Tail Index Estimation: Quantile-Driven Threshold Selection*

Jon Danielsson
London School of Economics

Lerby M. Ergun
Bank of Canada

Casper G. de Vries
Erasmus University Rotterdam
Tinbergen Institute

October 2024

Working Paper

Abstract

The selection of upper-order statistics in tail estimation is notoriously difficult. This paper advances a data-driven method that minimizes the maximum distance between the fitted Pareto type tail and the observed quantiles, the KS-quantile metric. Based on finite sample properties, the KS-quantile metric outperforms methods based on asymptotic arguments and heuristic approaches. Furthermore, by applying the various methods to existing financial market data applications, we show that the choice of method is economically important.

Bank Topics: Value-at-risk, Econometric and statistical methods.
JEL codes: C01, C14, C58

*Corresponding author Lerby M. Ergun. We would like to thank Jason Allen, Bruno Feunou, Laurens de Haan, Chen Zhou and an anonymous referee from the Bank of Canada for their helpful comments.

1 Introduction

In various research fields, e.g., biology, geography and economics, distributions exhibit heavy tails, e.g., the scaling behaviour described by the power law of Pareto. In this literature, the tail index α is the shape parameter in the power that determines how heavy the tail is (a higher α corresponds to a less heavy tail). The most widely used estimator for the tail index is by Hill (1975), which is the quasi-maximum likelihood estimator. There is an ongoing debate on the number of tail observations k that are used in the estimation of α. The choice of k involves a trade-off between the bias and variance of the estimator. Practitioners often use "Eye-Ball" methods (Resnick and Starica, 1997) to locate an initial series of k for which the estimates are stable. This heuristic rule is often criticized for being arbitrary. The typical approach suggested in the statistical literature is choosing k by minimizing the empirical analogue of the asymptotic mean squared error (AMSE) (Hall, 1990; Danielsson et al., 2001; Drees and Kaufmann, 1998). While these methods are asymptotically consistent, they often perform poorly with finite sample sizes.

The limitations of the currently available methods motivate our new approach. Our method leverages the fact that the tails for a large subset of heavy-tailed distributions, to a first-order expansion, correspond to the tail of a Pareto distribution. With a suitable benchmark distribution in hand, we minimize the distance between the scaled Pareto and empirical distribution. We take inspiration from Clauset et al. (2009) by penalizing deviation based on the Kolmogorov-Smirnov (KS) test statistic, but with a twist. Instead of minimizing the maximum difference between the Pareto and empirical probability distribution, we use the Weissman (1978) quantile estimator to minimize the distance in the quantile dimension. The Weissman (1978) estimator requires an estimate of the shape parameter α, which is typically obtained using the Hill estimator (Hill, 1975) based on the k highest order statistics. By varying k, our method simultaneously fits the scaled pareto to the empirical distribution and elicits k^*. We refer to this measure as the KS-quantile metric.

The choice of using the quantile dimension is supported by the results in De Haan and Ferreira (2007, page 138 and 154). The result shows that for $\alpha > 1$, i.e., when the first moment is bounded, the quantile estimator has a smaller bias and variance by a factor of $1/\alpha$.[1] Furthermore, small deviations in

[1] However, if in the case at hand even the mean is unbounded (i.e., $\alpha < 1$), we recommend using a probability-centric metric.

the probability dimension lead to increasingly large distortions in the quantile dimension. Deep into the tail, the difference in the probability dimension is of order $1/n$, but in the quantile dimension this is of order $(n/k)^{1/\alpha}$ for a distribution which has a bounded mean. This discrepancy becomes more pronounced as one moves further into the tail of the distribution. Therefore, focusing our metric on the quantile domain naturally emphasizes fitting the tail rather than central observations. Moreover, by concentrating on the maximum, the metric further avoids being influenced by the numerous central observations.

The arguments are supported by rigorous simulation analyses and tests. Among other results, we use an acid test, suggested by Drees et al. (2020), to assess the ability of the KS-quantile metric to identify a structural break in the distribution. We simulate tail data from the heavy-tailed Pareto distribution and past a certain threshold centre observations come from the thin-tailed exponential distribution. This acid test evaluates the metric's capability to distinguish between thin-tailed center observations and heavy-tailed tail observations. Methods that use many observations from the exponential distribution, outside of the domain of the Hill estimator, should be used with caution.

In the simulations the KS-quantile metric not only picks approximately 90% of the time a k^* from the heavy-tailed region, but the k^* are stacked against the threshold. There is a visual drop in k^* just beyond the threshold, clearly indicating the location of the structural break in the data. As noted by Drees et al. (2020), these properties are absent when minimizing the distance over the probability dimension. Additionally, among the other suggested methods, the "automated" Eye-Ball method demonstrates similar desirable properties to the KS-quantile metric in the acid test.

In additional results, we simulate from heavy-tailed distributions that conform the so-called Hall (1982) expansion. Hall and Welsh (1985) derives the optimal k^* for this expansion. For distributions such as the Student-t, symmetric stable, Fréchet, and ARCH processes, the competing methodologies fail to reproduce the emerging patterns for k^* as derived by Hall and Welsh (1985), unlike the KS-quantile metric. This discrepancy results in substantial biases in $\hat{\alpha}$ for the theory-based methods. Moreover, in terms of quantile estimation, the KS-quantile metric performs best beyond the 99% probability level, the region that matters most.

Furthermore, we demonstrate that the choice of k^* is impactful and eco-

nomically important. For a large panel of individual firm stock returns, the average absolute difference between estimated tail indexes (from the time series) varies between 0.19 and 6.11 depending on the methodology used. For instance, shifting the Hill estimate from 4 to 3 by using a different methodology suddenly implies that the fourth moment, which captures the variance of the volatility, does not exist. Furthermore, using the KS-quantile metric alters the results of an existing asset pricing application (Kelly and Jiang, 2014) where a fixed threshold is applied.

The paper first introduces the Hill estimator and the KS-quantile metric. This is followed by presenting the results from various simulations in section 3. Section 4 exhibits the estimates for the daily stock return data, followed by concluding remarks.

2 Building the KS-quantile metric

The first part of this section reviews of the main extreme value theory (EVT) results and the Hill estimator. These are the stepping stones for the semi-parametric component of our metric.

2.1 Extreme value theory

Consider a series $X_1, X_2, ..., X_n$ of i.i.d. random variables with cumulative distribution function (cdf) $\mathrm{F}(x)$. The sorted sample, i.e., order statistics, can be represented as

$$\max(X_1, ..., X_n) = X_{n,n} \geq X_{n-1,n} \geq ... \geq X_{1,n} = \min(X_1, ..., X_n).$$

Suppose one is interested in the probability that the maximum is not beyond a certain threshold x. This probability is given by

$$P\{X_{n,n} \leq x\} \overset{\text{iid}}{=} [F(x)]^n.$$

EVT gives conditions under which there exists sequences of norming constants a_n and b_n such that

$$\lim_{n\to\infty} [\mathrm{F}(a_n x + b_n)]^n \to \mathrm{G}(x),$$

where $\mathrm{G}(x)$ is a well-defined non-degenerate cdf. There are three possible $\mathrm{G}(x)$, depending on the tail shape of $\mathrm{F}(x)$. This paper concentrates on distributions that have regularly varying tails,

$$\frac{1-\mathrm{F}(x)}{x^{-\frac{1}{\gamma}} L(x)} = 1, \quad \text{as} \quad x \longrightarrow \infty, \quad \gamma > 0, \tag{1}$$

where L is a slowly varying function, i.e., $\lim_{t\to\infty} L(tx)/L(t) = 1$. Here $1/\gamma = \alpha$ is the index of regular variation, or the tail index. Since α corresponds to the number of bounded moments, results are often discussed in terms of α rather than γ. The limit distribution for the maximum of i.i.d. heavy-tailed observations is the Fréchet distribution (Balkema and De Haan, 1974):

$$\mathrm{G}_{\gamma>0}(x) = e^{-x^{-1/\gamma}}.$$

Note that $\mathrm{G}_{\gamma>0}(x)$ satisfies (1). Hence, the tail behaves approximately as a power function, $x^{-\frac{1}{\gamma}}$. This implies that the distribution for the maximum has a one-to-one relationship with the shape of the tail of $\mathrm{F}(x)$. As a consequence, the entire tail can be utilized for fitting instead of just using maxima; see Mandelbrot (1963) and Balkema and De Haan (1974).

Various estimators for α are proposed in the literature (Pickands, 1975; Hill, 1975; De Haan and Resnick, 1980; Hall, 1982; Mason, 1982; Davis and Resnick, 1984; Csörgo et al., 1985; Hall and Welsh, 1985). Among these, the Hill (1975) estimator is the most widely used for estimating the tail index,

$$\widehat{\gamma} = \frac{1}{\widehat{\alpha}} = \frac{1}{k}\sum_{i=0}^{k-1}\left(\log\left(X_{n-i,n}\right) - \log\left(X_{n-k,n}\right)\right), \tag{2}$$

where k is the number of upper-order statistics used in the estimation of α. Figure 1 shows the reciprocal of the Hill estimates for a sample drawn from a Student-t(4) distribution (where $\alpha = 4$), plotted against an increasing number of order statistics k. The estimate of α varies considerably with k, highlighting the importance of selecting an appropriate k to obtain an accurate estimate.

The pattern observed in Figure 1 can be attributed to the bias and variance of the Hill estimator. For small values of k, the variance of the Hill estimator is relatively high. As k increases, the variance decreases, but the bias begins to dominate. The bias and variance of the estimator can be derived for the subclass of distributions in (1) that satisfy the Hall (1982) expansion[2]

$$1 - \mathrm{F}(x) = Ax^{-1/\gamma}\left[1 + Bx^{-\beta} + o\left(x^{-\beta}\right)\right]. \tag{3}$$

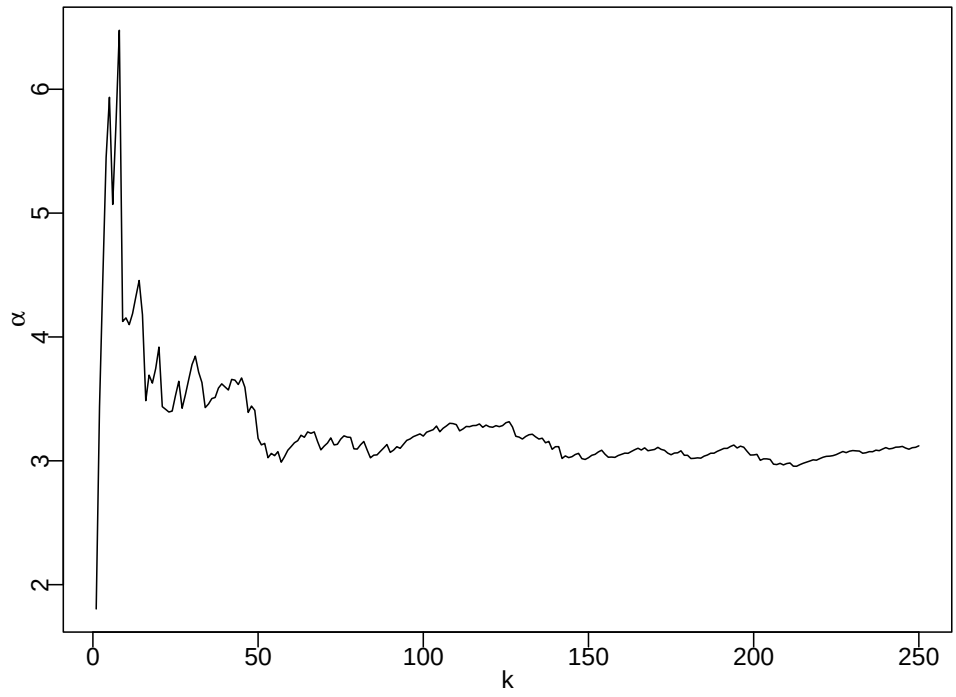
[2]For the Hall expansion $\alpha > 0$, $A > 0$, $\beta > 0$ and B is a real number. Here, A and B are the first and second-order scale parameters, while α and β are the first and second-order shape parameters. Most known heavy-tailed parametric distributions, like the Student-t, symmetric stable, Fréchet distribution and the distribution of the stationary solution to the ARCH process, conform to the Hall expansion. The parameter values A, α, B, and β for these distributions are listed in Table 6 in the Appendix, and in Appendix A.1, we derive the parameters for the stationary distribution of the ARCH process.

Using the Hall expansion one shows the asymptotic bias as

$$A\,\mathrm{E}\left[\frac{1}{\widehat{\alpha}}-\frac{1}{\alpha}\mid X_{n-i,n}>s\right]=\frac{-\beta B s^{-\beta}}{\alpha\left(\alpha+\beta\right)}+o\left(s^{-\beta}\right). \tag{4}$$

Equation (4) describes the relationship between the threshold s and the bias of the Hill estimator. As seen in (4), when s decreases, meaning the threshold moves closer to the center of the distribution, the bias increases

Figure 1: Hill plot for the Student-t (4) distribution



This graph depicts the estimate of α for various levels of k. The sample, consisting of 10,000 observations, is drawn from a Student-t(4) distribution so that $\alpha = 4$. This type of graph is commonly referred to as a Hill plot.

The asymptotic variance of the Hill estimator is[3]

$$A\,\mathrm{var}\left(\frac{1}{\widehat{\alpha}}\right)=\frac{s^{\alpha}}{nA}\frac{1}{\alpha^{2}}+o\left(\frac{s^{\alpha}}{n}\right).$$

The variance is also a function of the threshold s. As s decreases, the variance reduces. However, there is a trade-off between the squared bias and the variance. For large values of s, the bias is small, and the variance dominates.

[3]The expressions for bias and variance are based on the second-order expansion by Hall and Welsh (1985). For further details, see Appendix A.2.

Existing asymptotically consistent approaches, such as those by Danielsson et al. (2001) and Drees and Kaufmann (1998), leverage this trade-off by minimizing the AMSE. However, since these methods are based on asymptotic reasoning, their finite sample properties can be quite different. Methods based on heuristic rules often use the Hill plot to identify a stable region with a small k, graphically balancing bias and variance. Alternatively, many applications adopt a more straightforward approach, using a fixed percentage of the total sample (see, e.g., Van Oordt and Zhou (2016) and Davydov et al. (2021)). Going against this approach is the finding by Hall and Welsh (1985) that the optimal k^*, which minimizes the AMSE, varies across different distributions and exhibits different rates of convergence. For a more detailed exposition of existing methods, see Appendix A.3.

2.2 KS-quantile metric

In this paper we adopt a more data driven approach. We adopt the approach of Bickel and Sakov (2008), which involves matching the tail of the empirical cdf to a semi-parametric distribution. However, motivated by the results in De Haan and Ferreira (2007) and to more accurately fit the most extreme observations, where small differences in probability can lead to significant deviations in quantiles, we measure the distance over the quantile dimension.

2.2.1 The quantile metric

The original KS test statistic is defined as the supremum of the absolute difference between the empirical cdf, $\mathrm{F}_n(x)$ and a parametric cdf, $\mathrm{F}(x)$,

$$\sup_x \left|\mathrm{F}_n(x) - \mathrm{F}(x)\right|.$$

To convert this statistic to a quantile-based metric, we choose $\mathrm{F}(x)$ based on the first-order term from (3):

$$\mathrm{F}(x) = 1 - Ax^{-\alpha}. \tag{5}$$

This function resembles the scaled Pareto distribution when higher-order terms are ignored in (3). In fact, many known heavy-tailed distributions satisfy (3), making it an ideal candidate. The scaled Pareto distribution is unique in that the first-order term holds across the entire support. To obtain the quantile function, invert (5):

$$x = \left(\frac{1-F}{A}\right)^{-1/\alpha}.$$

Next substitute j/n for $1 - \mathrm{F}(x)$, which is the empirical distribution. The scale A is estimated with the Weissman (1978) estimator $\frac{k}{n}\left(X_{n-k+1,n}\right)^{\hat{\alpha}(k)}$, and α with the Hill estimator.[4] This results in the scaled Pareto quantile estimate:

$$q\left(j,k\right) = X_{n-k+1,n}\left(\frac{k}{j}\right)^{1/\hat{\alpha}_k}. \tag{6}$$

Here j denotes that the quantile estimate is measured at probability $(n-j)/n$.

Given the quantile estimator, the empirical quantile and the maximum absolute deviation as the penalty function, we get:

$$Q_T = \inf_{k=2,\ldots,T}\left[\sup_{j=1,\ldots,T}\left|X_{n-j,n} - q\left(j,k\right)\right|\right], \tag{7}$$

where the T^{th} order-statistic determines the region over which the metric is measured. Here, $X_{n-j,n}$ represents the empirical quantile, and $q\left(j,k\right)$ is the quantile estimate from (6). The value of k that minimizes the maximum horizontal deviation along the tail observations up to $X_{n-T,n}$ is chosen as k^* for the Hill estimator.[5]

Our main twist is that the distance is measured in the quantile dimension rather than the probability dimension. There are several reasons for the particular choice of metric. First, from a theoretical perspective, De Haan and Ferreira (2007, pages 138 and 145) provide the following results for the probability estimator:

$$\frac{\sqrt{k}}{\alpha\log\left(k/np_n\right)}\left(\frac{\hat{p}_n}{p_n} - 1\right) \xrightarrow{d} N\left(\frac{sign(B)}{\sqrt{2\alpha\beta}}, \frac{1}{\alpha^2}\right) \tag{8}$$

and similarly for the quantile estimator:

$$\frac{\sqrt{k}}{\log\left(k/np_n\right)}\left(\frac{\hat{x_{p_n}}}{x_{p_n}} - 1\right) \xrightarrow{d} N\left(\frac{sign(B)}{\sqrt{2\alpha\beta}}, \frac{1}{\alpha^2}\right). \tag{9}$$

[4] The estimate of A is obtained by inverting $P = Ax^{-\alpha}$ at threshold $X_{n-k+1,n}$.

[5] In Appendix A.5, we model the KS-quantile metric using a Brownian motion representation. This approach offers two key benefits. First, the selected parameters for a given distribution are not restricted to that distribution, allowing us to examine the KS-quantile metric's properties as a penalty function in a broader context. Second, since the representation is based parameters and normally distributed random variables, it enables simulation across a wide range of distributions using the same random draws. This reduces sampling noise when comparing performance under different parameter settings. Moreover, this representation highlights the challenge of proving asymptotic consistency for the metric.

Note that the probability estimate is scaled by a factor of $1/\alpha$. Consequently, for $\alpha > 1$, the bias and the variance of the probability estimator is larger compared to the quantile estimator. Conversely, for $\alpha \leq 1$, penalty functions based on the probability estimator may exhibit better properties.[6]

The second motivation is more intuitive. For tail observations, small mistakes in probability estimates lead to large deviations in the quantiles, the dimension most applications are ultimately interested in. By focusing on tail quantiles, we naturally emphasize fitting the tail of the distribution. This approach also eliminates the need for additional penalization, such as squaring the differences. To aggregate all the absolute differences along the tail into a single metric, we use the maximum of these absolute distances. Taking the maximum has the advantage of preventing the metric from being diluted by numerous central observations, making the choice of a large T innocuous. In contrast, averaging the differences would not have this benefit. For a comparison of different penalty functions, see Appendix A.4 for simulation analysis with other functions.

Providing a theoretical argument that this approach yields a consistent estimate for α is challenging. De Haan and Ferreira (2007) demonstrate that while the empirical distribution and the extreme value distribution do not converge in probability, they do converge in distribution. Therefore, it is reasonable to expect that statistics based on the difference between the empirical cdf and the tail cdf will exhibit similar behavior. To be cautious, we also examine the higher non-central moments of the simulation estimates.

3 Simulations

To assess the finite sample properties of the KS-quantile metric, we use extensive simulation studies. We highlight two key studies below and relegate additional results to the Appendix. The first simulation excersize evaluates the metric's ability to differentiate between heavy-tailed Pareto and thin-tailed exponential distributions. The second simulation conducts a horse race between competing methods found in the literature.

[6] In Appendix C figures 4 and 5, we simulate the probability and quantile criteria from De Haan and Ferreira (2007, page 138 and 154). The simulations confirm that for $\alpha > 1$, the quantile estimator has smaller bias and variance. This flips for around $\alpha < 1$ where the quantile estimator has smaller bias and variance. This is apparent from equations (8) and (9). Furthermore, for the Pareto distribution the bias is small and converges to zero as more observations are used.

3.1 Upper bound estimation

In the first simulation exercise, we draw samples from an exponential distribution with a Pareto tail beyond a threshold $X_{n-k^*,n}$. A candidate method should avoid selecting a threshold that includes many observations from the thin-tailed exponential part of the distribution. As Drees et al. (2020) points out, locating this smooth transition point is not an easy task. Furthermore, while the Hill estimator is the maximum likelihood estimator for the Pareto distribution and remains unbiased when fewer than k^* order statistics are used, the chosen threshold might slightly exceed k^* as it reduces variance.

Figure 2 depicts the ratio of estimated number of order-statistics with the KS-quantile metric, $\hat{k}$, to the number of observations drawn from the Pareto distribution, k^*. For $\alpha = 1$, the ratio, $\hat{k}/k^*$, is below 1 for 92.63% of the simulations. For α equals 2, 3 and 4 the percentages decrease to 92.00%, 87.83% and 59.09%, respectively. This indicates that, in most samples, the KS-quantile metric selects a threshold from the Pareto part of the tail. Additionally, all four panels reveal a notable concentration of cases just below 1, demonstrating that the KS-quantile metric captures the transition from the Pareto tail to the exponential center of the distribution. Figure 6 in the Appendix presents a similar analysis for the KS test statistic, based on the probability dimension, as the selection criterion. Here, almost all k^* values are chosen from the thin-tailed exponential part. To conclude, it appears that the KS-quantile method passes the acid test devised by Drees et al. (2020).[7]

3.2 Monte Carlo: Comparing existing methods

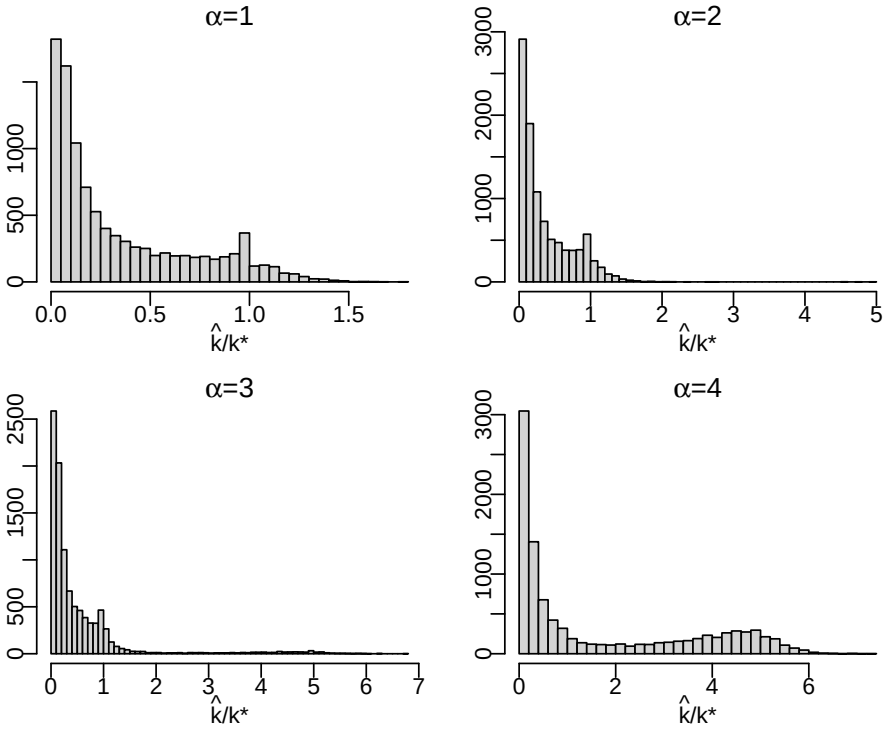
To compare existing methods in Monte Carlo (MC) simulations, we select a wide range of heavy-tailed distributions and processes. The chosen cdfs vary in their tail indices and second-order terms, leading to different rates of convergence as $n \to \infty$, and distinct bias and variance trade-offs.

The Student-t, symmetric stable, Fréchet distributions and the distribution of the stationary solution of the ARCH process all conform to the Hall expansion in (3).[8] Consequently, we know both α and k^*_{TH}, where k^*_{TH} minimizes

[7] In the Appendix (Figures 7, 8 and 9) report simulation results for various methods discussed in the following section. The asymptotically consistent approaches tend to overshoot into the exponential distribution region. The double bootstrap approach, on the other hand, tends to select smaller values for k^* when α is one or two.

[8] The tail index for the Student-t distribution corresponds to the degrees of freedom, for the symmetric stable distribution it is the characteristic exponent, and for the Fréchet

Figure 2: Pareto tail and exponential centre (KS - quantile dimension)



These figures present the performance of the KS-quantile metric in selecting a threshold within the heavy-tailed region of the distribution. The samples are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, for $X \geq x_c$. We set $\lambda = 1$ and $A = exp(x_c)x_c^{\alpha}$. The sample size is 10,000, and the threshold is set at $p(x_c) = 0.99$, so the average number of draws from the Pareto distribution is approximately 100. We use the KS-quantile metric to estimate k^*. The ratio $\hat{k}/k^*$ is used to normalize the choice of k relative to the benchmark k^*. The figures show results for different values of α, as indicated above the histograms. We draw 10,000 samples for each analysis.

the AMSE. Interestingly, α and k^*_{TH} are inversely related within the Student-t, symmetric stable distribution families, and ARCH processes. However, for the Fréchet distribution k^*_{TH} is independent of α. In the case of the ARCH process, the first and second-order tail indices for the tail expansion of the distribution of the stationary solution are known, as well as the second-order scale parameter. The first-order scale parameter, A, is more elusive. Goldie (1991) provides a method to obtain an explicit expression for A when α is

distribution it is the shape parameter. These distribution parameters are henceforth referred to as α. See Appendix A.2 for the derivation of the optimal theoretical asymptotic MSE based threshold.

an integer (see Appendix A.1 for details).

To provide perspective on the relative performance of the KS-quantile metric, we also apply other approaches to determine k^*. We include the double bootstrap method and the approach proposed by Drees and Kaufmann (1998), both of which are based on asymptotic arguments.[9] Furthermore, we use an automated version of the Hill plot 'eye-balling' method and a fixed sample proportion as heuristic rules. These methods are discussed in Appendix A.3. Additionally, we use the KS-probability statistic, motivated by Clauset et al. (2009), as an empirical benchmark. Lastly, estimates of α using the AMSE-minimizing threshold, k^*_{TH}, primarily serve as a theoretical benchmark, since the data-generating process (DGP) is typically unknown in practice.

Table 1 presents the first four non-central moments of the distribution of α estimates, along with the mean of the selected k^* for each method, using samples drawn from the Student-t distribution family. The results in the first set of rows reveal that most methods exhibit a downward bias, with the exception of the double bootstrap method, which consistently shows an upward bias across all degrees of freedom. This upward bias is due to instances where the smallest possible $\hat{k}^*$ is selected, leading to excessively large $\hat{\alpha}$ values. Among the methods with downward bias, this bias increases with the degrees of freedom and is particularly pronounced in the iterative method by Drees and Kaufmann (1998), the KS probability metric, and the fixed 5% threshold method.

The KS-quantile metric, the theoretical AMSE based threshold, and the automated Eye-Ball method produce estimates that are closest to the true tail index of the DGP. Based on these results for the Student-t distribution, we conclude that the KS-quantile metric demonstrates strong bias reduction performance relative to other implementable methods. However, the automated Eye-Ball method performs only marginally worse than the KS-quantile metric.

The fifth panel presents the mean choice of k^* across the different methods. The AMSE-minimizing k^*_{TH}, reported in the third column, serves as a benchmark for the other methods. For the Student-t distribution, the KS-quantile metric closely follows the decline in k^*_{TH} as a function of α. However, the average k^*_{KS} is generally higher than k^*_{TH}. Only the method by Drees and

[9]The double bootstrap method occasionally selects $k = 1$, leading to extremely large outliers in the α estimates. To address this, we exclude $k < 3$ from its domain in the simulations.

Kaufmann (1998) and the double bootstrap method exhibit a similar decreasing pattern, though the average k^*_{Drees} ranges from 6.7% to 10.3% of the sample fraction. The Eye-Ball method, on the other hand, shows an increasing pattern but selects a very low average k^*, resulting in the low bias observed in the first panel.

The second, third, and fourth panels display the respective non-central moments of the estimates. To benchmark the evaluation of these higher moments, the results for k^*_{TH} provide useful guidance. The KS-quantile metric demonstrates a similar pattern to that produced by the theoretical threshold across the second, third, and fourth moments. The KS-test statistic and the method by Drees and Kaufmann (1998) exhibit very low variances. However, as observed in the fifth panel, these methods tend to select high values for k^*, indicating a preference for k^* values that yield low variance but substantial bias. The double bootstrap method shows excessively high values for the higher moments. Due to its slow rate of convergence, the criterion function of the double bootstrap is relatively flat, resulting in significant variability in $\hat{k}^*$.

The simulation results for the symmetric stable, Fréchet distributions, and ARCH processes are presented in the Appendix. Overall, these results follow a pattern similar to those observed for the Student-t distribution: when the bias is relatively small, the higher moments closely align with the theoretical threshold. However, for the ARCH processes, different methods perform better depending on the value of α.

Quantile estimation

For many economic questions, quantile estimates are more relevant than the precise value of the tail index. We therefore compare Pareto quantile estimates from different methods with the simulated order statistics. Figure 10 in the Appendix illustrates the performance of these methods in estimating tail quantiles of the distribution.[10] Estimating quantiles beyond the 99.5% probability is notoriously difficult, and all methods introduce significant errors in this region. However, except for the KS-quantile metric for the Student-t(2) distribution, the KS-quantile metric, automated Eye-Ball method, double bootstrap method, and theoretical threshold generally produce smaller errors in the extreme tail region. Conversely, the method by

[10]The figures present median differences due to a small number of extreme outliers affecting tail quantiles near the center of the distribution. The qualitative results remain consistent when using average differences.

Table 1: Horse race **Student-t** distribution family

	α	KS qua	KS prob	TH	5%	Eye-Ball	Drees	Du bo
$E[\hat{\alpha}]$	2	2.02	1.60	1.92	1.85	1.99	1.70	2.76
	3	2.86	1.96	2.80	2.46	2.84	2.24	4.54
	4	3.53	2.17	3.59	2.87	3.48	2.63	6.18
	5	4.08	2.31	4.30	3.16	3.95	2.92	7.42
	6	4.51	2.40	4.95	3.38	4.29	3.13	9.32
$E[\hat{\alpha}^2]$	2	4.43	2.58	3.71	3.44	4.02	2.88	32.23
	3	8.63	3.84	7.87	6.04	8.14	5.05	176.85
	4	12.97	4.71	13.06	8.24	12.22	6.97	239.24
	5	17.13	5.33	18.86	10.01	15.72	8.55	233.73
	6	20.85	5.78	25.13	11.42	18.55	9.84	2909.77
$E[\hat{\alpha}^3]$	2	10.64	4.14	7.18	6.39	8.24	4.91	2760.66
	3	27.51	7.53	22.32	14.89	23.56	11.40	$8 * 10^4$
	4	49.48	10.23	48.05	23.72	43.18	18.49	$1 * 10^5$
	5	74.04	12.31	84.14	31.74	62.87	25.17	$4 * 10^4$
	6	98.48	13.92	130.58	38.67	80.60	31.09	$1.1 * 10^7$
$E[\hat{\alpha}^4]$	2	28.01	6.65	13.94	11.90	17.14	8.38	$4.6 * 10^5$
	3	92.62	14.77	63.76	36.77	68.87	25.85	$5.9 * 10^7$
	4	196.00	22.24	179.05	68.40	153.70	49.31	$8.7 * 10^7$
	5	328.84	28.46	382.47	100.81	253.05	74.45	$1.3 * 10^7$
	6	475.28	33.53	695.21	131.17	352.29	98.73	$5.1 * 10^{10}$
$E[\hat{k}^*]$	2	500.11	1300.63	281.00	500.00	19.37	1039.09	191.73
	3	339.27	1310.03	132.00	500.00	35.13	841.56	89.65
	4	237.13	1313.79	78.00	500.00	51.85	755.24	53.95
	5	169.82	1316.07	53.00	500.00	68.80	708.72	36.59
	6	133.37	1317.20	40.00	500.00	84.19	680.43	27.43

This table presents the first four non central moments of the distribution of estimated α values and the average k^* selected in the simulations across different methods. The samples are drawn from the Student-t distribution family, with the column labeled α indicating the degrees of freedom for each particular Student-t distribution. The various methods are listed in the first row. 'KS qua' refers to the Kolmogorov-Smirnov metric measured over the quantile dimension, as described in (7). 'KS prob' represents the Kolmogorov-Smirnov metric measured over the probability dimension. 'TH' denotes the theoretically derived optimal k from minimizing the AMSE for specific parametric distributions, as presented in Equation (17) in the Appendix. The 'Automated Eye-Ball' method refers to the heuristic approach aimed at identifying the first stable region in the Hill plot, as outlined in (18). For the column labeled 'Drees,' k^* is determined by the methodology described by Drees and Kaufmann (1998). 'Du bo' refers to the double bootstrap procedure by Danielsson et al. (2001). The sample size is $n = 10,000$, with 10,000 repetitions conducted for each method.

Drees and Kaufmann (1998), the 5% fixed sample fraction, and the KS-test statistic tend to make larger errors in the extreme tail beyond the 99% quantile but perform better for quantiles closer to the distribution center. Similar patterns are observed for the symmetric stable and Fréchet distributions (see Figures 11 and 12 in the Appendix).

Based on the MC simulation analysis, we conclude that both the KS-quantile metric and the automated Eye-Ball method exhibit superior performance compared to other implementable methods. These two methods perform well according to the first four moments of the estimated α. Notably, the KS-quantile metric aligns more closely with the theoretical optimal threshold in terms of k^*. However, the results for quantile estimation are more nuanced. While the KS-quantile metric and the automated Eye-Ball method excel in estimating quantiles deep in the tail of the distribution, they exhibit relatively larger bias closer to the central regions of the distribution.

4 Application: Financial return series

The estimation of tail indexes is important in many fields. Most applications use a single cross-section or time-series in their application. It is hard to tie conclusions to very few estimates. Financial market data are particularly advantageous for evaluating the impact of threshold selection on real-world applications. With their long time series and rich cross section, stock returns provide an excellent opportunity to empirically assess the effects of different threshold choices.

In this context, we apply the methods from the MC horse race to estimate tail indexes for the return distributions of individual U.S. stocks. Additionally, we replicate the tail risk factor analysis conducted by Kelly and Jiang (2014), who employed a 5% threshold level in their asset pricing study. By progressively lowering the threshold and ultimately applying the KS-quantile metric, we examine how these adjustments affect the economic implications of their constructed risk factor.

4.1 Data

The stock market data utilized in this study is sourced from the Centre for Research in Security Prices (CRSP). The CRSP database provides individual stock data from December 31, 1925, to December 31, 2015, covering NYSE, AMEX, NASDAQ, and NYSE Arca exchanges. We analyze data from a total of 17,918 stocks. To ensure data quality and the accuracy of the Hill estimator, we exclude stocks with fewer than 48 months of data and those with an average price below 5 dollars. For the accuracy of Hill estimator typically a large total sample size is required because only a small sample fraction is informative regarding the tail shape properties.

4.2 Empirical impact

To illustrate the impact of different threshold selection methods, we estimate the threshold for each stock using its time-series returns. Table 2 presents the average absolute differences in α estimates between various methods. The observed differences are substantial for both the left and right tails.[11] The KS-quantile metric and the automated Eye-Ball method show notable deviations compared to the method by Drees and Kaufmann (1998) and the KS-test statistic. The 5% fixed threshold method is closest to both. The double bootstrap method exhibits many large outliers, leading to a large average deviation relative to other methods.

Comparing the results from the MC horse race in Table 1 with this financial application reveals noticeable parallels. In both cases, the tail index estimates obtained using the KS-quantile metric and the Eye-Ball method are closely aligned, while both deviate from the iterative method. The double bootstrap method, consistent with the simulation results, is primarily characterized by large outliers. These similarities raise concerns about the applicability of fixed threshold and AMSE inspired approaches for real-world empirical estimation.

4.3 Cross-sectional Hill estimator

Most financial applications estimate the tail index α using time-series data on stock prices, see e.g., Jansen and De Vries (1991). Other applications use a single cross-section to analyze power law behavior in areas such as city size or income distribution, as seen in Gabaix (1999) and Reed (2003). Recent work by Kelly and Jiang (2014) examines the cross-section of individual U.S. stock returns to estimate a tail index conditional on time t. Their findings indicate that stocks more exposed to variations in the tail index demand a risk premium, while those with lower or negative covariance with the tail index offer a hedging opportunity and thus trade at a discount compared to stocks with high exposure. Furthermore, they observe persistence in the time-varying tail index estimates. If investors are averse to increasing tail risk, a positive shock to tail risk serves as a signal of future risk levels, leading to higher demanded compensation (i.e., risk premium). This persistence in tail risk introduces predictability in risk premia based on changes in conditional tail risk.

[11] For descriptive statistics on tail index estimates and thresholds for both tails, refer to Table 10 in the Appendix. Table 11 provides results for median absolute differences, which are similar to the mean results but smaller in magnitude.

Table 2: Mean absolute differences between different methods

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
KS qua	0	0.98	0.76	0.57	0.96	5.35	0	1.09	0.82	0.58	1.05	3.68
KS pr	0.98	0	0.36	0.88	0.22	6.11	1.09	0	0.39	0.92	0.19	4.56
5%	0.76	0.36	0	0.56	0.37	5.79	0.82	0.39	0	0.57	0.38	4.21
Eye-Ball	0.57	0.88	0.56	0	0.85	5.42	0.58	0.92	0.57	0	0.87	3.83
Drees	0.96	0.22	0.37	0.85	0	6.07	1.05	0.19	0.38	0.87	0	4.51
Du bo	5.35	6.11	5.79	5.42	6.07	0	3.68	4.56	4.21	3.83	4.51	0

(a) Estimates $\alpha\left(k_i^*\right)$

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
KS qua	0	238	136	89	216	92	0	243	134	83	227	88
KS pr	238	0	136	301	90	305	243	0	137	303	80	310
5%	136	136	0	167	123	173	134	137	0	167	126	176
Eye-Ball	89	301	167	0	277	45	83	303	167	0	285	47
Drees	216	90	123	277	0	280	227	80	126	285	0	288
Du bo	92	305	173	45	280	0	88	310	176	47	288	0

(b) Estimates k_i^*

This table presents mean absolute differences between $\hat{\alpha}\left(k_i^*\right)$ and k_i^* by applying the six different methods to choose k_i^* for the left and right tail of stock returns. The analysis uses data from the CRSP database, covering individual stock data from December 31, 1925, to December 31, 2015, across the NYSE, AMEX, NASDAQ, and NYSE Arca. The six different methods are the KS-quantile metric, KS test statistic, 5% threshold, automated Eye-Ball method, the iterative method by Drees and Kaufmann (1998) and the double bootstrap by Danielsson et al. (2001). Stocks with $\hat{\alpha} > 1,000$ sing any of the methods are excluded from the analysis. The maximum k^* is cut off at 15% of the total sample size. There are 17,918 stocks included in the analysis.

In this section, we reproduce the original results and explore the impact of selecting an appropriate threshold. The original study collected daily returns for all stocks within a month to estimate the tail index for month t. In Kelly and Jiang (2014), the threshold was set at 5% of the sample.[12] This approach generates a monthly series of γ_t estimates.[13] The covariation of each stock with $\hat{\gamma}_t$ is then measured, and stocks are ranked according to their co-movements. If γ_t is a priced risk factor, then the risk premium on different stocks and the exposure should be positively correlated.

We replicate the cross-sectional analysis and the predictability regressions. Table 3 presents the results for the cross-sectional pricing of tail risk using different thresholds for the γ_t estimates. The first row reproduces the approximately 4% risk premium that Kelly and Jiang (2014) originally identified for stocks with high covariance relative to those with low covariance. In

[12] On average, there are approximately 80,000 daily observations in a month (about 3,800 stocks over 21 trading days). The smallest cross-sectional sample size is around 35,000 at the start of the sample period, while the largest, around the year 2000, reaches about 130,000.

[13] In this part of the analysis, γ is used because an increase in indicates a heavier tail, which corresponds to higher risk.

Table 3: Cross-sectional pricing of tail risk

	Low	2	3	4	High	High-low	t-stat
$\hat{\gamma}_{5\%}$	7.02	8.29	9.33	10.12	11.81	4.78	2.07
$\hat{\gamma}_{2.5\%}$	7.17	8.34	9.15	10.25	11.67	4.50	2.12
$\hat{\gamma}_{1\%}$	8.18	8.33	8.67	10.13	11.26	3.09	1.40
$\hat{\gamma}_{0.5\%}$	8.16	8.58	8.84	9.89	11.11	2.95	1.49
$\hat{\gamma}_{KS}$	8.40	8.67	8.65	9.76	11.11	2.71	1.81

(a) Portfolio returns based on α_t from single-factor model

	Low	2	3	4	High	High-low	t-stat
$\hat{\gamma}_{5\%}$	8.72	8.60	8.71	9.12	11.42	2.70	1.79
$\hat{\gamma}_{2.5\%}$	8.62	8.55	8.48	9.50	11.43	2.81	1.94
$\hat{\gamma}_{1\%}$	9.27	7.90	8.57	9.26	11.58	2.30	1.75
$\hat{\gamma}_{0.5\%}$	9.60	8.28	8.45	9.18	11.06	1.46	1.04
$\hat{\gamma}_{KS}$	9.15	9.03	8.55	9.20	10.64	1.49	1.91

(b) Portfolio returns based on α_t from joint estimation with Fama-French factors

This table presents return statistics for portfolios formed based on their co-movement with cross-sectionally estimated tail indexes, using different thresholds. Each month, stocks are sorted into quintile portfolios according to tail betas, which are estimated from monthly data over the previous ten years. These equally weighted portfolios consist of NYSE/AMEX/NASDAQ common stocks with prices above 5 dollars at the time of portfolio formation. In the first four rows of each panel, the subscripts on γ indicate the sample fraction used to estimate γ_t. The last row uses the KS-quantile metric. Panels (a) and (b) report results where the tail risk beta is estimated using a single-factor model and the Fama-French three-factor model, respectively. The rightmost columns display the results for a high-minus-low zero net investment portfolio, which is long on quintile portfolio five and short on quintile portfolio one, along with the associated t-statistics. The t-statistics are calculated using Newey and West (1987) standard errors with 12 lags.

the second to fourth rows, the threshold is adjusted closer to the tail of the distribution for estimating γ_t. At these lower thresholds, the risk premium decreases by half and becomes insignificant at the 10% confidence level. The risk premium based on $\hat{\alpha}_{KS}$ also decreases by half relative to the estimates using the 5% threshold, but it remains statistically significant.[14]

Table 4 presents the results of predicting the risk premium across different horizons using $\hat{\gamma}_t$. The first row shows the results following the methodology in Kelly and Jiang (2014), while the subsequent rows progressively lower the threshold. As the threshold moves deeper into the tail of the distribution, the predictive power of $\hat{\gamma}_t$ diminishes across the various horizons. Both the R^2 and the t-statistic on the coefficient decrease as the threshold level low-

[14]The results are quantitatively similar when the exposures are jointly estimated with other known risk factors.

Table 4: Predictive regression U.S. stock index different horizons

	1 month			1 year			3 year		
	Coeff	t-stat.	R^2	Coeff	t-stat.	R^2	Coeff	t-stat.	R^2
$\hat{\gamma}_{5\%}$	0.10	2.60	1.02	1.01	2.35	7.17	3.02	2.36	19.90
$\hat{\gamma}_{2.5\%}$	0.10	2.44	0.90	0.98	2.11	5.84	3.07	2.23	17.60
$\hat{\gamma}_{1\%}$	0.08	1.98	0.60	0.78	1.71	3.63	2.62	1.92	12.21
$\hat{\gamma}_{0.5\%}$	0.08	1.93	0.57	0.57	1.39	2.14	1.97	1.61	7.53
$\hat{\gamma}_{KS}$	0.02	0.56	0.05	0.02	0.10	0.01	-0.09	-0.14	0.02

This table reports results from monthly predictive regressions. We use the cross-sectional tail index estimates to predict the CRSP value-weighted market index returns over one-month, one-year, and three-year horizons. The different rows report forecasting results based on cross-sectional tail index time series with different thresholds. The test statistics are calculated using Hodrick (1992) standard errors for overlapping data with lag length equal to the number of months in each horizon.

ers. Notably, when the KS-quantile metric is applied, the predictive power becomes insignificant.

This finding suggests that the predictability results are sensitive to the threshold choice. Given that the average cross-sectional sample size ranges between 35,000 and 120,000, the 5% threshold is significantly distant from the tail of the distribution.[15] Consequently, selecting an appropriate threshold becomes crucial in economic applications, particularly in distinguishing between rare, extreme events and more common occurrences.

5 Conclusion

In this paper, we propose a new data-driven approach for selecting the optimal number of order statistics for the Hill estimator. Our method uses the maximum absolute deviation over the quantile dimension to fit the tail with a scaled Pareto distribution. Rigorous simulation studies demonstrate that our metric outperforms existing asymptotically consistent methods, where the first moment is bounded.

We demonstrate the economic significance of choosing an appropriate threshold by contrasting the performance of various methods using individual financial stock return data. The variation between methods is substantial and

[15]Considering the use of a large sample fraction, one might question whether the observed predictability stems from moments of the data that depend on central observations. Additional regression results indicate that the predictive power of $\alpha_{5\%}$ is not explained by the variance, skewness, or kurtosis of the cross-sectional distribution.

alters the results in existing papers in an economically meaningful way. Although the choice of threshold might seem innocuous, it impacts the results, underscoring the importance of carefully selecting the threshold in empirical analysis.

References

Balkema, A. A., De Haan, L., 1974. Residual life time at great age. The Annals of Probability 2, 792–804.

Bickel, P. J., Sakov, A., 2008. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Statistica Sinica 18, 967–985.

Bingham, N. H., Goldie, C. M., Teugels, J. L., 1989. Regular variation, vol. 27. Cambridge University Press, New York.

Clauset, A., Shalizi, C. R., Newman, M. E., 2009. Power-law distributions in empirical data. Society for Industrial and Applied Mathematics Review 51, 661–703.

Csörgo, S., Deheuvels, P., Mason, D., 1985. Kernel estimates of the tail index of a distribution. The Annals of Statistics 13, 1050–1077.

Danielsson, J., Peng, L., De Vries, C., De Haan, L., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis 76, 226–248.

Davis, R., Resnick, S., 1984. Tail estimates motivated by extreme value theory. The Annals of Statistics 12, 1467–1487.

Davydov, D., Vähämaa, S., Yasar, S., 2021. Bank liquidity creation and systemic risk. Journal of Banking & Finance 123, 106031.

De Haan, L., Ferreira, A., 2007. Extreme value theory: An introduction. Springer, New York.

De Haan, L., Resnick, S. I., 1980. A simple asymptotic estimate for the index of a stable distribution. Journal of the Royal Statistical Society. Series B (Methodological) 42, 83–87.

Dietrich, D., De Haan, L., Hüsler, J., 2002. Testing extreme value conditions. Extremes 5, 71–85.

Drees, H., Janßen, A., Resnick, S. I., Wang, T., 2020. On a minimum distance procedure for threshold selection in tail analysis. SIAM Journal on Mathematics of Data Science 2, 75–102.

Drees, H., Kaufmann, E., 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and their Applications 75, 149–172.

Gabaix, X., 1999. Zipf's law for cities: an explanation. The Quarterly Journal of Economics 114, 739–767.

Goldie, C. M., 1991. Implicit renewal theory and tails of solutions of random equations. The Annals of Applied Probability pp. 126–166.

Hall, P., 1982. On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society. Series B (Methodological) 44, 37–42.

Hall, P., 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of Multivariate Analysis 32, 177–203.

Hall, P., Welsh, A., 1985. Adaptive estimates of parameters of regular variation. The Annals of Statistics 13, 331–341.

Hill, B. M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. The Review of Financial Studies 5, 357–386.

Jansen, D. W., De Vries, C. G., 1991. On the frequency of large stock returns: Putting booms and busts into perspective. The Review of Economics and Statistics 73, 18–24.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Review of Financial Studies 27, 2841–2871.

Mandelbrot, B. B., 1963. New methods in statistical economics. Journal of Political Economy 71, 421–440.

Mason, D. M., 1982. Laws of large numbers for sums of extreme values. The Annals of Probability 10, 754–764.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131.

Reed, W. J., 2003. The pareto law of incomes—an explanation and an extension. Physica A: Statistical Mechanics and its Applications 319, 469–486.

Resnick, S., Starica, C., 1997. Smoothing the Hill estimator. Advances in Applied Probability 29, 271–293.

Sun, P., De Vries, C. G., 2018. Exploiting tail shape biases to discriminate between stable and Student-t alternatives. Journal of Applied Econometrics 33, 708–726.

Van Oordt, M. R., Zhou, C., 2016. Systematic tail risk. Journal of Financial and Quantitative Analysis 51, 685–705.

Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association 73, 812–815.

A Appendix

A.1 The scale of the ARCH(1) process

For the ARCH(1) stochastic process, the first and second-order tail indices of the distribution's stationary solution are known, as well as the second-order scale parameter. However, the first-order scale parameter remains more elusive. Generally, only implicit expressions are available, as discussed in Goldie (1991). In cases where the implied first-order tail index is an integer, Goldie provides a method to obtain explicit expressions for the first-order scale parameter. This note implements Goldie's approach, which is valuable as it offers a benchmark for simulations in the tail area.

Construction of the first-order scale

Consider the ARCH(1) process:

$$Y_t = X_t H_t$$

and

$$H_t^2 = a + bY_{t-1}^2$$

and where X is a standard normal random variable. Combining the two equations after squaring the first yields the random difference equation

$$Y_t^2 = aX_t^2 + \left(bX_t^2\right) Y_{t-1}^2.$$

By the method outlined in Sun and De Vries (2018), it follows that the stationary solution to this process has a tail expansion

$$\Pr\left\{Y^2 > t\right\} \sim At^{-\alpha}\left[1 + Bt^{-\beta}\right]$$

for $t \to \infty$. The first-order tail index α follows from solving the Kesten equation

$$\Gamma\left(\alpha + \frac{1}{2}\right) = \sqrt{\pi}/\left(2b\right)^{\alpha}.$$

Note that $\Gamma\left(\alpha + \frac{1}{2}\right) = \frac{(2\alpha)!}{4^{\alpha}\alpha!}\sqrt{\pi}$ for α integer valued. In the following we state the parameter values for A, B and β when α is an integer. Giving,

$$b(\alpha) = \left[\frac{2\alpha!}{(2\alpha)!}\right]^{1/\alpha}.$$

Furthermore, necessarily $\beta = 1$ and B follows from

$$\begin{aligned} B &= a\alpha b(\alpha)^{\alpha} \frac{E\left[(X^2)^{\alpha+1}\right]}{1-\left(b(\alpha)^{\alpha+1}\right)E\left[(X^2)^{\alpha+1}\right]} \\ &= a\alpha b(\alpha)^{\alpha} \frac{(2\alpha+1)!!}{1-\left(b(\alpha)^{\alpha+1}\right)(2\alpha+1)!!,} \end{aligned} \tag{10}$$

where X is a standard normal random variable. Note that $E\left[X^p\right] = (p-1)!!$ for the even moments. Here $x!!$ is the double factorial.

We now turn to A. Goldie (1991) shows it requires several iterative steps to find

$$A = \frac{1}{\alpha m} E\left[\left(a + b(\alpha)Y^2\right)^{\alpha} - \left(b(\alpha)Y^2\right)^{\alpha}\right] E\left[|X|^{2\alpha}\right], \tag{11}$$

where

$$E\left[|X|^{2\alpha}\right] = 2^{\alpha}\frac{1}{\sqrt{\pi}}\Gamma\left(\alpha + \frac{1}{2}\right)$$

and

$$m = b(\alpha)^{\alpha}\ln\left(b(\alpha)\right) + b(\alpha)^{\alpha}2^{\alpha}\frac{1}{\sqrt{\pi}}\Gamma\left(\alpha+\frac{1}{2}\right)\left(\ln 2 + \psi\left(\alpha+\frac{1}{2}\right)\right)$$

and where $\psi(.) = \Gamma'(.)/\Gamma(.)$ is the digamma function. For integers $\alpha \geq 1$

$$\psi\left(\alpha+\frac{1}{2}\right) = -\gamma - 2\ln 2 + 2\left(1 + \frac{1}{3} + ... + \frac{1}{2\alpha-1}\right)$$

and where γ is Euler's constant.

The expression for A contains a binomial expansion with α as the exponent and can therefore be written as

$$\begin{aligned} A &= \frac{1}{\alpha m} E\left[\left(a + b(\alpha)Y^2\right)^{\alpha} - \left(b(\alpha)Y^2\right)^{\alpha}\right] E\left[|X|^{2\alpha}\right] \\ &= \frac{1}{\alpha m} E\left[\sum_{i=0}^{\alpha-1}\binom{\alpha}{i} a^{\alpha-i}\left(b(\alpha)Y^2\right)^{i}\right] E\left[|X|^{2\alpha}\right] \end{aligned} \tag{12}$$

For $\alpha = 1$ the expression for A boils down to a/m. Note that $E\left[|X|^2\right] = 1$. For $\alpha > 1$ the binomial expansion shows that we need the unconditional $2(\alpha-1)^{\text{th}}$ moment of Y_t which depends on the unconditional even moments of Y_t smaller than $2(\alpha-1)$. Therefore, we follow an iterative procedure to calculate A. Take for instance the expression of A for $\alpha = 3$

$$A = \frac{1}{3m} E\left[a^3 + 3a^2 b(\alpha)Y^2 + 3ab(\alpha)^2 Y^4\right] E\left[|X|^6\right]. \tag{13}$$

For the stationary distribution of Y, it has to hold that

$$\begin{aligned} E\left[Y^4\right] &= a^2 E\left[X^4\right] + 2ab(\alpha) E\left[Y^2\right] E\left[X^4\right] + b(\alpha)^2 E\left[Y^4\right] E\left[X^4\right] \\ E\left[Y^4\right] &= \frac{a^2 E\left[X^4\right] + 2ab(\alpha) E\left[Y^2\right] E\left[X^4\right]}{1 - b(\alpha)^2 E\left[X^4\right]} \end{aligned} \tag{14}$$

Also note that

$$\begin{aligned} E\left[Y^2\right] &= aE\left[X^2\right] + b(\alpha) E\left[Y^2\right] E\left[X^2\right] \\ E\left[Y^2\right] &= \frac{aE\left[X^2\right]}{1 - b(\alpha) E\left[X^2\right]}. \end{aligned} \tag{15}$$

By substituting the expression for $E\left[Y^2\right]$ into the expression for $E\left[Y^4\right]$ we have an explicit expression for A in case of $\alpha = 3$. For larger values of α more iterative steps, by the same logic, have to be taken to get an explicit expression for A. With the explicit expression for A, Table 5 provides all the parameters of the Hall expansion for the stationary solution of the ARCH(1) process up to $\alpha = 6$. For presentational purposes we choose $a = 1$.

Table 5: Parameters Hall expansion for ARCH(1)

α	1	2	3	4	5	6
b	1	0.58	0.41	0.31	0.25	0.21
β	1	1	1	1	1	1
B	-1.5	-5.3	-11.42	-19.87	-30.65	-43.75
A	1.37	4.61	28.65	298.02	4290.03	78071.7

We use these parameters to determine the optimal theoretical AMSE threshold in simulations with the ARCH process. Table 9 in the Appendix demonstrates that the performance of the various methods differs across the different ARCH processes. All methods have a region over α for which they perform well. The KS-quantile metric is particularly apt for larger values of α. For the more heavy-tailed processes the KS-probability statistic and the method by Drees and Kaufmann (1998) have the smallest bias. Consequently, as α increases, their estimates adjust marginally, especially the KS-test statistic estimates. Therefore, as α increases their bias increases and the KS-quantile and Eye-Ball method produce a smaller bias. This pattern is also mimicked in the higher moments of the $\hat{\alpha}$'s. The level of k^* for the KS-quantile and the Eye-Ball method are more in line with the AMSE optimal threshold. Furthermore, the KS-quantile metric also shows the decreasing pattern present in k^*_{TH}.

A.2 Optimal theoretical threshold

From the variance and the bias, the MSE $=$ var $+$ (bias)^2 is

$$\text{MSE} = \frac{s^{\alpha}}{nA}\frac{1}{\alpha^2} + \left(\frac{\beta B s^{-\beta}}{\alpha\left(\alpha+\beta\right)}\right)^2 + o\left(\frac{s^{\alpha}}{n}\right) + o\left(s^{-2\beta}\right).$$

For the AMSE the small terms go to 0,

$$\text{AMSE} = \frac{s^{\alpha}}{nA}\frac{1}{\alpha^2} + \left(\frac{\beta B s^{-\beta}}{\alpha\left(\alpha+\beta\right)}\right)^2.$$

Taking the derivative w.r.t. s and setting it to zero gives optimal threshold

$$s^{*} = \left[\frac{2AB^2\beta^3\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{\frac{1}{\alpha+2\beta}} n^{\frac{1}{\alpha+2\beta}}.$$

Substituting s^{*} back into the MSE gives

$$\text{MSE}^{*} = \frac{1}{A\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2AB^2\beta^3\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{-\frac{2\beta}{\alpha+2\beta}} + o\left(n^{-\frac{2\beta}{\alpha+2\beta}}\right). \tag{16}$$

Hall and Welsh (1985) show that there does not exist an estimator that can improve on the rate by which the AMSE of the Hill estimator disappears as n increases. Given s^{*} and noticing that $1-\text{F}\left(s\right) = As^{-\alpha}\left[1+s^{-\beta}\right]$ gives the following result for the number of upper-order statistics:

$$n^{\frac{-2\beta}{\alpha+2\beta}} M\left(s^{*}\right) \underset{n\to\infty}{\to} A\left[\frac{2AB^2\beta^3\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{-\frac{\alpha}{\alpha+2\beta}}. \tag{17}$$

Through the Hall expansion we have the functional forms for α, β, A and B for the Student-t, symmetric stable, Fréchet distribution and the distribution of the stationary solution to the ARCH process. See tables 5 and 6 for the parameter values for the specific distributions.

Table 6: Hall expansion parameters values

	Stable	Student-t	Fréchet
α	$(1,2)$	$(2,\infty)$	$(2,\infty)$
β	α	2	α
A	$\frac{1}{\pi}\Gamma\left(\alpha\right)\sin\left(\frac{\alpha\pi}{2}\right)$	$\frac{1}{\sqrt{\alpha\pi}}\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}\alpha^{(\alpha-1)/2}$	1
B	$-\frac{1}{2}\frac{\Gamma(2\alpha)\sin(\alpha\pi)}{\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}$	$-\frac{\alpha^2}{2}\frac{\alpha+1}{\alpha+2}$	$\frac{1}{2}$

A.3 Other methods

Danielsson et al.'s double bootstrap

The double bootstrap by Danielsson et al. (2001) minimizes the following criterium,

$$Q\left(n_1, k_1\right) := \mathrm{E}\left(\left[M^*_{n_1}\left(k_1\right) - 2\left(\gamma^*_{n_1}\left(k_1\right)\right)^2\right]^2\right),$$

where

$$M^*_{n_1}\left(k_1\right) = \frac{1}{k_1}\sum_{i=0}^{k_1}\left(\log\left(\frac{X_{n_1-i,n_1}}{X_{n_1-k_1,n_1}}\right)^2\right).$$

Here $n_1 = n^{1-\epsilon}$ is the smaller sub-sample for the bootstrap. The $\gamma^*_{n_1}$ is the Hill estimator for the bootstrapped sample. The Q function is minimized over two dimensions, namely: n_1 and k_1. Given the optimal n_1^* and k_1^*, a second bootstrap with a smaller sample size n_2 is executed to determine k_2^*. Here n_2 is typically chosen to be $n_2 = n_1^2/n$. The optimal number of order statistics is given by,

$$\widehat{k^*}_{db} = \frac{\left(k_2\right)^2}{k_1}\left[\frac{\log\left(k_1\right)^2}{\left(2\log\left(n_1\right) - \log\left(k_1\right)\right)^2}\right]^{\frac{\log(n_1)-\log(k_1)}{\log(n_1)}}.$$

Drees and Kaufmann (1998)'s sequential estimator

Drees and Kaufmann (1998) introduce a sequential procedure that yields stopping time,

$$\overline{k}_n\left(r_n\right) = \min\left\{k \in \{2, .., n\} \mid \max_{2\le i\le k_n} i^{1/2}\left|\widehat{\gamma}_{n,i} - \widehat{\gamma}_{n,k}\right| > r_n\right\},$$

where the threshold $r_n = 2.5\widetilde{\gamma}_n n^{1/4}$. Here $\widetilde{\gamma}_n$ is the initial estimator for γ with $k = 2\sqrt{n^+}$, where n^+ is the number of positive observations in the sample. This leads to the adaptive estimator

$$k^*_{DK} := \left[\left(2\widehat{\rho}_n + 1\right)^{-1/\widehat{\rho}_n}\left(2\widetilde{\gamma}_n^2\widehat{\rho}_n\right)^{1/(2\widehat{\rho}_n+1)}\left(\overline{k}_n\left(r_n^{\xi}\right)/\overline{k}_n\left(r_n\right)^{\xi}\right)^{1/(1-\xi)}\right]$$

with

$$\widehat{\rho}_{n,\lambda}\left(r_n\right) := \log\frac{\max\limits_{2\le i\le\left[\lambda\overline{k}_n(r_n)\right]} i^{1/2}\left|\widehat{\gamma}_{n,i} - \widehat{\gamma}_{n,\left[\lambda\overline{k}_n(r_n)\right]}\right|}{\max\limits_{2\le i\le\overline{k}_n(r_n)} i^{1/2}\left|\widehat{\gamma}_{n,i} - \widehat{\gamma}_{n,\overline{k}_n(r_n)}\right|}/\log\left(\lambda\right) - \frac{1}{2},$$

where $\lambda \in (0, 1)$.

Automated Eye-Ball method

The algorithms based on "Eye-Balling" the Hill plot aim to identify a significant drop in variance as k increases. To be able to use multiple simulations, we formalize an automated Eye-Ball method. To this end we employ a sequential procedure, as follows:

$$k^*_{eye} = \min\left\{k \in 2, ..., n^+ - w | h < \tfrac{1}{w}\sum\nolimits_{i=1}^{w} \mathrm{I}\left\{\widehat{\alpha}\left(k+i\right) < \widehat{\alpha}\left(k\right) \pm \varepsilon\right\}\right\}. \quad (18)$$

Here $\mathrm{I}\{\cdot\}$ is the indicator function and w is the size of the moving window, which is typically 1% of the full sample. This window is used to evaluate the volatility of the Hill estimate. The ε gives the range between which $[\hat{\alpha}\left(k+1\right), ..., \hat{\alpha}\left(k+w\right)]$ are within the permitted bound around $\hat{\alpha}\left(k\right)$. No less than h% of the estimates should be within the bound of $\hat{\alpha}\left(k\right)$ for k to be considered as a possible candidate. Here h is typically around 90%, and ε is chosen to be 0.3. The n^+ is the number of positive observations in the data.

A.4 Alternative penalty functions

We compare the performance of our metric to three other metrics: the mean squared deviations, the mean absolute deviations, and a discretized version of the metric used by Dietrich et al. (2002). The mean squared deviations metric is

$$Q_{2,n} = \frac{1}{T}\sum_{j=1}^{T}\left(x_{n-j,n} - q\left(j,k\right)\right)^2$$

and the mean absolute deviations

$$Q_{3,n} = \frac{1}{T}\sum_{j=1}^{T}\left|x_{n-j,n} - q\left(j,k\right)\right|$$

where T is the region over which the metric is measured. These types of penalty functions are often used in the econometric literature, but averaging tends to emphasize the numerous centre observations

The third metric we consider is inspired by the theoretical test statistic developed by Dietrich et al. (2002). They devised a statistic to test whether the extreme value conditions apply. Discretizing their statistic results in.

$$Q_{4,n} = \sum_{j=1}^{T}\frac{\left(x_{n-j,n} - q\left(j,k\right)\right)^2}{\left[q'\left(j,k\right)\right]^2} = \frac{1}{T}\sum_{j=1}^{T}\frac{\left(x_{n-j,n} - \left(\frac{k}{j}\right)^{\frac{1}{\widehat{\alpha}_k}} x_{n-k+1,n}\right)^2}{\left[-\frac{1}{\widehat{\alpha}_k}\left(\frac{j}{k}\right)^{-\left(1+\frac{1}{\widehat{\alpha}_k}\right)}\left(x_{n-k+1,n}\right)\frac{n}{k}\right]^2}.$$

We draw samples from the Student-t distribution family to demonstrate the properties of the different metrics. To conserve space, the analysis of the symmetric stable and Fréchet distribution families is available upon request.

In Figure 13, the level of $\alpha(k^*)$ is displayed against the threshold T over which the specified metric is optimized. These plots provide an indication of whether $\alpha(k^*)$ is at the correct level and remains insensitive to the nuisance parameter T. The upper left panel shows that the curves for the KS-quantile metric are relatively flat and close to the theoretical level of α. In contrast, based on the mean square distance, mean absolute distance, and the metric by Dietrich et al. (2002), the estimates of $\alpha(k^*)$ do not stabilize, except for the Student-t (2) distribution. The monotonic decline in the three graphs suggests that the level of k^* is dependent on the region over which the optimization occurs.

The lower four panels in Figure 13 depict the average k^* for the Student-t distribution family. These figures illustrate the properties of k^* as the interval $[X_{n,n}, X_{n-T,n}]$ is extended. For the KS-quantile metric, the average k^* as a function of T stabilizes once T is sufficiently large. However, for the Student-t(2) distribution, this stabilization occurs very slowly. The average mean squared distance displays roughly the same properties as the KS-quantile metric. Although the choice of k^* seems to stabilize, it does not necessarily lead to a stable and optimal estimation of $\alpha\,(k^*)$. This stabilization is not observed for the mean absolute difference or the metric by Dietrich et al. (2002). Additionally, we observe that for the KS-quantile metric, k^* is an increasing function of the degrees of freedom, which aligns with the pattern expected based on k^*_{TH} derived by minimizing the AMSE for the Student-t distribution family—a pattern not observed for the other criteria.

A.5 Brownian motion representation

There are various ways to study the behavior of the KS-quantile metric. In this context, we examine its properties by modeling the quantile process using a Brownian motion representation. This approach allows us to simulate under more general conditions than those provided by fully parametric distributions.

By Theorem 2.4.8 from De Haan and Ferreira (2007, 52) the KS-quantile metric in (7) can be written as

$$\underset{0<k<T}{\arg\min} \sup_{0<l<\frac{T}{k}} \left| x_{n-lk,n} - (l)^{-\hat{\gamma}}\, x_{n-k,n} \right|$$

when $x_{n-lk,n}$ are the logarithm of the order statistics. This is equal to

$$\underset{0<k<T}{\arg\min}\ \sup_{0<l<\frac{T}{k}} \left| \frac{\gamma}{\sqrt{k}} U\left(\frac{n}{k}\right) l^{-\hat{\gamma}} \left[l^{-1} w(l) - w(1) + A_0\left(\frac{n}{k}\right) \frac{\sqrt{k}}{\gamma} \frac{l^{-\rho}-1}{\rho} \right] \right|, \tag{19}$$

where $l = {}^{i}/_{k}$, $\rho \leq 0$, $U({}^{n}/_{k}) = \left(\frac{1}{1-F}\right)^{\leftarrow}$, $w(l)$ is a Brownian motion and $A_0({}^{n}/_{k})$ is a suitable normalizing function. We use the expectation of γ

$$\widehat{\gamma} = \gamma + \frac{\gamma}{\sqrt{k}} \int_0^1 \left(l^{-1} w(l) - w(1) \right) dl + \frac{A_0(n/k)}{1-\rho},$$

see De Haan and Ferreira (2007, 76). For the case that the cdf satisfies the Hall expansion in (3), $U\left(\frac{n}{k}\right)$ and $A_0\left(\frac{n}{k}\right)$ can be given further content. This is also needed for the simulations that are performed below. Suppose the cdf satisfies the Hall expansion (3). Then applying the De Bruijn inversion[16] we arrive at,

$$U\left(\frac{n}{k}\right) = A^{\gamma}(n/k)^{\gamma} \left[1 + \frac{B}{\alpha} A^{-\beta\gamma} (n/k)^{-\beta\gamma} \right]$$

and

$$A_0(n/k) = -\frac{\beta/\alpha}{\alpha B^{-1} A^{\beta/\alpha} \frac{n}{k}^{\beta/\alpha}}.$$

Simulating from (19) necessitates a choice of values for parameters α, β, A and B. For the robustness of the Monte Carlo simulations, we use distributions and processes that differ along the dimension of these parameters. The Student-t, symmetric stable and Fréchet distribution all satisfy the power expansion in (3).

The left plots in Figure 3 in Appendix C show, for a given k, at which order statistic the maximum quantile distance is observed for the Student-t, symmetric stable and Fréchet distributions for a given range of α family parameters. The right plot displays the value of this maximum distance for the given k. These two plots offer insight into how the KS-quantile metric selects k^* under relatively general conditions. It is evident that for large k, the largest deviations are almost always found at the most extreme observations. By choosing a large k, the Pareto distribution fits better as the Hill estimator in this case is unbiased, resulting in the largest deviation being observed at more extreme observations. Conversely, for smaller k, the largest deviations are more frequently found at less extreme observations.

The right panel shows that for the largest deviations given k, the smallest of these largest deviations are observed at $k = 2$. A small value of $k = 2$ is

[16] See Bingham et al. (1989, page 29).

desirable as it minimizes the bias of the Hill estimator. Since the modeled $\hat{\gamma}$ is based on the expectation of the Hill estimator, it is not surprising that the limit in (19) identifies $k = 2$ as the optimal threshold. This threshold minimizes the bias in the Hill estimator. This also holds for the parameters retrieved from the symmetric stable and Fréchet distribution in panels (b) and (c).

B Tables

Table 7: Horse Race **symmetric stable** distribution family

	α	KS qua	KS prob	TH	5%	Eye-Ball	Drees	Du bo
$E[\hat{\alpha}]$	1.1	1.21	1.08	1.10	1.11	1.11	1.07	1.34
	1.3	1.39	1.35	1.35	1.37	1.32	1.33	1.51
	1.5	1.57	1.68	1.60	1.72	1.54	1.67	1.82
	1.7	1.77	2.10	1.88	2.32	1.84	2.18	2.37
	1.9	2.30	2.62	2.36	3.55	3.35	3.13	3.66
$E[\hat{\alpha}^2]$	1.1	2.01	1.17	1.22	1.23	1.27	1.14	6.12
	1.3	2.32	1.82	1.83	1.87	1.78	1.78	5.64
	1.5	2.91	2.83	2.57	2.97	2.44	2.81	5.35
	1.7	3.58	4.40	3.58	5.41	3.47	4.75	9.32
	1.9	5.92	6.88	5.85	12.63	11.37	9.85	20.85
$E[\hat{\alpha}^3]$	1.1	9.77	1.26	1.35	1.37	1.49	1.22	355.53
	1.3	4.84	2.46	2.49	2.57	2.47	2.37	225.38
	1.5	6.43	4.76	4.17	5.13	3.94	4.72	86.07
	1.7	8.24	9.23	6.93	12.62	6.64	10.40	340.43
	1.9	16.54	18.06	15.23	45.04	38.98	31.16	1286.54
$E[\hat{\alpha}^4]$	1.1	239.68	1.37	1.49	1.54	1.80	1.30	$4.4 * 10^4$
	1.3	13.10	3.32	3.39	3.54	3.50	3.17	$1.9 * 10^4$
	1.5	17.40	8.01	6.81	8.89	6.48	7.93	3370.33
	1.7	21.54	19.39	13.64	29.54	12.94	22.79	$3.7 * 10^4$
	1.9	49.41	47.43	41.79	161.02	135.17	98.97	$2.4 * 10^5$
$E[\hat{k}^*]$	1.1	237.69	1116.87	817.00	500.00	7.86	1482.65	696.73
	1.3	181.82	1091.10	292.00	500.00	10.03	1467.80	904.31
	1.5	148.84	1158.05	146.00	500.00	12.68	1378.84	1014.46
	1.7	208.00	1262.27	74.00	500.00	18.73	1176.54	889.85
	1.9	662.72	1319.34	27.00	500.00	107.37	862.76	495.15

This table depicts for the different methods the first four moments of the distribution of estimated α's and the average k^* selected in the simulations. The samples are drawn from the symmetric stable distribution family. The column α indicates the stability parameter for the particular symmetric stable distribution. The different methods are stated in the first row. KS qua is the Kolmogorov-Smirnov metric measured over the quantile dimension, see (7). KS prob is the Kolmogorov-Smirnov over the probability dimension. TH is based on the theoretically derived optimal k from minimizing the AMSE for specific parametric distributions, presented in Equation (17) in the Appendix. The automated Eye-Ball method in (18) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column Drees, the k^* is determined by the methodology described by Drees and Kaufmann (1998). Du bo is the double bootstrap procedure by Danielsson et al. (2001). The sample size is $n = 10,000$ for $10,000$ repetitions.

Table 8: Horse Race **Fréchet** distribution family

	α	KS qua	KS prob	TH	5%	Eye-Ball	Drees	Du bo
	2	2.00	1.94	1.95	1.98	2.00	1.92	2.44
	3	2.89	2.91	2.93	2.97	3.00	2.88	3.64
$E[\hat{\alpha}]$	4	3.79	3.87	3.91	3.96	3.99	3.84	4.85
	5	4.69	4.84	4.88	4.95	4.99	4.80	6.08
	6	5.60	5.81	5.86	5.93	5.98	5.77	7.28
	2	4.46	3.75	3.82	3.92	4.08	3.70	22.87
	3	9.06	8.45	8.60	8.82	9.09	8.31	50.46
$E[\hat{\alpha}^2]$	4	15.39	15.02	15.28	15.68	16.06	14.78	89.72
	5	23.41	23.46	23.88	24.51	25.04	23.10	141.00
	6	33.25	33.79	34.38	35.29	35.99	33.26	201.86
	2	11.07	7.28	7.48	7.79	8.43	7.11	2023.09
	3	30.46	24.58	25.24	26.28	27.84	24.00	6753.49
$E[\hat{\alpha}^3]$	4	66.08	58.26	59.83	62.30	65.21	56.89	$1.6 * 10^4$
	5	122.80	113.79	116.86	121.69	126.71	111.13	$3.1 * 10^4$
	6	206.43	196.62	201.93	210.27	218.03	192.03	$5.4 * 10^4$
	2	30.47	14.14	14.66	15.50	17.68	13.69	$3.5 * 10^5$
	3	109.01	71.56	74.20	78.47	86.19	69.32	$1.8 * 10^6$
$E[\hat{\alpha}^4]$	4	297.94	226.17	234.51	248.00	267.06	219.10	$5.6 * 10^6$
	5	672.05	552.17	572.53	605.46	646.01	534.97	$1.4 * 10^7$
	6	1331.32	1144.99	1187.20	1255.48	1329.48	1109.32	$2.8 * 10^7$
	2	217.09	1028.10	928.00	500.00	18.98	1500.70	681.63
	3	220.58	1028.10	928.00	500.00	34.88	1500.70	682.25
$E[\hat{k}^*]$	4	223.81	1028.10	928.00	500.00	51.37	1500.85	682.25
	5	227.56	1028.10	928.00	500.00	67.34	1501.00	685.21
	6	229.79	1028.10	928.00	500.00	83.19	1501.00	682.25

This table depicts for the different methods the first four moments of the distribution of estimated α's and the average k^* selected in the simulations. The samples are drawn from the Fréchet distribution family. The column α indicates the shape parameter for the particular Fréchet distribution. The different methods are stated in the first row. KS qua is the Kolmogorov-Smirnov metric measured over the quantile dimension, see (7). KS prob is the Kolmogorov-Smirnov over the probability dimension. TH is based on the theoretically derived optimal k from minimizing the AMSE for specific parametric distributions, presented in Equation (17) in the Appendix. The automated Eye-Ball method in (18) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column Drees, the k^* is determined by the methodology described by Drees and Kaufmann (1998). Du bo is the double bootstrap procedure by Danielsson et al. (2001). The sample size is $n = 10,000$ for $10,000$ repetitions.

Table 9: Horse Race **ARCH** processes

	α	KS qua	KS prob	TH	5%	Eye-Ball	Drees	Du bo
	2	3.75	2.36	3.60	3.10	3.61	2.78	6.39
	3	4.77	2.61	5.07	3.73	4.61	3.38	9.34
$E[\hat{\alpha}]$	4	5.46	2.72	6.29	4.07	5.18	3.73	11.45
	5	5.94	2.79	7.31	4.26	5.45	3.94	13.35
	6	6.29	2.82	8.13	4.38	5.57	4.08	15.38
	2	14.63	5.60	13.12	9.64	13.16	7.77	177.09
	3	23.27	6.82	26.20	13.95	21.45	11.47	566.05
$E[\hat{\alpha}^2]$	4	30.33	7.43	40.77	16.55	26.99	13.95	495.71
	5	35.87	7.77	55.38	18.16	29.99	15.59	721.69
	6	40.06	7.97	68.94	19.19	31.29	16.70	2222.75
	2	59.42	13.25	48.35	30.04	48.56	21.78	$3.1 * 10^4$
	3	116.31	17.83	138.42	52.23	100.43	39.10	$2.9 * 10^5$
$E[\hat{\alpha}^3]$	4	171.51	20.26	272.14	67.54	141.63	52.51	$9.8 * 10^4$
	5	219.58	21.66	435.92	77.56	166.23	62.04	$1.7 * 10^5$
	6	258.40	22.53	611.46	84.27	177.58	68.81	$4.9 * 10^6$
	2	250.24	31.41	180.21	93.86	181.12	61.34	$1.1 * 10^7$
	3	594.13	46.63	746.77	196.03	473.62	133.96	$2.6 * 10^8$
$E[\hat{\alpha}^4]$	4	986.00	55.28	1872.84	276.02	748.26	198.59	$3.8 * 10^7$
	5	1363.92	60.46	3568.88	331.86	929.11	248.19	$6.6 * 10^7$
	6	1687.53	63.72	5694.45	370.66	1017.89	284.93	$1.6 * 10^{10}$
	2	242.85	1319.92	122.00	500.00	54.60	849.77	76.35
	3	157.57	1321.28	55.00	500.00	82.84	724.81	39.36
$E[\hat{k}^*]$	4	117.82	1322.05	36.00	500.00	115.79	669.00	24.35
	5	97.31	1322.23	28.00	500.00	147.39	639.94	17.96
	6	88.35	1322.48	24.00	500.00	175.83	622.50	14.66

This table depicts for the different methods the first four moments of the distribution of estimated α's and the average k^* selected in the simulations. The samples are drawn from ARCH processes. The column α indicates tail index of the unconditional distribution of the particular ARCH process, see Appendix A.1. The different methods are stated in the first row. KS qua is the Kolmogorov-Smirnov metric measured over the quantile dimension, see (7). KS prob is the Kolmogorov-Smirnov over the probability dimension. TH is based on the theoretically derived optimal k from minimizing the AMSE for specific parametric distributions, presented in Equation (17) in the Appendix. The automated Eye-Ball method in (18) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column Drees, the k^* is determined by the methodology described by Drees and Kaufmann (1998). Du bo is the double bootstrap procedure by Danielsson et al. (2001). The sample size is $n = 10,000$ for $10,000$ repetitions.

Table 10: Descriptive statistics stock data estimates

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
Mean	3.24	2.34	2.68	3.20	2.42	6.03	3.44	2.40	2.78	3.30	2.46	6.94
Median	3.16	2.31	2.65	3.19	2.36	3.63	3.40	2.39	2.76	3.28	2.41	3.88
St. Dev.	0.89	0.31	0.42	0.60	1.44	10.69	0.86	0.32	0.44	0.62	0.44	13.52
Min	0.58	0.99	0.95	0.52	0.31	0.34	0.65	1.00	1.00	0.33	0.36	0.94
Max	7.13	4.80	6.01	6.95	109.68	390.71	7.62	4.14	5.77	6.59	6.98	571.02
Skewness	0.34	0.69	0.64	0.21	66.70	14.42	0.34	0.26	0.42	0.23	1.64	17.66
Kurtosis	3.10	5.45	4.94	3.96	4,946.40	367.43	3.35	4.26	4.70	3.59	12.81	571.93

(a) Estimates $\alpha\left(k^*\right)_i$

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
Mean	107.91	337.05	203.04	36.11	312.47	34.88	100.54	339.05	203.04	36.40	320.43	35.00
Median	56	250	151	28	267	10	48	252	151	29	264	7
St. Dev.	135.95	254.27	145.79	30.52	190.56	64.45	138.82	254.12	145.79	28.32	200.39	78.26
Min	1	10	60	1	1	2	1	4	60	1	1	2
Max	1,334	1,293	692	296	1,385	729.20	1,381	1,297	692	324	1,385	1,236.85
Skewness	2.75	1.68	1.62	2.63	1.51	4.07	3.24	1.66	1.62	2.60	1.55	5.98
Kurtosis	13.44	5.58	5.29	13.08	6.09	25.22	17.66	5.49	5.29	14.06	5.83	54.22

(b) Estimates k_i^*

This table presents descriptive statistics for $\hat{\alpha}\left(k^*\right)_i$ and k_i^* by applying the six different methods to choose k_i^* for left and right tail of stock returns. The data are from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2015-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS-quantile metric, KS test statistic, 5% threshold, automated Eye-Ball method, the iterative method by Drees and Kaufmann (1998) and the double bootstrap by Danielsson et al. (2001). Different statistics are calculated for the distribution of $\hat{\alpha}$. The stocks for which one of the methods has $\hat{\alpha} > 1,000$ are excluded. The maximum k is cut off at 15% of the total sample size. There are 17,918 companies included in the analysis.

Table 11: Median absolute differences between different methods

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
KS qua	0	0.87	0.58	0.45	0.81	0.74	0	1.04	0.69	0.47	0.98	0.78
KS pr	0.87	0	0.34	0.86	0.12	1.26	1.04	0	0.37	0.90	0.13	1.41
5%	0.58	0.34	0	0.52	0.31	0.87	0.69	0.37	0	0.52	0.34	1.01
Eye-Ball	0.45	0.86	0.52	0	0.80	0.45	0.47	0.90	0.52	0	0.84	0.58
Drees	0.81	0.12	0.31	0.80	0	1.16	0.98	0.13	0.34	0.84	0	1.32
Du bo	0.74	1.26	0.87	0.45	1.16	0	0.78	1.41	1.01	0.58	1.32	0

(a) Estimates $\alpha\left(k^*\right)_i$

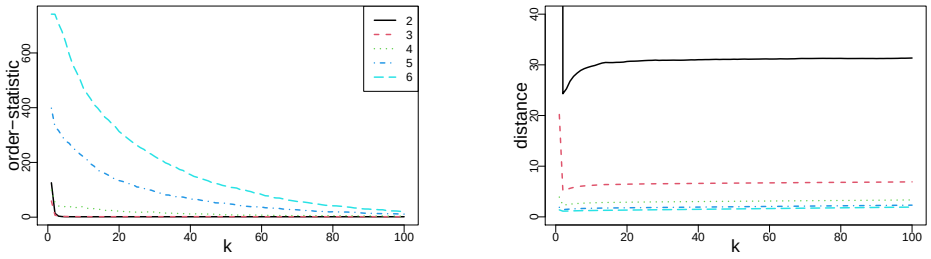
	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo	KS qua	KS pr	5%	Eye-Ball	Drees	Du bo
KS qua	0	158	95	37	169	56	0	178	97	32	181	47
KS pr	158	0	99	219	48	222	178	0	100	219	45	230
5%	95	99	0	122	103	128	97	100	0	120	104	135
Eye-Ball	37	219	122	0	232	29	32	219	120	0	227	30
Drees	169	48	103	232	0	231	181	45	104	227	0	238
Du bo	56	222	128	29	231	0	47	230	135	30	238	0

(b) Estimates k_i^*

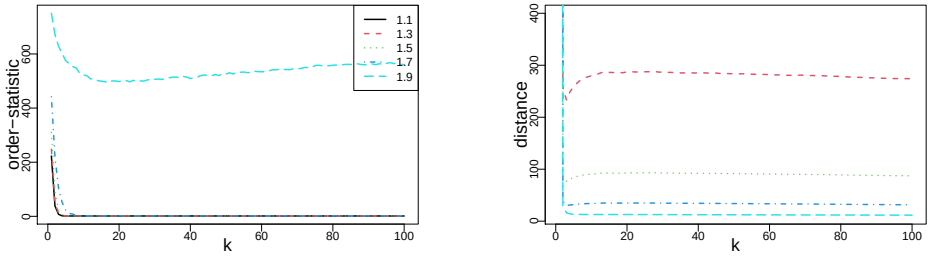
This table presents the median absolute difference between $\hat{\alpha}\left(k^*\right)_i$ and k_i^* by applying the six different methods to choose k_i^* for left and right tail of U.S. stock returns. The different methods are the KS-quantile metric, KS test statistic, 5% threshold, automated Eye-Ball method, the iterative method by Drees and Kaufmann (1998) and the double bootstrap by Danielsson et al. (2001). The stocks for which one of the methods has $\hat{\alpha} > 1,000$ are excluded. The maximum k is cut off at 15% of the sample size. There are 17,918 companies included in the analysis.

C Figures

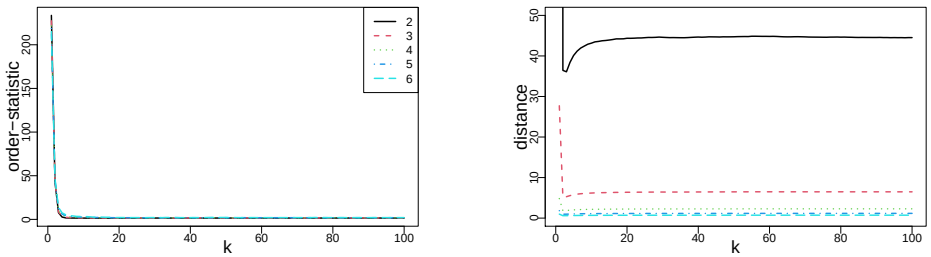
Figure 3: Simulations Brownian motion representation



(a) Student-t



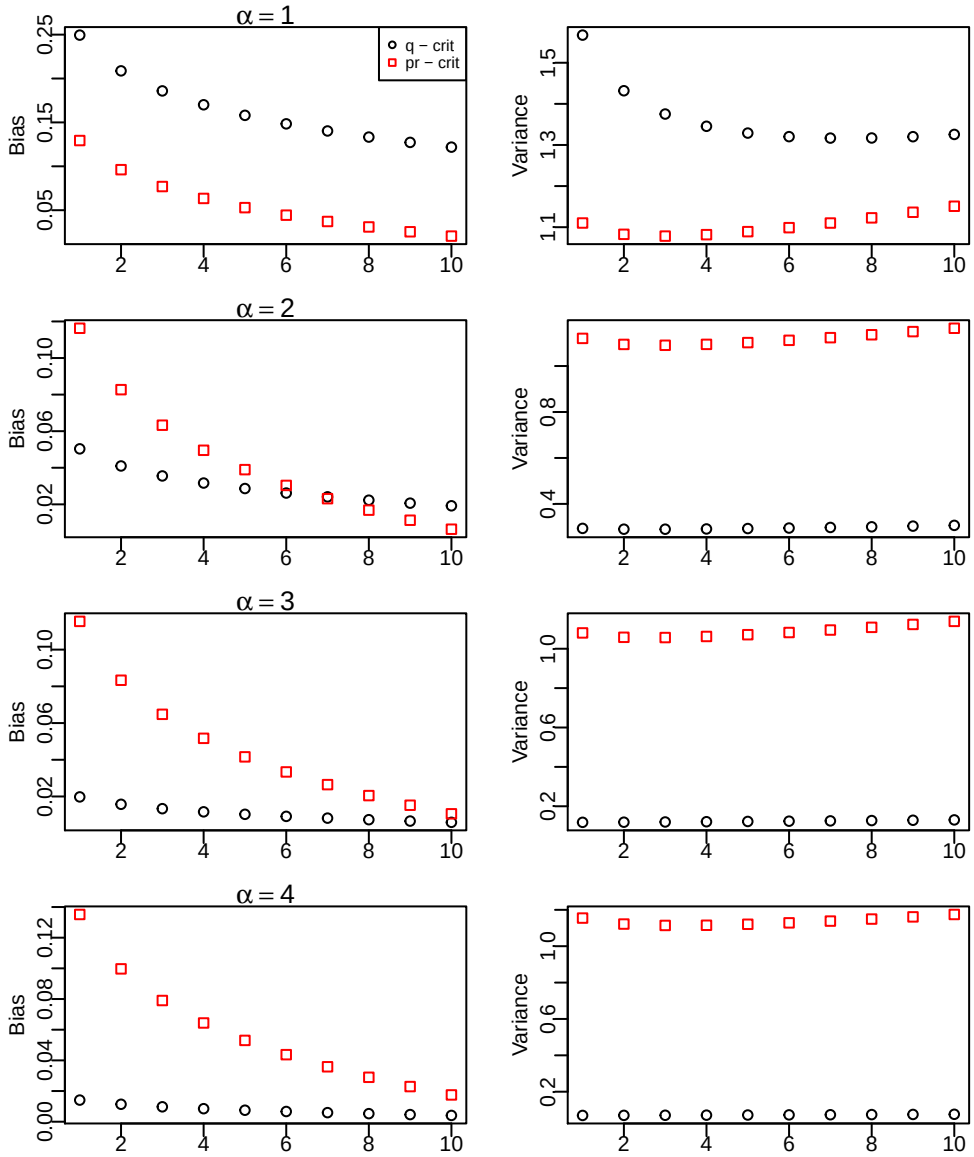
(b) Symmetric stable



(c) Fréchet

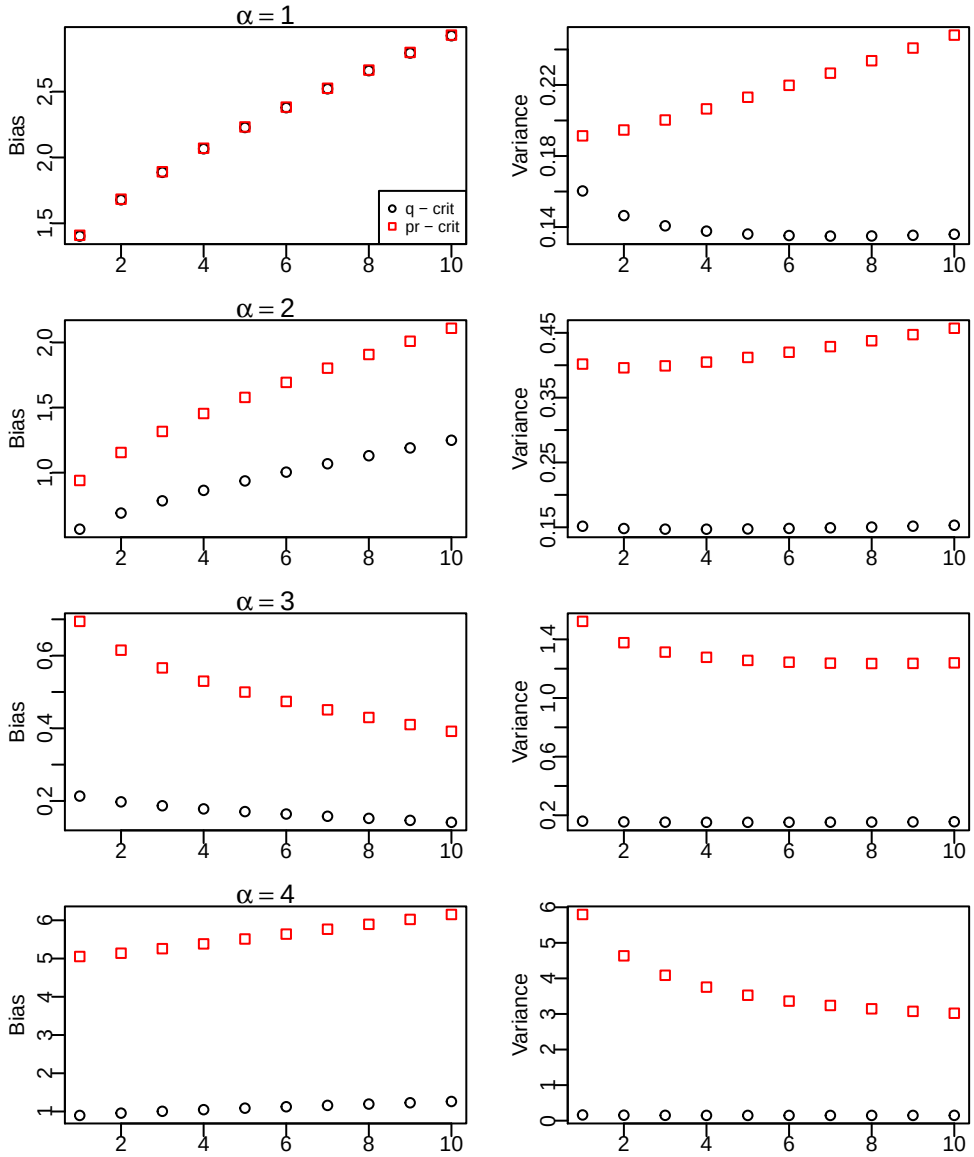
These plots show the simulations for the limit criterion function in (19). The parameters are for the **Student-t**, **symmetric stable** and **Fréchet** distribution. They can be found in Table 6 in Appendix B. The value of α for the different lines is stated in the legend. Here T is 1,500. The interval between $w(s_i) - w(s_{i+1})$ is normally distributed with mean 0 and variance $1/k$. The path of the Brownian motion is simulated 1,000 times. The left figures show the average number of order statistics at which the largest absolute distance is found for a given k (x-axis). The right figures depict the average distance found for the largest deviation at a given k (x-axis). The left and right figures are related by the fact that the right figures depict the distances found at the i^{th} observation found in the left figures for a given k.

Figure 4: Bias quantile vs probability estimator (Pareto)



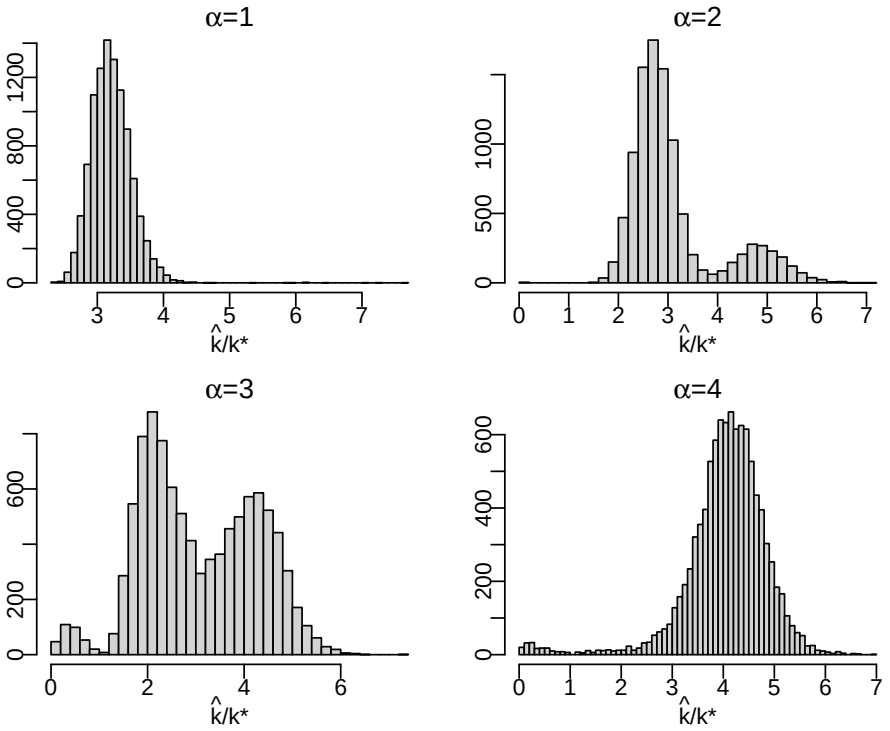
These figures illustrate the absolute bias and variance of the quantile estimator (on page 138) and the probability estimator (on page 145) from De Haan and Ferreira (2007). In each set of figures, the left column represents the bias, while the right column shows the variance of the estimators. The round black dots indicate the variance for the quantile criteria, whereas the red squares represent the bias for the probability criteria at the specified order statistic. We simulate 10,000 observations from a Pareto distribution with shape parameter α values of 1, 2, 3, and 4. The x-axis corresponds to the order statistic where the criteria are calculated, and the y-axis represents the bias of the simulated criteria.

Figure 5: Bias quantile vs probability estimator (Student-t)



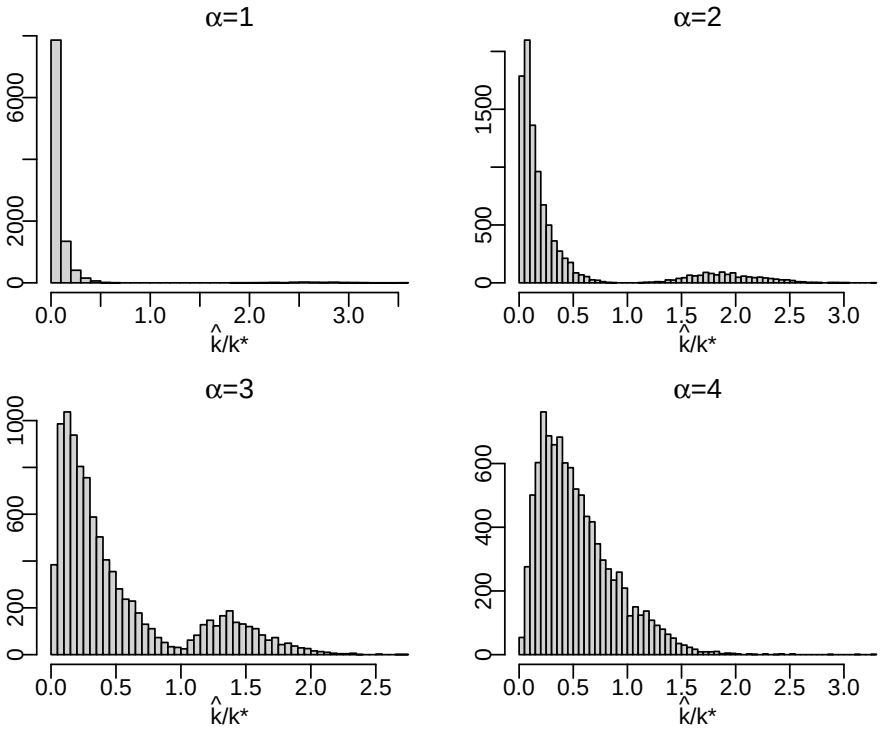
These figures illustrate the (absolute) bias and variance of the quantile estimator (on page 138) and the probability estimator (on page 145) from De Haan and Ferreira (2007). In each set of figures, the left column represents the bias, while the right column shows the variance of the estimators. The round black dots indicate the variance for the quantile criteria, whereas the red squares represent the bias for the probability criteria at the specified order statistic. We simulate 10,000 observations from a Student-t distribution with degrees of freedom (α) values of 1, 2, 3, and 4. The x-axis corresponds to the order statistic where the criteria are calculated, and the y-axis represents the bias of the simulated criteria.

Figure 6: Pareto tail and exponential centre (**KS - probability dimension**)



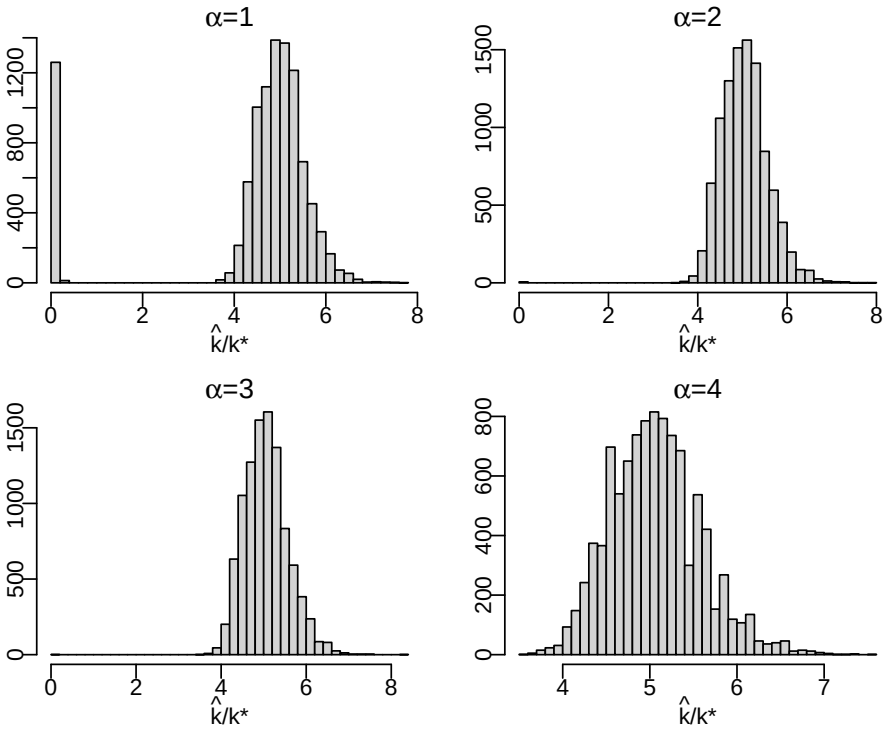
These figures depicts the ability of the KS test statistic to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, for $X \geq x_c$. We set $\lambda = 1$ and $A = exp(x_c)x_c^{\alpha}$. The sample size is 10,000 and the threshold is set to $p(x_c) = 0.99$, so that the average number of draws from the Pareto distribution is 100. We take the ratio $\hat{k}/k^*$ to normalize the choice of $\hat{k}$ relative to the benchmark, k^*. The figures are for different values of α for the Pareto part of the distribution. We draw 10,000 samples for each analysis.

Figure 7: Pareto tail and exponential centre (**Automated Eye-Ball**)



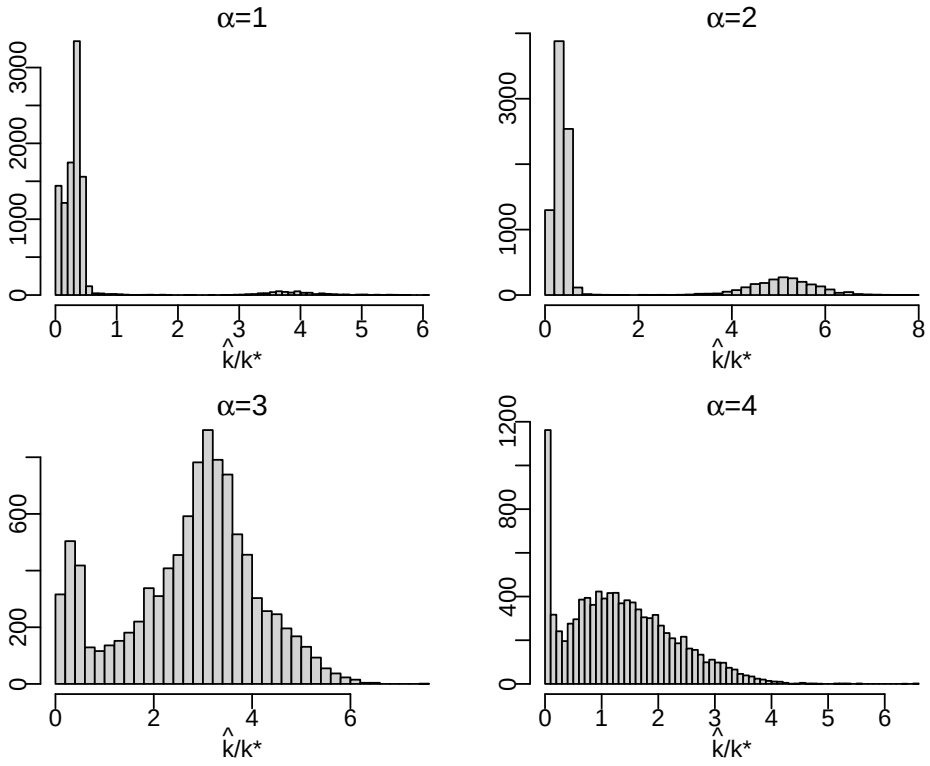
These figures depicts the ability of the Eye-Ball method to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, for $X \geq x_c$. We set $\lambda = 1$ and $A = exp(x_c)x_c^{\alpha}$. The sample size is 10,000 and the threshold is set to $p(x_c) = 0.99$, so that the average number of draws from the Pareto distribution is 100. We take the ratio $\hat{k}/k^*$ to normalize the choice of $\hat{k}$ relative to the benchmark, k^*. The figures are for different values of α for the Pareto part of the distribution. We draw 10,000 samples for each analysis.

Figure 8: Pareto tail and exponential centre (**Drees and Kaufmann**)



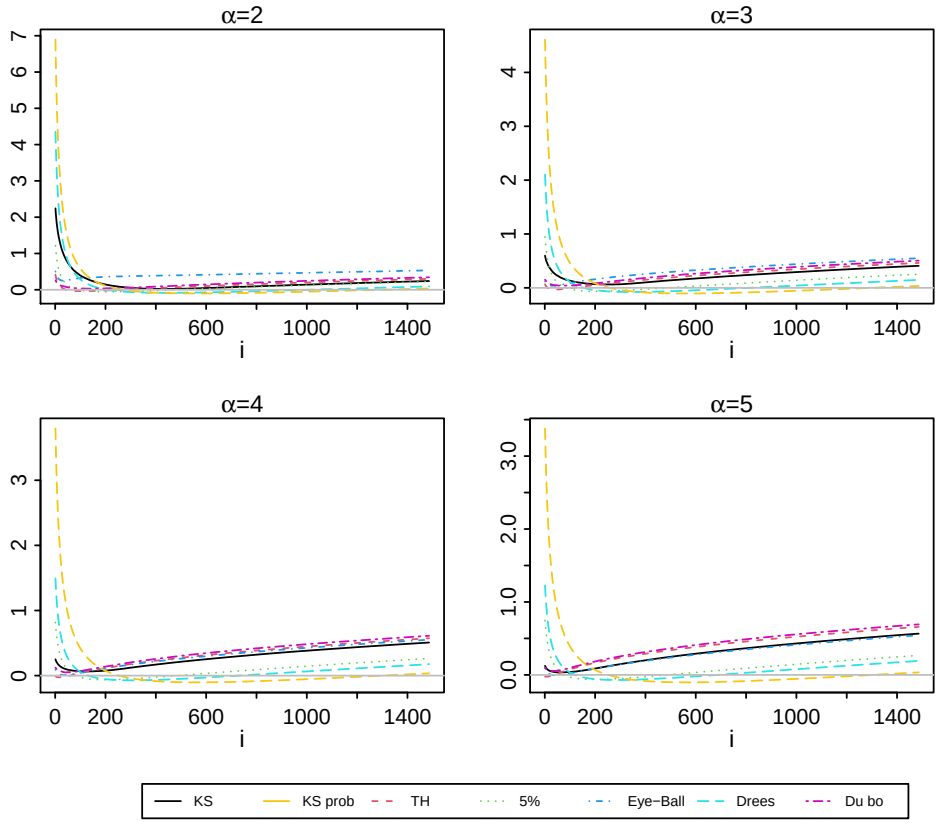
These figures depicts the ability of the method by Drees and Kaufmann to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, for $X \geq x_c$. We set $\lambda = 1$ and $A = exp(x_c)x_c^{\alpha}$. The sample size is 10,000 and the threshold is set to $p(x_c) = 0.99$, so that the average number of draws from the Pareto distribution is 100. We take the ratio $\hat{k}/k^*$ to normalize the choice of $\hat{k}$ relative to the benchmark, k^*. The figures are for different values of α for the Pareto part of the distribution. We draw 10,000 samples for each analysis.

Figure 9: Pareto tail and exponential centre (**Double bootstrap**)



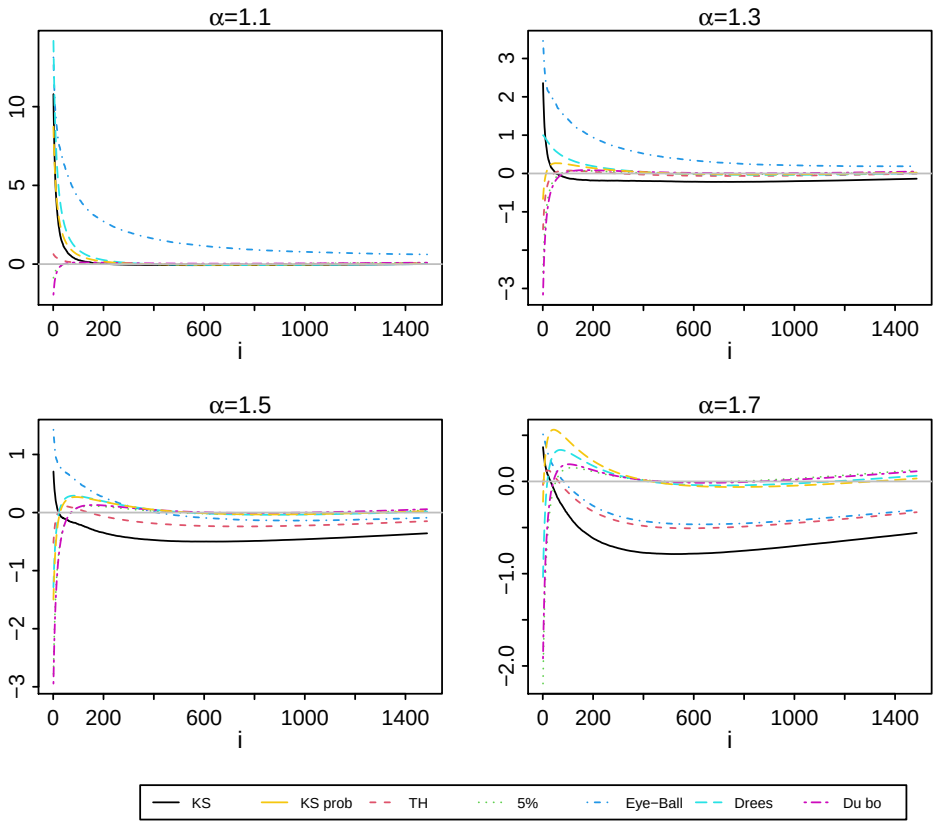
These figures depicts the ability of the double bootstrap to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, for $X \geq x_c$. We set $\lambda = 1$ and $A = exp(x_c)x_c^{\alpha}$. The sample size is 10,000 and the threshold is set to $p(x_c) = 0.99$, so that the average number of draws from the Pareto distribution is 100. We take the ratio $\hat{k}/k^*$ to normalize the choice of $\hat{k}$ relative to the benchmark, k^*. The figures are for different values of α for the Pareto part of the distribution. We draw 10,000 samples for each analysis.

Figure 10: Quantile estimation median difference (Student-t distribution)



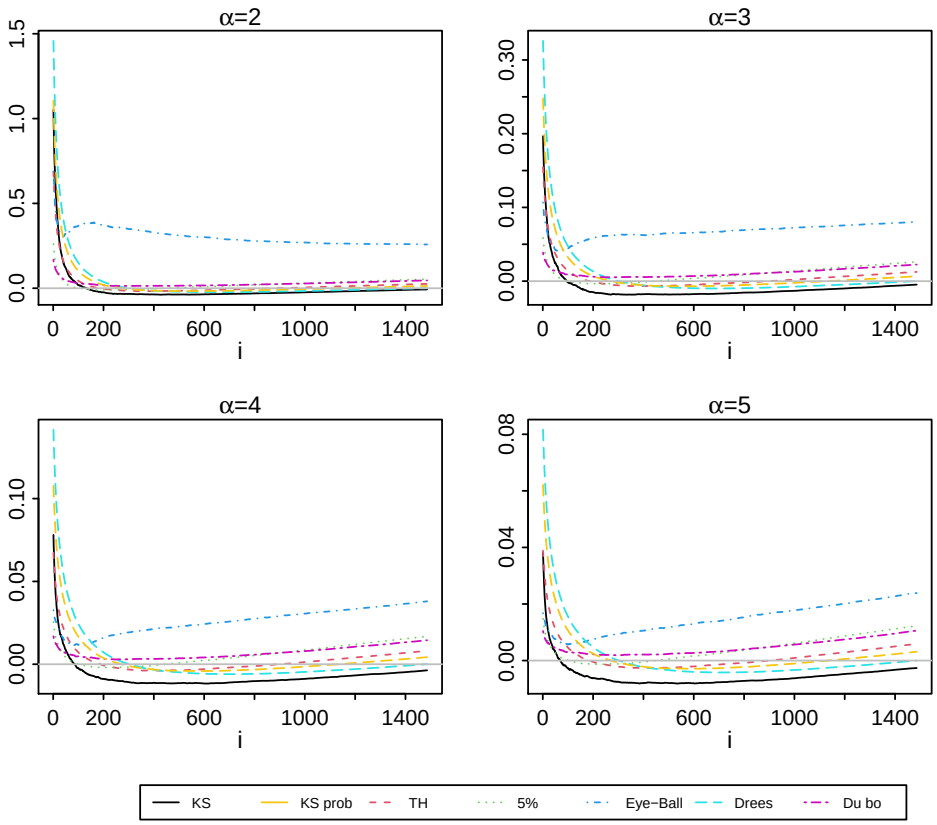
This figure shows the median deviation in quantile estimates, $X_{n-k+1}(k/j)^{1/\hat{\alpha}_k}$ presented in Equation (6). We use k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the Student-t distribution family with the shape parameter indicated at the top of the picture. The i on the horizontal axis gives the probability level $^{(n-i)}/_n$ at which the quantile is estimated. Moving rightwards along the x-axis represents a move towards the centre of the distribution.

Figure 11: Quantile estimation median difference (symmetric stable distribution)



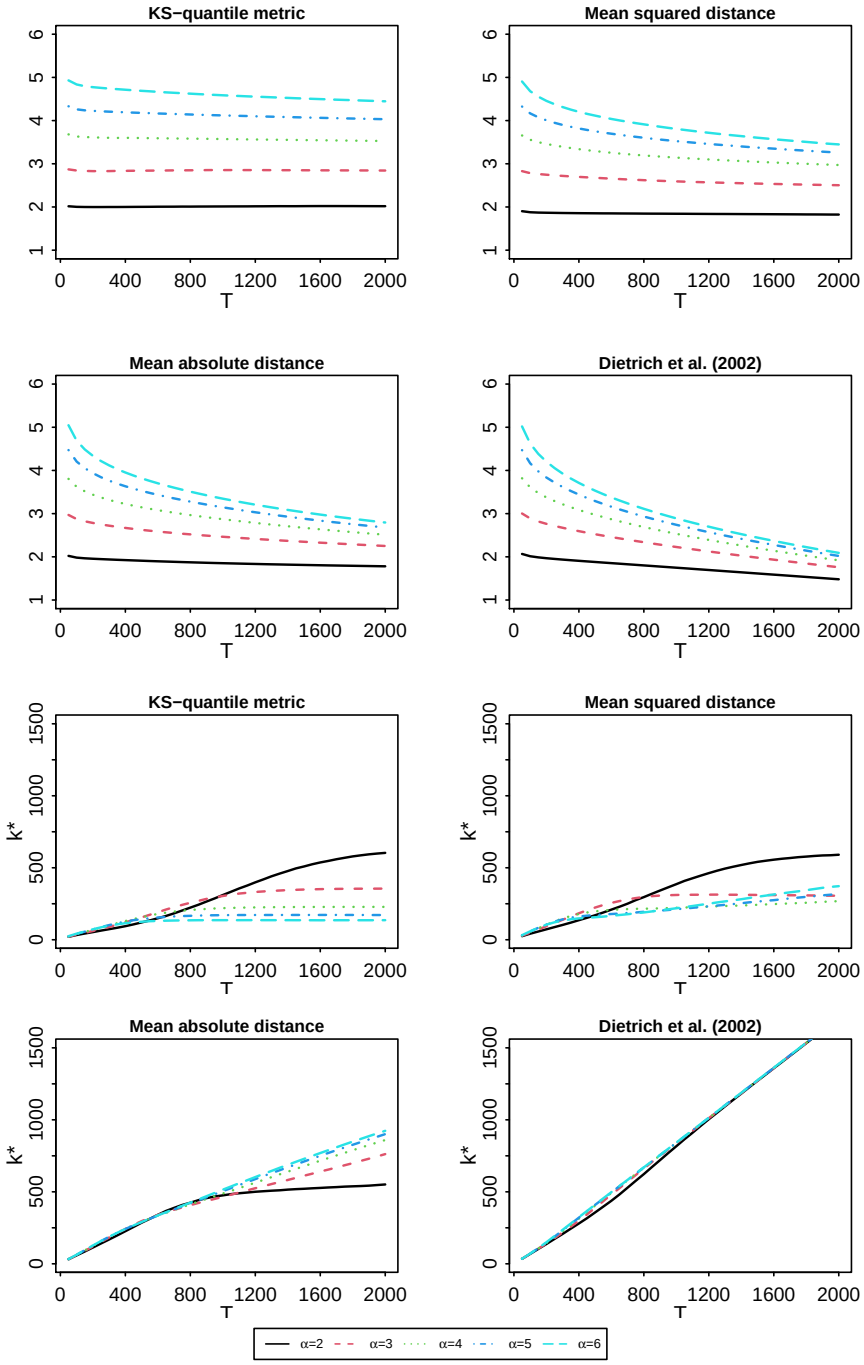
This figure show the median difference induced by using the quantile estimator presented in Equation (6). We use the k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the symmetric stable distribution family with the shape parameter indicated at the top of the picture. The i on the horizontal axis gives the probability level i/n at which the quantile is estimated. Moving rightwards along the x-axis represents a move towards the center of the distribution.

Figure 12: Quantile estimation median difference (Fréchet distribution)



This figure show the median difference induced by using the quantile estimator presented in Equation (6). We use the k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the Fréchet distribution family with the shape parameter indicated at the top of the picture. The i on the horizontal axis gives the probability level i/n at which the quantile is estimated. Moving rightwards along the X-axis represents a move towards the center of the distribution.

Figure 13: $\hat{\alpha}\,(k^*)$ and k^* for quantile metrics (Student-t distribution)



This figure depicts simulation results of the average k^* and $\alpha(k^*)$ for a given level of T for different metrics. Here T is the number of extreme-order statistics over which the metric is optimized. The upper four graphs depict the optimal $\alpha(k^*)$ and the lower four graphs show the choice of k^* for different values of T. We use the KS-quantile metric, mean squared distance, mean absolute distance and the criteria used by Dietrich et al. (2002). For the simulations we draw a sample of 10,000 from a **Student-t**(α) distribution.