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

April 2025

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.

Keywords: Tail index estimation, Power laws, 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. There are various practices to elicit k. We denote by k^* that k for which the particular method at hand selects their chosen threshold. 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 theory-based 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. Rather than minimizing the maximum probability difference, we use the Weissman (1978) quantile estimator to minimize 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 relative to its true counterpart has a smaller bias and variance by a factor of $1/\alpha$ vis-a-vis the probability estimator relative to the true probability.[1] Furthermore, as one moves deeper into the tail area an epsilon mistake in the probability dimension leads to an increasingly larger mistake in the quantile dimension. Therefore, focusing our metric on the quantile domain naturally emphasizes fitting the tail rather than central observations. Moreover, by concentrating on the maximum deviation in the quantile dimension, our metric further avoids being influenced by the numerous central observations.

The arguments are supported by rigorous simulation analyses and tests. We first subject each method to the acid test suggested by Drees et al. (2020). In the acid test we use a distribution that combines a Pareto distributed tail part with an exponentially distributed part in the centre. The threshold where the two parts meet is such that the distribution function is continuous at that point. In simulating from this distribution, data come from both parts. Since the Hill estimator is only applicable in the Pareto part, if part of the data in the estimator come from the exponential part this would give a large bias. The question is whether the proposed methodologies for selecting k^* are able to locate it close to the threshold without using exponentially distributed data from the centre. Thus, this acid test evaluates the capability of the metrics to distinguish between thin-tailed distributed center observations and heavy-tailed tail observations.

For the analysis below we compare the following methods for finding the optimal threshold for the Hill estimator. We use two methods that are motivated by balancing the asymptotic bias and variance of the Hill estimator. These are the double bootstrap method by Danielsson et al. (2001) and the iterative method by Drees and Kaufmann (1998). The second group of methods are heuristic in nature. We apply an automated form of the Eye-Ball (Resnick and Starica, 1997) method. For the other heuristic rule we pick a fixed

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

percentage of the total sample size, for instance 5% of the upper-order statistics as is often done in practice. Additionally, we use a metric over the probability dimension as in Clauset et al. (2009). Lastly, we introduce the theoretically AMSE minimizing threshold for the simulated distributions as a benchmark.

We find from the simulations that 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 simulated data. Since the Hill estimator is the maximum likelihood estimator for the Pareto distribution, using all the Pareto tail observations is optimal in terms of minimizing the variance. Drees et al. (2020) notes that 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 to 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 economically 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. Small differences in α estimates do matter. For example, if $\hat{\alpha} < 2$, the implied second moment is unbounded. But if the Hill estimate exceeds four, even the variance of the volatility is bounded.

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 key results from extreme value theory (EVT) and the Hill estimator. These serve as the foundation 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 starts to kick in. The bias and variance of the estimator can be derived for the subclass of distributions in (1) that satisfy the

Hall (1982) expansion[2]

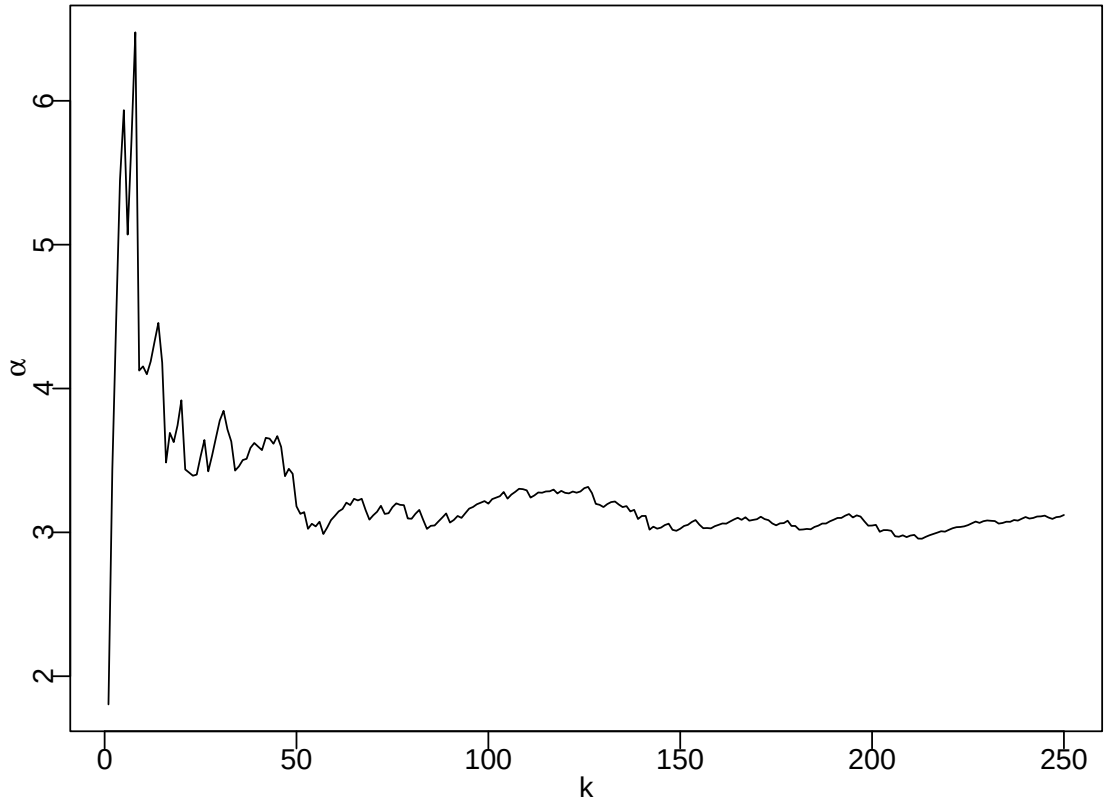
$$1 - \mathrm{F}(x) = Ax^{-1/\gamma}\left[1 + Bx^{-\beta} + o\left(x^{-\beta}\right)\right]. \tag{3}$$

For distributions that adhere to the Hall expansion, one shows that the asymptotic bias

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

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 shows α estimates across different k levels. The 10,000 sampled observations are drawn from a Student-t(4) distribution with $\alpha = 4$. This type of graph is commonly known as a Hill plot.

[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 3 in Appendix A.1.

The asymptotic variance of the Hill estimator is[3]

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

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.

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 theoretically optimal k which minimizes the AMSE, varies across different distributions and exhibits different rates of convergence. For details on existing methods, see Appendix A.2.

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. Following De Haan and Ferreira (2007) and to better capture the most extreme observations, where small probability differences yield large quantile shifts, we measure distance in the quantile dimension.

[3]The expressions for bias and variance are based on the second-order expansion by Hall and Welsh (1985) in (3). For further details, see Appendix A.1.

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 in (3) are ignored. 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:

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

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:

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

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

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 $q(j,k)$ 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:

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

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 we ultimately observe. 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

[5] In Appendix A.3, 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 on parameters in (3) 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 across parameter settings.

[6] For figures 4 and 5 in Appendix C, 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 if $\alpha < 1$ 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.

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.

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 concerns the acid test to see whether the metric is misled by the exponential part that is glued to the heavy-tailed Pareto part. The second simulation compares competing methods from the literature in a horse race.

3.1 Acid test

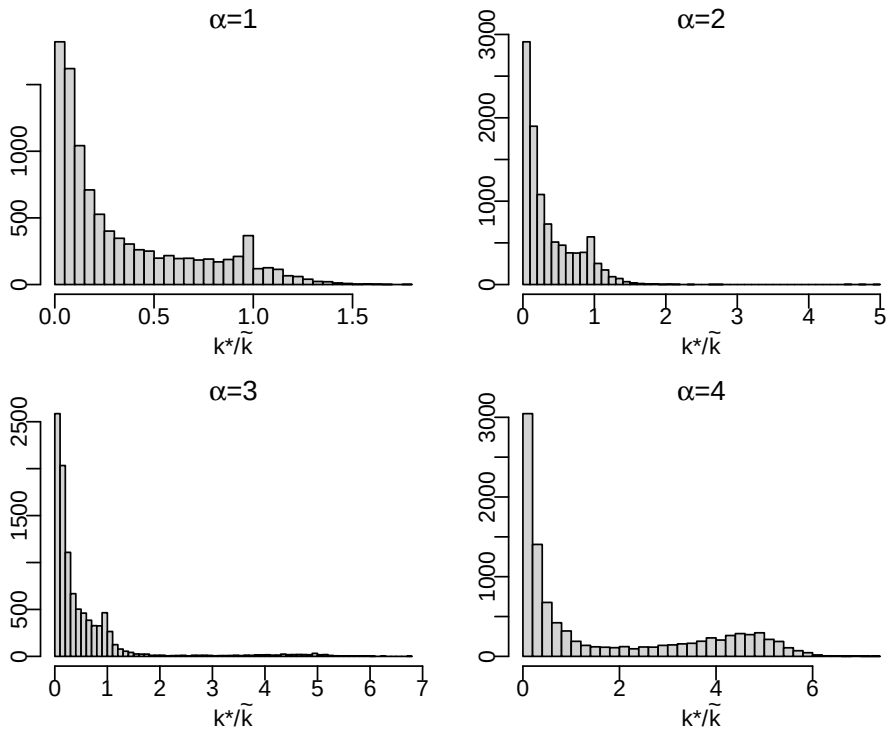
In the first simulation exercise, we draw samples from an exponential distribution with a Pareto tail beyond a threshold $X_{n-\tilde{k},n}$. A candidate method should avoid selecting a threshold that includes many observations from the thin-tailed exponential part of the distribution. As Drees et al. (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 $\tilde{k}$ order statistics are used, the chosen threshold optimally slightly exceeds $\tilde{k}$ as this reduces variance.

Figure 2 depicts the ratio of estimated number of order-statistics with the KS-quantile metric, k^*, to the number of observations drawn from the Pareto distribution, $\tilde{k}$. For $\alpha = 1$, the ratio, $k^*/\tilde{k}$, is below 1 for 92.63% of the simulations. For α equal to 2, 3 and 4 the percentages decrease to 92.00%, 87.83% and 59.09%, respectively. This indicates that, in the majority of 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. To conclude, it appears that the KS-quantile method passes the acid test devised by Drees et al. (2020).

We also confirm that applying the KS distance in the probability dimension does not perform well, as is shown by Drees et al. (2020). 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. 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.

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 $k^*/\tilde{k}$ is used to normalize the choice of k^* relative to the benchmark $\tilde{k}$. The figures show results for different values of α, as indicated above the histograms. We draw 10,000 samples for each analysis.

It is intuitive why metrics penalizing probability deviations are less suited for fitting heavy-tailed distributions. A simple Monte Carlo simulation illustrates this: simulating from a uniform distribution, the maximal vertical and horizontal deviations occur at the same order statistic. Deviations between the simulated and theoretical quantiles are proportional to differences in the probability dimension. However, introducing a kink at probability 1/4, where

$$F(x) = \begin{cases} \frac{x}{4}, & \text{for } x < 1 \\ \frac{3x-2}{4}, & \text{for } x > 1, \end{cases}$$

changes this pattern. As an abstraction, the slow increase in the left tail reflects a heavy tail, while the steeper slope beyond $x = 1$ mimics an exponential structure. The largest quantile deviations now occur at lower-order statistics, whereas the largest probability deviations appear at higher-order statistics. Because of this, fitting heavy tails is more effective when minimizing the horizontal distance between the empirical distribution and the theoretical tail shape.

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).[7] Consequently, we know both α and k^*_{TH}, where k^*_{TH} minimizes 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

[7]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 distribution it is the shape parameter. These distribution parameters are henceforth referred to as α. See Appendix A.1 for the derivation of the optimal theoretical asymptotic MSE based threshold.

the distribution of the stationary solution are known, as well as the second-order scale parameter. The first-order scale parameter A is harder to pin down, but Goldie (1991) derives an explicit expression when α is 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.[8] 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.2. 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 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 methods. However, the automated Eye-Ball method performs only slightly worse than the KS-quantile metric.

[8]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.

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 Kaufmann (1998) and the double bootstrap method exhibit a similar decreasing pattern, though the average k^*_{DK} 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 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 perfor-

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

	α	KS qua	KS prob	TH	5%	Eye-Ball	DK	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	$5.1 * 10^{10}$
$E[k^*]$	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 (A.1) 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 (10). For the column labeled 'DK', 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.

mance of these methods in estimating tail quantiles of the distribution.[9] 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 Drees and Kaufmann (1998), the 5% fixed

[9]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.

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 (results available on request).

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 a good 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.

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.[10] 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 the latter two methods. 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 cast doubt on the suitability of fixed-threshold and AMSE-based approaches for empirical estimation.

[10] For descriptive statistics on tail index estimates and thresholds for both tails, refer to Table 7 in the Appendix. Table 8 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	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	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
DK	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	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	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
DK	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 i's 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 $\hat{\alpha} > 1,000$ are excluded from the analysis for any of the methods. The maximum k_i^* is cut off at 15% of the total sample size. There are 17,918 stocks included in the analysis.

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, both in terms of chosen threshold and tail index estimate. Although the choice of threshold might seem innocuous, it impacts the interpretation of economic processes, underscoring the importance of carefully selecting it 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.

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.

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.

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

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

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 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

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

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:

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

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.[11] See Table 3 for distribution-specific parameters.

[11] A detailed derivation of the values of ARCH parameters is available on request.

Table 3: Parameter values Hall expansion

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

	ARCH					
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

(b) Parameters ARCH process

This table states the paprameter values for the hall expansion of the different distribution families and the ARCH process. Furthermore, b is the coefficient in the ARCH process on the squared error term.

A.2 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,

$$k_{\text{du bo}}^* = \frac{\left(k_2\right)^2}{k_1}\left[\frac{\log\left(k_1\right)^2}{\left(2\log\left(n_1\right)-\log\left(k_1\right)\right)^2}\right]^{\frac{\log(n_1)-\log(k_1)}{\log(n_1)}}.$$

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. Leading to the adaptive estimator

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

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:

$$k^*_{eye} = \min \left\{ k \in 2, ..., n^+ - w | h < \tfrac{1}{w} \sum\nolimits_{i=1}^{w} \mathrm{I}\left\{ \widehat{\alpha}\left(k+i\right) < \widehat{\alpha}\left(k\right) \pm \varepsilon \right\} \right\}. \quad (10)$$

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.

A.3 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, p. 52) the KS-quantile metric in (7) can be written as

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

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

$$\arg\min_{0<k<T} \sup_{0<l<\frac{T}{k}} \left| \frac{\gamma}{\sqrt{k}} U\left(\frac{n}{k}\right) l^{-\hat{\gamma}} \left[l^{-1} w(l) - w(1) + A_0\left(\frac{n}{k}\right) \frac{\sqrt{k}}{\gamma} \frac{l^{-\rho}-1}{\rho} \right] \right|, \tag{11}$$

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, p. 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, stated in Bingham et al. (1989, page 29),

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

and

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

Simulating from (11) 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 when using the parameters in (3) from the Student-t, symmetric stable and Fréchet distributions. 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 the extremes. Conversely, for smaller k, the largest deviations are 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 desirable as it minimizes the bias of the Hill estimator. Since the modeled $\hat{\gamma}$ is based on the expectation of the Hill estimator, where only the bias plays a role, it is not surprising that the limit in (11) 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).

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. Such penalty functions are common in econometrics, but averaging tends to overweight the many central observations.

he third metric is based on the test statistic by Dietrich et al. (2002), which assesses

whether extreme value conditions hold. A discretized version of their statistic yields:

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

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 11, 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 11 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 $\alpha\left(k^*\right)$. 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. This matches the pattern of k^*_{TH} that minimizes AMSE under the Student-t family, unlike the other criteria.

B Tables

Table 4: Horse Race **symmetric stable** distribution family

	α	KS qua	KS prob	TH	5%	Eye-Ball	DK	Du bo
	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
$E[\hat{\alpha}]$	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
	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
$E[\hat{\alpha}^2]$	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
	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
$E[\hat{\alpha}^3]$	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
	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	$1.9 * 10^4$
$E[\hat{\alpha}^4]$	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	$2.4 * 10^5$
	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
$E[k^*]$	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 (A.1) in the Appendix. The automated Eye-Ball method in (10) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column DK, 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 5: Horse Race **Fréchet** distribution family

	α	KS qua	KS prob	TH	5%	Eye-Ball	DK	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
$E[k^*]$	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 (A.1) in the Appendix. The automated Eye-Ball method in (10) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column DK, 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 6: Horse Race **ARCH** processes

	α	KS qua	KS prob	TH	5%	Eye-Ball	DK	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
$E[k^*]$	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 (A.1) in the Appendix. The automated Eye-Ball method in (10) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column DK, 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 7: Descriptive statistics stock data estimates

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	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	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	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 i's 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_i is cut off at 15% of the total sample size. There are 17,918 companies included in the analysis.

Table 8: Median absolute differences between different methods

	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	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
DK	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 $\alpha\left(k^*\right)_i$

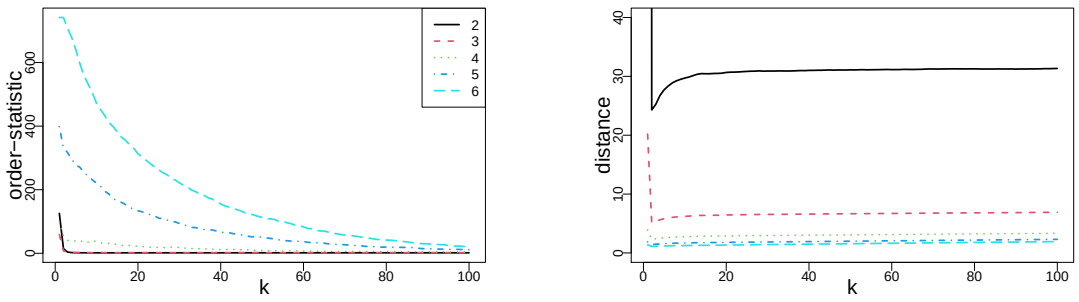
	Left Tail						Right Tail					
	KS qua	KS pr	5%	Eye-Ball	DK	Du bo	KS qua	KS pr	5%	Eye-Ball	DK	Du bo
KS qua	0	158	95	37	169	56	0	178	97	32	181	47
KS pr	158	0	99	219	48	222	178	0	100	219	45	230
5%	95	99	0	122	103	128	97	100	0	120	104	135
Eye-Ball	37	219	122	0	232	29	32	219	120	0	227	30
DK	169	48	103	232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

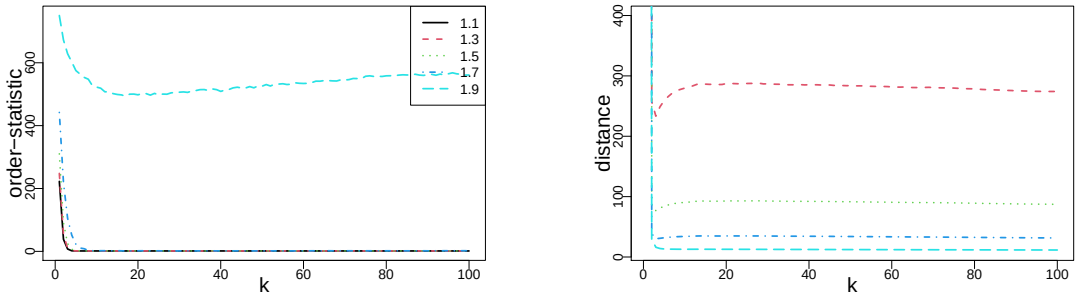
This table presents the median absolute difference between $\hat{\alpha}\left(k^*\right)_i$ and k_i^* by applying the six different methods to choose k_i^* for left and right tail of stock i's returns. The different methods are the KS-quantile metric, KS test statistic, 5% threshold, automated Eye-Ball 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_i is cut off at 15% of the sample size. There are 17,918 companies included in the analysis.

C Figures

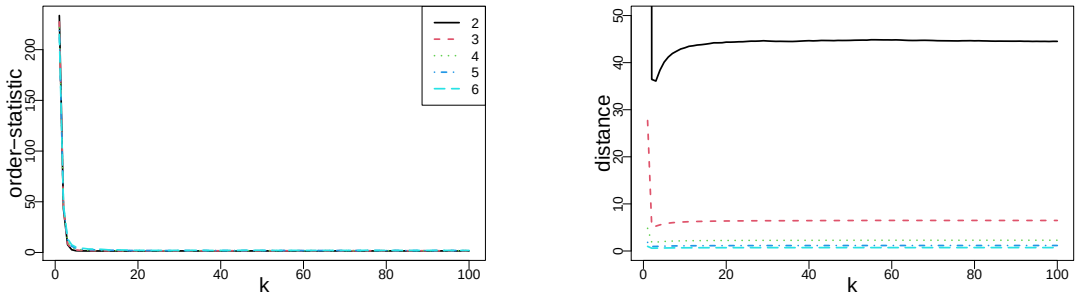
Figure 3: Simulations Brownian motion representation



(a) Student-t



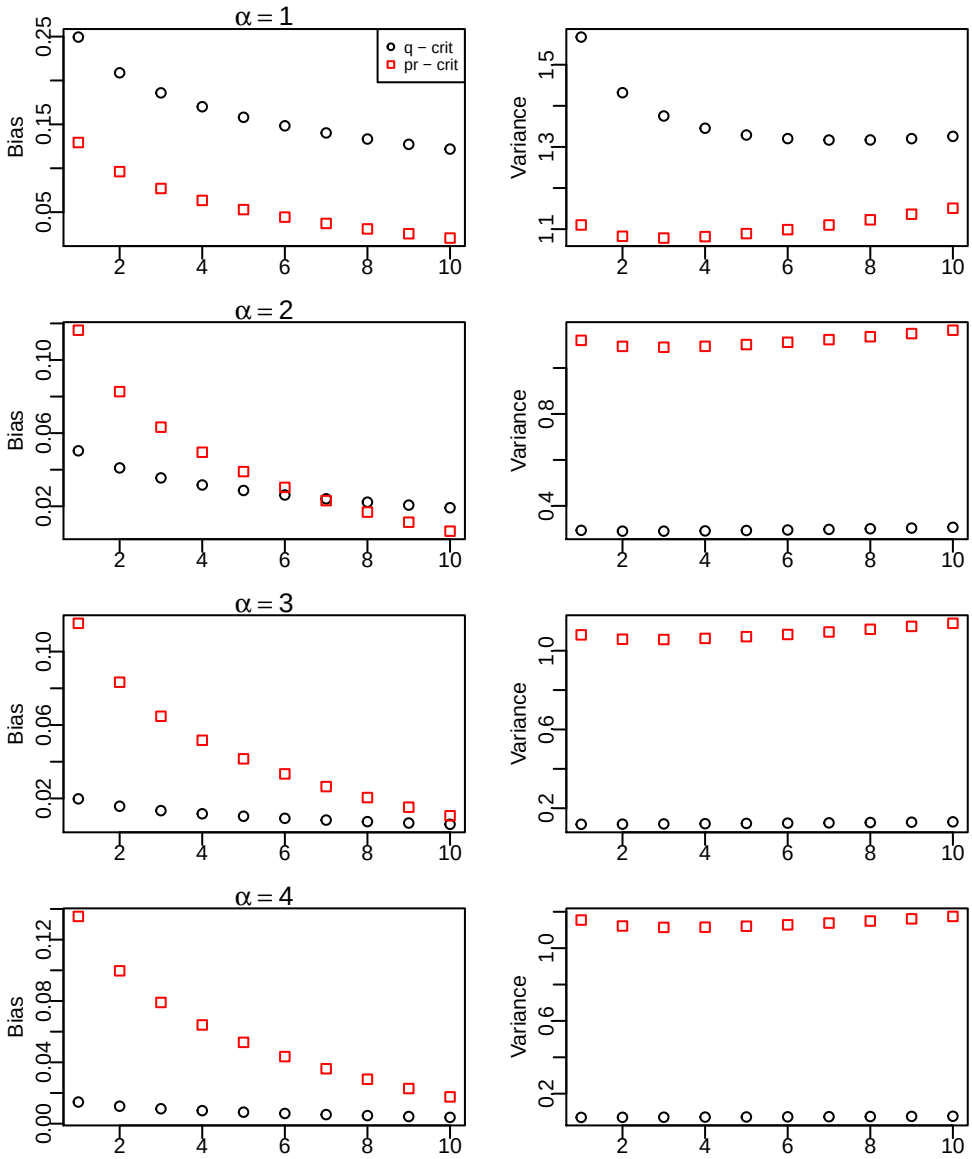
(b) Symmetric stable



(c) Fréchet

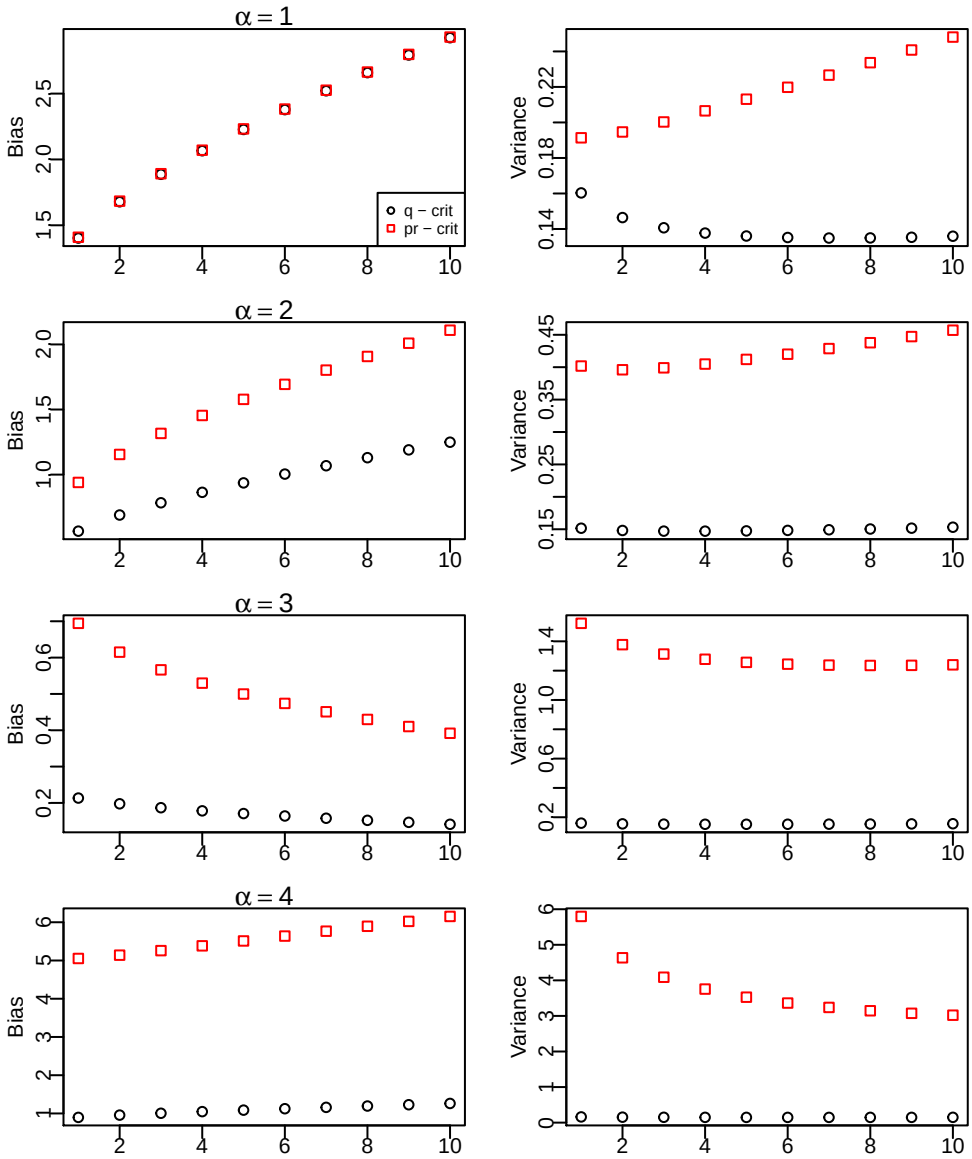
These plots show the simulations for the limit criterion function in (11). The parameters are for the **Student-t**, **symmetric stable** and **Fréchet** distribution. They can be found in Table 3 in Appendix A.1. The value of α for the different lines is stated in the legend. Here T is 1,500. The interval 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)



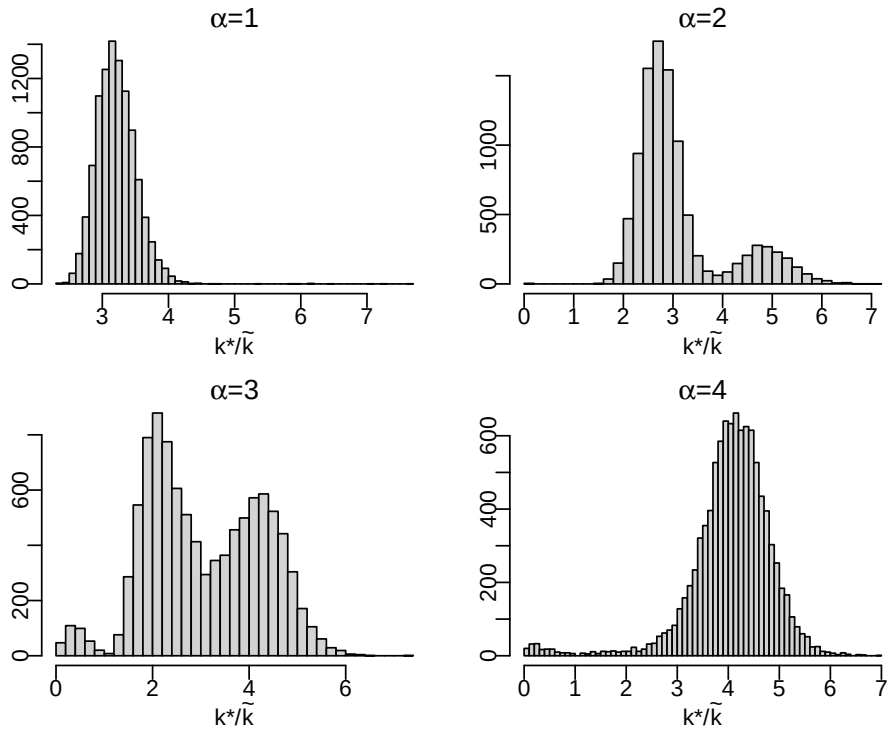
These figures illustrate the absolute bias and variance of the quantile estimator (on page 138) and the probability estimator (on page 145) from De Haan and Ferreira (2007). In each set of figures, the left column represents the bias, while the right column shows the variance of the estimators. The round 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)



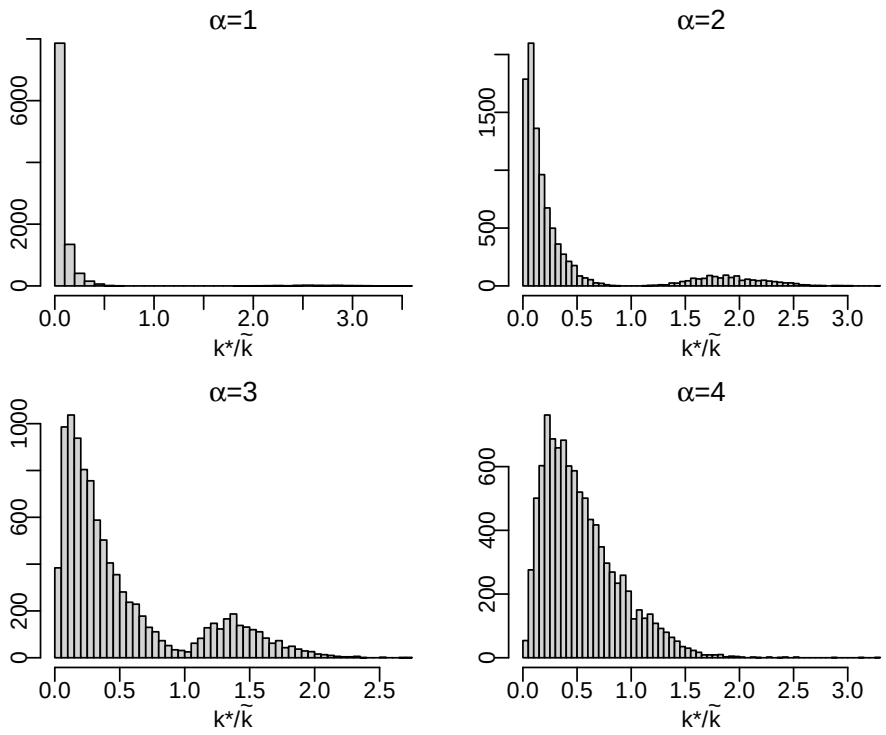
These figures illustrate the (absolute) bias and variance of the quantile estimator (on page 138) and the probability estimator (on page 145) from De Haan and Ferreira (2007). In each set of figures, the left column represents the bias, while the right column shows the variance of the estimators. The 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**)



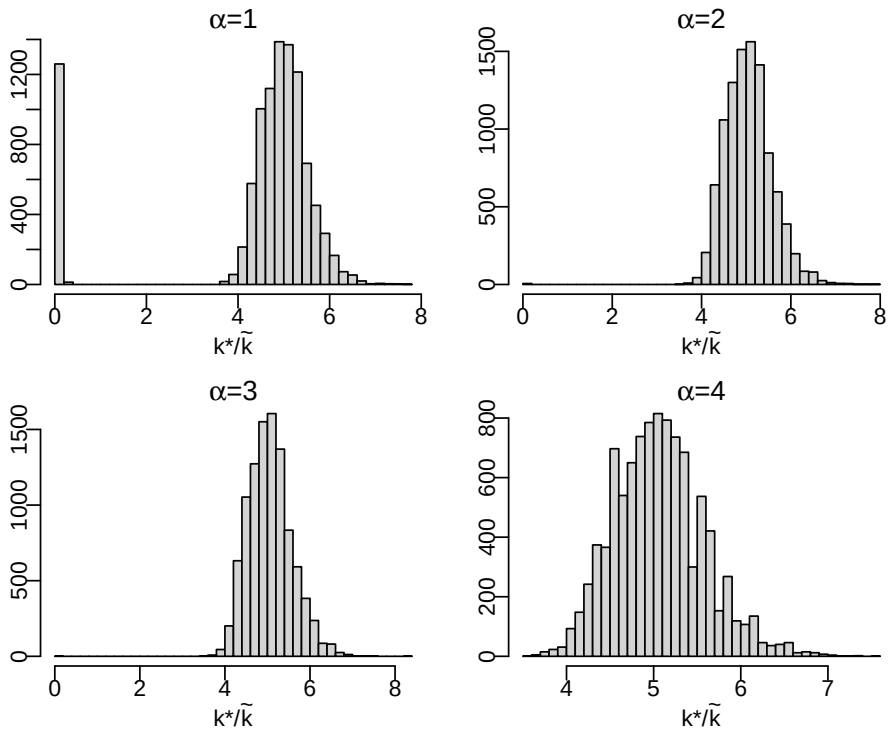
These figures depicts the ability of the KS test statistic to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, 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 $^{k^*}/_{\tilde{k}}$ to normalize the choice of k^* relative to the benchmark, $\tilde{k}$. 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**)



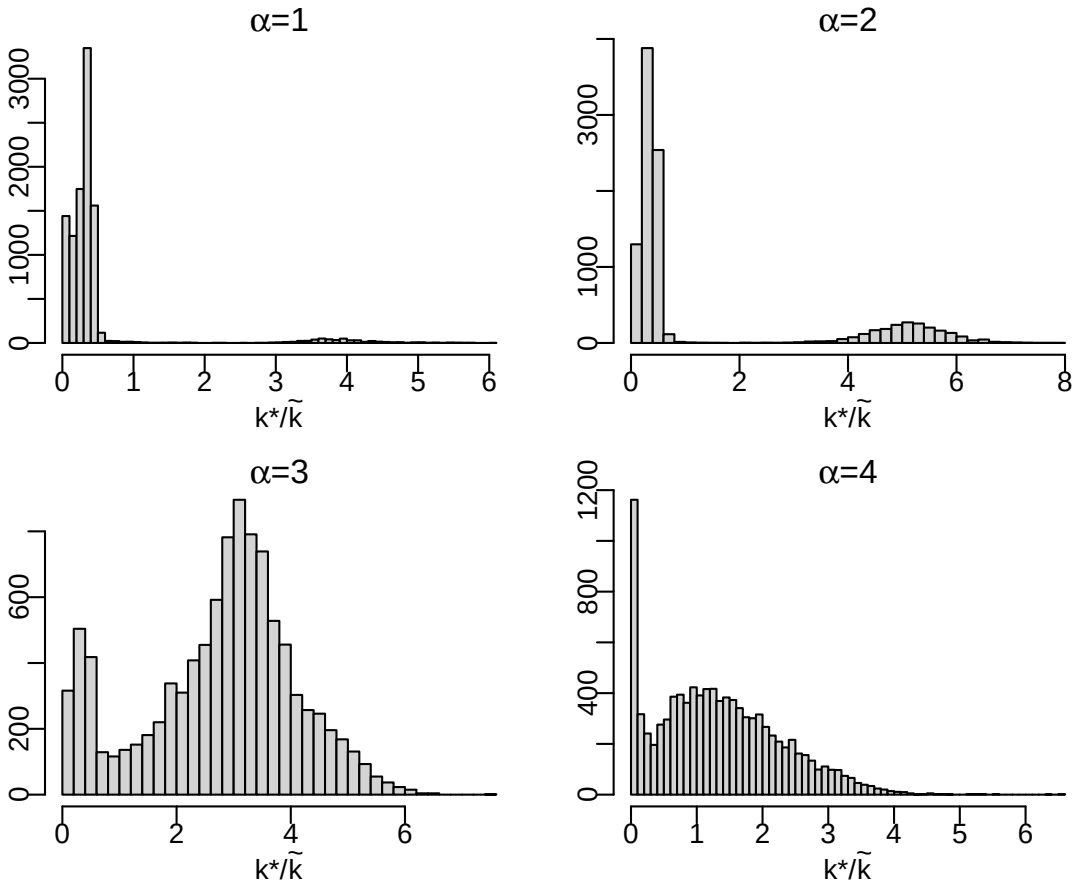
These figures depicts the ability of the Eye-Ball method to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, 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 $^{k^*}/_{\tilde{k}}$ to normalize the choice of k^* relative to the benchmark, $\tilde{k}$. 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**)



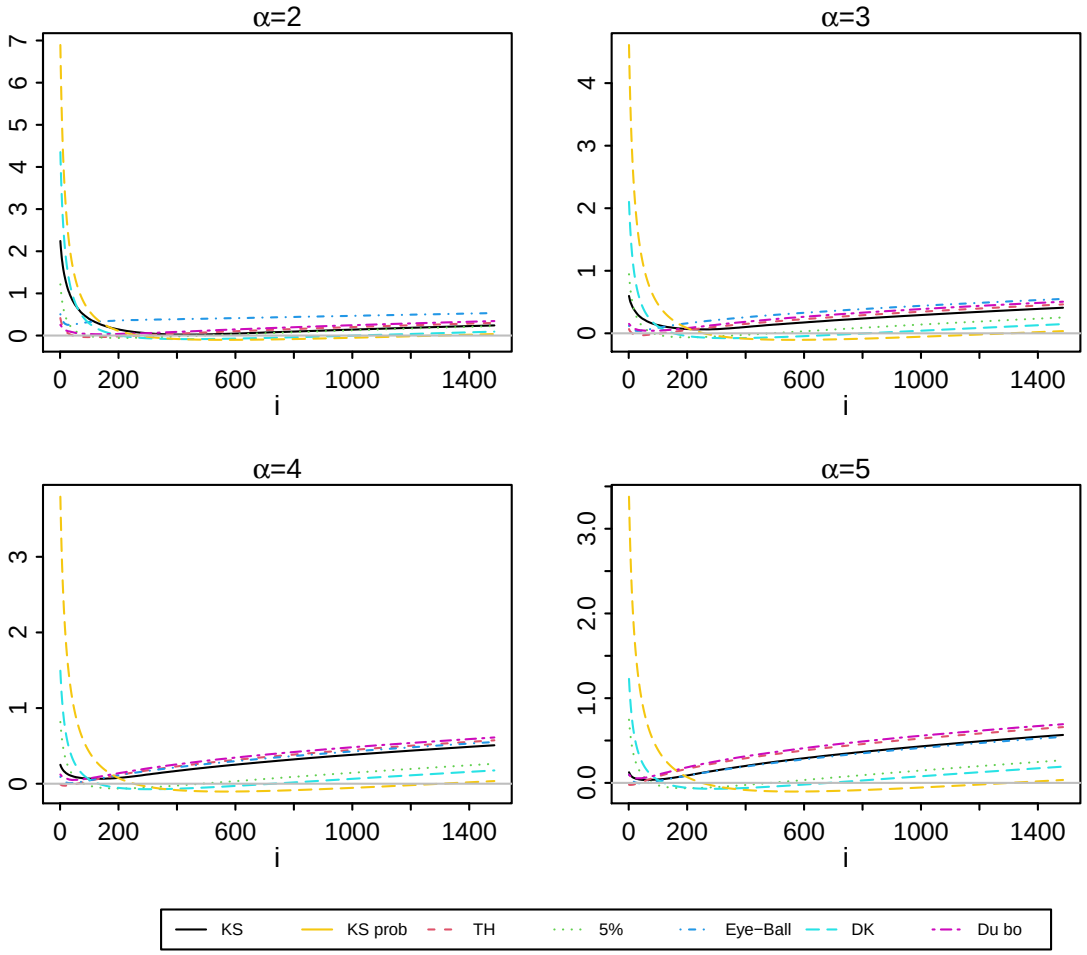
These figures depicts the ability of the method by Drees and Kaufmann to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, 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 $k^*/\tilde{k}$ to normalize the choice of $\hat{k}$ relative to the benchmark, $\tilde{k}$. 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**)



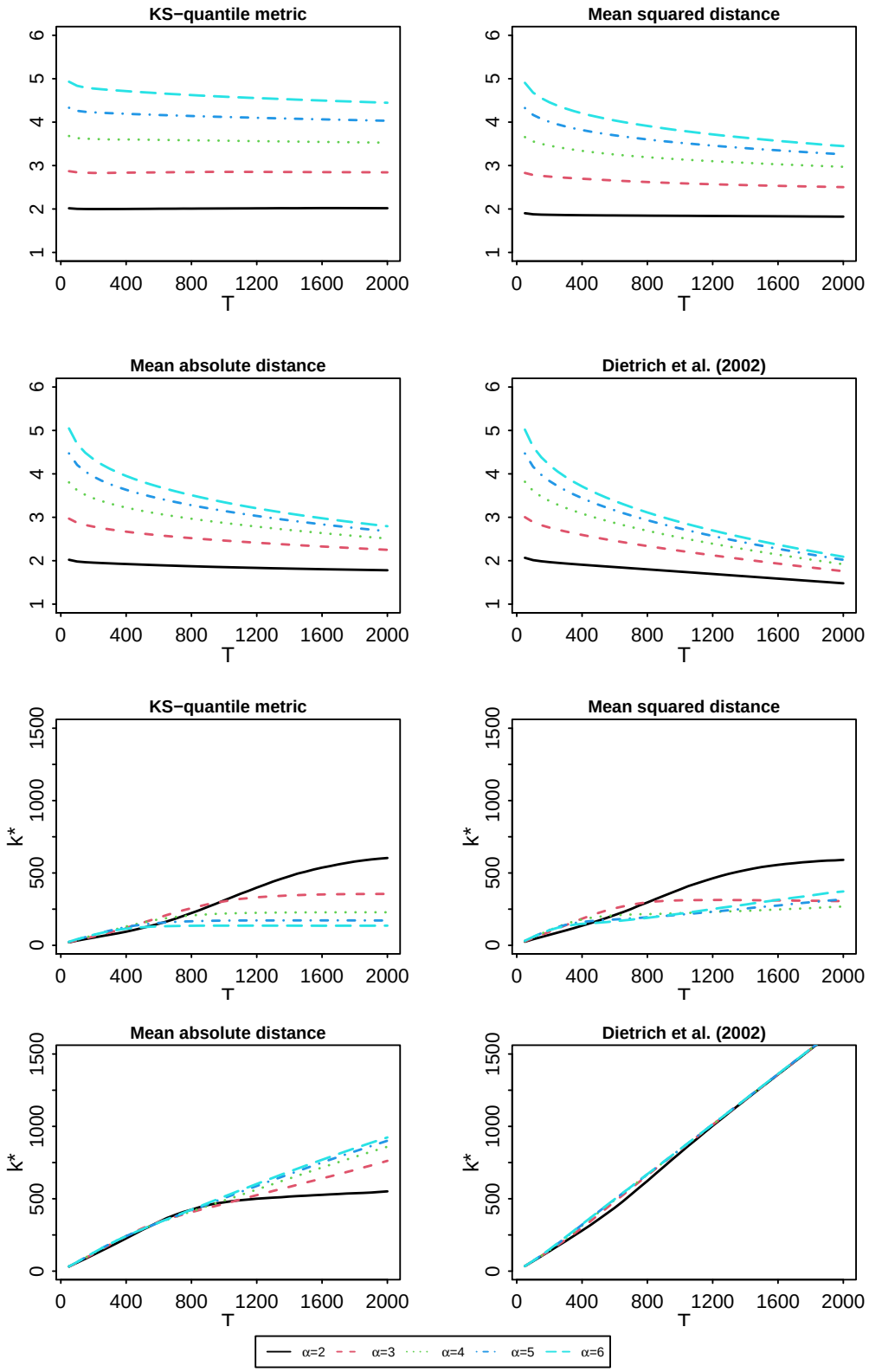
These figures depicts the ability of the double bootstrap to locate the starting point of a heavy tail in an empirical distribution. The samples in these figures are drawn from an exponential distribution, $p(x) = exp(-\lambda x)$, for $X < x_c$ and a scaled Pareto distribution, $p(x) = Ax^{-\alpha}$, 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 $^{k^*}/_{\tilde{k}}$ to normalize the choice of k^* relative to the benchmark, $\tilde{k}$. 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: $\hat{\alpha}(k^*)$ 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.