Fig. 2. It also holds that $k_{\text{IHS}}/k_{\text{opt}} \rightarrow 1$, as $\rho \rightarrow -\infty$, since both sample fractions converge to n in this case.

Although k_{IHS} is of smaller order than k_{opt} asymptotically, the simulation study in Section 4 shows that \hat{k}_{IHS} works quite well in small samples and in particular when used for quantile estimation of certain processes. We consider the following quantile estimator for the $(1 - p)$ -quantile,

$$\hat{q}_k(p) = X_{(n-k,n)} \left(\frac{k}{np} \right)^{\hat{\gamma}_k}, \tag{10}$$

as defined in Theorem 4.3.8 in de Haan and Ferreira (2006). This theorem shows that the sample fraction k_{opt} also minimizes the asymptotic relative MSE of $\hat{q}_k(p)$. For finite samples, however, the quantile estimator seems to benefit from k_{IHS} , since in this case IHS often works better than procedures estimating k_{opt} , see Section 4. This can have different reasons, two of which are illustrated by Fig. 3. Figure 3 displays the empirical expectation of IHS and of SAMSEE defined in Section 3, the empirical versions of the MSE of $\hat{\gamma}_k$ and the relative MSE of the quantile estimator,

$$\text{MSEQ} := \mathbb{E} \left[\left(\frac{\hat{q}_k(p)}{q(p)} - 1 \right)^2 \right] / \log \left(\frac{k}{np} \right), \tag{11}$$

as used in Theorem 4.3.8 in de Haan and Ferreira (2006). The estimator $\hat{q}_k(p)$ is more sensitive to changes in k than $\hat{\gamma}_k$ and thus the empirical version of MSEQ seems

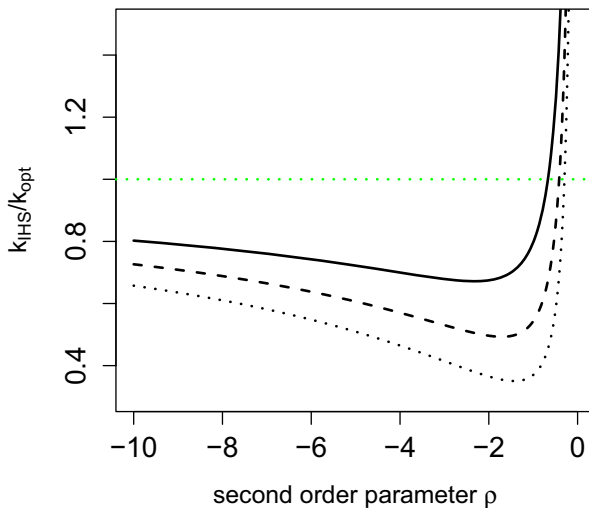


Fig. 2 The approximation of the quotient $k_{\text{IHS}}/k_{\text{opt}}$ in Eq. 9 is plotted as a function in the second-order parameter ρ with $\gamma = c = 1$ for $n = 500$ (solid), $n = 5000$ (dashed) and $n = 50000$ (dotted)

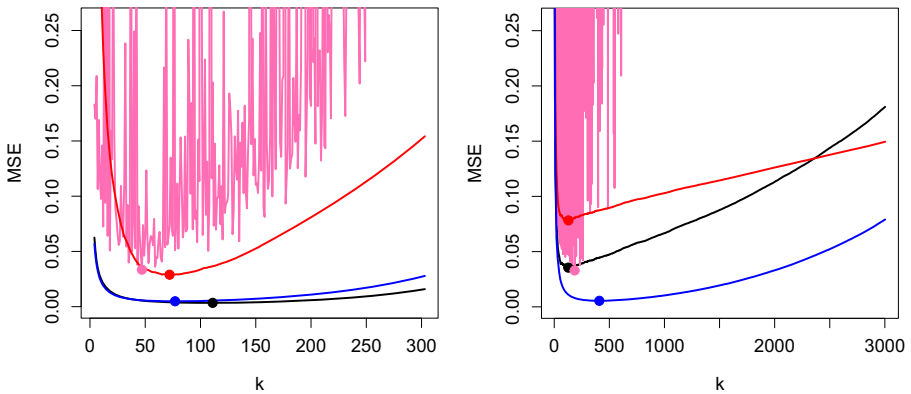


Fig. 3 Empirical expectations of IHS plus a constant (red), MSE (black), MSEQ (pink) and SAMSEE (blue, see Section 3) are presented. The left plot is based on 10,000 samples from a Fréchet(2) distribution of size 500. The graphic on the right is based on 500 samples of size 5000 from a Loggamma distribution

to need much more samples than the MSE of the Hill estimator to reach a stable estimate.

On the left of Fig. 3, we observe that k_{IHS} is indeed smaller than k_{opt} for a Fréchet distribution, where $\rho = -1$, but so is the minimizer of MSEQ. The graphic on the right highlights the similarities between MSE, MSEQ and IHS for the boundary case $\rho = 0$ of the Loggamma distribution. The method SAMSEE, which is defined in the next section, overestimates k_{opt} on average in the Loggamma case and its mean lies between k_{IHS} and k_{opt} for the Fréchet distribution.

3 Smooth AMSE estimator

In this section, we illustrate a way to estimate k_{opt} via minimizing a smooth estimator for the AMSE of the Hill estimator, called SAMSEE. From Theorem 1 it is easy to see that the AMSE of the Hill estimator is

$$A\mathbb{E}[(\hat{\gamma}_k - \gamma)^2] = \frac{\gamma^2}{k} + \frac{A(n/k)^2}{(1 - \rho)^2}. \tag{12}$$

To estimate the AMSE as a function of k , we employ two estimators, one for γ and one for the bias term as a combination of ρ and A . First we explain how we estimate γ and then we define the bias estimator.

For γ we consider the generalized Jackknife estimator $\hat{\gamma}_k^{GJ}$ introduced by Gomes et al. (2000) as γ_n^{G1} . This estimator is defined by

$$M_{n,k} := \frac{1}{k} \sum_{i=1}^k Y_{(i,k)}^2, \quad \hat{\gamma}_k^V := \frac{M_{n,k}}{2\hat{\gamma}_k}, \quad \text{and} \quad \hat{\gamma}_k^{GJ} := 2\hat{\gamma}_k^V - \hat{\gamma}_k, \tag{13}$$

where $Y_{(i,k)}$ denotes the log-spacings as in Eq. 3. Note, that $\hat{\gamma}_k^V$ is the de Vries estimator introduced under this name in de Haan and Peng (1998) and $\hat{\gamma}_k$ is the Hill

estimator as above. The generalized Jackknife estimator has a reduced bias compared to the Hill estimator and is even asymptotically unbiased if $\rho = -1$, see (2.11) in Gomes et al. (2000).

Danielsson et al. (2001) use $(\hat{\gamma}_k^V - \hat{\gamma}_k)$ to access the bias of $\hat{\gamma}_k$ and apply a double bootstrap procedure to stabilize this highly varying estimate. We also use the difference of two estimators for γ , but now consider the following averages of Hill estimators,

$$\bar{\gamma}_{k,K} := \frac{1}{K - k + 1} \sum_{i=k}^K \hat{\gamma}_i \quad \text{and} \quad \bar{\gamma}_k := \bar{\gamma}_{1,k} = \frac{1}{k} \sum_{i=1}^k \hat{\gamma}_i,$$

where $k < K$ is chosen to obtain a stable bias estimate. The idea to average the Hill estimator in order to reduce variation in the sample path is also studied in Resnick and Stărică (1997). These averages naturally contain a lot of structural information about the underlying asymptotic bias and inspire the definition

$$\hat{b}_{k,K} := \bar{\gamma}_{k,K} - \bar{\gamma}_K. \tag{14}$$

Theorem 3 *Under the conditions of Theorem 1 and for $k/K \rightarrow c$ with $0 < c < 1$ as $n \rightarrow \infty$, it holds for $\hat{b}_{k,K}$ defined in Eq. 14 that*

$$\sqrt{k} \cdot \hat{b}_{k,K} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{-\rho\lambda}{(1-\rho)^2} \delta_\rho(c), \gamma^2 v(c) \right),$$

where $\delta_\rho(c) = (c^\rho - 1)/(-\rho(c^{-1} - 1))$ and $v(c) = 2c^2/(1 - c)^2 \cdot (1 - c + c \log(c))$ with $0 \leq v(c) \leq 1$.

Proof The proof is given in Appendix A. □

By Theorem 3 the estimator $\hat{b}_{k,K}$ has the following property if $\rho = -1$, because in this case the function $\delta_{-1}(c)$ is equal to 1,

$$\mathbb{A}\mathbb{E} \left[\hat{b}_{k,K} \right] = \frac{-\rho A(n/k)}{(1-\rho)^2} = \frac{1}{2} \frac{A(n/k)}{(1-\rho)} = \frac{1}{2} \mathbb{A}\mathbb{E} [\hat{\gamma}_k - \gamma]. \tag{15}$$

It remains to choose an appropriate K that is large enough to allow for minimization over all relevant k and small enough to be an intermediate sequence itself, see Theorem 3. To find such a K we use the following relation between the estimators that holds if $\rho = -1$,

$$\mathbb{A}\mathbb{E}[\hat{\gamma}_k] = \mathbb{A}\mathbb{E}[\hat{\gamma}_k^V + \hat{b}_{k,K}]. \tag{16}$$

This provides us with a relatively stable function in k that has the same asymptotic expectation as the highly non-smooth Hill estimator. We want to find an intermediate sequence K for which Eq. 16 holds and thus define

$$\text{AD}(K) := \frac{1}{K} \sum_{k=1}^K \left(\hat{\gamma}_k^V + \hat{b}_{k,K} - \hat{\gamma}_k \right)^2 \tag{17}$$

to measure the deviation from approximation (16) uniformly over all $k \leq K$. We want to find the minimum of $\text{AD}(K)$, but not the one that occurs due to high variation

for small k . We look for a K such that the stabilized numerical approximation of the derivative of AD is closest to zero,

$$K^* := \arg \min_K \left\{ \sum_{i=-2, i \neq 0}^2 \left| \frac{AD(K) - AD(K + i)}{i} \right| \right\}. \tag{18}$$

Now we combine the previously described estimators to approach the AMSE in Eq. 12 under the assumption that $\rho = -1$. The idea of misspecifying ρ to simplify estimation – via avoiding the additional uncertainty through estimating ρ or selecting an influential tuning parameter – was already used, for example, by Gomes et al. (2000), Drees and Kaufmann (1998) and Goegebeur et al. (2008). It is also motivated by the simulations in Section 3.1.

Considering K^* in Eq. 18 and the property of $\hat{b}_{k,K}$ in Eq. 15, we define estimators for the AMSE of the Hill estimator and for k_{opt} by

$$\begin{aligned} \text{SAMSEE}(k) &:= \frac{(\hat{\gamma}_{K^*}^{\text{GJ}})^2}{k} + 4\hat{b}_{k,K^*}^2, \\ \hat{k}_{\text{SAMSEE}} &:= \underset{1 < k < K^*}{\text{argmin}} \text{SAMSEE}(k). \end{aligned} \tag{19}$$

The left panel of Fig. 4 displays SAMSEE for a Fréchet sample with parameters $\gamma = 1/2$ and $\rho = -1$. The Hill plot of the same sample and all $k \leq K^* = 388$ is shown in the right panel of Fig. 4.

This smooth estimate of the AMSE can be useful beyond the context of threshold selection. SAMSEE evaluates how well each k performs as a threshold in the peaks-over-threshold model, where the lowest values indicate a good fit. Thus, SAMSEE could be used to construct a transition function between bulk and tail distribution in extreme value mixture models or an empirical prior for the threshold in Bayesian threshold selection approaches, see Scarrott and MacDonald (2012) for a review on mixture models.

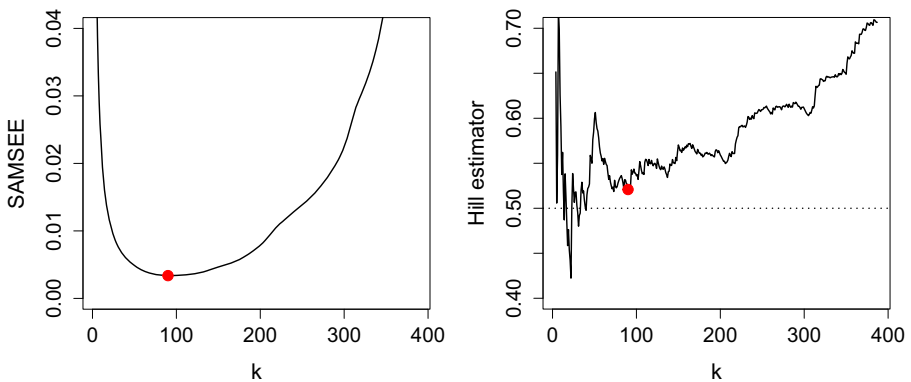


Fig. 4 SAMSEE with $K^* = 388$ on the left next to the Hill plot for the same Fréchet(2) random sample of size $n = 500$ for $k \leq K^*$. The red dot indicates the selected sample fraction \hat{k}_{SAMSEE} in the left plot and the adaptive Hill estimate $\hat{\gamma}_{\hat{k}_{\text{SAMSEE}}}$ in the right plot

3.1 SAMSEE if $\rho \neq -1$

In the broader context of an unknown second order parameter ρ , we note that the generalized Jackknife estimator is no longer unbiased and the behaviour of our bias estimator changes. From Theorem 3 follows that

$$\mathbb{A}\mathbb{E}[\hat{b}_{k,K}] = \frac{-\rho A(n/k)}{(1-\rho)^2} \cdot \delta_\rho(k/K).$$

If $\rho \neq -1$, we can observe that the function $\delta_\rho(k/K)$ influences the rate at which the bias estimator increases with k . We can still apply SAMSEE in this situation and select K^* from Eq. 18. However, approximation (16) does not hold anymore and instead it holds that

$$\mathbb{A}\mathbb{E}[\hat{\gamma}_k - \hat{\gamma}_k^V] - \mathbb{A}\mathbb{E}[\hat{b}_{k,K}] = \frac{-\rho A(n/k)}{(1-\rho)^2} (1 - \delta_\rho(k/K)). \tag{20}$$

The absolute value of the error described by Eq. 20 is high if δ_ρ strongly differs from 1 and the bias term A is large. If $\rho \neq -1$, δ_ρ indeed deviates from 1 and we minimize the difference in Eq. 20 by minimizing A . This is why applying (18) in this case leads to a small K^* . On the other hand, if $\delta_\rho = 1$, the approximation stays valid for an increasing bias and K^* will typically be larger.

An alternative to fixing $\rho = -1$ is to incorporate a consistent estimator $\hat{\rho}$ of the second order parameter. Estimators for ρ can be found in Caeiro and Gomes (2014), Gomes et al. (2002), and Fraga Alves et al. (2003) or Drees and Kaufmann (1998). Such an estimator can be included into the SAMSEE procedure via

$$K_{\hat{\rho}}^* := \arg \min_K \left\{ \sum_{i=-2, i \neq 0}^2 \left| \frac{\text{AD}_{\hat{\rho}}(K) - \text{AD}_{\hat{\rho}}(K+i)}{i} \right| \right\},$$

where $\text{AD}_{\hat{\rho}}(K) := \frac{1}{K} \sum_{k=1}^K \left(\hat{\gamma}_k^V + \hat{b}_{k,K} / \delta_{\hat{\rho}}(k/K) - \hat{\gamma}_k \right)^2$.

and

$$\text{SAMSEE}_{\hat{\rho}}(k) := \frac{(\hat{\gamma}_{K_{\hat{\rho}}^*}^{\text{GJ}})^2}{k} + \left((1 - \hat{\rho}) \frac{K_{\hat{\rho}}^*/k - 1}{(k/K_{\hat{\rho}}^*)^{\hat{\rho}} - 1} \cdot \hat{b}_{k, K_{\hat{\rho}}^*} \right)^2, \tag{21}$$

$$\hat{k}_{\hat{\rho}, \text{SAMSEE}} := \operatorname{argmin}_{1 < k < K^*} \text{SAMSEE}_{\hat{\rho}}(k).$$

In Table 1, we present the results of a simulation study indicating for which distributions it is beneficial to use $\hat{\rho}$ instead of $\rho = -1$. We estimate ρ using the estimator $\hat{\rho}^{(1)}$ suggested in Theorem 1 in Drees and Kaufmann (1998). The results indicate that for these processes it is sensible to fix $\rho = -1$ in SAMSEE, since only for the Cauchy distribution using $\hat{\rho}$ performs slightly better regarding both bias and root mean squared error (RMSE). This confirms the observations already made by others (Gomes et al. 2000; Drees and Kaufmann 1998; Goegebeur et al. 2008), that it is often recommendable to select $\rho = -1$ instead of allowing for further variability by including an additional estimator.

Table 1 The averages of adaptive γ estimates and their root mean squared error (in brackets) are presented for thresholds \hat{k} selected using SAMSEE or SAMSEE $_{\hat{\rho}}$ with the true ρ , $\rho = -1$ or $\hat{\rho} = \hat{\rho}^{(1)}$

	γ	ρ	$\mathbb{E}[\hat{\gamma}_{\hat{k}}]$ (RMSE)		
			true ρ	$\rho = -1$	$\hat{\rho}$
Student-t(6)	0.17	-1/3	0.21 (0.09)	0.26 (0.12)	0.28 (0.14)
Fréchet(2)	0.50	-1	0.51 (0.07)	0.51 (0.07)	0.51 (0.08)
Cauchy	1.00	-2	1.01 (0.13)	0.97 (0.17)	0.99 (0.16)
Burr(2,1)	2.00	-1	2.05 (0.34)	2.05 (0.34)	2.03 (0.40)

4 Simulation study

In the following, we numerically analyse the performance of ten threshold selection methods on heavy-tailed distributions with very different tail behaviours. The simulation study is based on the following distributions:

- the Student-t distribution with 6 degrees of freedom, which corresponds to $\gamma = 1/6$ and $\rho = -1/3$,
- the Fréchet distribution with parameter $\alpha = 2$ and distribution function $F(x) = \exp(-x^{-\alpha})$ for $x > 0$, which implies $\gamma = 1/2$ and $\rho = -1$,
- the standard Cauchy distribution leading to a tail behaviour with $\gamma = 1$ and $\rho = -2$,
- the Loggamma distribution with $\gamma = 1$ and $\rho = 0$ and density function

$$f(x) = \log(x)x^{-2} \mathbb{1}_{[1,\infty)}(x)$$

- the Burr distribution with a parametrisation such that $\gamma = 2$, $\rho = -1$ and distribution function

$$F(x) = 1 - (1 + \sqrt{x})^{-1}, \text{ for } x > 0,$$

- a logarithmically perturbed Pareto distribution of the random variable $g(U)$ with $\gamma = 1$ and $\rho = -1$, where $U \sim \text{Unif}(0, 1)$ and $g(x) = x^{-1}/\log(x^{-1})$. This distribution is denoted as negative Bias owing to the negative bias of the Hill estimator when applied to random samples from this distribution (Drees et al. 2000).

On these distributions, we evaluate the methods by their root median square error (RMedSE) when adaptively estimating γ with the Hill estimator relative to the RMedSE obtained using k_{opt} ,

$$\text{EFF}_{\gamma}(\hat{k}) := \sqrt{\frac{\mathbb{M}[(\hat{\gamma}_{\hat{k}} - \gamma)^2]}{\mathbb{M}[(\hat{\gamma}_{k_{\text{opt}}} - \gamma)^2]}}$$

where \mathbb{M} denotes the median. These efficiency quotients are also used with the mean instead of the median by, e.g., Guillou and Hall (2001) and Gomes et al. (2000) and Drees and Kaufmann (1998). We opted in favor of the median to reduce the influence

of few extreme outliers. The efficiency values based on the mean are additionally provided in Appendix B.

The smaller the quotient the better the threshold selection procedure performs compared to the optimal sample fraction k_{opt} . Values below 1 indicate that the selection method, that chooses a specific k for each sample, performs better than k_{opt} , which is fixed for all samples of a distribution. Furthermore, we study the efficiency in quantile estimation with the estimator defined in Eq. 10 for $p = 0.001$,

$$\text{EFF}_q(\hat{k}) := \sqrt{\frac{\mathbb{M}[(\hat{q}_{\hat{k}} - q)^2]}{\mathbb{M}[(\hat{q}_{k_{\text{opt}}} - q)^2]}}$$

Since the true minimizer k_{opt} of the AMSE is not known, we utilize an empirical version suggested by Gomes et al. (2000). Following their approach we approximate k_{opt} by the mean of 20 independent replicates of \bar{k}_{opt} , which is the minimizer of the empirical MSE based on 1000 samples, i.e. $\bar{k}_{\text{opt}} = \text{argmin}_k \mathbb{E}_{n=1000}[(\hat{\gamma}_k - \gamma)^2]$.

We compare these efficiency values for ten different threshold selection methods. If the choice of tuning parameters is required, we follow the recommendations given by the corresponding authors based on their numerical studies. Note that the choice of these parameters is not data-driven and may not be optimal.

Most of the considered approaches are constructed for adaptive estimation of γ applying the Hill estimator. This includes one procedure that looks for a stable region among the Hill estimates, while the others aim to estimate k_{opt} . We also evaluate the performance of the IHS approach discussed in Section 2, which aims to minimize the deviation from the exponential approximation and two methods relying on goodness-of-fit statistics. In total, the following methods are considered:

SAM: SAMSEE procedure with $\rho = -1$ as defined in Eq. 19 in Section 3,

GH: Method by Guillou and Hall (2001) utilizing $c_{\text{crit}} = 1.25$ and $p = 1$,

DK: Procedure by Drees and Kaufmann (1998) with fixed $\rho = -1$,

GO: Approach by Goegebeur et al. (2008) defined in their equation (3.3) with fixed $\rho = -1$,

DB: Double bootstrap approach by Danielsson et al. (2001) with n_1 minimizing a criterion suggested in their article, where $n_1 = n^{1-1/b}$ with $b \in \{4, 7, 12, 24, 38\}$,

B: Method by Beirlant et al. (2002) with $\rho = -1$ and taking the median of the estimates for $k_0 \in \{3, \dots, n/2\}$,

RT: Method by Reiss and Thomas (2007) with $\beta = 0$ as suggested by Neves and FragaAlves (2004),

C: Method defined in equation (3.9) of Clauset et al. (2009),

NC: Testing approach by Northrop and Coleman (2014) described in Section 4.2 using $\alpha = 0.2$ and 15 threshold suggestions from the 0.5 to the 0.95 empirical quantile,

sIHS: IHS smoothed by using the eBSc package, see Section 2.

Figures 5 and 6 illustrate the efficiency values obtained in our simulation study from random samples of size $n = 150, 500$ and 5000 drawn from the distributions above. The figures provide a convenient overview on the spread of the efficiency values for each method over the range of considered distributions. All values, including

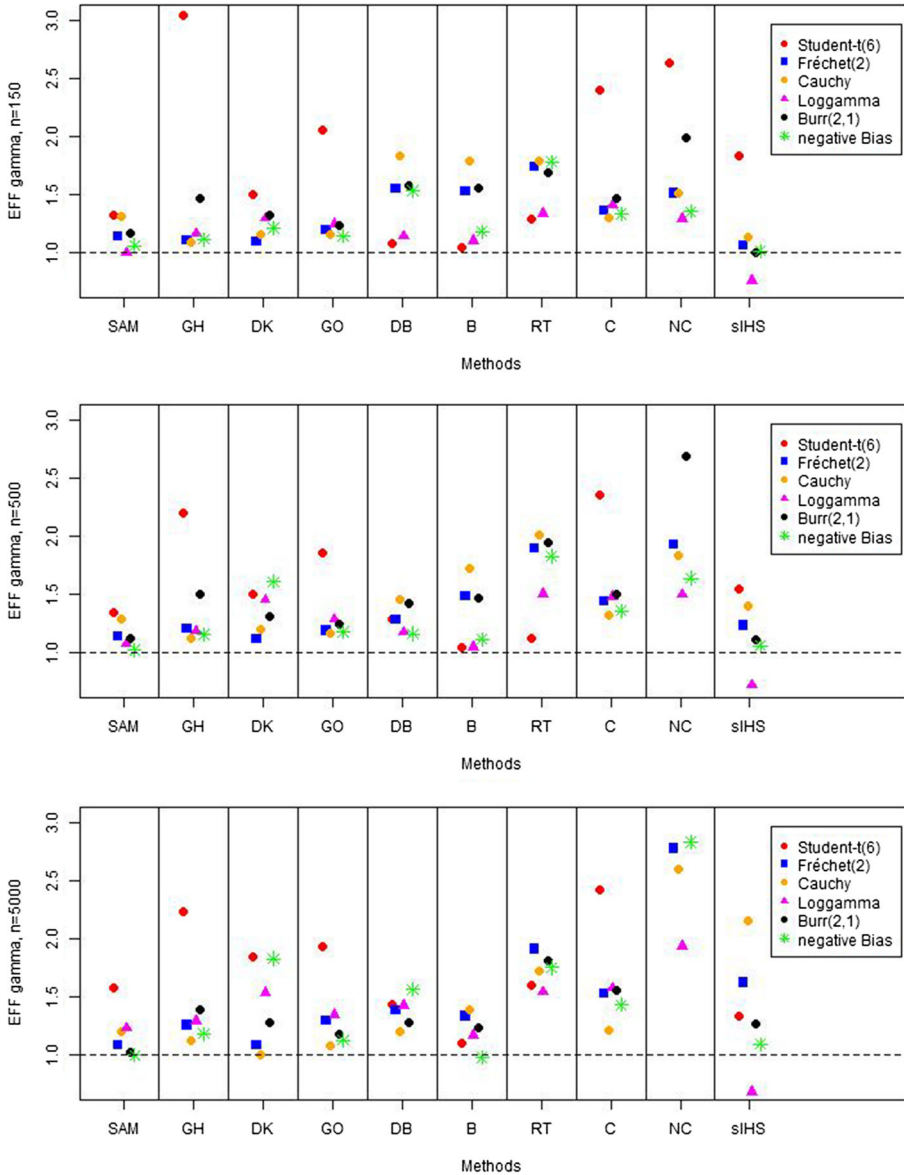


Fig. 5 Efficiency values EFF_γ based on 2000 samples if $n = 150$ and $n = 500$ and on 500 samples if $n = 5000$

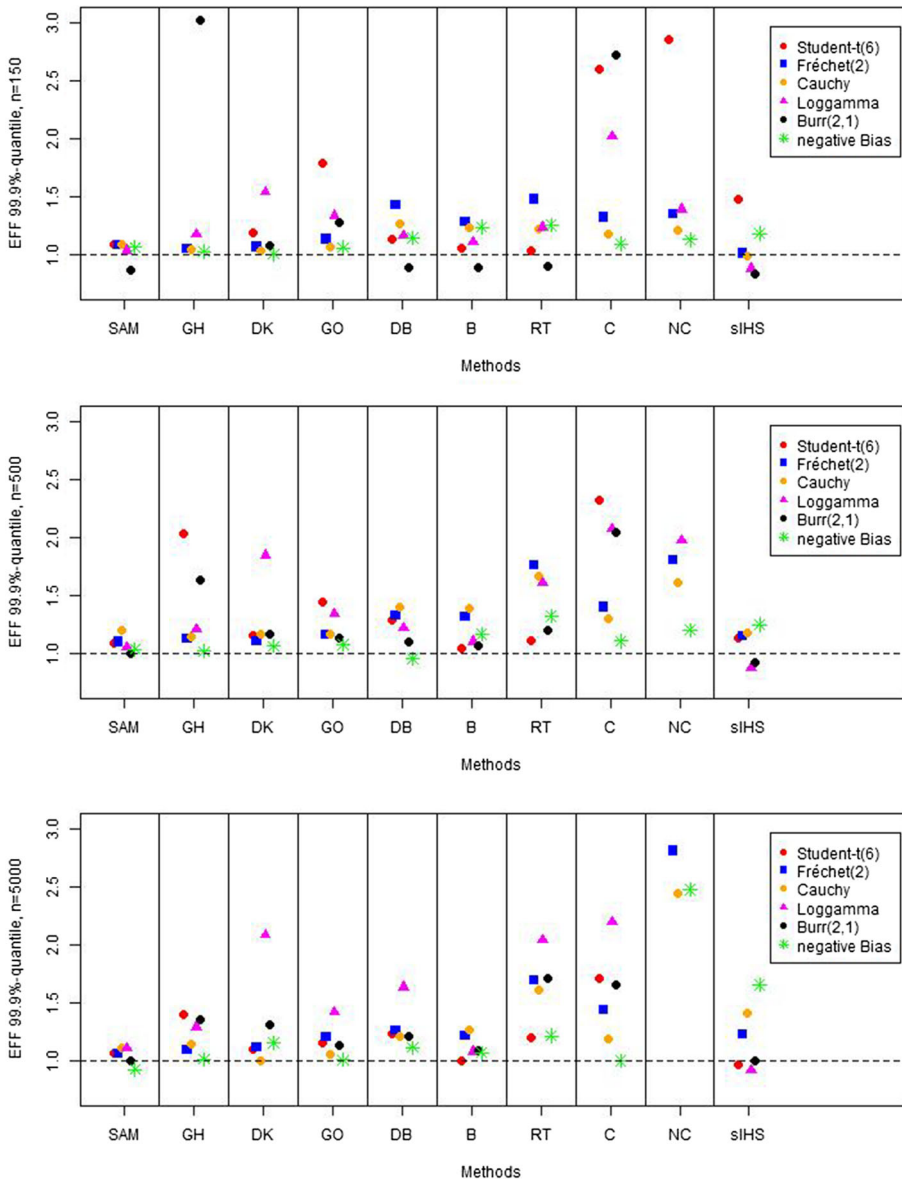


Fig. 6 Efficiency values EFF_q for $p = 0.001$ based on 2000 samples if $n = 150$ and $n = 500$ and on 500 samples if $n = 5000$

those that are too large for the detailed graphics, can be found in Tables 2 and 3 in Appendix B.

Overall, we see a very diverse picture of methods performing best depending on the distribution and the sample size. The graphs also show that SAMSEE is nicely concentrated close to one for all considered scenarios and especially for quantile

estimation. The good performance of SAMSEE stays valid for even more extreme quantiles like $p = 0.0001$, see Table 4.

When looking at adaptively estimating γ in Fig. 5, DK performs comparable well to SAMSEE on small samples, and for larger samples DB and B show efficiency values similarly concentrated as SAMSEE. Considering quantile estimation in Fig. 6 sIHS is an alternative to SAMSEE on small samples with even lower efficiency values for four of the six distributions. The approach B is almost as closely concentrated around 1 as SAMSEE when estimating quantiles from larger samples.

The performance of sIHS depends on the underlying distribution and deteriorates with growing sample size. The estimation of γ using sIHS still improves for larger samples but it performs weaker compared to k_{opt} , which confirms the theoretical findings in Section 2.1. However, sIHS performs impressively well, for example, for the loggamma distribution with efficiency values even below 1 in every scenario. In our study, sIHS performs well for smaller samples and for quantile estimation.

SAMSEE on the other hand shows impressively stable performance over the whole range of distributions, sample sizes and tasks. This makes SAMSEE a very good choice for threshold selection if no expert knowledge about the underlying distribution or about parameter tuning is available.

5 Application to varying extreme value index

In this section, we use the two introduced estimators for the sample fraction in a financial application on operational losses, where we are particularly interested in the distributional properties of very high losses. The observations of interest are operational losses from the Italian bank UniCredit collected from January 2005 to June 2014. The dataset provides scaled losses above 2000 Euro recorded with the corresponding date. The operational losses are grouped by the type of event that caused the specific loss. We consider the event type CPBP. The CPBP losses are caused by clients, products and business practices related to derivatives or other financial instruments. This group provides 16138 observations over the period from 2005 to 2014, which seems sufficient for our local estimation approach. The number of collected losses per quarter is presented on the left of Fig. 7 next to the logarithm of the raw data over time.

It has been discussed before that it is reasonable to assume that the distribution of such extreme losses is heavy-tailed (Chavez-Demoulin et al. 2016; Moscadelli 2004) and that it changes with the financial market over time (Hambuckers et al. 2018; Cope et al. 2012). In Fig. 12 in Appendix C, we provide some graphical justification that this assumption holds locally for the dataset of CPBP losses. In Hambuckers et al. (2018) the data is analysed in a regularized generalized Pareto regression approach including several firm-specific, macroeconomic and financial indicators as covariates. This approach describes the dependence of the GPD parameters on various covariates via parametric functions.

We consider the same data, but with a different focus. Our aim is to estimate the time dependent extreme value index $\gamma(t)$ non-parametrically with a simple ad hoc estimator that extends the estimator from de Haan and Zhou (2020) by utilizing the

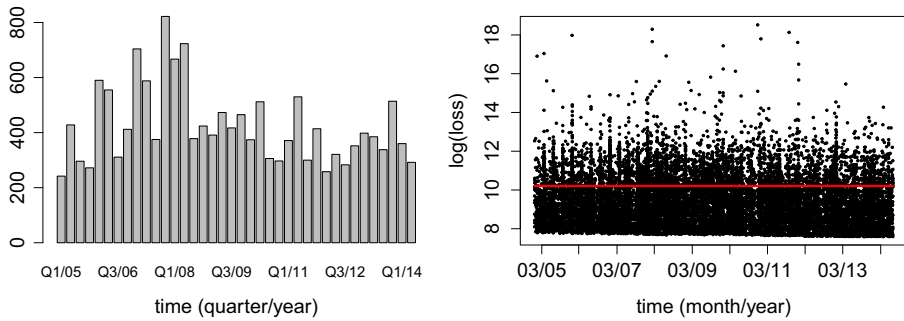


Fig. 7 The quarterly number of observed CPBP losses is illustrated on the left and on the right the logarithm of the losses is plotted over time with a red line marking the third quartile

approaches sIHS and SAMSEE presented in Sections 2 and 3 for the local selection of a threshold. We present the estimator for $\gamma(t)$ in Section 5.1 and the results we obtain when applying this estimator to the dataset of operational losses are shown in Section 5.2.

5.1 Estimating a varying extreme value index

In de Haan and Zhou (2020), the authors discuss testing for a trend in the extreme value index and estimating it non-parametrically. They consider n independent random variables $X_i \sim F_{i/n}$, where $F_s \in \text{DoA}(G_{\gamma(s)})$ for $s \in [0, 1]$, and they introduce the following estimator for $\gamma(s)$, which locally applies the Hill estimator and is based on a global sample fraction k ,

$$\hat{\gamma}_k(s) := \frac{1}{2kh} \sum_{i \in I_n(s)} (\log X_i - \log X_{(\lfloor 2nh \rfloor - \lfloor 2kh \rfloor, \lfloor 2nh \rfloor)})^+, \tag{22}$$

where $I_n(s)$ is the h -neighbourhood of s , i.e. $I_n(s) := \{i : |i/n - s| \leq h\}$. This estimator depends on the choice of the bandwidth h and the global sample fraction k , which is then rescaled to $2kh$ for the individual regions $I_n(s)$. A small bandwidth h leads to very high variability in $\hat{\gamma}_k(s)$ and a large value of h smooths out all interesting features. Thus, the choice of h should balance these two effects.

We suggest a modification of their estimator, where we locally estimate a threshold $\hat{k}(s)$, i.e.

$$\hat{\gamma}_{\hat{k}(s)}(s) := \frac{1}{\hat{k}(s)} \sum_{i \in I_n(s)} (\log X_i - \log X_{(\lfloor 2nh \rfloor - \hat{k}(s), \lfloor 2nh \rfloor)})^+. \tag{23}$$

To compare these two approaches, we repeat the simulation presented in de Haan and Zhou (2020), see their Figure 2(i), on samples of size $n = 5000$ with $X_i \sim \text{Fréchet}(1/\gamma(i/n))$, $\gamma(s) = 1 + c \cdot \sin(2\pi s)$ and $c = 1/4$. For both versions of the estimator the bandwidth is fixed to $h = 0.025$. Figure 8 illustrates the benefits of locally optimizing the threshold via SAMSEE from Section 3, as it strongly tightens

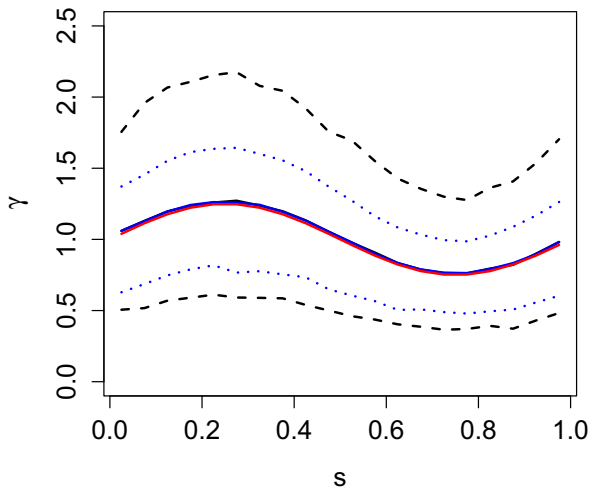


Fig. 8 The true extreme value index $\gamma(s) = 1 + 1/4 \cdot \sin(2\pi s)$ (red) next to the averaged estimators over 2000 samples. The estimator from de Haan and Zhou (2020) with fixed $k = 200$ is in black and its empirical 95% confidence interval is dashed. The localised estimator employing SAMSEE is blue with dotted empirical confidence bounds

the empirical confidence interval around the average, which is obtained from the 2.5% and 97.5% empirical quantiles among 2000 estimates.

5.2 Functional extreme value index of operational losses

First, we test if the extreme value index is constant over time. With the test T4 by Einmahl et al. (2016) we can reject the null hypothesis of constant γ for various k in the range of thresholds selected by our procedures. One concern in testing is the possible serial dependence in the financial losses, as discussed in Section 4.2 in de Haan and Zhou (2020), where it is suggested to repeat the test on subsets of the data to reduce the serial dependence. The test T4 still shows significant results if only considering every second or every 5th observation. Thus, we are confident that the extreme value index of the losses is indeed varying over time.

To estimate the time dependent $\gamma(t)$ we apply the estimator in Eq. 23. We select \hat{k} by using sIHS from Section 2 and the SAMSEE method from Section 3. It remains to choose the bandwidth h that defines the neighborhood $I_n(s)$. To our knowledge, no data-driven automated method for choosing an optimal bandwidth in the context of the peaks-over-threshold approach is available, and its investigation lies outside the scope of this article. Instead, we use a bandwidth of 170 days and thereby focusing on the long term trend of γ , since this bandwidth leads to a less varying extreme value index by smoothing over seasonal variations. Figure 9 shows the estimates of $\gamma(t)$ that we obtain for the event type CPBP. It is clearly visible that both procedures yield similar estimates for most time points and that the simple ad hoc estimators recover an increase of the severity of high losses during the financial and Euro crisis

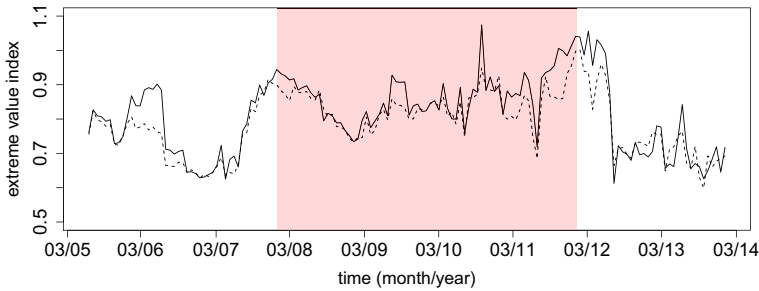


Fig. 9 The non-parametric estimate of the extreme value index of the operational losses of type CPBP is presented using \hat{k} from SAMSEE (solid) and siHS (dashed) and bandwidth $h = 170$ days. The red area indicates the time of the financial and Euro crisis

from 2008 to 2011. A similar overall trend in the extreme value index can also be identified in the estimates of Hambuckers et al. (2018) for CPBP.

For a more extensive discussion of the data and results of the more complex model including further covariates we refer to Hambuckers et al. (2018).

Appendix A: Theoretical results and proof of Theorem 3

Lemma 1 Let E_1, \dots, E_n be i.i.d. standard exponential random variables and k s.t. $1 \leq k \leq n$ and $k \rightarrow \infty$ as $n \rightarrow \infty$. We define the following random variables,

$$P_{k,n} := \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k E_i - 1 \right), \quad Q_{k,n} := \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k E_i^2 - 2 \right),$$

$$R_{k,n} := \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k e_{i+1}^k E_i - 1 \right),$$

where $e_i^k := \sum_{l=i}^k l^{-1} = \mathbb{E}[E_{(k-i+1,k)}]$. Then it holds for $n \rightarrow \infty$ that

$$(P_{k,n}, Q_{k,n}, R_{k,n})^T \xrightarrow{\mathcal{D}} \mathcal{N} \left((0, 0, 0)^T, \begin{pmatrix} 1 & 4 & 1 \\ 4 & 20 & 4 \\ 1 & 4 & 2 \end{pmatrix} \right).$$

Proof We use the Cramér-Wold device that gives us a joint normal limit distribution, if all linear combinations have an univariate normal limit distribution. For $a_1, a_2, a_3 \in \mathbb{R}$ we prove asymptotic normality for the sum $(a_1 P_{k,n} + a_2 Q_{k,n} + a_3 R_{k,n})$ by applying Liapounov’s central limit theorem (CLT), see Theorem 7.1.2. in Chung (1974).

We consider a sum $S_n := \sum_{i=1}^k X_{i,k}$ of independent random variables fulfilling the following three conditions,

- 1) $\mathbb{E}[X_{i,k}] = 0, \forall k \forall i,$
- 2) $\sum_{i=1}^k \text{Var}(X_{i,k}) = \sigma^2,$
- 3) $\Gamma(k) = \sum_{i=1}^k \mathbb{E}[|X_{i,k}|^3] \rightarrow 0, \text{ as } k \rightarrow \infty,$

Then the CLT proves a standard normal limit for S_n . We define

$$X_{i,k} := \frac{1}{\sqrt{k}} \left(a_1 E_i - a_1 + a_2 E_i^2 - 2a_2 + a_3 e_{i+1}^k E_i - a_3 e_{i+1}^k \right)$$

for $i = 1, \dots, k$ where $e_{k+1}^k := 0$, such that the approximation error between $\sum_{i=1}^k X_{i,k}$ and $(a_1 P_{k,n} + a_2 Q_{k,n} + a_3 R_{k,n})$ vanishes as $k \rightarrow \infty$, since

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} e_{i+1}^k &= \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} \frac{1}{k} \sum_{l=i+1}^k \frac{1}{l/k} = \lim_{k \rightarrow \infty} \int_{\frac{1}{k}}^1 \int_v^1 \frac{1}{u} du dv \\ &= - \lim_{k \rightarrow \infty} \int_{\frac{1}{k}}^1 \log(v) dv = 1. \end{aligned}$$

Now we have to check the three conditions. Condition 1) follows immediately from $\mathbb{E}[E_i] = 1$ and $\mathbb{E}[E_i^2] = 2$. For condition 2) we need to calculate the variance

$$\begin{aligned} \text{Var}(X_{i,k}) &= \text{Var} \left(\frac{1}{\sqrt{k}} \left(a_1 E_i - a_1 + a_2 E_i^2 - 2a_2 + a_3 e_{i+1}^k E_i - a_3 e_{i+1}^k \right) \right) \\ &= \frac{1}{k} \left(a_1^2 \text{Var}(E_i) + a_2^2 \text{Var}(E_i^2) + a_3^2 (e_{i+1}^k)^2 \text{Var}(E_i) \right. \\ &\quad \left. + 2(a_1 a_2 + a_2 a_3 e_{i+1}^k) \text{Cov}(E_i, E_i^2) + 2a_1 a_3 \text{Var}(E_i) \right) \\ &= \frac{1}{k} \left(a_1^2 + 20a_2^2 + a_3^2 (e_{i+1}^k)^2 + 8(a_1 a_2 + a_2 a_3 e_{i+1}^k) + 2a_1 a_3 \right), \end{aligned}$$

since $\text{Var}(E_i) = 1, \text{Var}(E_i^2) = 20$ and $\text{Cov}(E_i, E_i^2) = \mathbb{E}[E_i^3] - \mathbb{E}[E_i^2]\mathbb{E}[E_i] = 4$. With the approximation

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} (e_{i+1}^k)^2 = \lim_{k \rightarrow \infty} \int_{\frac{1}{k}}^1 \left(\int_v^1 \frac{1}{u} du \right)^2 dv = 2$$

follows that $\sum_{i=1}^k \text{Var}(X_{i,k}) = a_1^2 + 20a_2^2 + 2a_3^2 + 8a_1 a_2 + 8a_2 a_3 + 2a_1 a_3$. Condition 3) holds with

$$\Gamma(k) = \frac{1}{k\sqrt{k}} \sum_{i=1}^k \mathbb{E} \left[\left| (a_1 + a_3 e_i^k)(E_i - 1) + a_2(E_i^2 - 2) \right|^3 \right] = \frac{c}{\sqrt{k}} \rightarrow 0,$$

as $k \rightarrow \infty$ and for a constant $c > 0$, since the exponential distribution has finite moments and $\sum_{i=1}^k (e_{i+1}^k)^3/k \leq 6$. Thus, we obtain that

$$\sum_{i=1}^k X_{i,k} \xrightarrow{\mathcal{D}} \mathcal{N}(0, a_1^2 + 20a_2^2 + 2a_3^2 + 8a_1a_2 + 8a_2a_3 + 2a_1a_3).$$

This is the limiting distribution of all the linear combinations $(a_1 P_{k,n} + a_2 Q_{k,n} + a_3 R_{k,n})$ and gives us the joint normal distribution. \square

Lemma 2 *Let X_1, \dots, X_n be i.i.d. random variables with distribution function $F \in \text{DoA}(G_\gamma)$ with $\gamma > 0$ and T_1, T_2, \dots be i.i.d. random variables with distribution function $1 - 1/y$. We define*

$$\hat{\gamma}_k := \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}} \right), \quad M_n := \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}} \right)^2,$$

$$\overline{YE} := \frac{1}{k} \sum_{i=1}^k \log \left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}} \right) e_i^k, \quad \text{and} \quad e_i^k := \sum_{l=i}^k \frac{1}{l}.$$

If the second order condition in Eq. 7 holds for $\rho < 0$ and $x > 0$ and $P_{k,n}, Q_{k,n}$ and $R_{k,n}$ are as defined in Lemma 1, then

$$\hat{\gamma}_k \stackrel{\mathcal{D}}{=} \gamma + \gamma P_{k,n}/\sqrt{k} + \frac{A(T_{(n-k,n)})}{1 - \rho} + o_p(A(n/k)), \tag{24}$$

$$M_n \stackrel{\mathcal{D}}{=} 2\gamma^2 + \gamma^2 Q_{k,n}/\sqrt{k} + \frac{2\gamma(2 - \rho)}{(1 - \rho)^2} A(T_{(n-k,n)}) + o_p(A(n/k)), \tag{25}$$

$$\overline{YE} \stackrel{\mathcal{D}}{=} 2\gamma + \gamma(P_{k,n} + R_{k,n})/\sqrt{k} + \frac{2 - \rho}{(1 - \rho)^2} A(T_{(n-k,n)}) + o_p(A(n/k)), \tag{26}$$

holds for n and k s.t. $1 \leq k \leq n$ and $k \rightarrow \infty$ as $n \rightarrow \infty$.

Proof The results in Eqs. 24 and 25 are already stated in the proof of Theorem 1 in de Haan and Peng (1998).

To prove (26) we follow the proof of the asymptotic normality of the Hill estimator in de Haan and Ferreira (2006). Let A_0 be such that $A(t)/A_0(t) \rightarrow 1$, as $t \rightarrow \infty$. Then, for each $\epsilon > 0$ there exists a t_0 such that for $t \geq t_0$ and $x \geq 1$ the inequality in Theorem B.2.18 in de Haan and Ferreira (2006) holds. For $t = T_{n-k,n}$ and $x = T_{(n-i,n)}/T_{(n-k,n)}$ we obtain that

$$\begin{aligned} \overline{YE} \stackrel{\mathcal{D}}{=} & \frac{\gamma}{k} \sum_{i=1}^k \log \left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}} \right) e_i^k + A_0(T_{(n-k,n)}) \frac{1}{k} \sum_{i=1}^k \frac{\left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}} \right)^\rho - 1}{\rho} e_i^k \\ & + o_p(1) |A_0(T_{(n-k,n)})| \frac{1}{k} \sum_{i=1}^k \left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}} \right)^{\rho+\epsilon} e_i^k. \end{aligned}$$

The second term can be approximated by

$$\lim_{n \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \frac{\left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}}\right)^\rho - 1}{\rho} e_i^k = \int_0^1 \frac{v^{-\rho} - 1}{\rho} \int_v^1 \frac{1}{u} du dv = \frac{2 - \rho}{(1 - \rho)^2},$$

and for the third term holds

$$\lim_{n \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}}\right)^{\rho+\epsilon} e_i^k = \int_0^1 v^{-\rho-\epsilon} \int_v^1 \frac{1}{u} du dv = \frac{1}{(1 - \rho - \epsilon)^2}.$$

Note that for E_1, \dots, E_n i.i.d. standard exponential random variables follows by Rényi's representation that

$$\begin{aligned} \sum_{i=1}^k \log \left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}}\right) e_i^k &\stackrel{\mathcal{D}}{=} \sum_{i=1}^k e_i^k \sum_{j=i}^k \frac{E_j}{j} = \sum_{i=1}^k \frac{E_i}{i} \sum_{j=1}^i e_j^k \\ &= \sum_{i=1}^k \frac{E_i}{i} \left(i e_{i+1}^k + \sum_{j=1}^i \sum_{l=j}^i \frac{1}{l} \right) = \sum_{i=1}^k \frac{E_i}{i} \left(i + i e_{i+1}^k \right) = \sum_{i=1}^k E_i + \sum_{i=1}^k E_i e_{i+1}^k. \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\gamma}{k} \sum_{i=1}^k \log \left(\frac{T_{(n-i+1,n)}}{T_{(n-k,n)}}\right) e_i^k &\stackrel{\mathcal{D}}{=} 2\gamma + \gamma \left(\left(\frac{1}{k} \sum_{i=1}^k E_i - 1 \right) + \left(\frac{1}{k} \sum_{i=1}^k E_i e_{i+1}^k - 1 \right) \right) \\ &= 2\gamma + \gamma (P_{k,n} + R_{k,n}) / \sqrt{k}. \end{aligned}$$

Combining the above arguments as in the proof of Theorem 3.2.5 in de Haan and Ferreira (2006) gives (26). □

Lemma 3 *If $k \rightarrow \infty$, $k/n \rightarrow 0$ and $k/K \rightarrow c$ with $0 < c < 1$, as $n \rightarrow \infty$, then*

$$\lim_{n \rightarrow \infty} \text{Cov}(R_{K,n}, R_{k,n}) = \frac{2c - c \log(c)}{\sqrt{c}} \text{ and } \lim_{n \rightarrow \infty} \text{Cov}(R_{K,n}, P_{k,n}) = \frac{c - c \log(c)}{\sqrt{c}},$$

where $R_{k,n}$ and $P_{k,n}$ are defined in Lemma 1.

Proof Let E_1, E_2, \dots be i.i.d. standard exponential random variables, where $\text{Cov}(E_i, E_j)$ is equal to 1 if $i = j$ and 0 otherwise. Then

$$\lim_{k \rightarrow \infty} \text{Cov}(R_{K,n}, R_{k,n}) = \lim_{k \rightarrow \infty} \text{Cov} \left(\sum_{i=1}^k E_i \frac{e_{i+1}^k}{\sqrt{k}}, \sum_{i=1}^K E_i \frac{e_{i+1}^K}{\sqrt{K}} \right)$$

$$\begin{aligned}
 &= \lim_{k \rightarrow \infty} \sum_{i=1}^k \sum_{j=1}^K \frac{e_{i+1}^k}{\sqrt{k}} \frac{e_{j+1}^K}{\sqrt{K}} \text{Cov}(E_i, E_j) \\
 &= \lim_{k \rightarrow \infty} \frac{\sqrt{k}}{\sqrt{K}} \frac{1}{k} \sum_{i=1}^k \left(\frac{1}{k} \sum_{l=i+1}^k \frac{1}{l/k} \right) \left(\frac{1}{K} \sum_{l=i+1}^K \frac{1}{l/K} \right) \\
 &= \lim_{k \rightarrow \infty} \frac{\sqrt{k}}{\sqrt{K}} \frac{1}{k} \sum_{i=1}^k \left(\int_{(i+1)/k}^1 \frac{1}{u} du \right) \left(\int_{(i+1)/K}^1 \frac{1}{u} du \right) \\
 &= \sqrt{c} \int_0^1 \log(v)^2 - \log(v) \log(c) dv = 2\sqrt{c} - \sqrt{c} \log(c).
 \end{aligned}$$

In the same way we obtain

$$\begin{aligned}
 \lim_{k \rightarrow \infty} \text{Cov}(R_{K,n}, P_{k,n}) &= \lim_{k \rightarrow \infty} \sum_{i=1}^k \sum_{j=1}^K \frac{1}{\sqrt{k}} \frac{e_{j+1}^K}{\sqrt{K}} \text{Cov}(E_i, E_j) \\
 &= \lim_{k \rightarrow \infty} \frac{\sqrt{k}}{\sqrt{K}} \frac{1}{k} \sum_{i=1}^k \left(\int_{(i+1)/K}^1 \frac{1}{u} du \right) \\
 &= \lim_{k \rightarrow \infty} \frac{\sqrt{k}}{\sqrt{K}} \int_0^1 \left(\int_{cv}^1 \frac{1}{u} du \right) dv = \sqrt{c} - \sqrt{c} \log(c).
 \end{aligned}$$

□

Theorem 4 Let X_1, \dots, X_n be i.i.d. random variables with distribution function $F \in \text{DoA}(G_\gamma)$, $\gamma > 0$. If the second order condition in Eq. 7 holds and $k \rightarrow \infty$, $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) \rightarrow \lambda$, as $n \rightarrow \infty$,

$$\sqrt{k}(\bar{\gamma}_k - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{(1-\rho)^2}, 2\gamma^2\right).$$

Proof First we have to rewrite the average over the Hill estimator,

$$\begin{aligned}
 \bar{\gamma}_k &= \frac{1}{k} \sum_{i=1}^k \hat{\gamma}_k = \frac{1}{k} \sum_{i=1}^k \frac{1}{i} \sum_{j=1}^i \log\left(\frac{X_{(n-j+1,n)}}{X_{(n-i,n)}}\right) \\
 &= \frac{1}{k} \sum_{i=1}^k \log X_{(n-i+1,n)} \sum_{j=i}^k \frac{1}{j} - \frac{1}{k} \sum_{i=1}^k \log X_{(n-i,n)} \\
 &= \frac{1}{k} \sum_{i=1}^k \log\left(\frac{X_{(n-i+1,n)}}{X_{(n-k,n)}}\right) \sum_{j=i}^k \frac{1}{j} - \frac{1}{k} \sum_{i=1}^k \log\left(\frac{X_{(n-i,n)}}{X_{(n-k,n)}}\right) \\
 &= \overline{YE} - \hat{\gamma}_k + \frac{1}{k} \log\left(\frac{X_{(n,n)}}{X_{(n-k,n)}}\right),
 \end{aligned}$$

where \overline{YE} is defined in Lemma 2. Following the proof of Lemma 2 it holds that the last term above is in distribution equal to

$$\frac{\gamma}{k} \log \left(\frac{T_{(n,n)}}{T_{(n-k,n)}} \right) + \frac{A(T_{(n-k,n)})}{k} \frac{\left(\frac{T_{(n,n)}}{T_{(n-k,n)}} \right)^\rho - 1}{\rho} + o_p(1) \frac{|A(T_{(n-k,n)})|}{k} \left(\frac{T_{(n,n)}}{T_{(n-k,n)}} \right)^{\rho+\epsilon},$$

where $T_{(i,n)}$ is the i -th order statistic of T_1, \dots, T_n i.i.d. random variables with distribution function $1 - 1/y$. From Corollary 2.2.2 in de Haan and Ferreira (2006) it follows that, as $n \rightarrow \infty$, $\frac{k}{n} T_{(n-k,n)} \xrightarrow{\mathbb{P}} 1$ and $\frac{T_{(n,n)}}{kT_{(n-k,n)}} \xrightarrow{\mathbb{P}} 1$. Thus, $(\log(X_{(n,n)}) - \log(X_{(n-k,n)}))/k = O_p(\log(k)/k) + O_p(A(n/k)/k)$, and by Lemma 2 follows

$$\tilde{\gamma}_k \stackrel{\mathcal{D}}{=} \gamma + \gamma R_{k,n}/\sqrt{k} + \frac{A(n/k)}{(1-\rho)^2} + o_p(A(n/k)) + O_p(\log(k)/k). \quad \square$$

Table 2 Efficiency values EFF_γ , based on 2000 samples if $n = 500$ and on 500 samples if $n = 5000$

$n = 150$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.32	3.04	1.50	2.05	1.09	1.05	1.30	2.40	2.63	1.83
Fréchet(2)	1.14	1.11	1.10	1.20	1.55	1.53	1.74	1.37	1.51	1.07
Cauchy	1.31	1.10	1.16	1.16	1.84	1.79	1.79	1.30	1.52	1.13
Loggamma	1.00	1.16	1.30	1.25	1.14	1.10	1.34	1.41	1.29	0.76
Burr(2,1)	1.17	1.47	1.33	1.23	1.58	1.56	1.69	1.47	1.99	1.01
negative Bias	1.06	1.12	1.21	1.15	1.54	1.18	1.78	1.34	1.36	1.02
$n = 500$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.35	2.20	1.50	1.86	1.29	1.05	1.13	2.36	3.98	1.55
Fréchet(2)	1.14	1.21	1.12	1.19	1.29	1.48	1.90	1.44	1.93	1.24
Cauchy	1.29	1.13	1.20	1.17	1.46	1.73	2.01	1.32	1.84	1.41
Loggamma	1.07	1.19	1.46	1.29	1.18	1.05	1.50	1.48	1.50	0.72
Burr(2,1)	1.13	1.51	1.31	1.25	1.43	1.47	1.94	1.50	2.69	1.11
negative Bias	1.03	1.16	1.61	1.18	1.16	1.12	1.83	1.36	1.64	1.06
$n = 5000$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.58	2.23	1.85	1.93	1.44	1.10	1.60	2.42	4.67	1.33
Fréchet(2)	1.09	1.26	1.09	1.30	1.38	1.34	1.91	1.53	2.77	1.62
Cauchy	1.20	1.13	1.00	1.08	1.20	1.39	1.72	1.22	2.60	2.16
Loggamma	1.23	1.29	1.54	1.35	1.43	1.17	1.54	1.58	1.93	0.68
Burr(2,1)	1.03	1.39	1.29	1.18	1.28	1.23	1.81	1.56	3.65	1.27
negative Bias	1.00	1.19	1.83	1.13	1.57	0.98	1.75	1.43	2.83	1.09

Lowest (best) efficiency values are highlighted in blue

Table 3 Efficiency values EFF_q for $p = 0.001$ based on 2000 samples if $n = 500$ and on 500 samples if $n = 5000$

$n = 150$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.09	4.32	1.19	1.79	1.14	1.06	1.04	2.60	2.85	1.48
Fréchet(2)	1.09	1.06	1.07	1.14	1.43	1.29	1.48	1.32	1.35	1.02
Cauchy	1.09	1.05	1.04	1.07	1.27	1.23	1.23	1.18	1.22	0.99
Loggamma	1.04	1.18	1.54	1.34	1.16	1.11	1.24	2.02	1.39	0.88
Burr(2,1)	0.87	3.02	1.08	1.28	0.89	0.90	0.91	2.72	5.50	0.84
negative Bias	1.07	1.03	1.01	1.06	1.15	1.24	1.26	1.10	1.13	1.19
$n = 500$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.09	2.03	1.15	1.45	1.29	1.05	1.11	2.32	7.15	1.13
Fréchet(2)	1.10	1.13	1.11	1.17	1.33	1.32	1.77	1.40	1.81	1.15
Cauchy	1.20	1.15	1.17	1.17	1.40	1.39	1.66	1.31	1.61	1.19
Loggamma	1.06	1.21	1.85	1.34	1.22	1.10	1.61	2.07	1.97	0.88
Burr(2,1)	1.00	1.64	1.17	1.14	1.10	1.07	1.21	2.05	8.49	0.93
negative Bias	1.04	1.02	1.07	1.08	0.96	1.17	1.33	1.11	1.21	1.25
$n = 5000$	SAM	GH	DK	GO	DB	B	RT	C	NC	sIHS
Student-t(6)	1.07	1.40	1.11	1.16	1.23	1.00	1.20	1.72	7.87	0.97
Fréchet(2)	1.07	1.10	1.12	1.21	1.27	1.22	1.70	1.44	2.81	1.23
Cauchy	1.11	1.14	1.01	1.06	1.22	1.27	1.61	1.19	2.44	1.42
Loggamma	1.11	1.29	2.09	1.42	1.63	1.08	2.04	2.20	4.38	0.92
Burr(2,1)	1.00	1.36	1.31	1.14	1.21	1.10	1.71	1.66	7.68	1.00
negative Bias	0.93	1.02	1.16	1.01	1.12	1.07	1.22	1.00	2.48	1.66

Lowest (best) efficiency values are highlighted in blue

Table 4 Efficiency values EFF_q for $p = 0.0001$ based on 2000 samples and $n = 500$

$n = 500$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.18	2.89	1.36	2.01	1.32	1.08	1.12	3.27	10.58	1.45
Fréchet(2)	1.09	1.15	1.10	1.15	1.32	1.30	1.83	1.44	1.89	1.11
Cauchy	1.14	1.10	1.11	1.12	1.35	1.32	1.57	1.29	1.67	1.10
Loggamma	1.09	1.50	2.81	1.87	1.30	1.08	2.00	3.12	3.02	0.82
Burr(2,1)	0.80	2.62	1.46	1.43	0.83	0.83	1.18	3.37	21.03	0.76
negative Bias	1.01	1.03	1.12	1.08	1.20	1.14	1.31	1.13	1.22	1.21

Lowest (best) efficiency values are highlighted in blue

Theorem 5 Under the conditions of Theorem 4 and $k/K \rightarrow c$ with $0 < c < 1$ as $n \rightarrow \infty$, it holds that

$$\sqrt{k}(\bar{\gamma}_{k,K} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{(1-\rho)^2} \frac{c^\rho - c}{1-c}, \frac{2\gamma^2 c}{1-c} \left(1 + \frac{c \log(c)}{1-c}\right)\right).$$

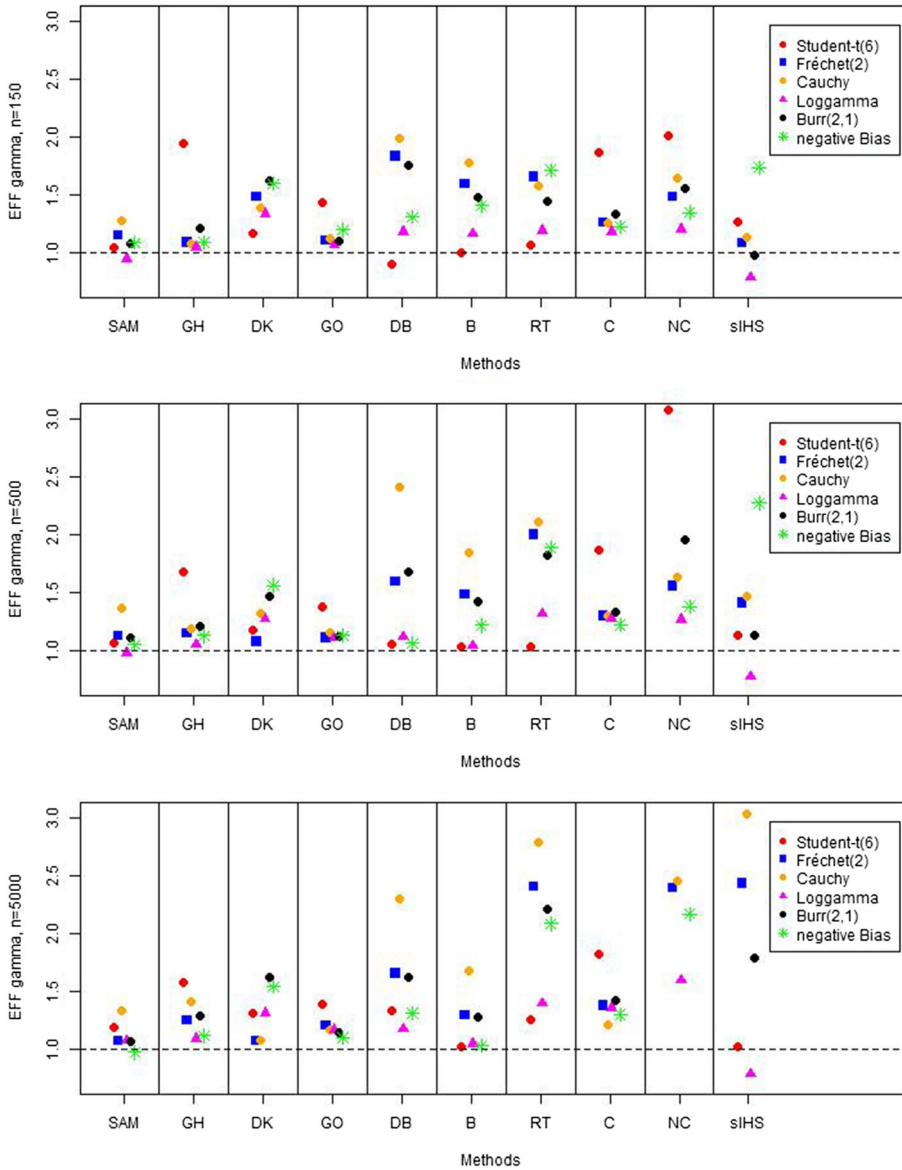


Fig. 10 Efficiency values $EFF_{mean, \gamma}$ based on 2000 samples if $n = 150$ and $n = 500$ and on 500 samples if $n = 5000$

Proof We can rewrite $\bar{\gamma}_{k,K} = \frac{K}{K-k+1}\bar{\gamma}_K - \frac{k}{K-k+1}\bar{\gamma}_k$ and apply Theorem 4,

$$\begin{aligned} \sqrt{k}(\bar{\gamma}_{k,K} - \gamma) &\stackrel{D}{=} \gamma \frac{K}{K-k+1} \frac{\sqrt{k}}{\sqrt{K}} R_{K,n} - \gamma \frac{k}{K-k+1} R_{k,n} \\ &\quad + \frac{K}{K-k+1} \frac{\sqrt{k}A(n/K)}{(1-\rho)^2} - \frac{k}{K-k+1} \frac{\sqrt{k}A(n/k)}{(1-\rho)^2} + o_p(1). \end{aligned}$$

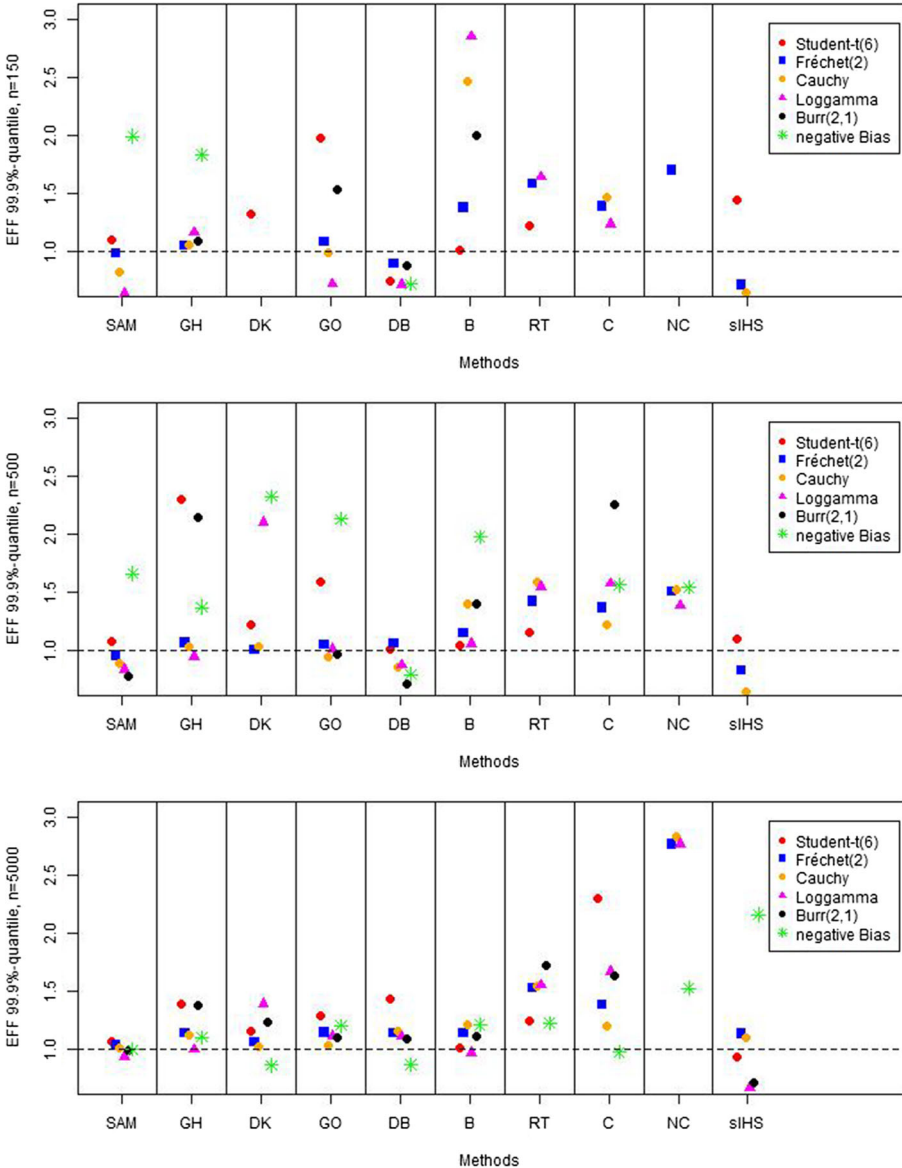


Fig. 11 Efficiency values EFF_{mean_q} for $p = 0.001$ based on 2000 samples if $n = 150$ and $n = 500$ and on 500 samples if $n = 5000$

Now we need $k/K \rightarrow c$ and the covariance between $R_{n,k}$ and $R_{n,K}$, see Lemma 3, to obtain the limit of the variance,

$$\begin{aligned} & \lim_{k \rightarrow \infty} \text{Var} \left(\gamma \frac{K}{K-k+1} \frac{\sqrt{k}}{\sqrt{K}} R_{K,n} - \gamma \frac{k}{K-k+1} R_{k,n} \right) \\ &= \lim_{k \rightarrow \infty} \gamma^2 \left(\left(\frac{\sqrt{c}}{1-c} \right)^2 \text{Var}(R_{K,n}) + \left(\frac{c}{1-c} \right)^2 \text{Var}(R_{k,n}) - \frac{2c\sqrt{c}}{(1-c)^2} \text{Cov}(R_{n,K}, R_{n,k}) \right) \\ &= \gamma^2 \frac{2c + 2c^2 - 4c^2 + 2c^2 \log(c)}{(1-c)^2} = \frac{2\gamma^2 c}{1-c} \left(1 + \frac{c \log(c)}{1-c} \right). \end{aligned}$$

The statement follows by applying this property of regular varying functions,

$$\sqrt{k}A(n/K) = \sqrt{k}A(n/k) \frac{A((k/K)n/k)}{A(n/k)} \rightarrow \lambda c^\rho, \text{ as } n \rightarrow \infty. \quad \square$$

Table 5 Efficiency values EFFmean_γ based on 2000 samples if $n = 500$ and on 500 samples if $n = 5000$

$n = 150$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.05	1.95	1.18	1.44	0.91	1.01	1.07	1.86	2.01	1.27
Fréchet(2)	1.15	1.09	1.48	1.11	1.83	1.60	1.66	1.26	1.49	1.09
Cauchy	1.28	1.08	1.39	1.13	1.99	1.78	1.58	1.26	1.65	1.13
Loggamma	0.95	1.05	1.34	1.07	1.18	1.17	1.20	1.18	1.21	0.79
Burr(2,1)	1.09	1.22	1.62	1.11	1.75	1.48	1.44	1.34	1.55	0.98
negative Bias	1.09	1.09	1.60	1.20	1.31	1.41	1.71	1.23	1.35	1.74
$n = 500$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.07	1.68	1.18	1.38	1.06	1.04	1.04	1.86	3.08	1.14
Fréchet(2)	1.13	1.15	1.08	1.12	1.60	1.49	2.00	1.30	1.56	1.41
Cauchy	1.37	1.19	1.32	1.16	2.41	1.85	2.11	1.30	1.63	1.47
Loggamma	0.98	1.06	1.27	1.11	1.12	1.04	1.32	1.28	1.27	0.78
Burr(2,1)	1.11	1.22	1.47	1.13	1.68	1.42	1.82	1.34	1.96	1.14
negative Bias	1.06	1.13	1.56	1.13	1.07	1.22	1.89	1.23	1.38	2.27
$n = 5000$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.20	1.58	1.31	1.39	1.34	1.03	1.26	1.83	4.86	1.03
Fréchet(2)	1.08	1.26	1.07	1.21	1.66	1.29	2.40	1.38	2.40	2.43
Cauchy	1.34	1.41	1.08	1.17	2.30	1.68	2.78	1.22	2.45	3.03
Loggamma	1.08	1.10	1.32	1.17	1.18	1.05	1.40	1.36	1.60	0.79
Burr(2,1)	1.07	1.29	1.62	1.14	1.63	1.29	2.21	1.43	3.72	1.79
negative Bias	0.98	1.12	1.54	1.10	1.31	1.04	2.08	1.30	2.17	3.98

Lowest (best) efficiency values are highlighted in blue

Proof of Theorem 3 To analyse the bias estimator $\hat{b}_{k,K} = \bar{\gamma}_{k,K} - \bar{\gamma}_K$, we utilize the Theorems 4 and 5. Following their proofs it holds that

$$\sqrt{k} \hat{b}_{k,K} \stackrel{D}{=} \gamma \frac{k-1}{K-k+1} \frac{\sqrt{k}}{\sqrt{K}} R_{K,n} - \gamma \frac{k}{K-k+1} R_{k,n} + \frac{k}{K-k+1} \frac{\sqrt{k}(A(n/K) - A(n/k))}{(1-\rho)^2} + \sqrt{k} (o_p(A(n/K)) + o_p(A(n/k))).$$

Here the random variable $R_{k,n}$ is defined in Lemma 1 and we know that $R_{k,n}$ has a normal limit distribution. Following the proof of Theorem 5 we can calculate the variance and bias, which is

$$\frac{k}{K-k+1} \frac{\sqrt{k}(A(n/K) - A(n/k))}{(1-\rho)^2} \rightarrow \frac{\lambda}{(1-\rho)^2} \frac{c(c^\rho - 1)}{1-c},$$

as $n \rightarrow \infty$, which follows from the regular variation of A and due to $\sqrt{k}A(n/k) \rightarrow \lambda$. Since

$$\sqrt{k} (o_p(A(n/K)) + o_p(A(n/k))) = o_p(1),$$

Table 6 Efficiency values EFF_{mean_q} for $p = 0.001$ based on 2000 samples if $n = 150$ and $n = 500$ and on 500 samples if $n = 5000$

$n = 150$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.11	3.45	1.33	1.98	0.75	1.02	1.22	7.96	5.27	1.45
Fréchet(2)	0.99	1.05	9.26	1.08	0.90	1.38	1.59	1.39	1.70	0.72
Cauchy	0.82	1.06	24.5	1.00	3.57	2.47	3.24	1.47	4.68	0.65
Loggamma	0.64	1.16	111.4	0.72	0.72	2.85	1.64	1.24	5.69	0.38
Burr(2,1)	0.47	1.10	120940	1.53	0.89	2.00	3.23	8.45	31.5	0.31
negative Bias	1.99	1.83	244.4	14.3	0.73	5.58	4.72	7.39	7.57	39.3
$n = 500$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.09	2.30	1.23	1.60	1.01	1.04	1.16	3.44	9.94	1.10
Fréchet(2)	0.96	1.07	1.01	1.06	1.07	1.16	1.42	1.37	1.51	0.83
Cauchy	0.89	1.04	1.03	0.95	0.86	1.40	1.59	1.22	1.52	0.65
Loggamma	0.84	0.95	2.10	1.02	0.88	1.06	1.55	1.57	1.39	0.50
Burr(2,1)	0.79	2.15	8.60	0.98	0.71	1.40	3.19	2.25	5.87	0.41
negative Bias	1.66	1.37	2.32	2.13	0.80	1.98	3.75	1.56	1.54	8.05
$n = 5000$	SAM	GH	DK	GO	DB	B	RT	CL	NC	sIHS
Student-t(6)	1.07	1.39	1.16	1.29	1.44	1.02	1.24	2.30	24.41	0.94
Fréchet(2)	1.04	1.14	1.07	1.15	1.14	1.14	1.53	1.39	2.76	1.14
Cauchy	1.01	1.13	1.03	1.04	1.16	1.22	1.55	1.20	2.83	1.10
Loggamma	0.94	1.00	1.39	1.12	1.11	0.97	1.55	1.67	2.77	0.67
Burr(2,1)	1.00	1.38	1.24	1.11	1.09	1.11	1.72	1.63	22.45	0.71
negative Bias	1.00	1.11	0.87	1.21	0.87	1.21	1.23	0.98	1.53	2.16

Lowest (best) efficiency values are highlighted in blue

the statement of the theorem follows. □

B Additional simulation results

This section provides in Tables 2 and 3 the exact efficiency values, which were presented graphically in the simulation study in Section 4. Further, the efficiency values for quantile estimation with $p = 0.0001$ and $n = 500$ are presented in Table 4. Additionally, efficiency values employing the mean instead of the median are presented for comparison. The mean efficiency values are defined as

$$EFF_{\text{mean}_\gamma}(\hat{k}) := \sqrt{\frac{\mathbb{E}[(\hat{\gamma}_{\hat{k}} - \gamma)^2]}{\mathbb{E}[(\hat{\gamma}_{k_{\text{opt}}} - \gamma)^2]}}$$

and can be found in Figs. 10 and 11 and Tables 5 and 6.

C Additional illustrations

In this section, we provide additional graphical illustrations for the dataset of financial losses studied in Section 5. Figure 12 visualizes the Pareto fit of the rescaled losses locally for three distinct time periods of four month each. These plots support the assumption that the data follows a heavy-tailed distribution.

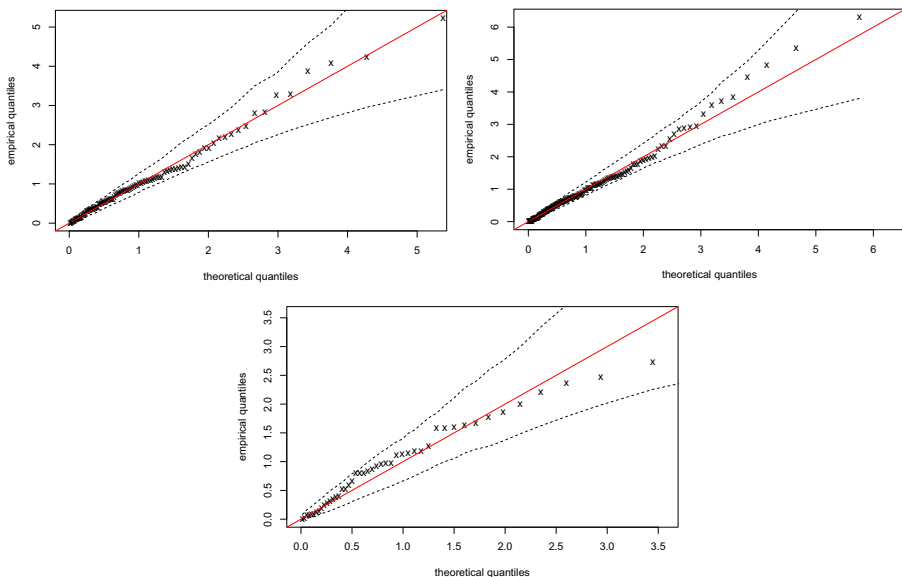


Fig. 12 Pareto QQ-plots with 95%-confidence intervals for the dataset of CPBP losses from Section 5 at three distinct four month time windows centered around 12/2007 (left), 11/2009 (right) and 09/2012 (below)

Acknowledgements This work has been prepared while all three authors were employed at the Georg-August University of Göttingen. Support of the DFG's (Deutsche Forschungsgemeinschaft) Research Training Group 2088 (B4) is gratefully acknowledged. We are grateful to Julien Hambuckers for providing us the dataset of operational losses from UniCredit. We acknowledge support by the Open Access Publication Funds of the Göttingen University.

Funding Open Access funding enabled and organized by Projekt DEAL.

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/>.

References

- Bader, B., Yan, J., Zhang, X.: Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Ann. Appl. Stat.* **12**, 310–329 (2018)
- Beirlant, J., Dierckx, G., Guillou, A., Stărică, C.: On exponential representations of log-spacings of extreme order statistics. *Extremes* **5**, 157–180 (2002)
- Beirlant, J., Kijko, A., Reynkens, T., Einmahl, J.H.J.: Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. *Nat. Hazards* **169**, 1–23 (2018)
- Caeiro, F., Gomes, M.I.: A semi-parametric estimator of a shape second-order parameter. In: *Studies in Theoretical and Applied Statistics*, pp. 137–144. Springer (2014)
- Carreau, J., Naveau, P., Neppel, L.: Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resour. Res.* **53**, 4407–4426 (2017)
- Chavez-Demoulin, V., Embrechts, P., Hofert, M.: An extreme value approach for modeling operational risk losses depending on covariates. *J. Risk Insur.* **83**, 735–776 (2016)
- Chung, K.L.: *A course in probability theory*. Academic Press (1974)
- Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. *SIAM Rev.*, 661–703 (2009)
- Cope, E.W., Piche, M.T., Walter, J.S.: Macroenvironmental determinants of operational loss severity. *J. Bank. Finance* **36**, 1362–1380 (2012)
- Danielsson, J., de Haan, L., Peng, L., de Vries, C.G.: Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* **76**, 226–248 (2001)
- de Haan, L., Peng, L.: Comparison of tail index estimators. *Stat. Neerl.* **52**, 60–70 (1998)
- de Haan, L., Ferreira, A.: *Extreme value theory - an introduction*. Springer (2006)
- de Haan, L., Zhou, C.: Trends in extreme value indices. *JASA* **0**(0), 1–15 (2020)
- Dey, D.Y., Yan, J.: *Extreme value modeling and risk analysis*. Taylor and Francis Group (2016)
- Drees, H., Kaufmann, E.: Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. Appl.* **75**, 149–172 (1998)
- Drees, H., Resnick, S., de Haan, L.: How to make a Hill plot. *Ann. Statist.* **28**, 254–274 (2000)
- Drees, H., Janßen, A., Resnick, S.I., Wang, T.: On a minimum distance procedure for threshold selection in tail analysis. *arXiv:1811.06433* (2018)
- DuMouchel, W.H.: Estimating the stable index α in order to measure tail thickness: A critique. *Ann. Statist.* **11**, 1019–1031 (1983)
- Einmahl, J.H.J., de Haan, L., Zhou, C.: Statistics of heteroscedastic extremes. *J. R. Statist. Soc B* **78**, 31–51 (2016)
- Ferreira, A., de Haan, L., Peng, L.: On optimising the estimation of high quantiles of a probability distribution. *Statistics* **37**, 401–434 (2003)

- Fraga Alves, M.I., Gomes, M.I., de Haan, L., Peng, L.: Semi-parametric estimation of the second order parameter in statistics of extremes. *Portugaliae Math.* **60**, 193–213 (2003)
- Goegebeur, Y., Beirlant, J., de Wet, T.: Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation. *REVSTAT - Stat. J.* **6**, 51–69 (2008)
- Gomes, M.I., Martins, M.J., Neves, M.: Alternatives to a semi-parametric estimator of parameters of rare events - the Jackknife methodology. *Extremes* **3**, 207–229 (2000)
- Gomes, M.I., de Haan, L., Peng, L.: Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes* **5**, 387–414 (2002)
- Gonzalo, J., Olmo, J.: Which extreme values are really extreme? *J. Financ. Economet.* **2**, 349–369 (2004)
- Guillou, A., Hall, P.: A diagnostic for selecting the threshold in extreme value analysis. *J. R. Statist. Soc. B* **63**, 293–305 (2001)
- Hall, P., Weissman, I.: On the estimation of extreme tail probabilities. *Ann. Statist.* **25**, 1311–1326 (1997)
- Hambuckers, J., Groll, A., Kneib, T.: Understanding the economic determinants of the severity of operational losses: a regularized generalized Pareto regression approach. *J. Appl. Economet.* **33**, 898–935 (2018)
- Hill, B.M.: A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174 (1975)
- Kratz, M.F., Resnick, S.I.: The qq-estimator and heavy tails. *Comm. Statist. Stoch. Model.* **12**, 699–724 (1996)
- Krivobokova, T., Kauermann, G.: A note on penalized spline smoothing with correlated errors. *J. Am. Statist. Ass.* **102**, 1328–1337 (2007)
- Lee, Y.K., Mammen, E., Park, B.U.: Bandwidth selection for kernel regression with correlated errors. *Statistics* **44**, 327–340 (2010)
- Moscadelli, M.: The modelling of operational risk: experience with the analysis of the data collected by the basel committee. Technical Report, Bank of Italy – Banking and Finance Department (2004)
- Neves, C., FragaAlves, M.I.: Reiss and Thomas’ automatic selection of the number of extremes. *Comput. Stat. Data Anal.* **47**, 689–704 (2004)
- Northrop, P.J., Coleman, C.L.: Improved threshold diagnostic plots for extreme value analyses. *Extremes* **17**, 289–303 (2014)
- Northrop, P.J., Attalides, N., Jonathan, P.: Crossvalidatory extreme value threshold selection and uncertainty with application to ocean storm severity. *J. R. Stat. Soc. C.* **66**, 93–120 (2017)
- Opsomer, J., Wang, Y., Yang, Y.: Nonparametric regression with correlated errors. *Statist. Sci.* **16**, 134–153 (2001)
- Pickands, J.: Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131 (1975)
- Reiss, R.-D., Thomas, M.: *Statistical analysis of extreme values*. Birkhäuser, Berlin (2007)
- Resnick, S., Stărică, C.: Smoothing the Hill estimator. *Adv. Appl. Probab.* **29**, 271–293 (1997)
- Scarrott, C., MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Stat. J.* **10**, 33–60 (2012)
- Serra, P., Krivobokova, T., Rosales, F.: Adaptive non-parametric estimation of mean and autocovariance in regression with dependent errors. [arXiv:1812.06948](https://arxiv.org/abs/1812.06948) (2018)
- Wadsworth, J.L.: Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics* **58**, 116–126 (2016)

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.