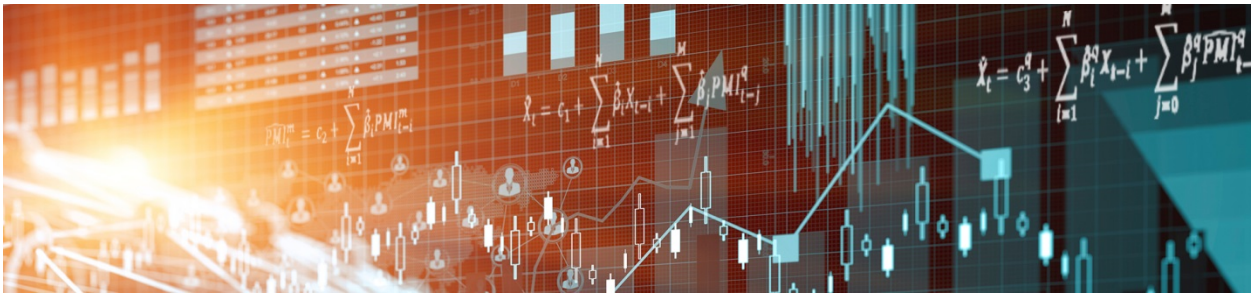
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2019-28

Tail Index Estimation: Quantile-Driven Threshold Selection



by Jon Danielsson, Lerby M. Ergun, Laurens de Haan and Casper G. de Vries

Bank of Canada Staff Working Paper 2019-28

August 2019

Tail Index Estimation: Quantile-Driven Threshold Selection

by

Jon Danielsson,[1] Lerby M. Ergun,[2] Laurens de Haan[3] and Casper G. de Vries[4]

[1] London School of Economics

[2] Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
lergun@bankofcanada.ca

[3] Erasmus University Rotterdam

[4] Erasums University Rotterdam
Tinbergen Institute

ISSN 1701-9397

Acknowledgements

Corresponding author Lerby M. Ergun. We thank the Netherlands Organisation for Scientific Research Mozaiek grant [grant number: 017.005.108] and the support of the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre [grant number ES/K002309/1]

Abstract

The selection of upper order statistics in tail estimation is notoriously difficult. Methods that are based on asymptotic arguments, like minimizing the asymptotic MSE, do not perform well in finite samples. Here, we advance a data-driven method that minimizes the maximum distance between the fitted Pareto type tail and the observed quantile. To analyze the finite sample properties of the metric, we perform rigorous simulation studies. In most cases, the finite sample-based methods perform best. To demonstrate the economic relevance of choosing the proper methodology, we use daily equity return data from the CRSP database and find economically relevant variation between the tail index estimates.

Bank topics: Financial stability; Econometric and statistical methods
JEL codes: C01, C14, C58

1 Introduction

In various research fields, like geography and economics, distributions exhibit heavy tails, e.g., the scaling behaviour described by the power laws 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 thinner tail). The most popular estimator for the tail index is by Hill (1975), which is the quasi-maximum likelihood estimator. In the statistical literature, there is an ongoing debate on the number of tail observations k that are used in the estimation of α. The choice of k leads to a trade-off between the bias and variance of the estimator. Although practitioners tend to use "Eye-Ball" methods (Resnick and Starica, 1997), the typical approach suggested in the statistical literature is choosing k by minimizing the asymptotic mean squared error (MSE). The methods that are used to find k are asymptotically consistent but have unsatisfactory finite sample properties. This paper proposes a novel approach for choosing the optimal k, labelled k^*. The methodology is based on fitting the tail of a heavy-tailed distribution by minimizing the maximum deviation in the quantile dimension. We show this method outperforms other methods put forth in the literature.

Hall (1990) and Danielsson et al. (2001) use a bootstrap procedure to minimize the MSE. Drees and Kaufmann (1998) exploit the same bias and variance trade-off but use the maximum random fluctuation of the estimator to locate the point where the trade-off is optimal. These methods are based on asymptotic arguments, but these methods may not perform very well in finite samples.

The shortcomings of the currently available methods motivate our new approach. In this paper, we propose using the penalty function of the Kolmogorov-Smirnov (KS) test statistic to fit the tail of the distribution. The KS statistic is measured as the maximum difference between the empirical distribution and a parametric distribution in the probability dimension. Clauset et al. (2009) use this metric to find the optimal k. In a response, Drees et al. (2018) find that the KS statistic chooses values that are too low for the optimal k and therefore introduces a large variances for the Hill estimate. In this paper, we use KS statistic, but with a twist. Instead of minimizing the distance between the empirical distribution and the parametric distribution in the probability dimension, we minimize the distance in the quantile dimension. This measure is henceforth referred to as the KS-distance metric. All heavy-tailed distributions in the sense of regular variation (see below) are characterized by ultimate power law behaviour. For a large subset of

distributions, the tails, to a first-order expansion at infinity, correspond to the tail of a Pareto distribution. The estimates of the shape parameter α use k of the highest order statistics, assuming that these order statistics fit well to a Pareto tail. By varying k, we are able to simultaneously fit the empirical distribution and elicit k^*.

The particular choice of the metric is mainly motivated by the fact that small deviations in the probability distribution 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 $(n/k)^{1/\alpha}$. This effect is amplified as one moves deeper into the tail of the distribution. Basing the metric in the quantile domain therefore puts a natural emphasis on fitting tail observations as opposed to the centre observations. Furthermore, by focusing on the maximum, the metric is not diluted by the numerous centre observations.

These intuitive arguments are backed by rigorous simulation analyses and tests. In the simulation study, we draw i.i.d. samples from the Fréchet, symmetric stable and the Student-t distribution. For these distribution families, Hall's second-order expansion of the cumulative distribution function (CDF) holds. Given this expansion, Hall and Welsh (1985) derive the theoretical k^*, which minimizes the asymptotic MSE. This allows us to study the relationship between the chosen k^* and the estimated tail index. Furthermore, we employ dependent time series in the simulation exercise. The autoregressive conditional heteroskedastic (ARCH) stochastic processes by Engle (1982) model the volatility clustering in time-series data. For the ARCH process, we do not know the variance and the bias of the Hill estimator. However, the tail index is known. Therefore, we are able to evaluate the performance of the methods under the clustering of extremes.

To test the performance of the KS-distance metric, we analyze it from four different perspectives. Firstly, we test the KS-distance metric against various other penalty functions. We use the MSE, mean absolute error and the metric used in Dietrich et al. (2002) to benchmark the performance of our penalty function. To evaluate the ability of different metrics to correctly penalize errors, we use results derived by Hall and Welsh (1985). Hall and Welsh derive the theoretical k^* for the class of distributions that adhere to Hall's second-order expansion of the CDF. These results stipulate the behaviour of k^* as a function of the sample size and α.

Secondly, we contrast the performance of the KS-distance metric with the other existing methods. We find that the KS-distance metric and the automated Eye-Ball method outperform other heuristic and statistical methodologies. For the Student-t, symmetric stable and the Fréchet distributions, the competing methodologies do not reproduce the expected patterns for k^* as derived by Hall and Welsh (1985). This leads to a larger bias in $\hat{\alpha}$ for less-heavy-tailed distributions. The theoretically motivated methodologies often choose a high value for k^* (too many order statistics close to the centre of the distribution), which corresponds to a large bias. For the ARCH processes, the methods by Drees and Kaufmann (1998), Danielsson et al. (2001) and the fixed sample fraction introduce a large bias in the Hill estimator. For the dependent time series, the automated Eye-Ball method and the KS-distance metric produce small biases. This is comforting as most economic time series exhibit volatility clustering.

In addition to estimating the tail index, we model the quantile estimates for the various competing methods, since the ultimate goal is to produce/obtain reliable estimates of the possible losses. When we analyze the quantile estimates, the KS-distance metric and the automated Eye-Ball method perform best for the quantiles greater than the 99% probability level. For the quantiles further towards the centre of the distribution, other methods produce smaller errors. Both the KS-distance metric and the automated Eye-Ball method have a tendency to pick a small k^* and therefore fit the tail close to the maximum. The other methodologies often use a larger number of order statistics and consequently fit better closer towards the centre of the distribution.

In the last section of this paper, we show that the choice of k^* can be economically important. For this purpose we use individual daily stock price information from the Centre for Research in Security Prices (CRSP) database. For 17,918 individual stocks, we estimate the left and right tail index of the equity returns. We measure the average absolute difference between the estimated α and find that the average difference between the methods ranges from 0.39 to 1.44. These differences are outside of the confidence interval of the Hill estimator. For example, 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.

To further demonstrate the impact of the choice of k, we use the KS-distance metric to estimate the conditional tail risk measure by Kelly and Jiang (2014). Kelly and Jiang (2014) use the cross-section of individual stock re-

turns to estimate tail risk in the economy at a given moment in time. They find that the exposure to the conditional (on time) tail index commences a risk premium. They find, furthermore, that the tail index predicts the excess return of the market portfolio. The threshold for this application is set to 5% of the daily returns collected in the given month.[1] We redo the analysis by applying a threshold of 2.5%, 1%, 0.5%, and the KS-distance metric. We find that the risk premium halves from 4% to 2% by applying the KS-distance measure. We find, moreover, that the return predictability by the tail risk measure disappears.

The paper first introduces the Hill estimator and the KS-distance metric along with the other methods from the literature. 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 Extreme value theory methodology

The first part of this section provides a review of the main extreme value theory (EVT) results. It is the stepping stone for the semi-parametric component of the metric. The second part introduces the alternative methods for determining the optimal number of order statistics.

2.1 Extreme value theory

Consider a series $X_1, X_2, ..., X_n$ of i.i.d. random variables with CDF F. Suppose one is interested in the probability that the maximum is not beyond a certain threshold x. This probability is given by

$$\mathrm{P}\left\{\max\left(X_1, ..., X_n\right) \leq x\right\} = \mathrm{P}\left\{X_1 \leq x, ..., X_n \leq x\right\} \overset{\text{iid}}{=} \left[\mathrm{F}\left(x\right)\right]^n.$$

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

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

where $\mathrm{G}\left(x\right)$ is a well-defined non-degenerate CDF. There are three possible $\mathrm{G}\left(x\right)$, depending on the tail shape of $\mathrm{F}\left(x\right)$. This paper concentrates on the

[1] The average number of daily observations in a month is approximately 80,000. The lowest cross-sectional sample is about 35,000 in the beginning of the sample period and the highest around the year 2000 with a sample size of 130,000.

distributions that have a regularly varying tail,

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

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. The α determines how heavy the tail is. Since α corresponds to the number of bounded moments, we often discuss results in terms of α rather than γ. The heavy-tailed limit distribution is the Fréchet distribution (Balkema and De Haan, 1974):

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

Note that $\mathrm{G}_{\gamma>0}\left(x\right)$ 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 $\mathrm{F}\left(x\right)$. 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).

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

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

where k is the number of upper-order statistics used in the estimation of α. Figure 1 depicts the reciprocal of the Hill estimates for a sample drawn from a Student-t(4) distribution plotted against an increasing number of order statistics k. The estimate of α varies with k quite substantially. This shows that the choice of k matters for obtaining the proper estimate.

The pattern in Figure 1 can be decomposed in the variance and the bias of the Hill estimator. For small k, the variance of the Hill estimator is relatively high. As k increases, the volatility subsides and the bias kicks in. One can find the bias and variance of the estimator for parametric distributions for the subclass of distributions in (1) that satisfy the so-called Hall expansion[2]

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

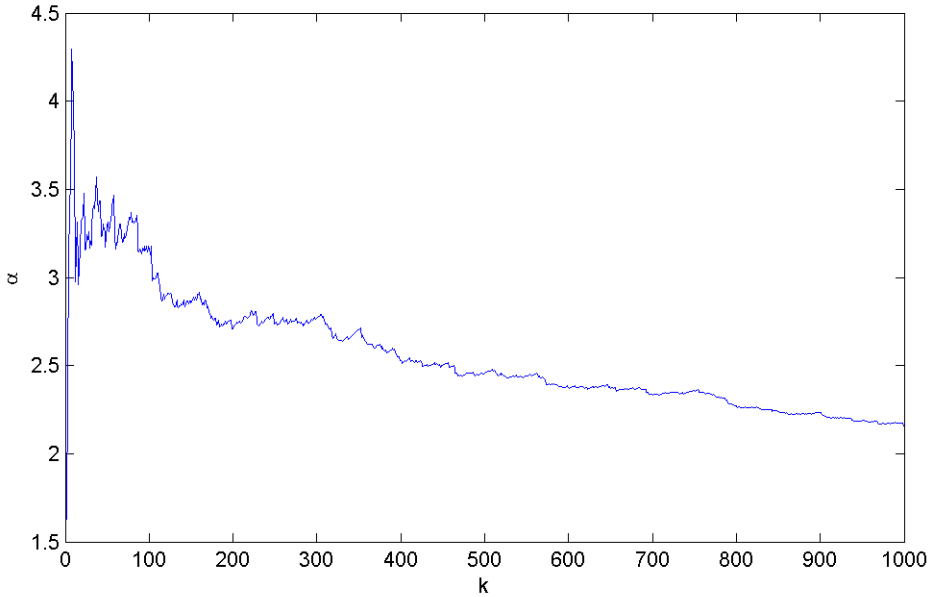
[2]Most known heavy-tailed parametric distributions, like the Student-t, symmetric stable and Fréchet distribution, conform to the Hall expansion. The parameter values A, α, B and β for these distributions are presented in Table 5 of the Appendix.

Using the Hall expansion one shows the asymptotic bias as

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

Equation (4) provides the relationship between the threshold s and the bias of the Hill estimator. From (4) one notices that as s becomes smaller, i.e., the threshold moves towards the centre of the distribution, the bias increases.

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



This graph depicts the estimate of α for different levels of k. The sample is drawn from a Student-t distribution with 4 degrees of freedom so that $\alpha = 4$. The sample size is 10,000. This graph is known as the Hill plot.

The asymptotic variance of the Hill estimator is,[3]

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

The variance is also a function of s. As s decreases, the variance becomes smaller. When comparing the bias squared and the variance, one notices a trade-off. For large s, the bias is small, and the variance dominates. In contrast, for small s the bias dominates. Suppose one likes to choose a k^{*} that balances the two vices. Given this objective, how can we extract the asymptotic MSE from the data?

[3]The bias and the variance expressions are based on the second-order expansion by Hall and Welsh (1985). Given the bias and the variance, the optimal threshold, s^{*}, which minimizes the MSE can be derived. See Appendix A.1 for further details.

2.2 Finding k^*

Various methods exist for choosing k^*. These methods can be roughly divided into two groups. The first group of methods comes from the theoretical statistics literature and is based on asymptotic arguments. The second group of methods stems from suggestions by practitioners. The latter are more heuristic in nature, but some perform surprisingly well.

2.2.1 Theoretical-based methods

Hall (1990) and Danielsson et al. (2001) utilize the bias and the variance to minimize the asymptotic mean squared error (AMSE). They propose a bootstrap procedure that minimizes the AMSE by choosing k appropriately. Hall devises a sub-sample bootstrap to find the k^* under the restrictive assumption $\alpha = \beta$ in (3). To obtain the optimal rate[4] in the bootstrap, the assumption of $\alpha = \beta$ is crucial.

In general, β differs from α, and one is faced with eliciting the optimal rate from the data. To this end, Danielsson et al. (2001) propose a double bootstrap procedure to estimate

$$\lim_{n\to\infty} \text{mse} = \text{E}\left[(\hat{\gamma} - \gamma)^2\right].$$

For the AMSE, γ is unknown. To tackle this problem the theoretical γ in the AMSE expression is replaced with a control variate. However, a simple bootstrap is inconsistent in the tail area. Consequently, a sub-sample bootstrap is applied. Furthermore, to be able to scale the sub-sample MSE back to the original sample size, a second, even smaller sub-sample bootstrap is needed. As a by-product of their procedure the ratio of α/β is also estimated.

A second approach is proposed by Drees and Kaufmann (1998). They introduce a sequential procedure that yields an asymptotically consistent estimator of k^*. Their estimator relies on the fact that the maximum random fluctuation of the estimator is of the order $(\log\log n)^{1/2}$ for all intermediate sequences k_n. This procedure also yields a consistent estimate of α/β.

The theoretical methods by Danielsson et al. (2001) and Drees and Kaufmann (1998) are asymptotically consistent methods. As the arguments are based on asymptotic reasoning, the question is how well these methods perform in finite samples.[5]

[4] The sub-sample bootstrap size needs to increase slower than n to achieve asymptotic optimality in the bootstrap procedure.

[5] For more details, see Appendix A.3.

2.2.2 Heuristics

Applications in the economic literature frequently resort to heuristic rules for choosing k. These rules are based on finding the regions where, as k increases, the variance substantially decreases. The algorithms based on "Eye-Balling" the Hill plot seek a substantial drop in the variance as k is increased. To formalize an automated Eye-Ball method, we use a sequential procedure. This leads to the following estimator,

$$k^*_{eye} = \min \left\{ k \in 2, ..., n^+ - w | h < \frac{1}{w} \sum\nolimits_{i=1}^{w} \mathrm{I} \left\{ \widehat{\alpha}\left(k+i\right) < \widehat{\alpha}\left(k\right) \pm \varepsilon \right\} \right\}. \quad (5)$$

Here 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 $\left[\hat{\alpha}\left(k+1\right), ..., \hat{\alpha}\left(k+w\right)\right]$ 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.[6]

Many applications take a more blunt approach and use a fixed percentage of the total sample. Heuristic rules are easy to apply but are somewhat arbitrary. This has consequences for the application in which these are used. In accordance with the theoretical k^*, put forth by Hall and Welsh (1985), different distributions have different optimal regions and different rates of convergence. Furthermore, the optimal sample fraction depends on the sample size. Therefore, choosing a fixed portion of the sample is not appropriate.

2.3 KS-Distance metric

The shortcomings of the existing methods outlined above motivate our alternative approach. Our approach is based on minimizing the distance between the empirical distribution and a semi-parametric distribution. This procedure is partially inspired by Bickel and Sakov (2008). Bickel and Sakov show that a sub-sample bootstrap is consistent in many cases but may fail in some important examples. Through an adaptive rule based on the minimization of the KS test statistic they find the proper sub-sample bootstrap size.[7] We use

[6] In the Monte Carlo studies, we choose n^+ to be a prespecified threshold, which also applies to the other methods. This threshold will later be defined as T.

[7] The KS statistic is the supremum of the absolute difference between the empirical CDF and a parametric CDF, i.e., $\sup_x \left|\mathrm{F}_n\left(x\right) - \mathrm{F}\left(x\right)\right|$.

their idea of matching the tail of the empirical CDF to a theoretical distribution for finding $\alpha(k^*)$. This matching process requires a semi-parametric form for the theoretical distribution. The scaled Pareto distribution is the ideal candidate for matching the empirical tail. After all, by definition, all distributions in this class satisfy (1) and the Pareto distribution is the only distribution for which (1) holds over the entire support, as it does not contain a second-order term.

2.3.1 The distance metric

The starting point for locating k^* is the first-order term in (3):

$$\mathrm{P}(X \leq x) = \mathrm{F}(x) = 1 - Ax^{-\alpha}[1 + o(1)]. \tag{6}$$

This function is identical to a Pareto distribution if the higher-order terms are ignored. By inverting (6), we get the quantile function

$$x \approx \left[\frac{\mathrm{P}(X \geq x)}{A}\right]^{\frac{1}{-\alpha}}. \tag{7}$$

To turn the quantile function into an estimator, the empirical probability j/n is substituted for $\mathrm{P}(X \geq x)$. A is replaced with the Weissman (1978) estimator $\frac{k}{n}(X_{n-k+1,n})^{\alpha}$, and α is estimated by the Hill estimator.[8] The quantile is thus estimated by

$$q(j,k) = X_{n-k+1,n}\left(\frac{k}{j}\right)^{1/\widehat{\alpha}_k}. \tag{8}$$

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

Given the quantile estimator, the empirical quantile and the penalty function, we get:

$$Q_T = \inf_k \left[\sup_j |X_{n-j,n} - q(j,k)|\right], \quad \text{for} \quad j = 1, ..., T, \tag{9}$$

where $T > k$ is the region over which the KS-distance metric is measured. Here $X_{n-j,n}$ is the empirical quantile and $q(j,k)$ is the quantile estimate from (8). k, which produces the smallest maximum horizontal deviation along all the tail observation up to T, is k^* for the Hill estimator.[9]

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

[9] In Appendix A.2, we model the KS-distance metric in terms of a Brownian motion representation. This allows us to study the properties of the KS-distance metric as an appropriate penalty function in a more general setting.

2.3.2 Motivation for the KS-distance metric

The metric in (9) deviates from the classic Kolmogorov-Smirnov statistic. The adjustment is that the distance is measured in the quantile dimension rather than the probability distribution dimension. There are several reasons for this choice. The first reason is that most economic variables, such as gains and losses, are concepts in the quantile dimension rather than the probability dimension. Various risk measures, such as value at risk and expected shortfall, are concepts related to quantiles at a given probability level.

The second motivation is more technical. Our analysis is solely focused on the tail of the distribution, rather than the centre observations. For tail observations, small changes in probabilities lead to large changes in quantiles. Consequently, small mistakes in estimating probabilities lead to large deviations in the quantiles. We therefore prefer to minimize the mistakes made in the quantile dimension rather than the probability distribution dimension.

Given the decision to measure over the quantile dimension, a function is needed to penalize deviations from the empirical distribution. The inner part of the penalty function takes the absolute difference between the quantiles instead of, for instance, the squared difference. A small error in the tail is automatically magnified. Therefore, fitting the tail quantiles already introduces a natural way to put emphasis on the larger deviations. It consequently does not necessitate additional penalizing, like the squared differences do.

To translate all the absolute differences along the tail into one metric, we use the maximum over the absolute distances. Taking the maximum has as a benefit that the metric is not diluted by the numerous centre observations. This, for instance, is the case when the differences are averaged.

3 Simulations

For the simulations, we choose a wide range of heavy-tailed distributions and processes. One prerequisite is that the tail index for these distributions is known. We choose CDFs that differ with respect to the tail index and second-order terms. These imply different rates of convergence as $n \to \infty$, and differences in the bias and variance trade-offs.

For the simulation study, we choose distribution families that adhere to the Hall expansion in Equation (3). Therefore, we know α and k^*, where k^* min-

imizes the AMSE. We use three distribution families to draw i.i.d. samples from: The Fréchet, symmetric stable and the Student-t distribution.[10] We also employ dependent time series. The ARCH stochastic processes by Engle (1982) model the volatility clustering in time-series data. For the ARCH process, we do not know the variance and the bias of the Hill estimator. However, the tail index is known.[11] Therefore, we are able to evaluate the performance of the methods under the clustering of extremes.

In the simulation study, estimates of α and k^* for different penalty functions are analyzed. We evaluate six properties. The first property is the bias in the estimate of α. Secondly, we compare the α estimates for different members within the same distribution family, like the Student-t distribution.

Thirdly, from Hall and Welsh (1985) we have the level of k^* that minimizes the AMSE for a given parametric distribution and sample size.[12] This allows us to evaluate how close the different criteria come to the k^*. The fourth property addresses the theoretical relationship between α and k^*. It turns out that α and k^* are inversely related within most distribution families. The fifth property of k^* is that for $n \to \infty$ and $k(n) \to \infty$ that $k/n \to 0$. This entails that, for the same distribution, a larger sample size implies that a smaller proportion of observations is used for the estimation of α.

Lastly, in order to locate the optimal k^* in practice, we consider a range of higher-order statistics up to a threshold $X_{n-T,n}$. The selected k^* should be insensitive to the choice of the nuisance threshold T.

3.1 Alternative penalty functions

We contrast the performance of the KS-distance metric to three other metrics presented in Appendix A.4. These metrics are the mean squared, mean absolute deviations and the discretized version of the metric used by Dietrich et al. (2002). For a thorough analysis, we draw samples from the Student-t, symmetric stable and Fréchet distribution families.[13]

[10] The tail index for the Student-t distribution is equal to the degrees of freedom. For the symmetric stable distribution it is the characteristic exponent. For the Fréchet distribution it is the shape parameter. Henceforth we refer to these distribution parameters as α.

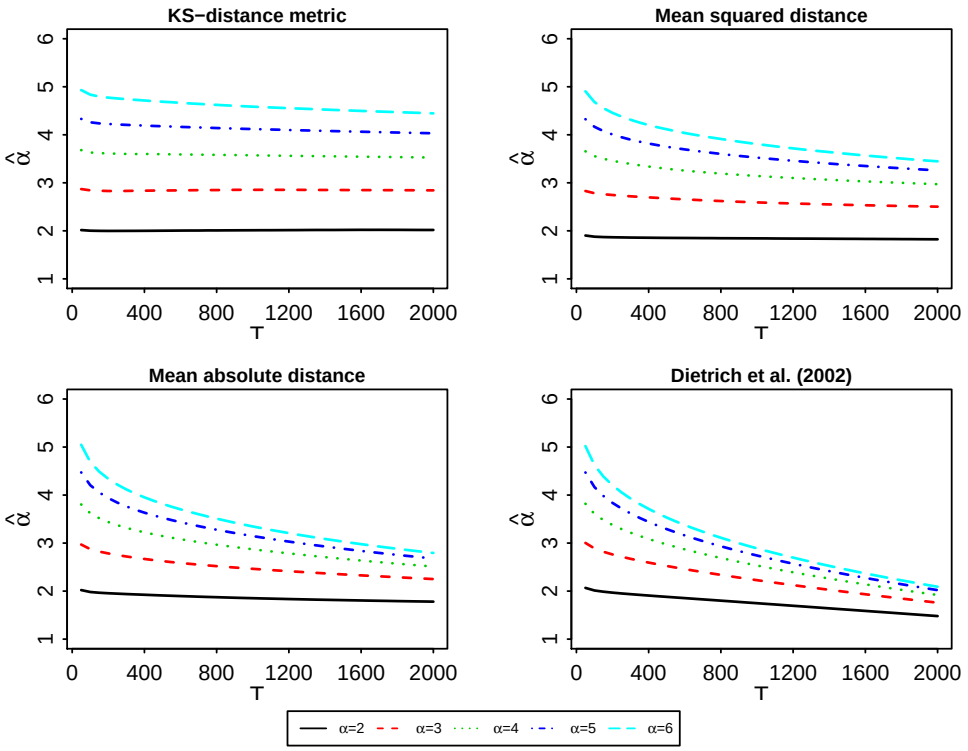
[11] See De Haan, Resnick, Rootzén, and De Vries (1989).

[12] See Appendix A.1 for the derivation of the optimal threshold.

[13] The parameter values of these distribution families for the Hall expansion in (3) are given in Table 5 in the Appendix.

In Figure 2, the level of $\alpha(k^*)$ is displayed against the threshold T over which the specified metric is optimized. These plots give an indication whether the $\alpha(k^*)$ is at the right level and is insensitive to the nuisance parameter T. The first fact to notice from Figure 2 is that the curves are relatively flat for the KS-distance metric. More importantly, the curves come closest to the theoretical level of α. On the basis of 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 more or less monotonic decline in the three graphs indicates that the level of k^* is dependent on the area you optimize over for these three criteria.

Figure 2: Optimal $\hat{\alpha}$ for the quantile metrics (Student-t distribution)



This figure depicts simulation results of the average optimally chosen $\alpha(k)$ for a given level of T. Here T is the number of extreme-order statistics over which the metric is optimized. In the upper left graph this is done for the KS-distance metric for different Student-t distributions with degrees of freedom α. This is also done for the mean squared distance, mean absolute distance and the criteria used by Dietrich et al. (2002). The simulation experiment has 10,000 iterations for sample size n=10,000.

In Figure 8 in the Appendix, we show that the stability of the KS metric with regards to the choice of the nuisance parameter T also holds for the symmetric stable distributions. The estimates of the level of α are biased.

There is a positive bias of 0.1 for the α between 1.1 and 1.7. For the specific case of $\alpha = 1.9$, all methods have trouble finding the correct α. This is because the symmetric stable distribution with $\alpha = 1.9$ comes close to $\alpha = 2$. At the characteristic exponent 2 there is a switch from distributions that exhibit power law behaviour with infinite variance to the normal distribution with exponentially declining tails and all moments bounded. The normal distribution falls in the domain of attraction of the Gumbel distribution and is therefore outside of the domain where the Hill estimator is applicable.

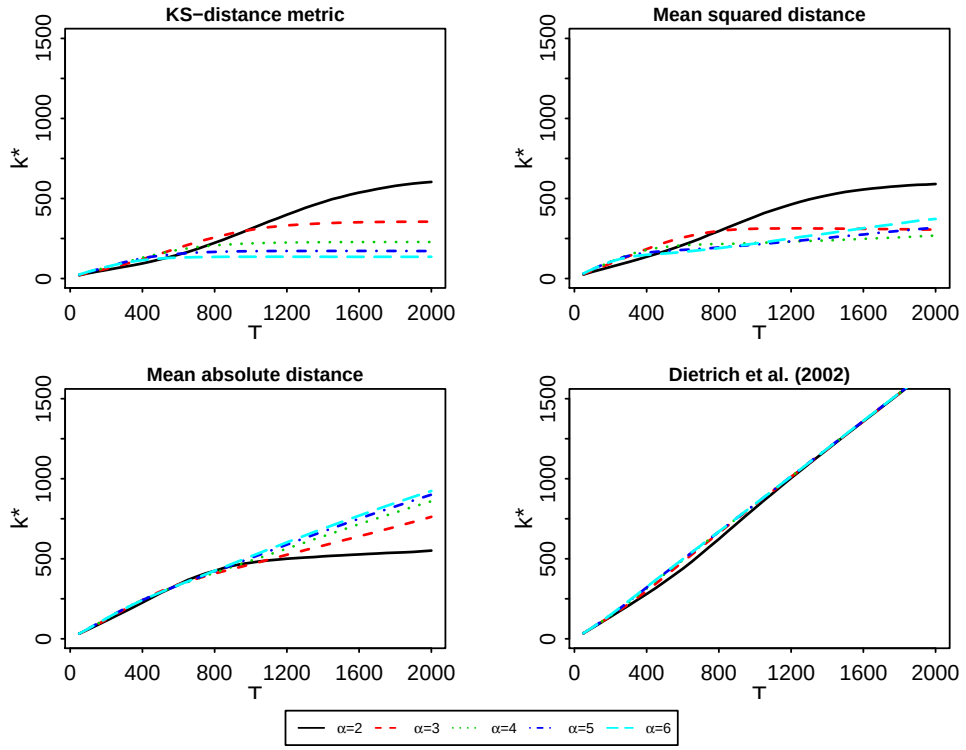
As can be observed from Figure 9 in Appendix C, all the estimates of α have a small bias when the samples are drawn from the Fréchet distribution family. Therefore the choice of k^* is less crucial. In contrast to α, k^* appears to be sensitive to the choice of T. Nevertheless, this has little effect on $\hat{\alpha}$ due to the particular structure of the AMSE for the Fréchet CDF.[14] From Figure 10 we see that the bias is relatively small for the Fréchet distribution. This makes the choice of k^* less important in contrast to the other distributions.

Figure 3 depicts the average k^* for the Student-t distribution family. These figures show the properties of k^* as the interval $[X_{n,n}, X_{n-T,n}]$ over which the metric is optimized changes. We observe that the average k^* as a function of T stabilizes for the KS-distance metric. This indicates that the choice of k^* is stable once T is sufficiently large. For the Student-t (2) distribution no such stabilization occurs. Judging from Figure 8, however, the precise choice of k^* does not seem very important for the case of two degrees of freedom. The average mean squared distance displays roughly the same properties as the KS-distance metric. Although the choice of k seems to roughly stabilize, this does not automatically translate into a stable and optimal estimation of $\alpha(k)$. This stabilization does not occur for the mean absolute difference and the metric by Dietrich et al. (2002).

Next, we study the relationship of the average level of k^* for the different members within the distribution family. In Figure 3 we observe that for the KS-distance metric k^* is an increasing function of the degrees of freedom for the Student-t distribution. This is the pattern that we expect based on k^* derived by minimizing the AMSE. This pattern is not observed for the other criteria. Additionally, from Figure 10 we see that the bias is relatively small for the Fréchet distribution. This makes the choice of k^* less important in contrast to the other distributions.

[14] The results are based on the second-order expansion. This might be different when higher-order terms are used.

Figure 3: Optimal k for quantile metrics (Student-t distribution)



This figure depicts simulation results of the average optimally chosen k for a given level of T. Here T is the number of extreme-order statistics over which the metric is optimized. In the upper left graph this is done for the KS-distance metric for different Student-t distributions with degrees of freedom α. This is also done for the mean squared distance, mean absolute distance and the criteria used by Dietrich et al. (2002). The simulation experiment has 10,000 iterations for sample size $n = 10,000$.

Figure 8 in the Appendix depicts the results for the symmetric stable distribution. There is less stability in these plots compared with the plots for the Student-t distribution. Until approximately $T = 600$, the chosen levels k^* are coherent relative to one another for the symmetric stable distributions with different values of α. For larger values of T, the levels of k^* start to cut across one another.

The symmetric stable distribution is relatively difficult to analyze. Figure 10 in the Appendix shows a hump shape for the Hill plot of the symmetric stable distribution. Sun and De Vries (2018) show that the positive sign of the scale parameter of the third-order term in the Hall expansion explains the hump shape. This convexity can lead to two intersections of the Hill plot with the true value of α. The region where k is small, the high volatility

region, has an intermediate probability containing the best estimate. As k increases and moves into the hump, the volatility subsides; however, the bias kicks in. These estimates are biased and therefore have a low probability of containing the best estimate. As k increases further towards T, the Hill estimates move back in the range of the true α. These estimates have a lower variance and possibly a better estimate than the initial volatile part. This is a possible explanation for the shape of k^* as a function of T in the KS-distance metric plotted in Figure 8. The increase in k^* after a flat line between 500 and 1,000 is an indication that the effect of the third-order term is kicking in.

In MC simulation studies none of the metrics attain the optimal level of k^*. Based on the other desirable attributes described at the start of this section, the KS-distance metric outperforms the other metrics. Furthermore, the chosen k^* should not change as the interval for the metric changes, i.e., change in T. The KS-distance metric is the only metric that is robust to changes in T. This alleviates the concern of arbitrarily chosen parameters driving the results.[15]

3.2 Monte Carlo: Comparing existing methods

Given the choice of the KS-distance metric as the appropriate penalty function, the literature offers competing methods for choosing k^*. These are the double bootstrap and the method by Drees and Kaufmann (1998) reviewed in Appendix A.3. Additionally, we use the automated Eye-Ball method, fixed sample proportion and the theoretical threshold (TH) in an MC horse race.[16] In the horse race, we judge the methods on their ability to estimate the tail index and to reproduce the patterns in α and k^*. In addition, we evaluate the ability to estimate the different quantiles. Even though the methodologies are focused on the Hill estimator, estimating the quantiles can give an interesting new dimension to the performance of these methodologies. For the quantile estimator, both the shape and scale parameters need to be estimated. These are both dependent on k^*.

Table 1 presents the results from the MC horse race for the heavy-tailed distributions and processes with the mean estimates of α for the different methodologies. From Table 1 it is clear that all methods for the Student-t distribution exhibit an increasing bias as the degrees of freedom increase.

[15]For additional simulation results with different sample sizes n, consult the tables and figures from Monte Carlo simulations in the online Appendix.

[16]The k^* based on the minimization of the theoretical AMSE is only reported for comparison as a benchmark. In practice one does not know the TH optimal k^*.

This problem is most prominent for the double bootstrap method and the iterative method of Drees and Kaufmann (1998). The KS-distance metric, TH and the automated Eye-Ball method give estimates that are closest to the true value of the tail index. Based on these results for the Student-t distribution, we conclude that the KS-distance metric performs better than other implementable methods. However, the automated Eye-Ball method is only performing marginally inferior to the KS-distance metric.

Table 1: Estimates of $\alpha\,(k^*)$ for different methods

	α	KS Dis	TH	5%	Eye-Ball	Drees	Du Bo
Student-t	2	2.01	1.92	1.85	1.98	1.70	1.71
	3	2.85	2.79	2.45	2.83	2.24	2.20
	4	3.53	3.58	2.87	3.48	2.64	2.52
	5	4.10	4.32	3.16	3.96	2.92	2.75
	6	4.49	4.96	3.38	4.29	3.14	2.92
Stable	1.1	1.21	1.11	1.11	1.10	1.07	1.09
	1.3	1.39	1.33	1.37	1.32	1.33	1.36
	1.5	1.58	1.57	1.72	1.54	1.68	1.71
	1.7	1.78	1.84	2.32	1.84	2.18	2.19
	1.9	2.31	2.55	3.55	3.36	3.13	2.90
Fréchet	2	2.01	1.99	1.98	2.00	1.92	1.93
	3	2.93	3.01	2.97	3.00	2.88	2.90
	4	3.79	4.05	3.96	3.99	3.85	3.87
	5	4.71	5.09	4.95	4.99	4.81	4.84
	6	5.63	6.14	5.94	5.98	5.77	5.81
ARCH	2.30	2.59		2.13	2.34	1.93	1.88
	2.68	2.87		2.39	2.66	2.16	2.05
	3.17	3.22		2.69	3.04	2.42	2.22
	3.82	3.66		3.02	3.50	2.71	2.39
	4.73	4.18		3.38	4.03	3.04	2.55

This table depicts the mean for the estimated α for the different methodologies. The samples are drawn from four different heavy-tailed distribution families. The samples are drawn from the Student-t, symmetric stable, Fréchet distribution and ARCH process. The different methods are stated in the first row. KS Dis is the Kolmogorov-Smirnov distance metric in (9). TH is based on the theoretically derived optimal k from minimizing the MSE for specific parametric distributions, presented in Equation (11) in the Appendix. The automated Eye-Ball method in (5) 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). Here, α indicates the corresponding theoretical tail exponent for the particular distribution which the sample is drawn from. The sample size is $n = 10,000$ for $10,000$ repetitions.

The simulation results for the symmetric stable distribution do not point

towards a method that is clearly superior. For $\alpha = 1.1$ and $\alpha = 1.3$, most of the other methods perform better in terms of the mean estimate than the KS-distance metric. For $\alpha \geq 1.5$, the KS-distance metric and the automated Eye-Ball method start outperforming the competing methods. For $\alpha = 1.9$, the competing methods completely miss the mark. The same is true for the KS-distance metric, but the bias is the smallest among all the methods. As previously discussed, the difficulty for relatively high values of α stem from the fact that the symmetric stable distribution becomes the thin-tailed normal distribution at the boundary.

The bias of the Hill estimator for the Fréchet distribution is relatively small compared to the Student-t and symmetric stable distribution. Therefore, all the methods perform relatively well. The automated Eye-Ball method has the best performance for this family of distributions. The bias in the KS-distance metric is large relative to the other metrics for higher α's. Because the bias in the Fréchet distribution is small, the bias due to the KS-distance metric is still limited in absolute terms.

The Hill plot for the ARCH process is similar to that of the Student-t distribution. Therefore, we expect that methods that performed well for the Student-t distribution will also do well for the ARCH process. Table 1 shows that, for the ARCH process, the KS-distance metric and the automated Eye-Ball method indeed outperform the other methods. For the very heavy-tailed processes the automated Eye-Ball method has a smaller bias. For $\alpha \geq 3.172$, the KS-distance metric shows a smaller bias. The other methods show a substantial bias over the whole range of α. To conclude, the KS-distance metric and the automated Eye-Ball method are the preferred methods since these perform best across a wide range of α values.

Table 6 in Appendix B gives k^* for the different distributions and criteria. The patterns in k^* for the various distributions give a mixed picture. The automated Eye-Ball method chooses low values of k^* across the different distributions. For the Student-t distribution the average number of observations used for the estimation increases with α. This goes against the results for k^*_{TH}. The same holds true when the sample is drawn from the symmetric stable distribution.

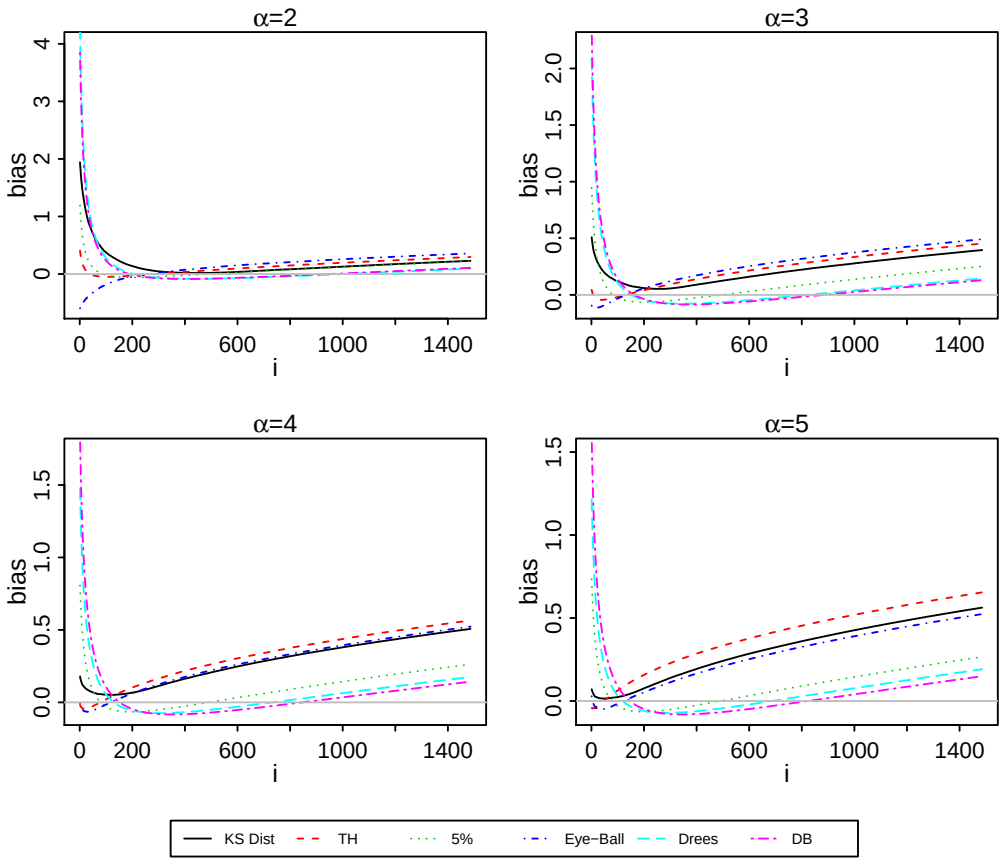
The method by Drees and Kaufmann (1998) shows the negative relationship between the α implied from the parametric distribution and the choice of k^* within the family of distributions. However, the levels for the Student-t and symmetric stable distribution family are far higher than desired. In

part, this is because the practical criterion is based on asymptotic arguments. In addition, the bootstrap has a slow rate of convergence. In practice this leads the criterion function to be flat and volatile near the optimum. As a consequence, often no clear global minimum is found.

3.3 Simulation results for the quantiles

We also analyze how the different metrics influence the quantile estimates. For many of the economic questions this is more relevant than the precise value of the tail index.

Figure 4: Quantile estimation median difference (Student-t distribution)



This figure shows the bias induced by using the quantile estimator presented in Equation (8). 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 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 i/n at which the quantile is estimated. Moving rightwards along the X-axis represents a move towards the center of the distribution.

Figure 4 depicts how well the various methods perform in estimating the quantiles of the distribution.[17] Estimating the quantile beyond the 99.5% quantile is notoriously difficult. Therefore all methods introduce large mistakes. With the exception of the KS-distance metric for the Student-t distribution with two degrees of freedom, the KS-distance metric, the automated Eye-Ball method and the theoretical MSE produce smaller mistakes in the extreme tail region. For the Student-t distribution, the method by Drees and Kaufmann (1998), the double bootstrap and the 5% fixed sample size approach generate a comparatively large mistake in the 99% to 100% quantile region. The flip side is that these methods have small mistakes for the quantiles further towards the centre of the distribution.

Figure 11 in the Appendix depicts the results for the symmetric stable distribution. A very similar picture to the Student-t distribution emerges. The KS-distance metric introduces a relatively small mistake in the very extreme tail quantiles but performs weakly for the quantiles closer to the centre of the distribution. The results for the Fréchet distribution are presented in Figure 12 in the Appendix. All methods seem to produce small mistakes for the less extreme quantiles. It is only in the extreme tail quantiles that the KS-distance metric sets itself apart from the other methods. This reconfirms the results for the Student-t and symmetric stable distribution, that the KS-distance metric is most effective in modelling the most extreme part of the distribution.

Based on the analysis of the Monte Carlo horse race, we conclude that both the KS-distance metric and the automated Eye-Ball method have a superior performance over the other implementable methods. Both methods perform well based on $\hat{\alpha}$. However, based on the analysis of the choice of k^*, the KS-distance metric shows a better pattern. This translates into a smaller bias in the simulation study for the Student-t and symmetric stable distribution for higher values of α. The conclusions for quantile estimation are more sobering. Since the KS-distance metric and the automated Eye-Ball method perform well deep in the tail of the distribution, they have a relatively large bias towards the centre.

The performance of the methods of Drees and Kaufmann (1998) and Danielsson et al. (2001) in finite samples is inferior to the other methods, notwith-

[17] In these figures we study the median difference. Due to a small number of extreme outliers in the tail quantiles closer to the centre of the distribution, we opt to depict the median instead of the average. The qualitative results holds for the average difference.

standing the proofs showing that the methods are asymptotically consistent. Using a fixed percentage of observations, such as 5%, ignores the information that can be obtained from the Hill plot. For larger samples this often means that α is estimated with a relatively large bias, as can be observed from the Monte Carlo simulation. This leads to the conclusion that the KS-distance metric overall comes out as the preferred approach.

4 Application: Financial return series

We now take the KS-distance metric to real data. We use the methods from the horse race to estimate the tail indexes for the return distribution of individual U.S. stocks. The various methods used in the horse race are used to estimate the tail exponent for returns on U.S. stocks.

4.1 Data

The stock market data is obtained from the Centre for Research in Security Prices (CRSP). The CRSP database contains individual stock data from 1925-12-31 to 2015-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. In total, 17,918 stocks are used. Every stock included in the analysis needs to trade on one of the four exchanges during the measurement period.[18] We exclude stocks with less than 48 months of data. For the accuracy of EVT estimators 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

The mean absolute differences between the different methods' tail exponent estimates are displayed in Table 2. The differences are quite substantial for both the left and right tail.[19] On the basis of the results the methods can be divided into two groups. The first group consists of the KS-distance metric and the automated Eye-Ball method. The second group is the theory-based methods. The KS-distance metric and the automated Eye-Ball method show a relatively large deviation from the estimates that are obtained with the double bootstrap and the iterative method by Drees and Kaufmann (1998).

[18] In the CRSP database exchange codes -2, -1 and 0 indicate that a stock is not traded on one of the four exchanges and thus no price data is recorded for these days.

[19] For descriptive statistics on the tail estimates of the left and right tail consult Table 7 in the Appendix. In Table 8 we also report the results for the median difference. The results for the median are similar to the mean results but smaller.

This result is in line with the Monte Carlo simulations, where the KS-distance metric and automated Eye-Ball method estimates are relatively close to one another. Both methodologies use a low fraction of the total sample for $\hat{\alpha}$ in the simulations. The same holds for the return data. The lower panel of Table 2 shows that the double bootstrap procedure has the smallest average difference in choosing k^* with the method by Drees and Kaufmann (1998). Given that the Hill estimator is normally distributed with a standard deviation of $1/\alpha$, it implies that a large portion of the estimates of the two different groups are significantly different from one another.[20]

Table 2: Mean absolute differences between different methodologies

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	0	0.84	0.59	1.11	1.44	0	0.75	0.54	0.88	1.22
5%	0.84	0	0.56	0.49	0.74	0.75	0	0.55	0.39	0.63
Eye-Ball	0.59	0.56	0	0.92	1.22	0.54	0.55	0	0.77	1.12
Drees	1.11	0.49	0.92	0	0.49	0.88	0.39	0.77	0	0.50
DB	1.44	0.74	1.22	0.49	0	1.22	0.63	1.12	0.50	0

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

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	0	127	78	224	361	0	128	85	195	350
5%	127	0	142	140	268	128	0	149	120	264
Eye-Ball	78	142	0	267	409	85	149	0	244	413
Drees	224	140	267	0	168	195	120	244	0	191
DB	361	268	409	168	0	350	264	413	191	0

(b) Estimates k_i^*

This table presents mean absolute difference between $\hat{\alpha}\left(k^*\right)_i$ and k^* by applying the five different methods to choose k^* for the 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 five different methods are the KS-distance metric, 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.

Comparing the results between the horse race and this financial application does show parallels. In the horse race the KS-distance metric and the automated Eye-Ball method perform well and have estimates close to one another. This is also the case for the analysis on the equity return data. The methods by Drees and Kaufmann (1998) and Danielsson et al. (2001) generate estimates of α which are close to one another for the financial data. In

[20]The results for the right tail of the empirical distribution are somewhat proportionally smaller, but the same relative differences are preserved among methodologies.

the Monte Carlo horse race, they show poor finite sample properties. Even though these patterns might be coincidental, it does cast doubt on the applicability of these methods for real-world empirical estimations.

4.3 Cross-sectional Hill estimator

Most applications estimate α and quantiles from time-series data on stock prices, see e.g., Jansen and De Vries (1991). Other applications use a single cross-section to determine the power law for city size or income distribution, e.g., Gabaix (1999) and Reed (2003). Recent work by Kelly and Jiang (2014) use the cross-section of individual U.S. stock returns to estimate a tail index conditional on time t. They find that stocks that are more exposed to the time variation in the tail index command a risk premium. Stocks that covary less or negatively with the tail index provide a hedging opportunity. These stocks sell at a discount relative to the stocks with high exposures. Additionally, they find persistence in their time varying tail index estimates. If investors are averse to an increase in tail risk, then a shock to tail risk is informative about the future level of risk. Therefore, a shock in tail risk increases their demand for compensation (for holding that risk), i.e., the risk premium, for investing in the stock market. The persistence induces predictability of the risk premium by changes in conditional tail risk.

Here we reproduce the results of the original paper and investigate the influence picking an appropriate threshold has. In the paper, daily returns on all the stocks within a month are collected and used to estimate the tail index for month t. In Kelly and Jiang (2014), the threshold is chosen at 5% of the sample. This provides a monthly series of α_t estimates. They subsequently measure the covariation of each stock with $\hat{\alpha}_t$ and rank the stocks according to their exposure. If α_t is a priced risk factor, then the risk premium on different stocks and the exposure should be positively correlated.

Table 3 shows the results for the cross-sectional pricing of tail risk based on α_t estimated with different thresholds. The first row reproduces the approximate 4% risk premium that Kelly and Jiang (2014) originally find on stocks with a high covariance relative to low covarying stocks. In the second to fourth rows, we move k closer to the tail of the distribution to estimate α_t. For these lower thresholds, the risk premium halves and becomes insignificant at a 10% confidence level. The risk premia based on α_{KS} is also halved relative to the estimates based on the 5% threshold, but is significant.[21]

[21]The results are quantitatively similar when the exposures are jointly estimated with

Table 3: Cross-sectional pricing of tail risk

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

(a) Loading on α_t from single-factor model

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

(b) Loading on α_t from joint estimation with Fama-French factors

This table reports return statistics for portfolios formed on the basis of tail risk beta based on different thresholds for the tail index estimation. Each month, stocks are sorted into quintile portfolios based on predictive tail loadings that are estimated from monthly data over the previous ten years. These equally weighted portfolios are based on NYSE/AMEX/NASDAQ stocks with CRSP share codes 10 and 11. The rows indicate which threshold is used to estimate the cross-sectional tail index for each month. The subscripts on α in the first four rows of each table indicate the sample fraction used to estimate α_t. The last row uses the KS-distance metric to estimate α_t. Panel (a) reports results where the tail risk beta is estimated in a single-factor model. In panel (b) a multi-factor model, including the Fama-French factors, is estimated to retrieve the tail risk beta. The rightmost columns report results for the high minus low zero net investment portfolio that is long quintile portfolio five and short quintile one and associated t-statistics. For 12-month returns, t-statistics use Newey and West (1987) standard errors based on 12 lags. Stocks with prices below \$5 at the portfolio formation date are excluded.

Table 4 presents the results predicting the risk premium at different horizons with $\hat{\alpha}_t$. The results show that as the threshold moves further down into the tail of the distribution, the predictive power of $\hat{\alpha}_t$ drops for the different horizons. Both the R^2 and the t-statistic on the coefficients drop with the threshold level. When we apply the KS-distance metric, the predictive result disappears. The KS-distance metric picks a different sample fraction as a threshold to find the best fit for the tail index each month.

This implies that the predictive results are most likely not driven by the variation in the tail index when modelling the most extreme observations. Given that the average cross-sectional sample size is between 35.000 and

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
$\hat{\alpha}_{5\%}$	0.10	2.60	1.02	1.01	2.35	7.17	3.02	2.36	19.90
$\hat{\alpha}_{2.5\%}$	0.10	2.44	0.90	0.98	2.11	5.84	3.07	2.23	17.60
$\hat{\alpha}_{1\%}$	0.08	1.98	0.60	0.78	1.71	3.63	2.62	1.92	12.21
$\hat{\alpha}_{0.5\%}$	0.08	1.93	0.57	0.57	1.39	2.14	1.97	1.61	7.53
$\hat{\alpha}_{KS}$	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 of CRSP value-weighted market index returns over one-month, one-year, and three-year horizons. The different rows report forecasting results based on our estimated tail-risk time series with different thresholds. Because overlapping monthly observations are used, test statistics are calculated using Hodrick (1992) standard errors for overlapping data with lag length equal to the number of months in each horizon.

120.000, the 5% threshold is far removed from the tail of the distribution.[22] It therefore seems that picking the threshold is an important choice in economic applications. This is especially true for the applications that are truly interested in the region of the distribution where EVT is applicable.

[22]In additional regression results, we show that the predictive power of $\alpha_{5\%}$ cannot be explained by the variance, skewness or kurtosis of the cross-sectional distribution.

5 Conclusion

In this paper we propose a new approach to choose the optimal number of order statistics for the Hill estimator. We employ the KS distance over the quantile dimension to fit the Pareto quantile estimator to the empirical distribution. The scale and shape coefficients of the Pareto quantile estimator are dependent on the tail index estimate and therefore on the number of order statistics used for the estimation. By fitting the tail of the distribution, we find the optimal sample fraction for the Hill estimator.

To study its properties we perform rigorous simulation studies. We contrast this new approach with methods from the theoretical statistical literature and heuristic methods used in the applied literature. We show that the KS-distance metric overcomes the problems asymptotic valid methods have in finite samples. The KS-distance metric and automated Eye-Ball method outperform the competing theoretical methods based on the size of the bias. Additionally, we benchmark the quantile estimates produced by the different methodologies. The various methodologies have different areas in the tail where they outperform other competing methods. The KS-distance metric is best suited for estimating the quantiles deep in the tail, where the theoretical methods do better in the tail region towards the centre of the distribution.

To show that the choice of the proper number of order statistics matters, we estimate the tail indexes for the universe of daily CRSP U.S. stock returns. From the results, the methods can be divided into two groups. We find that the heuristic methods that perform well in the simulation studies have estimates close to one another in our financial application. Applying the correct methodology to choose the threshold for the Hill estimator can therefore have substantial consequences for the perception of the riskiness of the asset. With the KS-distance metric we also estimate the conditional tail risk measure by Kelly and Jiang (2014). They use a 5% threshold for the Hill estimator to estimate a conditional tail index from the cross-section of individual stock returns. With the KS-distance metric choosing the threshold, tail risk is a robust systematic risk factor in determining the cross-section of expected returns. However, tail risk no longer predicts the risk premium on the market index.

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.

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

De Haan, L., Resnick, S., Rootzén, H., De Vries, C., 1989. Extremal behavior of solutions to a stochastic difference equation, with applications to arch processes. Stochastic Processes and their Applications pp. 213–224.

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., 2018. On a minimum distance procedure for threshold selection in tail analysis. arXiv preprint arXiv:1811.06433 .

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

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 981–1008.

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

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.

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 $\text{mse} = \text{var} + (\text{bias})^2$ is

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

For the AMSE the small terms go to 0,

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

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

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

Substituting s^* back into the MSE gives

$$\text{amse}^* = \frac{1}{A\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{-\frac{2\beta}{\alpha+2\beta}} + o\left(n^{-\frac{2\beta}{\alpha+2\beta}}\right). \quad (10)$$

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

$$n^{\frac{-2\beta}{\alpha+2\beta}} M(s^*) \underset{n\to\infty}{\to} A\left[\frac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{-\frac{\alpha}{\alpha+2\beta}}. \quad (11)$$

Through the Hall expansion we have the functional forms for α, β, A and B for the Student-t, Stable and Fréchet distribution.[23]

A.2 Brownian motion representation

There are various ways to study the behaviour of the KS-distance metric. Here we study the properties of the KS-distance metric by modelling the quantile process with a Brownian motion representation. This allows us to simulate under more general conditions than simulating from fully parametric distributions. Even though this does not allow us to verify the correct

[23] See Table 5 for the parameter values for the specific distributions.

estimation of the tail index, it does facilitate the study of the behaviour of the penalty function.

By Theorem 2.4.8 from De Haan and Ferreira (2007, page 52 and 76) the KS-distance metric in (9) can be written as

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

$$\arg\min_{0<k<T} \sup_{0<l<\frac{T}{k}} \left| \frac{\gamma}{\sqrt{k}} U\left(\frac{n}{k}\right) l^{-\hat{\gamma}} \left[l^{-1} w(l) - w(1) - A_0\left(\frac{n}{k}\right) \frac{\sqrt{k}}{\gamma} \frac{l^{-\rho}-1}{\rho} \right] \right|, \quad (12)$$

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

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

For the case that the CDF satisfies the Hall expansion (6) the functions $U\left(\frac{n}{k}\right)$ and $A_0\left(\frac{n}{k}\right)$ can be given further content. This is also needed for the simulations that are performed below. Applying the De Bruijn inversion[24] 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 the limit function (12) allows us to study the metric under relatively general conditions. Simulating the function in Equation (12) 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 upper panel in Figure 5 in Appendix C shows for a given k at which order statistic the maximum quantile distance is found for the parameters of the Student-t distribution family. The lower panel shows what the value

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

is of that distance for the given k. These two panels provide insight into how the KS-distance metric picks k^* under relatively general conditions. It is clear that the largest deviations for k large are almost always found at the most extreme observation. By choosing a large k, the Pareto distribution is better fitted towards the centre of the distribution. As a consequence, the largest deviation is found towards the more extreme observations. By the same logic, for k small, the largest deviations are more frequently found at the less extreme observations.

The lower panel shows that the smallest largest deviation is found for $k = 2$. A small value of k is desirable as this minimizes the bias of the Hill estimator. Given that the modeled $\hat{\gamma}$ is based on the expectation of the Hill estimator, it is therefore not surprising the limit in (12) picks $k = 2$ as the optimal threshold. This threshold minimizes the bias present in the Hill estimator. This also holds for the parameters retrieved from the symmetric stable and Fréchet distribution.

A.3 Theory-based methods

Hall (1990) and Danielsson et al. (2001) utilize the bias and the variance to minimize the AMSE. They propose a bootstrap method that minimizes the AMSE by choosing k appropriately. For the distributions that satisfy the second-order expansion by Hall, the sample fraction at which the AMSE is minimized can be determined. Hall devises a sub-sample bootstrap to find the k^* under the restrictive assumption $\alpha = \beta$ in (3). To obtain the optimal rate[25] in the bootstrap, the assumption of $\alpha = \beta$ is crucial.

In general, β differs from α, and one is faced with eliciting the optimal rate from the data. To this end, Danielsson et al. (2001) propose a double bootstrap to estimate

$$\lim_{n\to\infty} \text{mse} = \text{E}\left[(\hat{\gamma} - \gamma)^2\right].$$

In the AMSE the value of γ is unknown. To tackle this problem the theoretical γ value in the MSE expression is replaced with a control variate. For the control variate an alternative estimator to the Hill estimator is used, namely $\hat{\gamma}^*$. The control variate has an AMSE with the same rate of convergence as the AMSE of $\hat{\gamma}$.

[25]The sub-sample bootstrap size needs to increase slower than n to achieve asymptotic optimality in the bootstrap procedure.

Due to the use of this control variate, the true value 0 is known. Therefore, a bootstrap procedure can be used to construct an estimate of the MSE of $\hat{\gamma} - \hat{\gamma}^*$. However, a simple bootstrap is inconsistent in the tail area. Consequently, a sub-sample bootstrap is applied. Furthermore, to be able to scale the sub-sample MSE back to the original sample size, a second even smaller sub-sample bootstrap is performed as well. As a by-product of their procedure the ratio of α/β is also estimated. This bypasses the restrictive assumption made in Hall (1990). The AMSE of the control variate is

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

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 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 find k_2^*. Here n_2 is typically chosen to be $n_2 = n_1^2/n$. The optimal number of order statistics is,

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

A second approach is the method by Drees and Kaufmann (1998). Drees and Kaufmann (1998) rely on the results by Hall and Welsh (1985). They show that if the underlying CDF satisfies the Hall expansion, the AMSE of the Hill estimator is minimal for

$$k_{DK}^* \sim \left(\frac{A^{2\rho}\left(\rho+1\right)^2}{2\beta^2\rho^3}\right)^{1/(2\rho+1)} n^{2\rho/(2\rho+1)},$$

with $\rho > 0$, where, for convenience, $\rho = \alpha/\beta$, $A > 0$ and $\beta \neq 0$. Drees and Kaufmann (1998) show that for the estimation of the second-order tail index

$$\widehat{\rho} := \left|\log\left|\frac{\widehat{\gamma}_{n,t_1}^{-1} - \widehat{\gamma}_{n,s}^{-1}}{\widehat{\gamma}_{n,t_2}^{-1} - \widehat{\gamma}_{n,s}^{-1}}\right| / \log\left(\frac{t_1}{t_2}\right)\right|$$

and

$$\widehat{\lambda}_0 := \left|\left(2\widehat{\rho}\right)^{-1/2}\left(\frac{n}{t_1}\right)^{\widehat{\rho}}\frac{\widehat{\gamma}_{n,t_1}^{-1} - \widehat{\gamma}_{n,s}^{-1}}{\widehat{\gamma}_{n,s}}\right|^{2/(2\widehat{\rho}+1)}$$

that

$$\widehat{k}_n := \left[\widehat{\lambda}_0 n^{2\widehat{\rho}/(2\widehat{\rho}+1)}\right] \tag{13}$$

is a consistent estimator of k^*_{DK}.

Drees and Kaufmann (1998) introduce a sequential procedure that yields an asymptotically consistent estimator of k^*. Their estimator relies on the fact that the maximum random fluctuation $i^{1/2}(\widehat{\gamma}_{n,i} - \gamma)$, with $2 \leq i \leq k_n$, is of the order $(\log\log n)^{1/2}$ for all intermediate sequences k_n. This property is used to define the stopping time,

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

where the threshold $r_n = 2.5\widetilde{\gamma}_n n^{1/4}$ is a sequence larger than $(\log\log n)^{1/2}$ and smaller than $n^{1/2}$. 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. Given that $|\widehat{\gamma}_{n,i} - \widehat{\gamma}_{n,k}|$ is composed of a variance and a bias, the bias dominates if the absolute difference exceeds the $(\log\log n)^{1/2}$. Under conditions $r_n = o\left(n^{1/2}\right)$ and $(\log\log n)^{1/2} = o(r_n)$ one shows that $\overline{k}_n(r_n) \sim const.\,(r_n n^{\rho})^{2/(2\rho+1)}$ so that $\left(\overline{k}_n\left(r_n^{\xi}\right)/\overline{k}_n(r_n)^{\xi}\right)^{1/(1-\xi)}$ with $\xi \in (0,1)$ has the optimal order $\widehat{k}_n$ defined in (13). This leads to the adaptive estimator

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

with

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

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

A.4 Alternative penalty functions

To benchmark the relative performance of the penalty function of the KS-distance metric, we introduce four additional metrics for the simulation exercise. The following three are introduced for a comparative MC study to serve as benchmarks for the relative performance of the specific penalty function in (9). The last distance metric in this section is used in the MC horse race.

The following two metrics average the difference measured over the region indicated by T. The first alternative penalty function is the average squared distance in the quantile dimension,

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

The second alternative measure is the average absolute distance

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

These two penalty functions are intuitive and are often used in the econometric literature.

The third metric we consider is motivated by the theoretical test statistic by Dietrich et al. (2002). They develop a statistic to test whether the extreme value conditions apply. We take the discrete form of this statistic and adjust it for our own purpose, resulting in

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

B Tables

Table 5: Hall expansion parameters values

	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	$-\frac{1}{2}\frac{\Gamma(2\alpha)\sin(\alpha\pi)}{\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}$	$-\frac{\alpha^2}{2}\frac{\alpha+1}{\alpha+2}$	$\frac{1}{2}$

Table 6: Estimates of k^* under different methods

	α	KS Dif	TH	5%	Eye-Ball	Drees	Du Bo
Student-t	2	509.89	281.00	500.00	19.67	1036.73	968.18
	3	343.13	132.00	500.00	35.21	841.59	895.65
	4	227.99	78.00	500.00	51.48	754.68	859.72
	5	164.88	53.00	500.00	69.02	708.41	837.61
	6	140.07	40.00	500.00	84.55	677.95	823.24
Stable	1.1	240.82	817.00	500.00	8.00	1481.88	1180.98
	1.3	172.74	292.00	500.00	10.14	1466.24	1218.68
	1.5	137.18	146.00	500.00	12.66	1376.89	1214.89
	1.7	200.45	74.00	500.00	18.88	1176.03	1153.69
	1.9	667.03	27.00	500.00	108.09	861.59	1061.44
Fréchet	2	217.71	928.00	500.00	19.26	1500.70	1305.65
	3	231.47	928.00	500.00	34.99	1501.00	1304.65
	4	226.54	928.00	500.00	51.35	1501.00	1305.28
	5	227.16	928.00	500.00	67.51	1501.00	1303.90
	6	229.31	928.00	500.00	84.04	1501.00	1304.10
ARCH	2.30	290.39		500.00	31.32	1131.36	1244.62
	2.68	300.24		500.00	36.21	1036.93	1244.78
	3.17	290.97		500.00	42.90	947.32	1245.28
	3.82	246.72		500.00	52.81	864.97	1246.05
	4.73	202.79		500.00	64.75	791.26	1247.14

This table depicts the mean for the estimated k^* for the different methodologies. The samples are drawn from four different heavy-tailed distribution families. The samples are drawn from the Student-t, symmetric stable, Fréchet distribution and ARCH process. The different methods are stated in the first row. KS Dis is the Kolmogorov-Smirnov distance metric in (9). TH is based on the theoretically derived optimal k from minimizing the MSE for specific parametric distributions, presented in Equation (11) in the Appendix. The automated Eye-Ball method in (5) is the heuristic method aimed at finding the first stable region in the Hill plot. For the column Drees, k^* is determined by the methodology described by Drees and Kaufmann (1998). Du Bo is the double bootstrap procedure by Danielsson et al. (2001). Here α indicates the corresponding theoretical tail exponent for the particular distribution from which the sample is drawn. The sample size is $n = 10,000$ for $10,000$ repetitions.

Table 7: Descriptive statistics stock data estimates

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
Mean	3.40	2.70	3.19	2.41	1.98	2.97	2.37	2.87	2.22	1.77
Median	3.35	2.72	3.19	2.32	2.05	2.90	2.39	2.87	2.13	1.83
St. Dev.	0.81	0.58	0.65	0.94	0.53	0.81	0.52	0.63	0.76	0.55
Min	0.27	0.16	0.16	0.48	0.19	0.54	0.12	0.11	0.32	0.12
Max	7.79	15.04	8.52	53.00	10.09	7.48	7.22	7.09	45.42	34.91
Skewness	0.40	0.71	0.09	18.34	-0.58	0.42	-0.22	0.15	17.52	12.51
Kurtosis	3.07	19.53	4.56	717.86	11.60	3.01	6.79	4.11	779.80	740.25

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

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
Mean	93.36	185.05	43.85	307.37	452.71	103.07	185.07	35.99	277.36	449.36
Median	47	147	33	259	355	54	147	27	237	352
St. Dev.	125.86	122.21	40.02	190.35	313.67	132.95	122.28	31.94	177.98	314.10
Min	1	50	1	1	97	1	50	1	1	1
Max	1,455	623	773	1,869	1,628	1,819	623	643	1,869	1,655
Skewness	3.45	1.54	3.96	1.56	1.58	3.17	1.54	3.42	1.54	1.58
Kurtosis	20.81	5.18	34.61	6.46	5.34	18.70	5.18	28.55	6.90	5.33

(b) Estimates k_i^*

This table presents descriptive statistics for estimates of $\hat{\alpha}\left(k^*\right)_i$ and k^* by applying the five different methods to choose k^* 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 five different methods are the KS-distance metric, 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 8: Median absolute differences between different methodologies

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	0	0.70	0.46	1.02	1.40	0	0.59	0.42	0.74	1.16
5%	0.70	0	0.50	0.40	0.68	0.59	0	0.50	0.30	0.59
Eye-Ball	0.46	0.50	0	0.86	1.19	0.42	0.50	0	0.73	1.10
Drees	1.02	0.40	0.86	0	0.27	0.74	0.30	0.73	0	0.31
DB	1.40	0.68	1.19	0.27	0	1.16	0.59	1.10	0.31	0

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

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	0	94	35	183	272	0	94	36	159	256
5%	94	0	107	122	207	94	0	114	102	204
Eye-Ball	35	107	0	220	312	36	114	0	206	318
Drees	183	122	220	0	90	159	102	206	0	112
DB	272	207	312	90	0	256	204	318	112	0

(b) Estimates k_i^*

This table presents the median absolute difference between $\hat{\alpha}\left(k^*\right)_i$ and k^* by applying the five different methods to choose k^* for left and right tail of U.S. stock returns. The five different methods are the KS-distance metric, 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 sample size. There are 17,918 companies included in the analysis.

Table 9: The correlation of estimates between different methodologies

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	1	0.38	0.49	0.24	0.24	1	0.41	0.55	0.32	0.26
5%	0.38	1	0.68	0.33	0.65	0.41	1	0.72	0.39	0.62
Eye-Ball	0.49	0.68	1	0.26	0.49	0.55	0.72	1	0.37	0.45
Drees	0.24	0.33	0.26	1	0.24	0.32	0.39	0.37	1	0.25
DB	0.24	0.65	0.49	0.24	1	0.26	0.62	0.45	0.25	1

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

	Left Tail					Right Tail				
	KS Dis	5%	Eye-Ball	Drees	Du Bo	KS Dis	5%	Eye-Ball	Drees	Du Bo
KS Dis	1	0.35	-0.04	0.34	0.35	1	0.30	-0.02	0.31	0.30
5%	0.35	1	0.33	0.75	1.00	0.30	1	0.42	0.71	1.00
Eye-Ball	-0.04	0.33	1	0.08	0.33	-0.02	0.42	1	0.17	0.41
Drees	0.34	0.75	0.08	1	0.75	0.31	0.71	0.17	1	0.71
DB	0.35	1.00	0.33	0.75	1	0.30	1.00	0.41	0.71	1

(b) Estimates k_i^*

This table presents the correlation matrix for the estimates of $\hat{\alpha}\left(k^*\right)_i$ and k^* by applying the five different methods for left and right tail of U.S. stock returns. The different methods are the KS-distance metric, 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

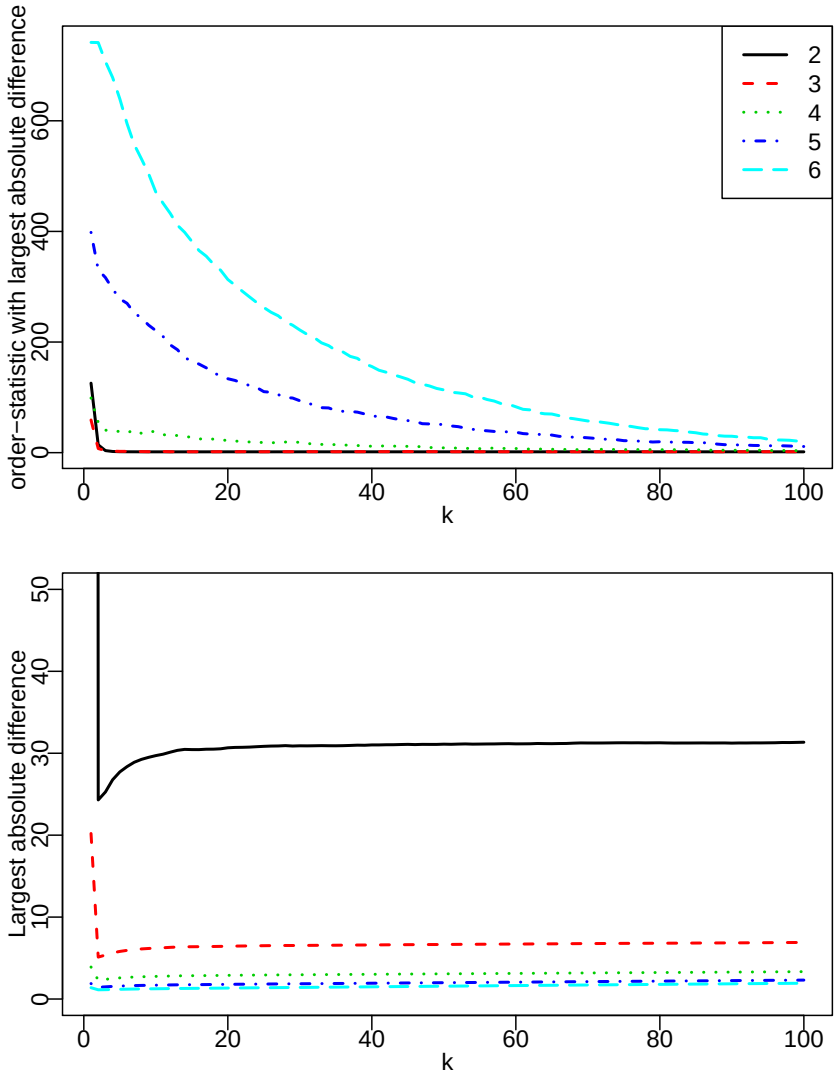


Figure 5: These two figures show the simulations for the limit criterion function in (12). The parameters are for the **Student-t** distribution and are from Table 5 in Appendix B. The value for α for the different lines is stated in the legend. Here T is 1,500; therefore, the interval between $w(s_i) - w(s_{i+1})$ is normally distributed with mean 0 and variance $1/k$. The path of the Brownian motion is simulated 1,000 times. The top figure shows the average number of order statistics at which the largest absolute distance is found for a given k. The bottom graph depicts the average distance found for the largest deviation at a given k. The top and bottom graphs are related by the fact that the bottom graph depicts the distances found at the i^{th} observation found in the top graph.

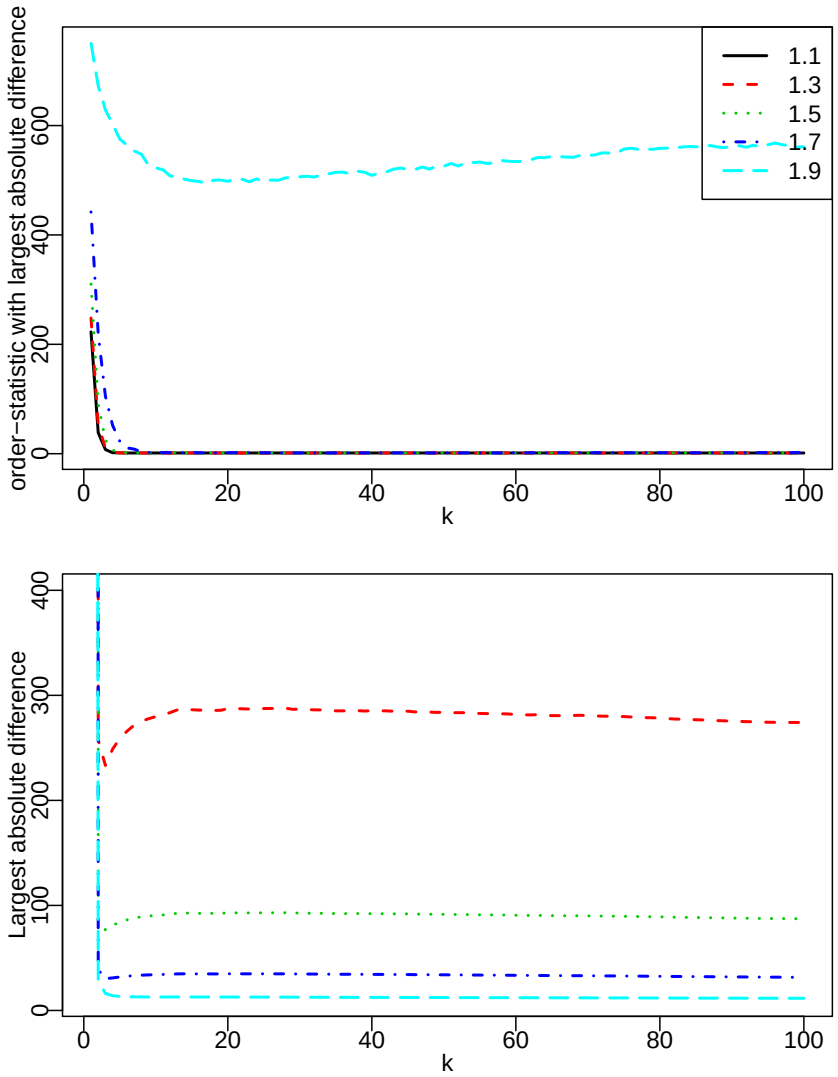


Figure 6: These two figures show the simulations for the limit criterion function in (12). The parameters are for the **symmetric Stable** distribution and are from Table 5 in Appendix B. The value for α for the different lines is stated in the legend. Here T is 1,500; therefore, the interval between $w(s_i) - w(s_{i+1})$ is normally distributed with mean 0 and variance $1/k$. The path of the Brownian motion is simulated 1,000 times. The top figure shows the average number of order statistics at which the largest absolute distance is found for a given k. The bottom graph depicts the average distance found for the largest deviation at a given k. The top and bottom graphs are related by the fact that the bottom graph depicts the distances found at the i^{th} observation found in the top graph. **Note**: For $\alpha = 1.1$ the line in the lower panel stabilizes around 1150.

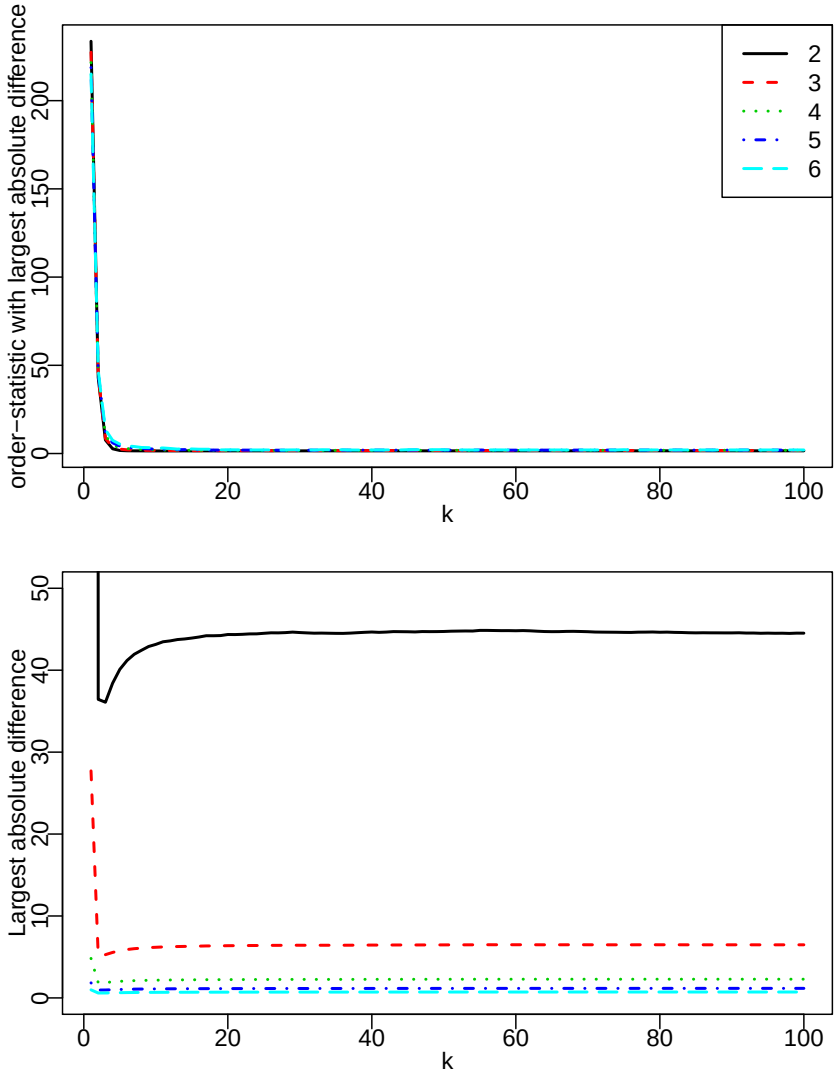
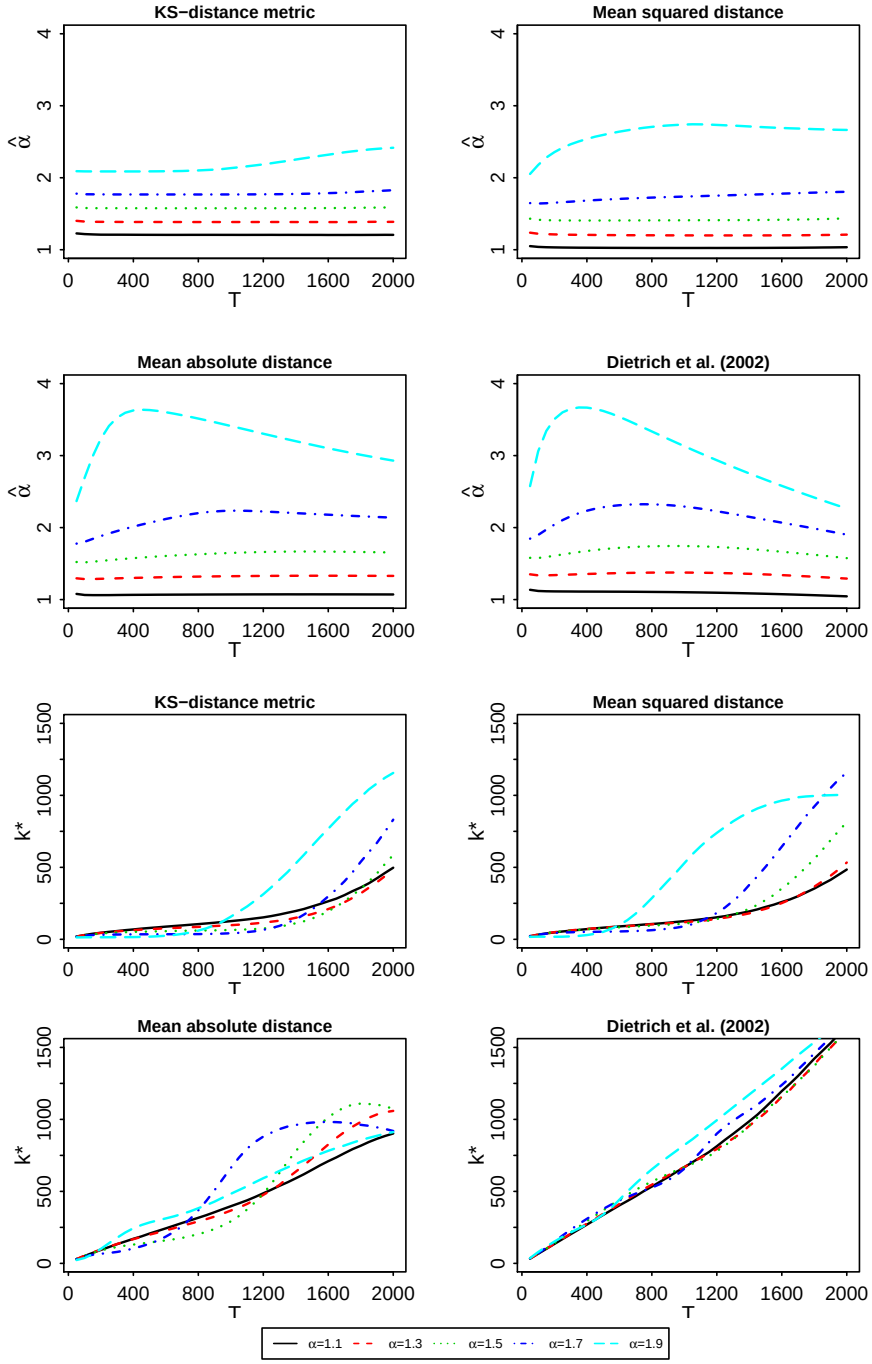


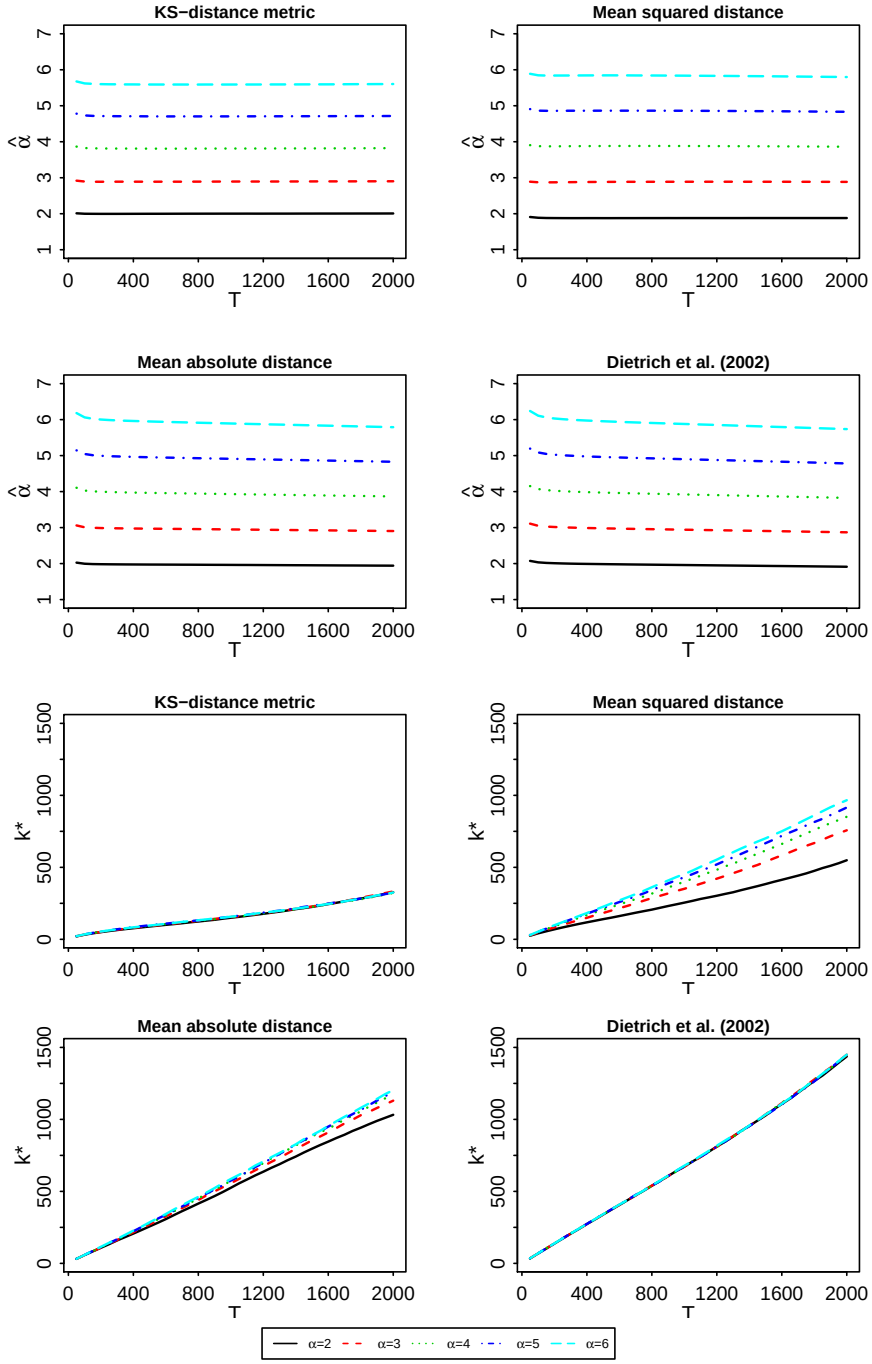
Figure 7: These two figures show the simulations for the limit criterion function in (12). The parameters are for the **Fréchet** distribution and are from Table 5 in Appendix B. The value for α for the different lines is stated in the legend. Here T is 1,500; therefore, the interval between $w(s_i) - w(s_{i+1})$ is normally distributed with mean 0 and variance $1/k$. The path of the Brownian motion is simulated 1,000 times. The top figure shows the average number of order statistics at which the largest absolute distance is found for a given k. The bottom graph depicts the average distance found for the largest deviation at a given k. The top and bottom graphs are related by the fact that the bottom graph depicts the distances found at the i^{th} observation found in the top graph.

Figure 8: $\hat{\alpha}\,(k^*)$ for quantile metrics (symmetric stable Distribution)



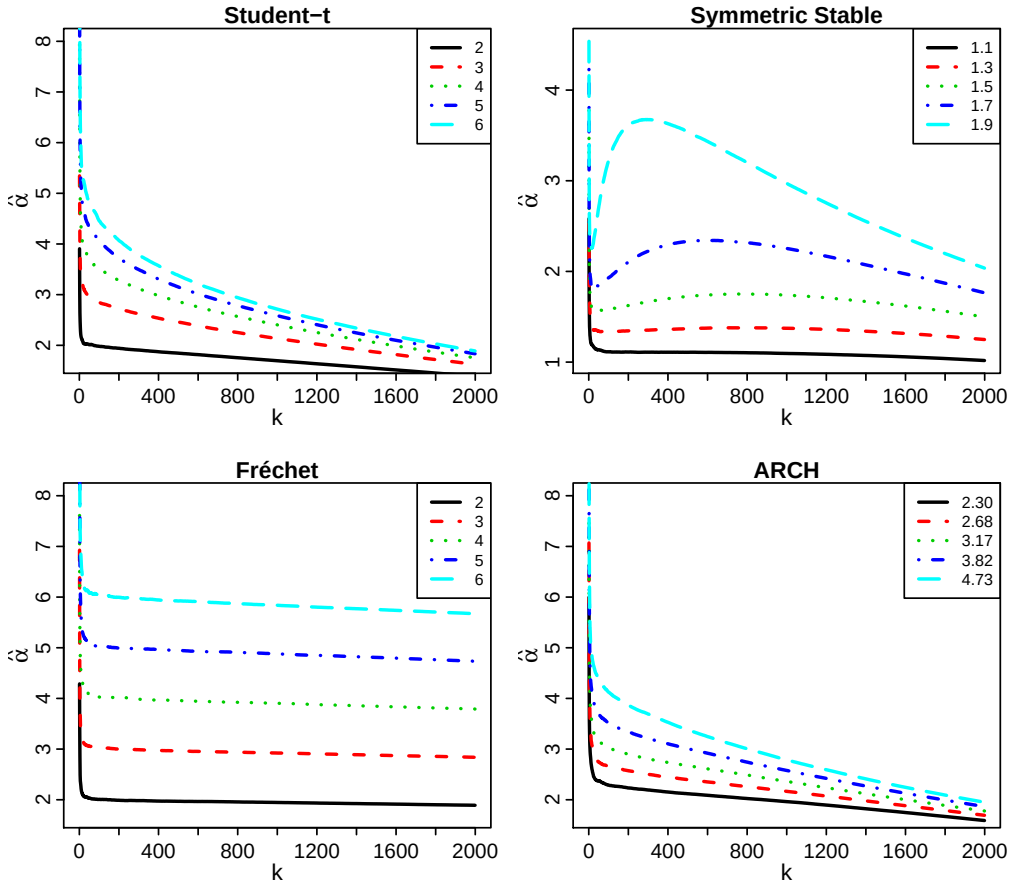
This figure depicts simulation results of the average optimally chosen k^* and $\alpha(k*)$ for a given level of T. Here T is the number of extreme-order statistics over which the metric is optimized. The upper four graphs depict the optimal $\alpha(k^*)$ and the lower four graphs show the choice of k^* for different values of T. This is also done for the 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 **symmetric stable**(α) distribution.

Figure 9: $\hat{\alpha}$ for quantile metrics (Fréchet Distribution)



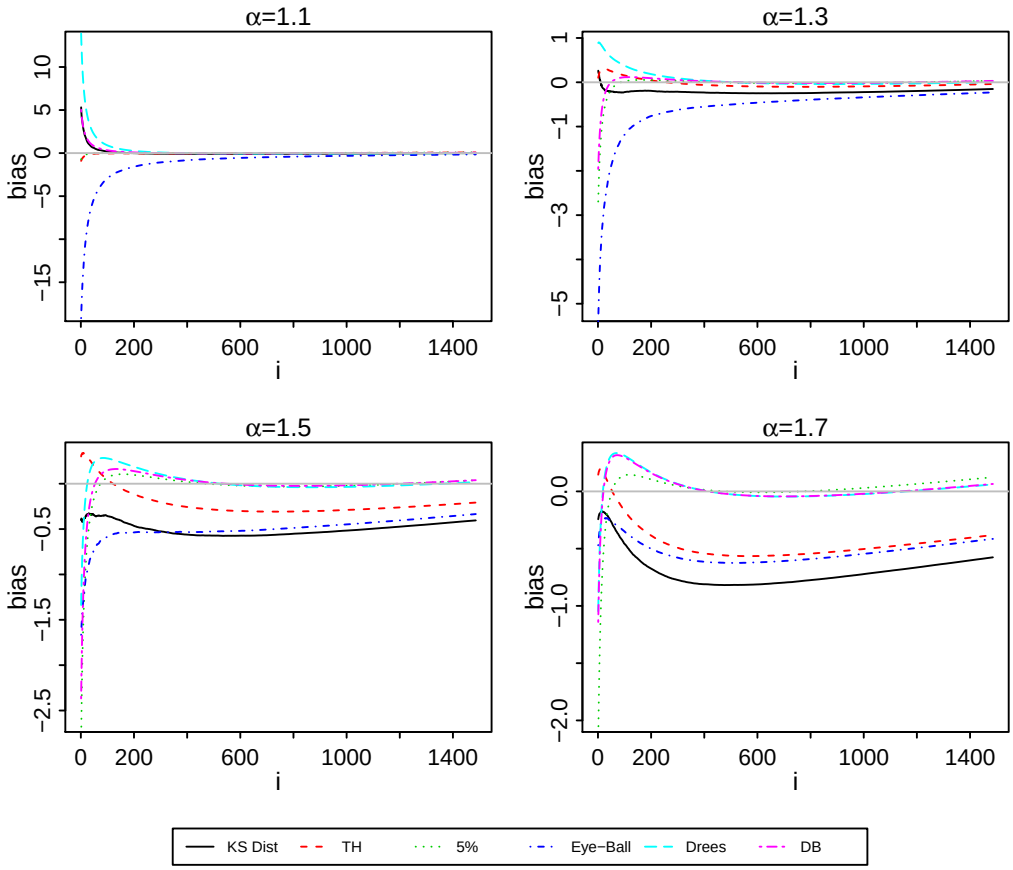
This figure depicts simulation results of the average optimally chosen k^* and $\alpha(k*)$ for a given level of T. Here T is the number of extreme-order statistics over which the metric is optimized. The upper four graphs depict the optimal $\alpha(k^*)$ and the lower four graphs show the choice of k^* for different values of T. This is also done for the 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 **Fréchet**(α) distribution.

Figure 10: Shape of the Hill plot for different distributions



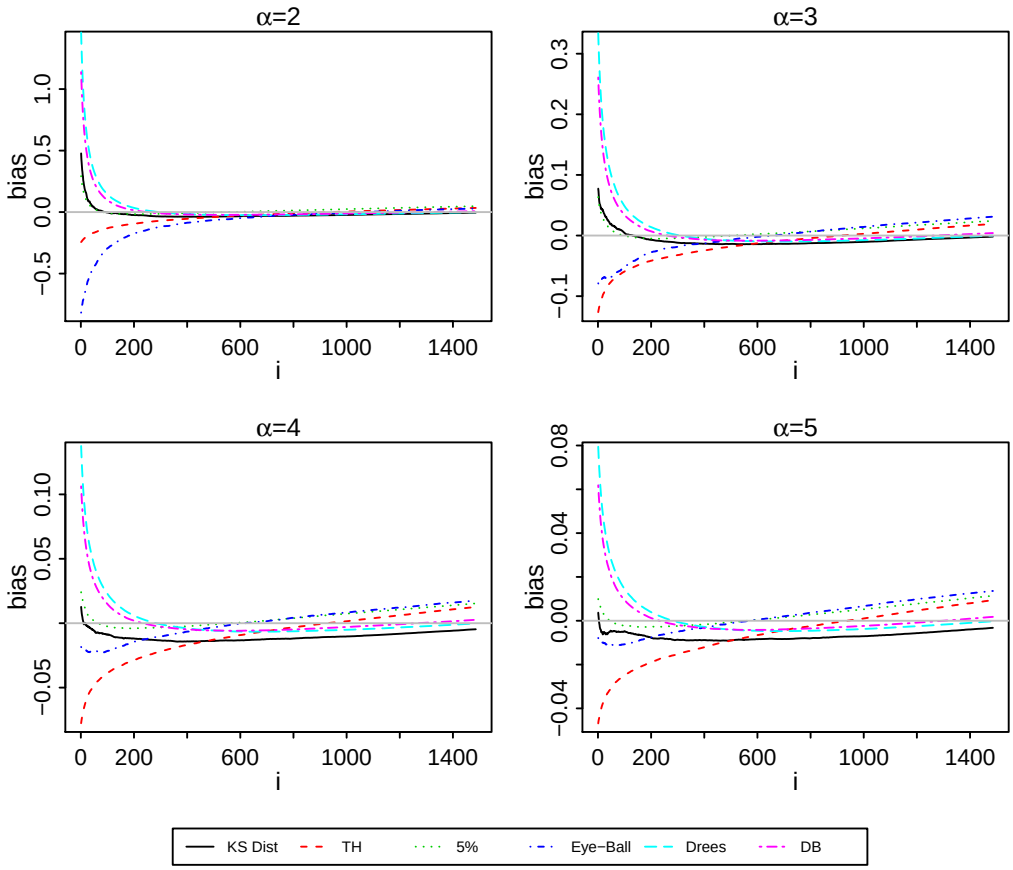
The figure depicts the Hill estimates for the Student-t, symmetric stable, Fréchet(3) distribution and ARCH process against the number of order-statistics used in the estimation. The legend indicates the value of the tail index given the distributional family. For the ARCH process, we use the Kesten theorem to derive the tail index for a range of ar(1) coefficients. These graphs are constructed by taking the average Hill estimates over 500 simulations.

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



This figure show the median difference induced by using the quantile estimator presented in Equation (8). 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 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 (8). 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.