Tail Index Estimation: Quantile Driven Threshold Selection.*

Jon Danielsson
London School of Economics
Systemic Risk Centre

Lerby M. Ergun
London School of Economics
Systemic Risk Centre

Laurens de Haan
Erasmus University Rotterdam

Casper G. de Vries
Erasmus University Rotterdam
Tinbergen Institute

January 2016

Working Paper

Abstract

The selection of upper order statistics in tail estimation is notoriously difficult. Most methods are based on asymptotic arguments, like minimizing the asymptotic mse, that 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 analyse the finite sample properties of the metric we organize a horse race between the other methods. 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 economic relevant variation between the tail index estimates.

*Corresponding author Lerby M. Ergun. We thank NWO Mozaiek grant and the Economic and Social Research Council (UK) [grant number: ES/K002309/1] for supporting the research. All errors are ours. Additional Tables and Figures are available in the online appendix at www.lerbyergun.com/research

1 Introduction

In various research fields the tails of distributions are characterized as being heavy tailed, e.g. the scaling behaviour described by Zipf's Law and Gibrat's law (Reed, 2001). In the statistical literature there is an ongoing debate on the number of tail data that have to be used in the estimation of the tail index. The tail index is the shape parameter of these heavy tailed distributions. The most popular estimator for the tail index of heavy tailed distributions is the Hill (1975) estimator. This estimator necessitates a choice of the number of order statistics utilized in the estimation of the tail index. This number is referred to as k. The choice of k leads to a trade-off between the bias and variance of the estimator. The literature to date utilizes the minimization of the asymptotic mean squared error (mse) as the criterion on which k is based. The methods that are used to find the k that minimizes the mse are asymptotically consistent, but have unsatisfactory finite sample properties. This paper proposes a novel methodology to pick the optimal k, labeled as 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 that the metric outperforms the methods put forth by the theoretical statistical literature.

The theoretical statistical literature and applied literature offer different methods to choose an optimal k. These methods can be roughly divided into two groups. The first group consists of heuristic approaches. These methods are often used in applications and focus on analysing the plot of k against the estimates of the tail index. Examples of these methods are the Eye-Ball method and the automated form of the Eye-Ball method (Resnick and Starica, 1997). Another heuristic rule is picking a fixed percentage of the total sample size, for instance 5% of the upper order statistics. These methods have a weak theoretical foundation and might therefore not be robust.

The second group of methods derives from the theoretical statistical literature. These are based on the minimization of the mean squared error (mse) of the estimator. Hall (1990) and Danielsson, Peng, De Vries, and De Haan (2001) utilize 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 their finite sample properties are subject to improvement.

The shortcomings of the currently available methods motivated our new approach. In this paper we utilize the penalty function of the Kolmogorov-Smirnov statistic to fit the tail of the distribution. From the fitted tail we can subsequently determine the optimal k^*. This procedure is partially inspired by Bickel and Sakov (2008). Bickel and Sakov (2008) consider a bootstrap procedure for inference regarding the tail of a distribution. Since a full sample bootstrap is known to be inconsistent for replicating tail properties, they use a subsample bootstrap. Their procedure aims to determine the proper sub-sample bootstrap size for heavy tailed distributions. Their procedure exploits the difference between subsequently smaller subsample bootstrap distributions. The optimal subsample size has the smallest distance to its smaller subsample neighbor. In their adaptive rule the Kolmogorov-Smirnov (KS) test statistic functions as the distance metric. The Kolmogorov-Smirnov statistic is measured as the maximum probability difference between the empirical distribution and a parametric distribution.

In this paper we utilize Bickel and Sakov's (2008) minimal maximum deviation criterion, but with a twist. Instead of minimization in the probability dimension, we minimize in the quantile dimension. This measure will hence forth be referred to as the KS distance metric. The benchmark is the Pareto distribution. The tail similarity of the heavy tailed EVT distributions allows us to model the tails of these distributions with the Pareto distribution. The estimates of the scaling constant and the tail index of the Pareto distribution depend on k. By varying k we are able to simultaneously fit the empirical distribution and elicit k^*.

The particular choice of the metric is motivated by problems that are specific for fitting the tail of the distribution. In the tail region, small mistakes in the probability domain lead to a large distortion in the quantile domain. However, the quantile is the domain that economist care about. Therefore, we choose to base our metric in the quantile dimension rather than the probability dimension. Given the choice for the quantile domain, the choice of penalty function is specific to the problem of fitting the tail. In the tail of the distribution a small mistake in probability leads to an increasing error for subsequently higher quantile levels. Consequently, no extra emphasis has to be put on larger deviations as these naturally occur the further we move into the tail. We therefore choose to represent the errors in absolute values rather than squared deviations. Furthermore, deeper in the tail the size of the expected deviations become larger in the quantile domain. By focusing on the maximum, the metric is not diluted by the numerous center observations.

These intuitive arguments are backed up by a rigorous simulation analysis

and tests. To test the performance of the KS distance metric we analyze it from four different perspectives. Firstly, to get a better idea of the behavior of the KS distance, the metric is placed in a more general setting by considering a Brownian motion representation. This allows us to analize the metric without any strong parametric assumptions. We analize the properties of the limit behaviour via extensive simulation studies. The simulation study shows that our methodology locates an interior k^* for various specifications. Other metrics often locate k^* at the boundaries of the admissible area and as a result select a very low or very large number of order statistics.

Subsequently, we test the KS distance metric against various other penalty functions. These penalty functions are often used in econometric applications. For instance, the OLS estimator is based on minimizing the average squared errors. To evaluate the different metrics ability to correctly penalize errors we use the results derived by Hall and Welsh (1985). They derive the theoretical k^* for the class of distributions that has a second order term which is also hyperbolic. These results stipulate the behaviour of k^* as a function of the sample size. Also, k^* is a function of the heaviness of the tail within a family of heavy tailed distributions. For example, for the Student-t distribution with $\alpha > 1$, there is a negative relationship between k^* and the degrees of freedom. Furthermore, as the metric is only measured over the tail of the distribution a desired property of the metric is that k^* does not change as the measurement area is altered.

To test the performance of the KS distance metric we contrast the metric with several other penalty functions in Monte Carlo simulations. We use the mean squared error, mean absolute error and the discrete version of the metric used in Dietrich, De Haan, and Hüsler (2002) to benchmark the performance. The KS distance metric shows promising results. For the Student-t, Symmetric Stable and Fréchet distribution the patterns mentioned above observed for k^* are as predicted by the Hall and Welsh (1985). The other metrics fail to reproduce stable results. This corroborates with the patterns for k^* as mentioned above. This translates in more unstable and biased estimates for the tail index than the KS distance metric.

To test the finite sample properties of the different methods we perform various Monte Carlo simulation studies. In this horse race we simulate from various families of heavy tailed distributions which conform the Hall expansion or the conditions of the Kesten Theorem. For distributions that fit these requirements, the underlying value of the tail index is known. As in the previous section, this allows us to benchmark the results. The primary focus of

the simulations is the estimates of the tail index although the prime interest in the end is quantile estimation. As a result, we also evaluate the methods in this directions. Furthermore, we evaluate k^* chosen by the various methods. Studying the behavior of k^* helps us to better understand the estimates of α.

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 distribution, 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 α for less heavy tailed distributions. The alternative methodologies often choose a high value for k^* which corresponds to a large bias.

We also consider dependent stochastic processes. For example, an ARCH process captures the volatility clustering in financial data. It turns out that 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 this type of stochastic process. For the dependent time series, the automated Eye-Ball method and the KS distance metric produce small biases.

In addition to estimating the tail index, we model the quantile function for the various competing methods. We also compare the quantile estimates at different probability levels to evaluate the performance at different regions of the tail. We find that the KS distance metric produces relatively skewed and volatile quantile estimates. Therefore, the bias of the KS distance metric is relatively large. The distribution of the errors is skewed and consequently the mean difference criterion produces a distorted image. When we analyse the median of the errors, the KS distance metric performs well for the quantiles beyond the 0.995 probability level. For the quantiles further towards the center of the distribution, a large bias arises. The automated Eye-Ball method produces less volatile estimates, but a similar bias in the estimates to the KS distance metric. These results can be explained in light of the method to choose k^*. These methods have a tendency to pick a small k^* and therefore fit the tail close to the maximum. The other methodologies often utilize a larger number of order statistics and consequently fit well closer towards the center of the distribution.

The Monte Carlo studies are limited to parametric distributions. In real world applications the underlying stochastic process is unknown. Therefore, it is difficult to assess the importance of the choice of methodology. In the last section of this paper we show that the choice of k^* is an economically

important choice. For this purpose we use individual daily stock price information from the Center 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 estimates. We find that the average difference between the methods ranges from 0.13 to 1.80. 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 (relevant for the confidence bounds around the VIX index), does not exist. This shows that the choice of methodology to choose k^* is economically important and impacts the tail estimate significantly.

The paper first introduces the EVT framework and the Hill estimator. This is followed by an overview of the theoretical and heuristic methodologies from the literature. We then explain the framework of the distance metric and give the intuition behind the KS distance metric. Section 4 explains the set-up of the Monte Carlo horse race and analyses the simulation results. Section 5 presents the results of the different methodologies for daily stock return data, followed by concluding remarks.

2 Extreme Value Theory methodology

The Extreme Value Theory (EVT) methodology employed in this paper comes in two parts. The first part provides a review of the main EVT results. It is the stepping stone for the semi-parametric approach. 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 $Y_n = \max(X_1, ..., X_n)$ is not beyond a certain threshold x. This probability is given by

$$\mathrm{P}\{Y_n \leq x\} = \mathrm{P}\{\max(X_1, ..., X_n) \leq x\} =$$

$$\mathrm{P}\{X_1 \leq x, X_2 \leq x..., X_n \leq x\} = \prod_{i=1}^{n} \mathrm{P}\{X_i \leq x\} = [\mathrm{F}(x)]^n .$$

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. EVT gives the forms of $\mathrm{G}\left(x\right)$ that can occur as a limit, first derived by Fisher and Tippett (1928) and Gnedenko (1943). There are three possible $\mathrm{G}\left(x\right)$, depending on the shape of the tail of $\mathrm{F}\left(x\right)$. This paper concentrates on the case of 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 γ 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. This property characterizes the distributions that fall in the domain of attraction of the heavy tailed EVT limit distribution. This 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 γ are proposed in the literature (Hill, 1975; Pickands, 1975; De Haan and Resnick, 1980; Hall, 1982; Mason, 1982; Davis and Resnick, 1984; Csorgo, Deheuvels and Mason 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 are the number of upper order statistics used in the estimation of γ. 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 $1/\gamma$ 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.[1] 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]

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

Using the Hall expansion one shows the asymptotic bias as

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

Equation (4) provides the relationship between the threshold s and the bias of the Hill estimator.[3] From (4) one notices that as s becomes smaller, i.e. the threshold moves towards the center of the distribution, that the bias increases.[4]

The asymptotic variance of the Hill estimator is,[5]

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

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^* which balances the two vices. Given this objective, how to elicit the minimum mse from the data? This is the topic of the next section.

[1] See Appendix A.1.

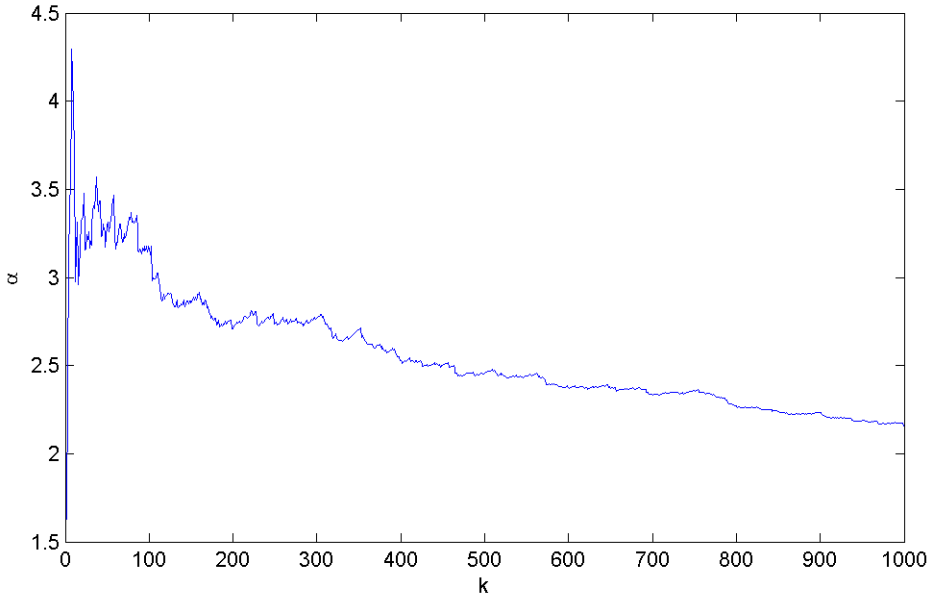
[2] Heavy tailed parametric distributions like the Student-t, Symmetric Stable and Fréchet distribution all conform the Hall expansion. The parameter values for these distributions are presented in Table 4 of the Appendix.

[3] Here s is the quantile at which the threshold is set.

[4] This result is based on the second order expansion by Hall and Welsh (1985).

[5] See Appendix A.2.

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.

2.2 Finding k^*

Various methods exist for choosing k^*. These methods can be roughly divided into two groups. The first group of methods come from the theoretical statistics literature and are based on asymptotic arguments. The second group of methods stem from suggestions by practitioners. The later are more heuristic in nature, but some perform surprisingly well. The next section elaborates further on these approaches.

2.2.1 Theoretical based methods

Hall (1990) and Danielsson et al. (2001) utilize the bias and the variance to minimize the asymptotic mean square error (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 subsample bootstrap to find the k^* under the restrictive assumption $\alpha = \beta$ in

(3). To obtain the optimal rate[6] 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 $\hat{\gamma}$.

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 subsample bootstrap is applied. Furthermore, to be able to scale the subsample mse back to the original sample size, a second even smaller subsample 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,

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

where

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

Here $n_1 = n^{1-\epsilon}$ is the smaller subsample 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,

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

[6]The subsample bootstrap size needs to increase slower than n to achieve asymptotic optimality in the bootstrap procedure.

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

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

that

$$\widehat{k}_n := \left[\widehat{\lambda}_0 n^{2\widehat{\rho}/(2\widehat{\rho}+1)} \right] \tag{5}$$

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 $i^{1/2} \left(\widehat{\gamma}_{n,i} - \gamma\right)$, with $2 \leq i \leq k_n$, is of the order $\left(\log\log n\right)^{1/2}$ for all intermediate sequences k_n. This property is used to define the stopping time,

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

where the threshold $r_n = 2.5\widetilde{\gamma}_n n^{1/4}$ is a sequence larger than $\left(\log\log n\right)^{1/2}$ 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 $\left(\log\log n\right)^{1/2}$. 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 (5). 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

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

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

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.

2.2.2 Heuristics

Applications in the economic literature frequently resort to heuristic rules. These rules are based on finding the region where the Hill plot, as in Figure 1, becomes more stable. This is the region where, as k increases, the variance is subsiding, and the bias of the Hill estimators has not become dominant yet. The method of finding the stable region in the Hill plot by observation is referred to as the "Eye-Balling technique".

This method might be practical for a single experiment, but other applications require a more automated approach. Automated approaches are often based on an algorithm which tracks the variance of the Hill plot as k increases. These algorithms 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,

$$k_{eye}^{*} = \min\left\{k \in 2,...,n^{+}-w | h < \frac{1}{w}\sum\nolimits_{i=1}^{w} \mathrm{I}\left\{\widehat{\alpha}(k+i) < \widehat{\alpha}(k) \pm \varepsilon\right\}\right\}. \quad (6)$$

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 plot. The ε gives the range between which $[\hat{\alpha}(k+1), ..., \hat{\alpha}(k+w)]$ are within the permitted bound around $\hat{\alpha}(k)$. No less than $h\%$ of the estimates should be within the bound of $\hat{\alpha}(k)$ 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.[7]

The Eye-Ball method and the corresponding automatized algorithm, attempt to find a proper trade-off between the variance and the bias. To find this stable region, the information in the high variance region that precedes is ignored. Similarly, once the bias causes the Hill plot to fall off sharply, this variance bounds method ignores such a region as well.There is a possibility that there is an optimal choice of k nested in the high variance region of the Hill plot.

Other heuristic methods are more blunt and take a fixed percentage of the total sample. Kelly and Jiang (2014), for instance, use the 5% sample fraction to estimate the tail index for the cross-section of the US stock market returns to price disaster risk.

The 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. Therefore, choosing a fixed portion of the sample is not appropriate. The optimal sample fraction also depends on the sample size.

3 Alternative framework

The shortcomings of the existing methods outlined above motivated our alternative approach. This alternative 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). Bikkel and Sakov show that a sub-sample bootstrap is consistent in many cases, but may fail in some important examples. They show that an adaptive rule based on the minimization of the Kolmogorov-Smirnov (KS) test statistic

[7] In the Monte Carlo studies we choose n^+ to be a prespecified threshold which also applies to the other methods. Later to be defined as T.

finds the proper sub-sample bootstrap size.[8] For instance, the Gumbel, Exponential and the Uniform distribution need relatively small sub-samples for convergence. Therefore, the choice of the subsample size in the bootstrap procedure is essential.

Bickel and Sakov (2008) find the proper subsample size by matching the empirical and theoretical distribution. We use 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, all distributions in this class by definition 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.[9] The choice of the supremum, rather than other well known penalty functions, is validated via a simulation study.

3.1 Motivation for the KS distance metric

The new metric deviates from the classic Kolmogorov-Smirnov distance. The difference lies in the fact that the distance is measured in the quantile dimension rather than the probability 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, Expected Shortfall, and the variance are concepts related to quantiles at a given probability level in the horizontal dimension. The second motivation is more technical. Our analysis is solely focused on the tail of the distribution, rather than the center 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 dimension.

Given the decision to measure over the quantile dimension, a function is needed to penalize deviations from the empirical distribution. Some examples of penalty functions are the mean squared error, and the mean absolute error in addition to various others that weigh the deviations differently. Different penalty functions put emphasis on minimizing a specific array of

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

[9] $L(1)$ is constant.

mistakes. For instance, the mean squared error punishes large deviations disproportionally more than small errors. The penalty function we opt for is the maximum absolute deviation. The maximum absolute deviation has specific benefits which makes it suitable for fitting the quantiles in the tail.

The inner part of the penalty function takes the absolute difference between the quantiles instead of, for instance, the squared difference. The reason is that our application focuses on fitting tail quantiles. 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 center observations. This, for instance, is the case when the differences are averaged.

3.2 The distance metric

The starting point for locating k^* is the first order term of the power expansion:

$$\mathrm{P}\left(X \leq x\right) = \mathrm{F}(x) = 1 - Ax^{-\alpha}[1 + o(1)]. \tag{7}$$

This function is identical to a Pareto distribution if the higher order terms are ignored. By inverting (7), we get the quantile function

$$x = \left(\frac{\mathrm{P}\left(X \geq x\right)}{A}\right)^{\frac{1}{-\alpha}}. \tag{8}$$

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

$$q\left(j,k\right) = \left(\frac{\mathrm{P}\left(X > x\right)}{A}\right)^{\frac{1}{-\alpha}} = \left[\frac{k}{j}\left(x_{n-k+1,n}\right)^{\widehat{\alpha}_k}\right]^{\frac{1}{\widehat{\alpha}_k}}. \tag{9}$$

Here j is the $(n-j)^{th}$ order statistic $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n-j,n} \leq ... \leq X_{n,n}$ such that j/n comes closest to the probability level $\mathrm{P}\left(X > x\right)$.

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

$$Q_{1,n} = \inf_k \left[\sup_T \left| x_{n-j,n} - q\left(j,k\right) \right| \right], \quad for\ j = 1, ..., T \tag{10}$$

where $T > k$ is the region over which the KS distance metric is measured. Here $x_{n-j,n}$ is the empirical quantile and $q\left(j,k\right)$ is the estimated quantile from (9). This is done for different levels of k. The k, which produces the smallest maximum horizontal deviation along all the tail observation till T, is the k^* for the Hill estimator.

3.3 Brownian motion representation

There are various ways to study the behavior of this metric. Our first approach to study the properties of the KS distance metric is to model the quantile process with a Brownian motion representation. This helps in devising the Monte Carlo experiments. By Theorem 2.4.8 from De Haan and Ferreira (2006, page 52) the KS distance metric in (10) can be written as[10]

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

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

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

For the case that the cdf satisfies the Hall expansion (7) 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[11] we arrive at,

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

and

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

Below we report the results of an extensive elaborate simulation study on the performance of the minimization in (11) and whether this renders a k for

[10] For the derivation see Appendix A.5.

[11] See Bringham, Goldie, and Teugels (1989, p. 29).

which the Hill estimator performs well. Thus far, we have not been able to prove on the basis of (11) that the resulting choice for k yields a consistent Hill estimator of α, nor whether k^* minimizes the asymptotic mse. However, the simulation studies at least seem to suggest that the new criterion performs better than existing approaches.

3.4 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 (10). 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.[12]

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

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

[12]The vast literature on ordinary least square- and least absolute deviation regressions demonstrates this.

For the purpose of benchmarking the KS distance metric in the Monte Carlo horse race, we introduce a slightly altered version of the metric. We normalize the distance by dividing the difference by the threshold quantile $x_{n-k+1,n}$.[13] This results in

$$Q_{5,n} = \inf_k \left[\sup_T \left| \frac{x_{n-j,n}}{x_{n-k+1,n}} - \frac{k}{j}^{\frac{1}{\hat{\alpha}_k}} \right| \right]. \tag{12}$$

The relative distance leads to a slight alteration of the KS distance metric. This ratio metric will be introduced in the Monte Carlo horse race as an alternative to the $Q_{1,n}$ measure.

3.5 Simulation approach

We use a Monte Carlo simulation to study the properties of the KS distance metric. The theoretical derivation of the optimal number of order statistics for the Hill estimator by Hall and Welsh (1985) gives some guidelines on how the optimal threshold behaves. This provides us the opportunity to analyse the properties of k^* across different distributions. For the simulation study, we choose distribution families which adhere to the Hall expansion in Equation (3). These distributions therefore have a known α and k^*, where k^* minimizes the amse.

In the simulation study, estimates of α and k^* for different penalty functions are analyzed. There are six properties which we evaluate. The first property is the bias in the estimate of α. Secondly, we compare the estimates for α of different members within the same distribution family, like the Student-t. This helps us to isolate the specific performance of metrics keeping the distribution family constant.

Thirdly, the results derived by Hall and Welsh (1985) give us the level of k^* for a given parametric distribution and sample size.[14] 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^* which are inversely related for most distributions. We evaluate whether this is born out in the simulations. 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 should lead to a smaller proportion of observations being used for the estimation of α. The methods should capture this decline.

[13] The normalized difference lends itself better for proving theoretical results.

[14] See Appendix A.3.

The sixth property we pay attention to is the fact that the choice of k^* does not depend on the tail region we optimize over, i.e. T. We have defined above that the metric is optimized over a region $[x_{n-T,n}, \infty]$, where $T > k$. Here T is an arbitrarily chosen parameter, but in the simulation study we investigate the sensitivity of k^* with respect to T.

3.6 MC Brownian motion representation

To study the behavior of the KS distance metric, we perform Monte Carlo simulation studies. Firstly, the simulations are done on the basis of the stochastic process representation of the distance criterion in (11). This allows us to study the metric under relatively general conditions. The second step is to test the relative performance of the KS distance metric.

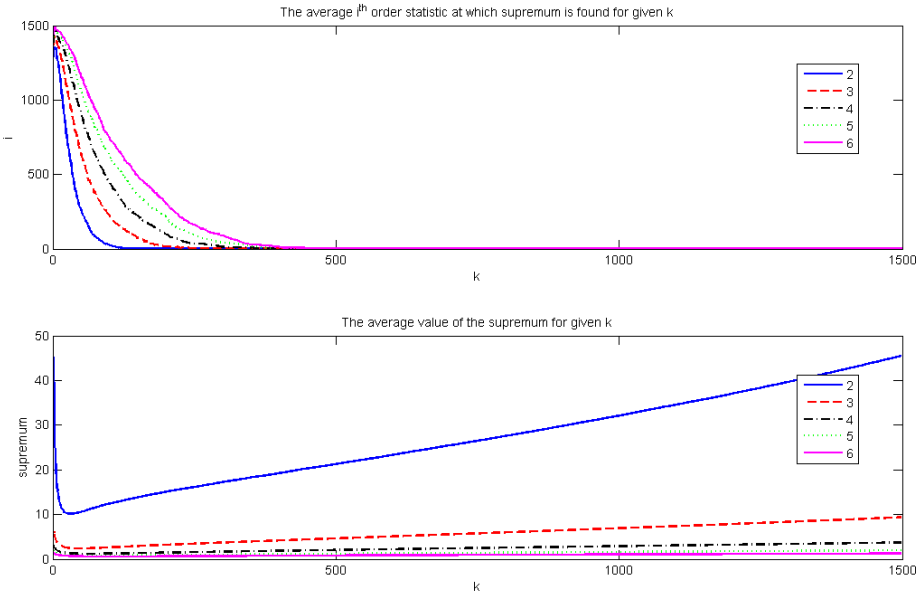
Simulating the function in Equation (11) necessitates a choice of values for parameters α, β, A and B. For the robustness of the Monte Carlo simulations, we use distributions and processes that differ along the dimension of α, A and the second order terms in (3). These parameters are extracted using the power expansion by Hall for the Student-t, Symmetric Stable and Fréchet distribution.

Figure 2 shows whether the KS distance metric finds an interior k^* which has the smallest maximum absolute difference. The upper graph in Figure 2 shows where the largest deviations are found for a given k. This illustrates that the largest deviations are often found close to the extremes of the optimization area. By using few observations, a small k, the tail is fitted towards the largest observations. As a consequence, the largest deviation is found towards the center of the distribution. This logic also holds for when k is fixed towards the center. For k large, the largest deviation is found deep in the tail of the distribution. Combined, the two results imply that there is an optimum somewhere between the two endpoints. Given the upper graph of Figure 2, the lower graph shows how large on average these corresponding deviations are for a fixed k. In addition, it tells us which k on average gets the best fit by producing the smallest maximum deviation.

The lower graph shows a U-shape. This implies that the KS distance metric does not provide an undesirable corner solution. This corresponds to a k^* which is not in the volatile or extremely biased area of the Hill plot.[15]

[15]The figures for the Symmetric Stable and the Fréchet distribution are in Appendix B.2.

Figure 2: Simulations Brownian motion for Student-t parameters



These two figures show the simulations for the limit function in (11). The parameters are for the Student-t distribution, which are found in Table 4 of the Appendix. 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.

The simulation results for the Fréchet distribution are presented in Figure 7 in the Appendix. The results for the Fréchet distribution show a similar pattern to the Student-t distribution. A U-shaped pattern emerges for the value of the supremum for increasing values of k.

For the Symmetric Stable distribution, the results are less clear. Figure 6 in the Appendix does not show the same pattern as the previously discussed distributions. For $\alpha = 1.7$, there is no clear pattern in which the largest deviations are found for a given k. For $\alpha = 1.9$, the largest deviations are found at the observations closest to the center of the distribution for almost any given k. For $k > 1,400$, the largest deviations are on average found further towards the more extreme tail observations. For the other shape parameter values the largest deviations are found close to either the largest or smallest observations in the optimization area. This is similar to the patterns

found for the Student-t distribution. In the lower graph the supremum fails to produce the desired U-shape, suggesting no clear interior k^* is found for this distribution family. We therefore suspect that the performance of the KS distance metric is less clear for the Symmetric Stable distribution.

3.7 MC penalty functions

Next, the Monte Carlo simulation study of the relative performance of the KS distance metric is presented. We contrast the performance of the KS distance metric with the other three metrics presented in Section 3.5. For a thorough analysis, we draw samples from the same three families of heavy tailed distributions as in the previous section.

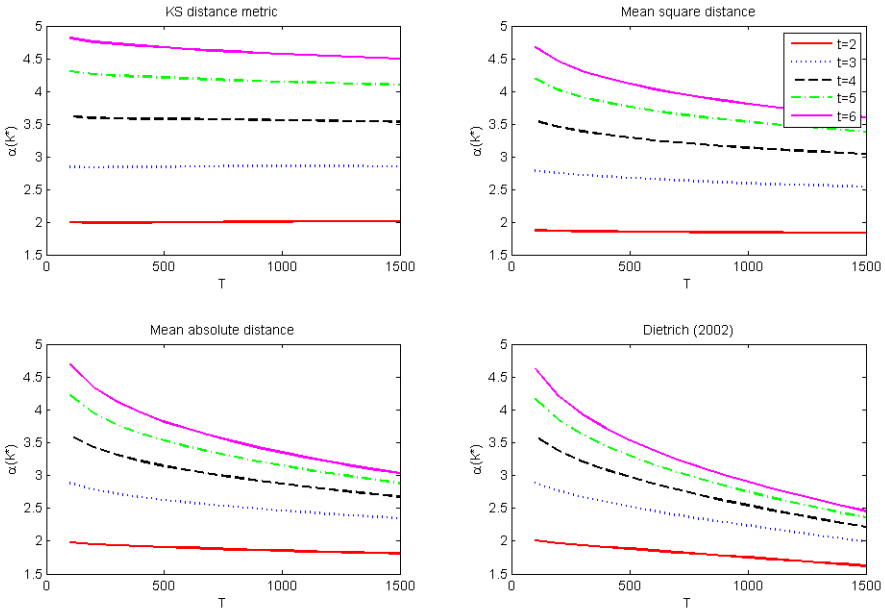
In Figure 3, the level of $\alpha(k^*)$ is displayed against the area over which the specified metric is optimized, i.e. T. These plots give an indication whether the $\alpha(k^*)$ is at the right level and stabilizes as a function of T.

The first fact to notice in Figure 3 is that for the KS distance metric the curves are relatively flat. More importantly, the curves come closest to the theoretical α levels. 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 it is hard to choose an optimal k-level on the basis of the three criteria.

In Figure 17, in the Appendix, the curves for the Symmetric Stable distribution family are depicted. In the upper left graph, the curves for the KS distance metric are relatively horizontal. This indicates that the inability to generate a U-shape in section 3.6 does not have a strong influence on the estimate of α. The estimates of the level of α are not unbiased. 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 $\alpha = 2$ which is the thin tailed normal distribution. The normal distribution falls in the domain of attraction of the Gumbel distribution and is therefore outside of the domain of the Hill estimator.

All the estimates of α have a small bias when the samples are drawn from the Fréchet distribution family, as can be observed from Figure 18 in Appendix B.2. The Fréchet distribution has a small bias and therefore the choice of k^* is less crucial. In addition, the k^* which minimizes the amse, as derived by

Figure 3: Optimal $\hat{\alpha}$ for 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.

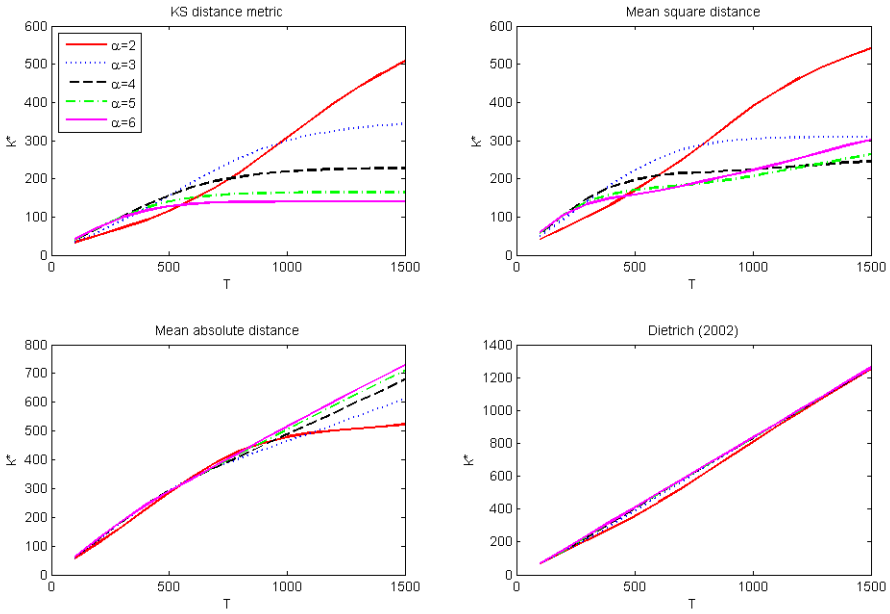
Hall and Welsh (1985), for the Fréchet distribution does not depend on the shape parameter α. This implies that the same k^* minimizes the amse for different members within the Fréchet distribution family.[16]

Figure 4 depicts the average k^* over 10,000 simulations for the Student-t distribution family. Via these figures we can study the properties of k^* as the interval $[0, T]$ 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 given that the area you optimize over is sufficiently large. For the Student-t (2) distribution no such stabilization occurs. The average mean square 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

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

optimal estimation of $\alpha(k)$. This stabilization does not occur for the mean absolute difference and the metric by Dietrich et al. (2002).

Figure 4: 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 $n = 10,000$.

Next, we study the relationship of the average level of k^* for the different members within the distribution family. In Figure 4 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 the k^* derived by minimizing the amse. This pattern is not observed for the other criteria.

Figure 19 in the Appendix depicts the results for the Symmetric Stable distribution. The plots do not show the same stabilizing results that are found for the Student-t distribution. The choices of k^* are the expected level relative to one another until approximately $T = 600$. For larger values of T, the level

of the estimates start to shift relative to one another. The Symmetric Stable distribution is relatively difficult to analyse. This is due to several reasons.

Figure 14 in the Appendix shows a hump shape for the Hill plot of the Symmetric Stable distribution. Sun and De Vries (2014) 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 first 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 19. The increase in k^* after a flat line between 500 and 1,000 is an indication that the before suggested effect is kicking in.

The results for the Fréchet distribution are depicted in Figure 20. The value of k^* in the plots is an increasing function of T. This is explained by the relatively large value of k^*. For a total sample size of 10,000, the mse optimal threshold is approximately at 10%. This implies that T needs to be large to reach a stable region. The pattern of the lines being close together is as expected. For the Fréchet distribution, the k^* is independent of α. Therefore, we do not expect to see a particular pattern between the estimates of the different members of the Fréchet distribution. Additionally, from Figure 15 we see that the bias is relatively small for the Fréchet distribution. This makes the choice of k^* less important in contrast with the other distributions.

In Monte Carlo simulation studies none of the metrics used attain the optimal level of k^*. Based on the other desirable attributes described in Section 3.5, the KS distance metric outperforms the other metrics. The simulation results show that as the tail becomes heavier the number of observations chosen for the Hill estimator increases for the Student-t and Symmetric Stable distribution. One of the concerns where the other metrics fail considerably is the robustness of their results. Ideally, 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.

4 Monte Carlo: The big horse race

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 Section 2.2.1. Additionally, we use the automated Eye-Ball method, fixed sample proportion and $Q_{5,n}$ in (12) in the MC horse race. The k^* as derived by Hall and Welsh (1985) is not useful for real world applications, but it can function as a useful benchmark in a MC study. 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^* described in Section 3.5. 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 dependent on k^*.

4.1 Distributions and processes for MC horse race

To do the Monte Carlo horse race properly, we have chosen a wide range of heavy tailed distributions and processes. One prerequisite is that the tail index for these distributions is known. Although this restriction is not necessary, it allows the analysis to be more robust and informative. Given this limitation, the distributions vary in their tail index, in addition to their rate of convergence as $n \to \infty$, and their bias and variance trade-offs.

As before we use three distribution families to draw i.i.d. samples from: The Fréchet, Symmetric Stable and the Student-t distribution. We also employ dependent time series. The ARCH and the GARCH models by Engle (1982) and Bollerslev (1986), respectively, model the volatility clustering in financial data. Therefore, the simulations also use non-stationary times series in the form of ARCH volatility models to evaluate the performance of the methods under the clustering of extremes. For the ARCH process we do not know the variance and the bias of the Hill estimator. However, due to the Kesten Theorem, we are able to derive the tail index.[17]

[17] See Appendix A.4.

4.2 Results of the MC horse race for the tail index

Table 1 presents the results from the Monte Carlo 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. This problem is most prominent for the Double Bootstrap method and the iterative method of Drees and Kaufmann (1998). The KS ratio metric introduces a large bias for the more heavy tailed distributions of the Student-t family. The KS distance metric, Theoretical threshold and the automated Eye-Ball method give estimates that are closest to the true value of the tail index.[18] 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.

The simulation results for the Symmetric Stable distribution do not point towards a method that is clearly superior. With the exception of the KS ratio metric, for $\alpha = 1.1$ and $\alpha = 1.3$, the other methods perform better in terms of the mean estimate than the KS distance metric. The bias of the KS distance metric is 0.11 and 0.09, respectively. For $\alpha = 1.5$ and $\alpha = 1.7$, the bias is around 0.08 for the KS distance metric. The performance of the other methods starts to worsen at these levels of α. The estimates of the other methods are upwards biased. This bias is even more severe for $\alpha = 1.9$. 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. At $\alpha = 2$ the tail of the Symmetric Stable distribution becomes the thin tailed normal distribution and is therefore outside of the domain of where the Hill estimator applies.

The bias of the Hill estimator for the Fréchet distribution is relatively small compared to the Student-t and Symmetric Stable distribution.[19] Therefore, all of the methods perform relatively well except for the KS ratio metric. 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. As the bias in the Fréchet distribution is small, the bias due to the KS distance metric is still limited in absolute terms. Due to the small bias

[18] The k^* chosen by the results of Hall and Welsh (1985) does not have any empirical application. The true data generating process needs to be known in order to determine k^*, but as a benchmark the comparison can be insightful.

[19] See Figure 15 in the Appendix.

Table 1: Estimates of α under different methods for the four families of processes

	α	KS dif	KS rat	TH	5%	Eye-Ball	Drees	Du Bo
	2	2.01	3.33	1.92	1.85	1.98	1.70	1.71
	3	2.85	3.56	2.79	2.45	2.83	2.24	2.20
Student-t	4	3.53	3.92	3.58	2.87	3.48	2.64	2.52
	5	4.10	4.37	4.32	3.16	3.96	2.92	2.75
	6	4.49	4.71	4.96	3.38	4.29	3.14	2.92
	1.1	1.21	4.33	1.11	1.11	1.10	1.07	1.09
	1.3	1.39	3.30	1.33	1.37	1.32	1.33	1.36
Stable	1.5	1.58	3.74	1.57	1.72	1.54	1.68	1.71
	1.7	1.78	3.63	1.84	2.32	1.84	2.18	2.19
	1.9	2.31	3.63	2.55	3.55	3.36	3.13	2.90
	2	2.01	3.63	1.99	1.98	2.00	1.92	1.93
	3	2.93	3.89	3.01	2.97	3.00	2.88	2.90
Fréchet	4	3.79	4.25	4.05	3.96	3.99	3.85	3.87
	5	4.71	5.16	5.09	4.95	4.99	4.81	4.84
	6	5.63	5.82	6.14	5.94	5.98	5.77	5.81
	2.30	2.59	15.15		2.13	2.34	1.93	1.88
	2.68	2.87	3.72		2.39	2.66	2.16	2.05
ARCH	3.17	3.22	3.95		2.69	3.04	2.42	2.22
	3.82	3.66	4.49		3.02	3.50	2.71	2.39
	4.73	4.18	4.50		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, Frechet distribution and ARCH process. The different methods are stated in the first row. KS dif is the Kolmogorov-Smirnov distance metric in (10). The KS rat is the Kolmogorov-Smirnov distance in (12). TH is based on the theoretically derived optimal k from minimizing the mse for specific parametric distributions, presented in Equation (17) in the Appendix. The automated Eye-Ball method in (6) 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.

for the Fréchet distribution, the choice of method for k^* is less important.

The Hill plot of the ARCH process is similar to that of the Student-t dis-

tribution. Therefore, we expect that the methods which performed well in the Student-t simulation to also perform well for the ARCH process. Table 1 shows that for the ARCH simulations 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 $\alpha = 3.172$ and larger values of α, the KS distance metric shows a smaller bias. The other methods show a substantial bias over the whole range of α. To conclude, the automated Eye-Ball method and KS distance metric are the preferred methods since these perform best across a wide range of α values.

In Table 5 of the Appendix the patterns in k^* for the various distributions give a mixed picture. The KS distance metric, as previously discussed, shows the patterns derived by Hall and Welsh (1985). The automated Eye-Ball method offers a more confusing picture. 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.

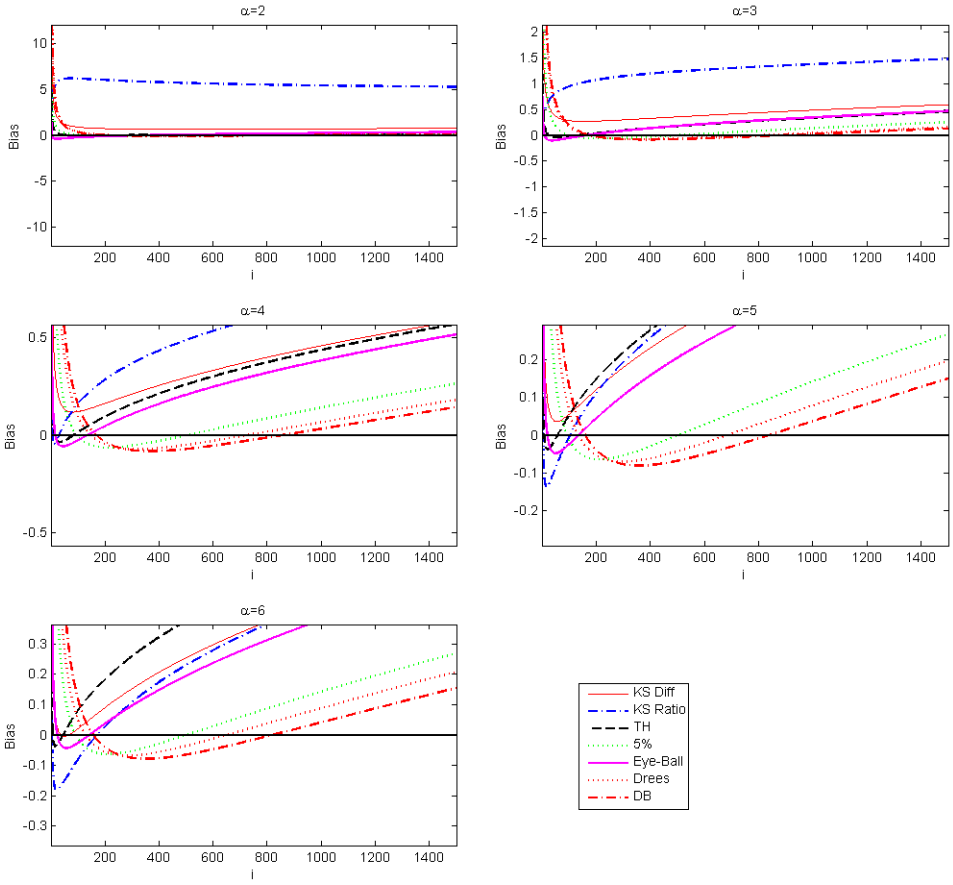
The Double Bootstrap method shows the right patterns in the choice of k^*, but the levels for the Student-t and Symmetric Stable distribution are far higher than desired. In part, this is due to the fact that for the Double Bootstrap method the practical criterion is based on asymptotic arguments. This means that the asymptotic results might not hold in finite samples. 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.

4.3 Simulation results for the quantiles

We also included an analysis on how the different metrics perform in estimating the quantiles of the distribution. For many of the economic questions this is more relevant than the precise value of the tail index. Figure 5 depicts the bias of the quantile estimator for the different methodologies.

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 bias in the 99% to 100% quantile region. The flip side is that these methods have a small bias for the quantiles further towards the center of the distribution compared to the other competing methodologies. With 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

Figure 5: Quantile estimation bias different methodologies (Student-t distribution)



This Figure show the bias induced by using the quantile estimator presented in Equation (9). 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.

and the theoretical mse produce a smaller bias in the tail region. These three methods exhibit a larger bias towards the center of the distribution. Given that one intends to model the quantiles deep in the tail of the distribution, the KS distance metric and the automated Eye-Ball method are the preferred methods.

Figure 21 depicts the results for the Symmetric Stable distribution. The KS distance metric introduces a large bias over the entire region. Comparing these results with Table 1 suggests that the bias in the quantiles stems from the estimated scale parameter A. The methods by Drees and Kaufmann (1998) and Danielsson et al. (2001) perform consistently well over the region for $\alpha \leq 1.5$. As α increases, the ranking of the different methodologies alters and no clear choice of methodology arises.

The results for the Fréchet distribution are presented in Figure 22 in the Appendix. The 5% fixed sample fraction produces a consistently small bias. The theoretical mse and automated Eye-Ball method produce a relatively small bias for the quantiles larger than 99.5% region. Their bias increases relative to other methodologies for quantiles that are closer towards the center of the distribution.

The bias might be subject to asymmetric outliers in the error distribution. Therefore, we also produce results for the median difference to obtain an outlier robust centrality measure. The outcomes for the Student-t distribution do not change dramatically. For the quantiles towards the center of the distribution, the KS distance metric improves its performance compared to the automated Eye-Ball and theoretical mse criteria. The median error for the Symmetric Stable distribution also produces a different picture for the performance of KS distance metric. The KS distance metric performs relatively better compared to the analysis based on the mean estimation error. This is throughout the whole region of the tail. The same results emerge for the samples drawn from the Fréchet distribution. From this we infer that the KS distance metric is rather susceptible to outliers.

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 center.

The performance of the methods of Drees and Kaufmann (1998) and Danielsson et al. (2001) in finite samples is inferior to the other methods. This notwithstanding the proofs that asymptotically the methods are consistent.

Picking a fixed percentage of observations, such as 5%, ignores the information which 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.[20]

5 Application: Financial return series

We now take the KS distance metric to the real data. We estimate the tail index for the return data on individual U.S. stocks. The various methods that are used in the horse race are used to estimate the tail exponent for returns on U.S. stocks.

5.1 Data

The stock market data is obtained from the Center for Research in Security Prices (CRSP). The CRSP database contains individual stock data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. In total 17,918 stocks are used. For every stock included in the analysis, it needs to trade on one of the four exchanges during the whole measurement period.[21] Only stocks with more than 48 months of data are used for estimation. 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.

5.2 Empirical impact

The average absolute differences between the estimates of the tail exponent are compared to one another in the tables below.[22]

[20]For additional simulation results the reader can consult the Tables and Figures from Monte Carlo simulations in the online Appendix.

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

[22]For the descriptive statistics on the tail estimates of the left and right tail for the different methods consult Tables 6 and 7 in the Appendix.

Table 2: The mean absolute difference of the estimates of α in the **left** tail for the 6 different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	0.71	0.84	0.59	1.11	1.44
KS rat	0.71	0.00	1.45	1.14	1.73	2.05
5%	0.84	1.45	0.00	0.56	0.49	0.74
Eye-Ball	0.59	1.14	0.56	0.00	0.92	1.22
Drees	1.11	1.73	0.49	0.92	0.00	0.49
Du Bo	1.44	2.05	0.74	1.22	0.49	0.00

This table presents the results of applying the six different methods to estimate α for left tail of stock returns. The data is from the CRSP database that contains all individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the mean absolute difference between the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 3: The mean absolute difference of the estimates of α in the **right** tail for the 6 different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	0.78	0.75	0.54	0.88	1.22
KS rat	0.78	0.00	1.44	1.15	1.57	1.90
5%	0.75	1.44	0.00	0.55	0.39	0.63
Eye-Ball	0.54	1.15	0.55	0.00	0.77	1.12
Drees	0.88	1.57	0.39	0.77	0.00	0.50
Du Bo	1.22	1.90	0.63	1.12	0.50	0.00

This table presents the results of applying the six different methods to estimate α for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the mean absolute difference between the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Tables 2 and 3 present the results of the estimation of the tail exponents for the left and right tail. The thresholds are estimated with the six methodologies from the horse race.[23] The absolute difference between the different methods is quite substantial for both the left and right tail.[24] Note from the results that the methods can be divided into two groups. These groups have relatively similar estimates. 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). 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 utilize a low fraction of the total sample for $\hat{\alpha}$ in the simulations. The same holds for the return data. Tables 12 and 13 in the Appendix show that the Double Bootstrap procedure has the smallest average difference in choosing the k^* with the method by Drees and Kaufmann (1998). This is in line with the Monte Carlo horse race where both methods picked similar optimal sample fractions. The results for the right tail of the empirical distribution are somewhat proportionally smaller, but the same relative differences are preserved between methodologies.

Comparing the results between the horse race and the 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) also generate estimates of α which are close to one another for the financial data. In 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.

6 Conclusion

In this paper we propose a new approach to choose the optimal number of order statistics for the Hill estimator. We employ the Kolmogorov-Smirnov distance over the quantile dimension to fit the Pareto quantile estimator to

[23]The theoretical threshold is not applicable for these applications. Only with strong parametric assumptions is the theoretical threshold applicable.

[24]The average absolute difference can easily be dominated by large estimates of α. In Tables 8 and 9 we also report the results for the median difference. The results for the median are similar to these results, but smaller in size.

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 utilized for the estimation. By fitting the tail of the distribution we try to find the optimal sample fraction for the Hill estimator.

Validating the performance of the methodology we model the KS distance metric in terms of a Brownian motion. Modelling the KS distance metric as a Brownian motion allows us to study the properties of the KS distance metric in a more general setting. In addition, we use the properties derived by Hall and Welsh (1985) on the optimal number of order statistics to run further simulation studies. A comparison is drawn between the KS distance metric and other commonly applied penalty functions. Guided by the results by Hall and Welsh (1985), we find that among an array of penalty functions the KS distance metric has the best properties.

The literature to date provides methods from the theoretical statistical literature. These methods are backed by asymptotic arguments. Although these methods are asymptotically consistent, their finite sample properties are often unstable and inaccurate. In the applied literature, heuristic methods are frequently applied as well. This can range from picking a fixed percentage of order statistics to Eye-Ball the Hill plot. These methods are somewhat arbitrary and subjective. To test the performance of the KS distance metric we consider a horse race between the different methodologies. In the horse race we use various parametric heavy tailed distributions and processes to estimate the tail index with the Hill estimator. The KS distance metric and the automated Eye-Ball method outperform the competing methods based on the size of the bias. Both methods come close to the theoretically optimal threshold as the threshold derived by Hall and Welsh (1985). Although the theoretical optimal threshold is unknown in empirical application, it does give confidence that the KS distance metric and the automated Eye-Ball method estimate the tail index properly in empirical applications.

We also estimated quantiles in the simulation studies. The various methodologies have different areas in the tail where they outperform other methods. The KS distance metric and the automated Eye-Ball method have better quantile estimates for the very high probabilities. The Double Bootstrap and the method by Drees and Kaufmann (1998) preform relatively better for the quantiles towards the center of the distribution. This can be explained by the high values of k^* for these methods. A high value for k^* normally services a better fit towards the center observations. This is a possible ex-

planation for the outcomes of the simulations for the Double Bootstrap and method by Drees and Kaufmann (1998).

To show that the choice of the proper number of order statistics matters, we estimate the tail index for various securities. We apply the different methodologies to the universe of daily CRSP US stock returns. We estimate α for the tail of the empirical distribution and compare the absolute differences between the estimates of the different methods for the individual stocks. We do this both for the left and right tail. The methods can be divided into two groups. We find that the methods that perform well in the simulation studies have estimates close to one another in financial applications. The methods that ranked low in the Monte Carlo simulations also have estimates which are highly correlated. The variation in estimates holds for both sides of the tail. This gives the impression that the KS distance metric and the automated Eye-Ball method should be preferred for empirical applications.

The automated Eye-Ball method has the advantage over the theoretical methods that it locates in a direct way the trade-off between the bias and the variance of the estimator. The region where the volatility firstly subsides is directly observed from the Hill plot. The maximum absolute distance in the KS distance metric focuses on minimizing the largest deviations. For the horizontal dimension, these naturally occur deep in the tail of the distribution. This is also the region for which EVT is intended. The conclusion therefore is that the two empirically driven methods are best suited for the estimation of $\alpha\left(k^{*}\right)$ and the quantiles deep in the tail.

This has important implications for the risk management of investors which for instance hold these stocks. For example, Value-at-Risk estimates heavily depend on the thickness of the left tail. The Value-at-Risk determines the allowable risk taking in many financial institutions via internal risk management or regulations. This makes the choice of methodology economically important.

References

Balkema, A. A., L. De Haan, 1974. Residual life time at great age. The Annals of Probability 2(5), 792–804.

Bickel, P. J., A. Sakov, 2008. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Statistica Sinica 18(3), 967–985.

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

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3), 307–327.

Csorgo, S., P. Deheuvels, D. Mason, 1985. Kernel estimates of the tail index of a distribution. The Annals of Statistics 13(3), 1050–1077.

Danielsson, J., L. Peng, C. De Vries, L. De Haan, 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis 76(2), 226–248.

Davis, R., S. Resnick, 1984. Tail estimates motivated by extreme value theory. The Annals of Statistics 12(4), 1467–1487.

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

De Haan, L., S. I. Resnick, 1980. A simple asymptotic estimate for the index of a Stable distribution. Journal of the Royal Statistical Society. Series B (Methodological) 42(1), 83–87.

Dietrich, D., L. De Haan, J. Hüsler, 2002. Testing extreme value conditions. Extremes 5(1), 71–85.

Drees, H., E. Kaufmann, 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and their Applications 75(2), 149–172.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4), 981–1008.

Fisher, R. A., L. H. C. Tippett, 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. 24(02), 180–190.

Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of Mathematics 44(1), 423–453.

Hall, P., 1982. On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society. Series B (Methodological) 44(1), 37–42.

——— , 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of Multivariate Analysis 32(2), 177–203.

Hall, P., A. Welsh, 1985. Adaptive estimates of parameters of regular variation. The Annals of Statistics 13(1), 331–341.

Hill, B. M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3(5), 1163–1174.

Kelly, B., H. Jiang, 2014. Tail risk and asset prices. Review of Financial Studies 27(10), 2841–2871.

Mandelbrot, B. B., 1963. New methods in statistical economics. Journal of Political Economy 71(1), 421–440.

Mason, D. M., 1982. Laws of large numbers for sums of extreme values. The Annals of Probability 10(3), 754–764.

Pickands, J., 1975. Statistical inference using extreme order statistics. the Annals of Statistics 3(1), 119–131.

Reed, W. J., 2001. The Pareto, Zipf and other power laws. Economics Letters 74(1), 15–19.

Resnick, S., C. Starica, 1997. Smoothing the Hill estimator. Advances in Applied Probability 29(1), 271–293.

Sun, P., C. De Vries, 2014. Hill bias under Stable and Student-t alternatives. Unpublished working paper. Tinbergen discussion paper.

A Appendix

A.1 The bias of the Hill estimator

$$\mathrm{E}\left[\frac{1}{\widehat{\alpha}}\right] = \mathrm{E}\left[\frac{1}{k}\sum_{i=0}^{k-1}\left(\log\left(X_{n-i,n}\right) - \log\left(X_{n-k,n}\right)\right)\right] \overset{ind.}{=}$$

$$\mathrm{E}\left[\frac{1}{k}\left(k\log\left(X_{n-i,n}\right) - k\log\left(X_{n-k,n}\right)\right)\right] = \mathrm{E}\left[\log\left(\frac{X_{n-i,n}}{X_{n-k,n}}\right)\right].$$

The expectation is conditional on that $X_{n-k,n} = s$ is in the tail of the distribution, i.e.

$$\mathrm{E}\left[\log\left(\frac{X_{n-i,n}}{s}\right) \mid X_{n-i,n} > s\right] = \frac{\int_s^\infty \log\left(\frac{X_{n-i,n}}{s}\right) f\left(X_{n-i,n}\right) dX_{n-i,n}}{1 - F\left(s\right)}.$$

The numerator is a gamma 2 function, given that first order expansion is taken for the cdf,

$$\alpha A\int_s^\infty \log\left(\frac{X_{n-i,n}}{s}\right)\left(X_{n-i,n}\right)^{-\alpha-1} dX_{n-i,n} \overset{X_{n-i,n}/s=y}{=}$$

$$\alpha A\left(s\right)^{-\alpha}\int_1^\infty \log\left(y\right)\left(y\right)^{-\alpha-1} dy \overset{\log(y)=t}{=}$$

$$A\alpha s^{-\alpha}\int_0^\infty t\left(e^t\right)^{-\alpha-1} e^t dt \overset{\alpha t=x}{=} A\alpha^{-1}s^{-\alpha}\int_o^\infty xe^{-x}dx = \frac{s^{-\alpha}\Gamma\left(2\right)}{\alpha}.$$

Given that we take a second order expansion for the tail of the density function of the numerator

$$\mathrm{E}\left[\log\left(\frac{X_{n-i,n}}{s}\right) \mid X_{n-i,n} > s\right] = \frac{\int_s^\infty \log\left(\frac{X_{n-i,n}}{s}\right) f\left(X_{n-i,n}\right) dX_{n-i,n}}{1 - \mathrm{F}\left(s\right)}$$

$$= \frac{\int_s^\infty \log\left(\frac{X_{n-i,n}}{s}\right) - \alpha A\left(X_{n-i,n}\right)^{-\alpha-1} - \left(\alpha+\beta\right)AB\left(X_{n-i,n}\right)^{-(\alpha+\beta)-1} dX_{n-i,n}}{1 - \mathrm{F}\left(s\right)} =$$

$$\frac{\frac{A\Gamma(2)}{\alpha}s^{-\alpha} + \frac{BA\Gamma(2)s^{-(\alpha+\beta)}}{\alpha+\beta}}{As^{-\alpha}\left(1 + Bs^{-\beta}\right)} = \frac{\frac{1}{\alpha} + \frac{Bs^{-\beta}}{\alpha+\beta}}{\left(1 + Bs^{-\beta}\right)}.$$

The last equation is obtained from applying the Taylor expansion to the denominator. The bias is thus

$$\mathrm{E}\left[\frac{1}{\widehat{\alpha}} - \frac{1}{\alpha} \mid X_{n-i,n} > X_{n-k,n}\right] = \frac{\int_s^\infty \left(\log\left(\frac{X_{n-i,n}}{s}\right) - \frac{1}{\alpha}\right) f\left(X_{n-i,n}\right) dX_{n-i,n}}{1 - \mathrm{F}\left(s\right)}$$

$$= \frac{\frac{1}{\alpha} + \frac{Bs^{-\beta}}{\alpha+\beta} - \frac{1}{\alpha}\left(1+Bs^{-\beta}\right)}{\left(1+Bs^{-\beta}\right)} = \frac{\frac{(\alpha+\beta)}{\alpha(\alpha+\beta)} + \frac{\alpha Bs^{-\beta}}{\alpha(\alpha+\beta)} - \frac{(\alpha+\beta)}{\alpha(\alpha+\beta)}\left(1+Bs^{-\beta}\right)}{\left(1+Bs^{-\beta}\right)}$$

$$= \frac{\frac{\alpha Bs^{-\beta} - (\alpha+\beta)Bs^{-\beta}}{\alpha(\alpha+\beta)}}{\left(1+Bs^{-\beta}\right)} = \frac{\frac{-\beta Bs^{-\beta}}{\alpha(\alpha+\beta)}}{\left(1+Bs^{-\beta}\right)} = \frac{-\beta Bs^{-\beta}}{\alpha\left(\alpha+\beta\right)} + o\left(s^{-\beta}\right).$$

A.2 Variance of the Hill estimator

For the variance of the Hill estimator the variance is written out in the form of expectations

$$\mathrm{var}_\tau\left(\frac{1}{\widehat{\alpha}}\right) = \mathrm{E}_\tau\left[\left(\frac{1}{\widehat{\alpha}}\right)^2\right] - \left[\mathrm{E}_\tau\left(\frac{1}{\widehat{\alpha}}\right)\right]^2 =$$

$$\mathrm{E}_\tau\left[\left(\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{X_{n-i,n}}{s}\right)\right)^2 \mid X_{n-i,n} > s\right] - \mathrm{E}_\tau\left[\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{X_{n-i,n}}{s}\right) \mid X_{n-i,n} > s\right]^2.$$

The first term is worked out first,

$$\mathrm{E}_\tau\left[\left(\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{X_{n-i,n}}{s}\right)\right)^2 \mid X_{n-i,n} > s\right] =$$

$$\frac{1}{\mathrm{P}\left[M \geq 1\right]}\sum_{k=1}^{n}\mathrm{P}\left[M=k\right]\mathrm{E}\left[\left(\frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{X_{n-i,n}}{s}\right)\right)^2 \mid X_{n-i,n} > s\right] =$$

$$\frac{1}{\mathrm{P}\left[M \geq 1\right]}\sum_{k=1}^{n}\left\{\frac{\mathrm{P}\left[M=k\right]}{k^2}k\,\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right)^2\right] + k\left(k-1\right)\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right) \mid X_{n-i,n} > s\right]^2\right\}.$$

Substituting in the variance expression

$$\frac{1}{\mathrm{P}\left[M \geq 1\right]}\sum_{k=1}^{n}\frac{\mathrm{P}\left[M=k\right]}{k^2}k\,\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right)^2 \mid X_{n-i,n} > s\right] +$$

$$\sum_{k=1}^{n}k\left(k-1\right)\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right) \mid X_{n-i,n} > s\right]^2 - k\,\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right) \mid X_{n-i,n} > s\right]^2 =$$

$$\left(\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right)^2 \mid X_{n-i,n} > s\right] - \mathrm{E}\left[\log\left(\frac{X_{(1)}}{s}\right) \mid X_{n-i,n} > s\right]^2\right)$$
$$\sum_{k=1}^{n}\left\{\frac{\mathrm{P}\left[M=k\right]}{k}\right\}\frac{1}{P\left[M \geq 1\right]}.$$

All parts are known except for

$$\sum_{k=1}^{n}\left\{\frac{\mathrm{P}\left[M=k\right]}{k}\right\}\frac{1}{\mathrm{P}\left[M \geq 1\right]} = \mathrm{E}_{\tau}\left[\frac{1}{M}\right].$$

From the binomial distribution it is known that

$$\mathrm{E}\left(M\right) = n\,\mathrm{P} = n\left(1-\mathrm{F}\left(s\right)\right) = nAs^{-\alpha}\left[1+Bs^{-\beta}\right].$$

The conditional expectation of M is,

$$\mathrm{E}\left[M \mid M \geq 1\right] = \frac{\sum_{k=1}^{n} k\,\mathrm{P}\left[M=k\right]}{\mathrm{P}\left[M \geq 1\right]} = \frac{\sum_{k=1}^{n} k\,\mathrm{P}\left[M=k\right] + 0\cdot\mathrm{P}\left[M=k\right]}{1-\mathrm{F}\left(s\right)^n} =$$

$$\frac{\mathrm{E}\left[M\right]}{1-\mathrm{F}\left(s\right)^n} = \mathrm{E}\left[M\right] + \frac{\mathrm{F}\left(s\right)^n}{1-\mathrm{F}\left(s\right)^n}\,\mathrm{E}\left[M\right].$$

As n becomes very large the second term goes to zero and $\mathrm{E}\left[M\right] = nAs^{-\alpha} + o\left(ns^{-\alpha}\right)$. For $\mathrm{E}\left[\frac{1}{M} \mid M \geq 1\right]$ we first take a second order Taylor expansion of $\frac{1}{n}$ around $\mathrm{E}\left[M\right]$,

$$\mathrm{E}\left[\frac{1}{M} \mid M \geq 1\right] \approx \frac{1}{\mathrm{E}\left[M\right]} - \frac{1}{\mathrm{E}\left[M\right]^2}(n - \mathrm{E}\left[M\right]) + ... =$$

$$\frac{1}{\mathrm{E}\left[M\right]} - O\left(\frac{1}{\mathrm{E}\left[M\right]^2}\right) = \frac{1}{np} - O\left(\frac{1}{(np)^2}\right).$$

Taking the Pareto approximation for the probability,

$$\mathrm{E}\left[\frac{1}{M} \mid M \geq 1\right] \approx \frac{s^{\alpha}}{nA}\frac{1}{1+Bs^{-\beta}+o\left(s^{-\beta}\right)} - O\left(\frac{s^{2\alpha}}{n^2A^2}\right) =$$

$$\frac{s^{\alpha}}{nA} - \frac{s^{\alpha}}{nA}\frac{Bs^{-\beta}+o\left(s^{-\beta}\right)}{1+Bs^{-\beta}+o\left(s^{-\beta}\right)} - O\left(\frac{s^{2\alpha}}{n^2A^2}\right) =$$

$$\frac{s^{\alpha}}{nA} - \left[O\left(\frac{s^{\alpha-\beta}}{n}\right) + O\left(\frac{s^{2\alpha}}{n^2A^2}\right)\right] =$$

$$\frac{s^{\alpha}}{nA} - \left[O\left(\frac{s^{\alpha-\beta}}{n} \right) + O\left(\frac{s^{2\alpha}}{n^2 A^2} \right) \right] = \frac{s^{\alpha}}{nA} - o\left(\frac{s^{\alpha}}{n} \right).$$

With the results that are found

$$\mathrm{E}\left[\frac{1}{M} \mid M \geq 1 \right] = \frac{s^{\alpha}}{nA} - o\left(\frac{s^{\alpha}}{n} \right)$$

$$\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s} \right)^2 \mid X_{n-i,n} > s \right] =$$

$$\frac{\Gamma(3)}{1 + Bs^{-\beta}} \left[\frac{1}{\alpha^2} + \frac{Bs^{-\beta}}{(\alpha+\beta)^2} \right] = 2! \left[\frac{1}{\alpha^2} + \frac{Bs^{-\beta}}{(\alpha+\beta)^2} \right] + o\left(s^{-\beta} \right).$$

Taking all the individual results together,

$$\mathrm{E}\left[\log\left(\frac{X_{(1)}}{s} \right) \mid X_{n-i,n} > s \right]^2 = \left[\frac{1}{\alpha^2} + \frac{Bs^{-\beta}}{\alpha+\beta} \right]^2$$

$$\mathrm{var}_{\tau}\left(\frac{1}{\widehat{\alpha}} \right) = \sum_{k=1}^{n} \left\{ \frac{\mathrm{P}[M=k]}{k} \right\} \frac{1}{\mathrm{P}[M \geq 1]} \mathrm{E}\left[\log\left(\frac{X_{(1)}}{s} \right)^2 \mid X_{n-i,n} > s \right] -$$

$$\sum_{k=1}^{n} \left\{ \frac{\mathrm{P}[M=k]}{k} \right\} \frac{1}{\mathrm{P}[M \geq 1]} \mathrm{E}\left[\log\left(\frac{X_{(1)}}{s} \right) \mid X_{n-i,n} > s \right]^2 =$$

$$\frac{s^{\alpha}}{nA} \left(\left[\frac{2}{\alpha^2} + \frac{2Bs^{-\beta}}{(\alpha+\beta)^2} \right] - \left[\frac{1}{\alpha} + \frac{Bs^{-\beta}}{\alpha+\beta} \right]^2 \right) + o\left(\frac{s^{\alpha}}{n} \right) =$$

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

Giving the variance of the Hill estimator,

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

A.3 Optimal theoretical threshold

From the variance and the bias the $\mathrm{mse} = \mathrm{var} + (\mathrm{bias})^2$ is

$$\mathrm{mse} = \frac{s^{\alpha}}{nA} \frac{1}{\alpha^2} + \left(\frac{\beta Bs^{-\beta}}{\alpha(\alpha+\beta)} \right)^2 + o\left(\frac{s^{\alpha}}{n} \right) + o\left(s^{-2\beta} \right). \tag{13}$$

For the amse the small terms go to 0,

$$\text{amse} = \frac{s^{\alpha}}{nA}\frac{1}{\alpha^2} + \left(\frac{\beta B s^{-\beta}}{\alpha\left(\alpha+\beta\right)}\right)^2 \tag{14}$$

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

$$s^* = \left[\frac{2AB^2\beta^3\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{\frac{1}{\alpha+2\beta}} n^{\frac{1}{\alpha+2\beta}} \tag{15}$$

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

Substituting s^* back into the mse gives

$$\begin{aligned}
&\tfrac{1}{\alpha^2}\left(\tfrac{s^{\alpha}}{nA} + \left(\tfrac{\beta B s^{-\beta}}{(\alpha+\beta)}\right)^2\right) + o\left(\tfrac{s^{\alpha}}{n}\right) + o\left(s^{-2\beta}\right) = \\
&\tfrac{1}{\alpha^2}\left(\frac{\left[\frac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{\frac{\alpha}{\alpha+2\beta}}}{nA} + \frac{(\beta B)^2\left[\frac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}} n^{\frac{-2\beta}{\alpha+2\beta}}}{(\alpha+\beta)^2}\right) \\
&+o\left(-\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{\frac{-2\beta}{\alpha+2\beta}}\right) + o\left(\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{n(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right) \\
&= \tfrac{1}{\alpha^2}\left(\left(\left[\tfrac{2B^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} + \frac{(\beta B)^2\left[\frac{2B^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}}{(\alpha+\beta)^2}\right)(nA)^{\frac{-2\beta}{\alpha+2\beta}}\right) \\
&+o\left(-\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{\frac{-2\beta}{\alpha+2\beta}}\right) + o\left(\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{n(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right) \\
&= \tfrac{1}{\alpha^2}\left[(\alpha+\beta)^2\right]^{\frac{-\alpha}{\alpha+2\beta}}(nA)^{\frac{-2\beta}{\alpha+2\beta}}\left(\left[2B^2\beta^3\alpha^{-1}\right]^{\frac{\alpha}{\alpha+2\beta}} + (\beta B)^2\left[2B^2\beta^3\alpha^{-1}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right) \\
&+o\left(-\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{\frac{-2\beta}{\alpha+2\beta}}\right) + o\left(\left[\tfrac{2AB^2\beta^3\alpha^{-1}}{n(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right)
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\alpha^2}\left[\frac{2\beta^3B^2\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{\frac{\alpha}{\alpha+2\beta}}(nA)^{\frac{-2\beta}{\alpha+2\beta}}\left(1+\frac{\alpha}{2\beta}\right) + \\
&+o\left(-\left[\frac{2AB^2\beta^3\alpha^{-1}}{\left(\alpha+\beta\right)^2}\right]^{\frac{\alpha}{\alpha+2\beta}} n^{\frac{-2\beta}{\alpha+2\beta}}\right) + o\left(\left[\frac{2AB^2\beta^3\alpha^{-1}}{n\left(\alpha+\beta\right)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right) =
\end{aligned}$$

$$\frac{1}{A\alpha}\left(\frac{1}{\alpha}+\frac{1}{2\beta}\right)\left[\frac{2\beta^3B^2\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}}(nA)^{\frac{-2\beta}{\alpha+2\beta}}+$$
$$o\left(-\left[\frac{2AB^2\beta^3\alpha^{-1}}{(\alpha+\beta)^2}\right]^{\frac{\alpha}{\alpha+2\beta}}n^{\frac{-2\beta}{\alpha+2\beta}}\right)+o\left(\left[\frac{2AB^2\beta^3\alpha^{-1}}{n(\alpha+\beta)^2}\right]^{\frac{-2\beta}{\alpha+2\beta}}\right)$$

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

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

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

with probability 1.

A.4 Kesten Theorem

Given stochastic difference equation (SDE)

$$Y_t = A_t Y_{t-1} + B_t,$$

where (A_t, B_t) are i.i.d. random vectors and they are not necessarily independent from each other.

Kesten Theorem:

Suppose $\mathrm{E}\left(A_1^k\right) = 1, \mathrm{E}\, A_1^k \left(\log A_1\right)_+ < \infty$ and $0 < \mathrm{E}\left(B_1^k\right) < \infty$ for some $k > 0$. Suppose $B_1/\left(1 - A_1\right)$ is non-degenerate and $\log A_1 | A_1 \neq 0$ is non-lattice. Then the stationary solution of the SDE must be heavy tailed with tail index k.

Given:

$$X_t = z_t \sigma_t$$

$$\sigma_t^2 = \gamma + \beta X_{t-1}^2,$$

where $z_t \sim N\left(0, 1\right)$.

For ARCH process we have

$$X_t^2 = z_t^2 \left(\gamma + \beta X_{t-1}^2\right),$$

or

$$X_t^2 = z_t^2 \gamma + \beta z_t^2 X_{t-1}^2.$$

So that we may write

$$X_t^2 = B_t + A_t X_{t-1}^2,$$

where $A_t = \beta z_t^2, B_t = z_t^2 \gamma$.

We can see that conditions $\mathrm{E}\, A_1^k \left(\log A_1\right)_+ < \infty$ and $0 < \mathrm{E}\left(B_1^k\right) < \infty$ are met for the ARCH process with normally distributed shocks.

Given that $z_t \sim N\left(0, 1\right)$ and the central moments of the normal distribution are

$$\mathrm{E}\left[z^p\right] = \sigma^p \frac{2^{p/2}\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}},$$

the expectation works out to be

$$1 = \mathrm{E}\left(\beta^k z_t^{2k}\right)$$

$$1 = \beta^k \mathrm{E}\left(z_t^{2k}\right) = \beta^k \sigma^{2k} \frac{2^k \Gamma\left(k+\frac{1}{2}\right)}{\sqrt{\pi}}$$

$$1 = \frac{(\beta 2)^k}{\sqrt{\pi}} \Gamma\left(k+\frac{1}{2}\right).$$

Here β is the AR component of the ARCH process. Solving this equation gives the tail exponent of X_t^2. Assuming that X is Pareto distributed,

$$\mathrm{P}\left(X^2 > x\right) = \mathrm{P}\left(X > x^{0.5}\right) = Ax^{-\frac{\alpha}{2}}.$$

Making the tail index of an ARCH random variable $2k$.

A.5 Limit function KS distance metric

The distance to be minimized is,

$$x_{n-i,n} - \left(\frac{i}{k}\right)^{-\gamma} x_{n-k,n}.$$

Using from page 52 Theorem 2.4.8 in De Haan and Ferreira (2007) we get, where we write $s = {}^{i}\!/_{k}$,

$$x_{n-i,n} = U\left(\frac{n}{k}\right)\left[\left(\frac{i}{k}\right)^{-\gamma} + \frac{\gamma}{\sqrt{k}}\left(\frac{i}{k}\right)^{-\gamma-1} w\left(\frac{i}{k}\right) - A_0\left(\frac{n}{k}\right)\left(\frac{i}{k}\right)^{-\gamma} \frac{\left(\frac{i}{k}\right)^{-\rho}-1}{\rho}\right],$$

where $0 < k < T < n$ and $0 < i < T$. Here i, k, T and n are positive integers. Furthermore, $\rho \leq 0$, $\gamma > 0$, and n is the sample size. The T are the number of order statistics over which the distance is minimized (tail region). The $w(\cdot)$ is Brownian motion. Here $A_0(\cdot)$ is a suitable function. As an example we have the following function for the case that the Hall expansion applies

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

Here $\beta \geq 0, B < 0,\ A > 0$ and

$$U\left(i/k\right) = \left(\frac{1}{1-\mathrm{F}\left(i/k\right)}\right)^{\leftarrow},$$

and where $(\cdot)^{\leftarrow}$ denotes the inverse. The right hand part of the distance can be modelled as,

$$\left(\frac{i}{k}\right)^{-\gamma} x_{n-k,n} = U\left(\frac{n}{k}\right)\left(\frac{i}{k}\right)^{-\gamma}\left[1+\frac{\gamma}{\sqrt{k}}w\left(1\right)\right].$$

Putting the two parts together gives

$$\begin{aligned} & x_{n-i,n} - \left(\frac{i}{k}\right)^{-\gamma} x_{n-k,n} \\ & \quad = U\left(\frac{n}{k}\right)\left(\frac{i}{k}\right)^{-\gamma}\left[\frac{\gamma}{\sqrt{k}}\left(\frac{k}{i}w\left(\frac{i}{k}\right)-w\left(1\right)\right) - A_0\left(\frac{n}{k}\right)\frac{\left(\frac{i}{k}\right)^{-\rho}-1}{\rho}\right] \\ & \quad = \frac{\gamma}{\sqrt{k}}U\left(\frac{n}{k}\right)\left(\frac{i}{k}\right)^{-\gamma}\left[\frac{k}{i}w\left(\frac{i}{k}\right) - w\left(1\right) - A_0\left(\frac{n}{k}\right)\frac{\sqrt{k}}{\gamma}\frac{\left(\frac{i}{k}\right)^{-\rho}-1}{\rho}\right]. \end{aligned}$$

For $i/k = s$ we get

$$x_{n-sk,n} - (s)^{-\gamma} x_{n-k,n} = \frac{\gamma}{\sqrt{k}}U\left(\frac{n}{k}\right)s^{-\gamma}\left[s^{-1}w\left(s\right) - w\left(1\right) - A_0\left(\frac{n}{k}\right)\frac{\sqrt{k}}{\gamma}\frac{s^{-\rho}-1}{\rho}\right].$$

This gives the limit function for the KS distance metric,

$$\begin{aligned} & \underset{0<k<T}{\arg\min}\ \sup_{0<s<\frac{T}{k}} \left|x_{n-sk,n} - (s)^{-\gamma} x_{n-k,n}\right| = \\ & \underset{0<k<T}{\arg\min}\ \sup_{0<s<\frac{T}{k}} \left|\frac{\gamma}{\sqrt{k}}U\left(\frac{n}{k}\right)s^{-\gamma}\left[s^{-1}w\left(s\right) - w\left(1\right) - A_0\left(\frac{n}{k}\right)\frac{\sqrt{k}}{\gamma}\frac{s^{-\rho}-1}{\rho}\right]\right|, \end{aligned}$$

where $T \geq k$.

B Tables and Figures

B.1 Tables

Table 4: 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 5: Estimates of k^* under different methods for the four families of distributions

	α	KS dif	KS rat	TH	5%	Eye-Ball	Drees	Du Bo
	2	509.89	5.32	281.00	500.00	19.67	1036.73	968.18
	3	343.13	9.63	132.00	500.00	35.21	841.59	895.65
Student-t	4	227.99	13.92	78.00	500.00	51.48	754.68	859.72
	5	164.88	17.93	53.00	500.00	69.02	708.41	837.61
	6	140.07	20.25	40.00	500.00	84.55	677.95	823.24
	1.1	240.82	2.43	817.00	500.00	8.00	1481.88	1180.98
	1.3	172.74	2.97	292.00	500.00	10.14	1466.24	1218.68
Stable	1.5	137.18	3.54	146.00	500.00	12.66	1376.89	1214.89
	1.7	200.45	4.34	74.00	500.00	18.88	1176.03	1153.69
	1.9	667.03	4.13	27.00	500.00	108.09	861.59	1061.44
	2	217.71	5.39	928.00	500.00	19.26	1500.70	1305.65
	3	231.47	9.58	928.00	500.00	34.99	1501.00	1304.65
Frechet	4	226.54	14.53	928.00	500.00	51.35	1501.00	1305.28
	5	227.16	19.49	928.00	500.00	67.51	1501.00	1303.90
	6	229.31	25.70	928.00	500.00	84.04	1501.00	1304.10
	2.30	290.39	8.53		500.00	31.32	1131.36	1244.62
	2.68	300.24	10.29		500.00	36.21	1036.93	1244.78
ARCH	3.17	290.97	12.06		500.00	42.90	947.32	1245.28
	3.82	246.72	14.90		500.00	52.81	864.97	1246.05
	4.73	202.79	17.84		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, Frechet distribution and ARCH process. The different methods are stated in the first row. KS dif is the Kolmogorov-Smirnov distance metric in (10). The KS rat is the Kolmogorov-Smirnov distance in (12). TH is based on the theoretically derived optimal k from minimizing the mse for specific parametric distributions, presented in Equation (17) in the Appendix. The automated Eye-Ball method in (6) 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.

Table 6: Different statistics for $\hat{\alpha}$ in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
mean	3.40	3.99	2.70	3.19	2.41	1.98
median	3.35	3.78	2.72	3.19	2.32	2.05
st. dev.	0.81	5.01	0.58	0.65	0.94	0.53
min	0.27	0.49	0.16	0.16	0.48	0.19
max	7.79	427.86	15.04	8.52	53.00	10.09
skewness	0.40	55.10	0.71	0.09	18.34	-0.58
kurtosis	3.07	3835.31	19.53	4.56	717.86	11.60

This table presents the results of applying the six different methods to estimate α for left tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from the 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The six different methods are the KS distance metric, KS ratio method, 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 which are included in the analysis.

Table 7: Different statistics for $\hat{\alpha}$ in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
mean	2.97	3.58	2.37	2.87	2.22	1.77
median	2.90	3.40	2.39	2.87	2.13	1.83
st. dev.	0.81	3.91	0.52	0.63	0.76	0.55
min	0.54	0.32	0.12	0.11	0.32	0.12
max	7.48	357.56	7.22	7.09	45.42	34.91
skewness	0.42	53.65	-0.22	0.15	17.52	12.51
kurtosis	3.01	4196.10	6.79	4.11	779.80	740.25

This table presents the results of applying the six different methods to estimate α for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from the 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The six different methods are the KS distance metric, KS ratio method, 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 which are included in the analysis.

Table 8: The median absolute difference of the estimates of α in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	0.22	0.70	0.46	1.02	1.40
KS rat	0.22	0.00	1.14	0.77	1.45	1.80
5%	0.70	1.14	0.00	0.50	0.40	0.68
Eye-Ball	0.46	0.77	0.50	0.00	0.86	1.19
Drees	1.02	1.45	0.40	0.86	0.00	0.27
Du Bo	1.40	1.80	0.68	1.19	0.27	0.00

This table presents the results of applying the six different methods to estimate α for left tail of stock returns. The data is from the CRSP database that contains all individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the median absolute difference between the KS distance metric, KS ratio method, 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 cutt-off at 15% of the total sample size. There are 17,918 companies which are included in the analysis.

Table 9: The median absolute difference of the estimates of α in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	0.28	0.59	0.42	0.74	1.16
KS rat	0.28	0.00	1.12	0.78	1.25	1.64
5%	0.59	1.12	0.00	0.50	0.30	0.59
Eye-Ball	0.42	0.78	0.50	0.00	0.73	1.10
Drees	0.74	1.25	0.30	0.73	0.00	0.31
Du Bo	1.16	1.64	0.59	1.10	0.31	0.00

This table presents the results of applying the six different methods to estimate α for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the median absolute difference between the KS distance metric, KS ratio method, 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 cutt-off at 15% of the total sample size. There are 17,918 companies which are included in the analysis.

Table 10: The correlation matrix of $\hat{\alpha}$ in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	0.10	0.38	0.49	0.24	0.24
KS rat	0.10	1.00	0.05	0.06	0.04	0.03
5%	0.38	0.05	1.00	0.68	0.33	0.65
Eye-Ball	0.49	0.06	0.68	1.00	0.26	0.49
Drees	0.24	0.04	0.33	0.26	1.00	0.24
Du Bo	0.24	0.03	0.65	0.49	0.24	1.00

This table presents the correlation matrix for the estimates of α by applying the six different methods for left tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 11: The correlation matrix of $\hat{\alpha}$ in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	0.13	0.41	0.55	0.32	0.26
KS rat	0.13	1.00	0.05	0.06	0.05	0.04
5%	0.41	0.05	1.00	0.72	0.39	0.62
Eye-Ball	0.55	0.06	0.72	1.00	0.37	0.45
Drees	0.32	0.05	0.39	0.37	1.00	0.25
Du Bo	0.26	0.04	0.62	0.45	0.25	1.00

This table presents the correlation matrix for the estimates of α by applying the six different methods for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 12: The mean absolute difference of the estimates of k^* in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	80.51	126.78	78.18	224.47	360.98
KS rat	80.51	0.00	172.30	34.54	294.60	439.86
5%	126.78	172.30	0.00	142.01	140.18	267.66
Eye-Ball	78.18	34.54	142.01	0.00	266.68	408.87
Drees	224.47	294.60	140.18	266.68	0.00	168.47
Du Bo	360.98	439.86	267.66	408.87	168.47	0.00

This table presents the results of applying the six different methods to estimate α for left tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the mean absolute difference between the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 13: The mean absolute difference of the estimates of k^* in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	0.00	92.13	128.39	85.37	195.12	349.54
KS rat	92.13	0.00	174.16	28.10	266.57	438.43
5%	128.39	174.16	0.00	149.29	120.49	264.38
Eye-Ball	85.37	28.10	149.29	0.00	243.99	413.43
Drees	195.12	266.57	120.49	243.99	0.00	191.04
Du Bo	349.54	438.43	264.38	413.43	191.04	0.00

This table presents the results of applying the six different methods to estimate k^* for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The matrix presents the mean absolute difference between the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 14: The correlation matrix of $\hat{\alpha}(k^*)$ in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	0.10	0.38	0.49	0.24	0.24
KS rat	0.10	1.00	0.05	0.06	0.04	0.03
5%	0.38	0.05	1.00	0.68	0.33	0.65
Eye-Ball	0.49	0.06	0.68	1.00	0.26	0.49
Drees	0.24	0.04	0.33	0.26	1.00	0.24
Du Bo	0.24	0.03	0.65	0.49	0.24	1.00

This table presents the correlation matrix for the estimates of α by applying the six different methods for left tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 15: The correlation matrix of $\hat{\alpha}(k^*)$ in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	0.13	0.41	0.55	0.32	0.26
KS rat	0.13	1.00	0.05	0.06	0.05	0.04
5%	0.41	0.05	1.00	0.72	0.39	0.62
Eye-Ball	0.55	0.06	0.72	1.00	0.37	0.45
Drees	0.32	0.05	0.39	0.37	1.00	0.25
Du Bo	0.26	0.04	0.62	0.45	0.25	1.00

This table presents the correlation matrix for the estimates of α by applying the six different methods for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

Table 16: The correlation matrix of k^* in the **left** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	-0.04	-0.00	-0.10	0.19	-0.11
KS rat	-0.04	1.00	-0.01	0.17	0.16	-0.27
5%	-0.00	-0.01	1.00	0.00	0.01	0.01
Eye-Ball	-0.10	0.17	0.00	1.00	-0.08	-0.26
Drees	0.19	0.16	0.01	-0.08	1.00	-0.29
Du Bo	-0.11	-0.27	0.01	-0.26	-0.29	1.00

This table presents the correlation matrix for the estimates of k^* by applying the six different methods for left tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 5% threshold, automated Eye-Ball method, the 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 total sample size. There are 17,918 companies which are included in the analysis.

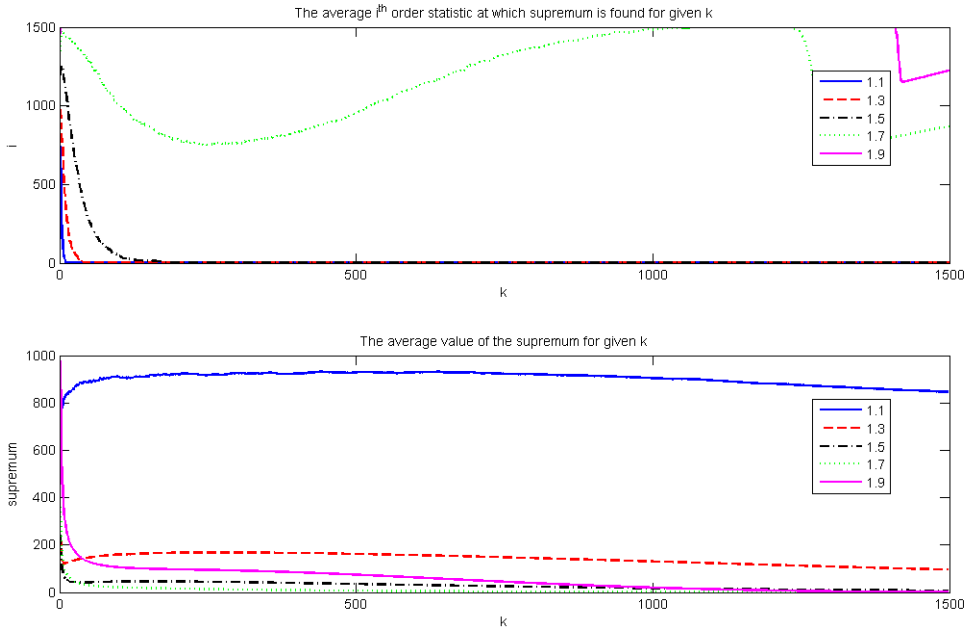
Table 17: The correlation matrix of k^* in the **right** tail for the six different methods

	KS dif	KS rat	5%	Eye-Ball	Drees	Du Bo
KS dif	1.00	-0.13	-0.01	-0.11	0.21	-0.06
KS rat	-0.13	1.00	0.02	0.20	0.13	-0.16
5%	-0.01	0.02	1.00	0.00	-0.00	0.02
Eye-Ball	-0.11	0.20	0.00	1.00	-0.06	-0.22
Drees	0.21	0.13	-0.00	-0.06	1.00	-0.10
Du Bo	-0.06	-0.16	0.02	-0.22	-0.10	1.00

This table presents the correlation matrix for the estimates of k^* by applying the six different methods for right tail of stock returns. The data is from the CRSP database that contains all the individual stocks data from 1925-12-31 to 2013-12-31 for NYSE, AMEX, NASDAQ and NYSE Arca. The different methods are the KS distance metric, KS ratio method, 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 total sample size. There are 17,918 companies which are included in the analysis.

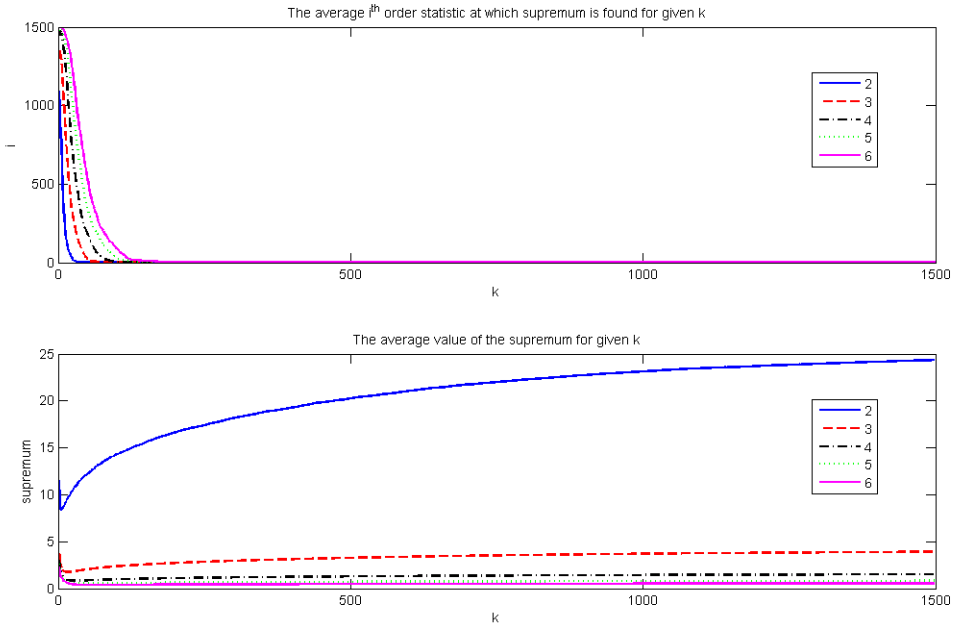
B.2 Figures

Figure 6: Simulations Brownian motion for Symmetric Stable parameters



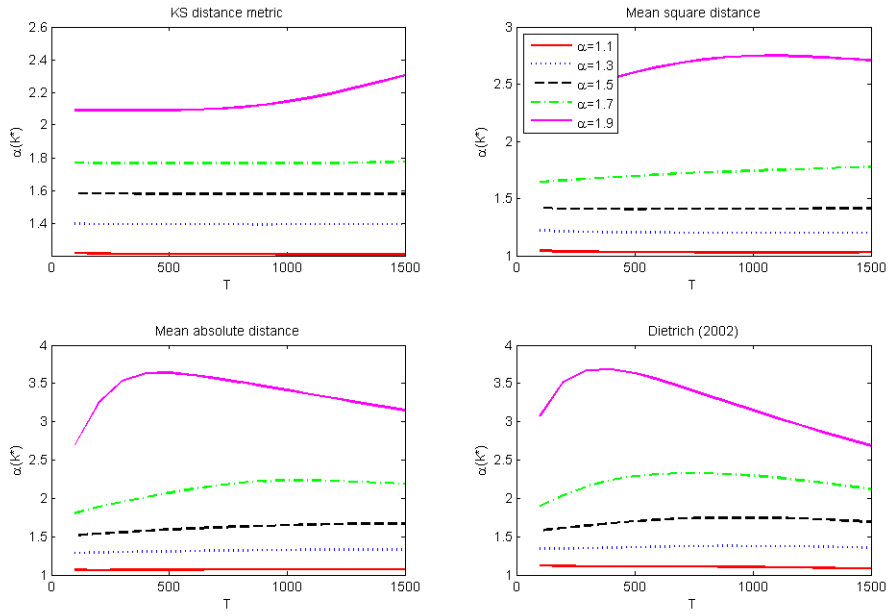
These two figures show the simulations for the limit function in (11). The parameters are for the Symmetric Stable distribution, which are found in Table 4 of the Appendix. The value for α for the different lines is stated in the legend. Here T is 1,500, therefore, the interval between $w(s_i) - w(s_{i+1})$ 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 7: Simulations Brownian motion for Fréchet parameters



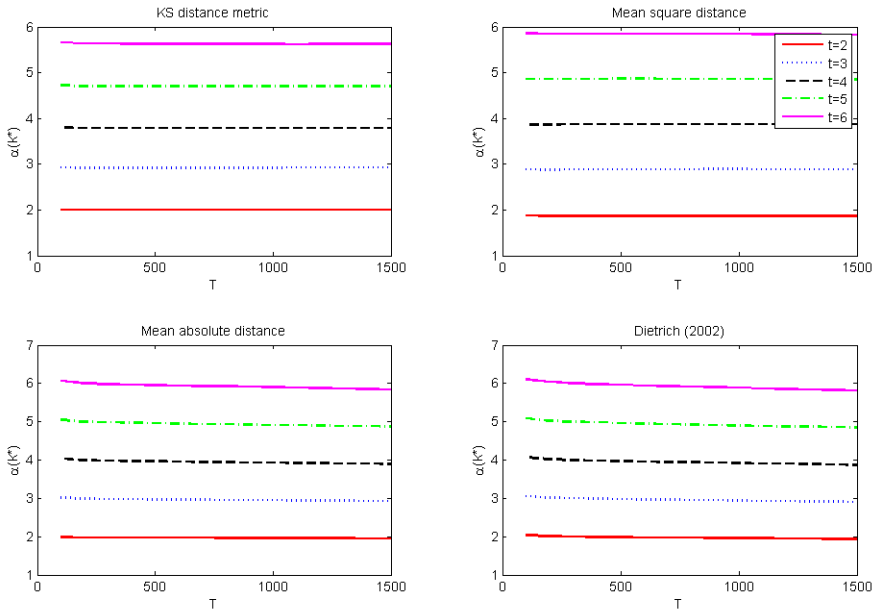
These two figures show the simulations for the limit function in (11). The parameters are for the Fréchet distribution, which are found in Table 4 of the Appendix. The value for α for the different lines is stated in the legend. Here T is 1,500, therefore, the interval between $w(s_i) - w(s_{i+1})$ is 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: Optimal $\hat{\alpha}$ for quantile metrics (Symmetric Stable distribution)



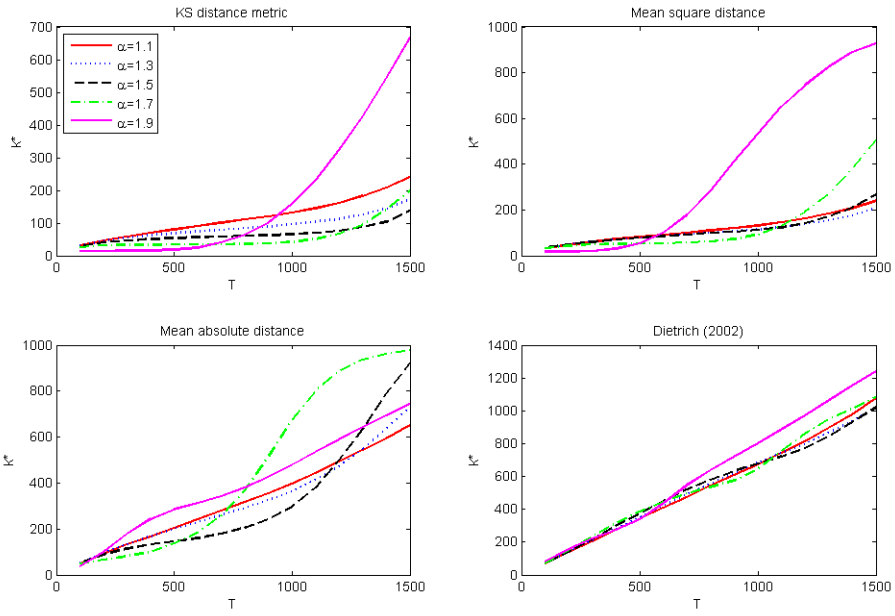
This Figure depicts simulation results of the average optimally chosen $\alpha(k)$ for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Symmetric Stable distributions 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 9: Optimal $\hat{\alpha}$ for quantile metrics (Fréchet distribution)



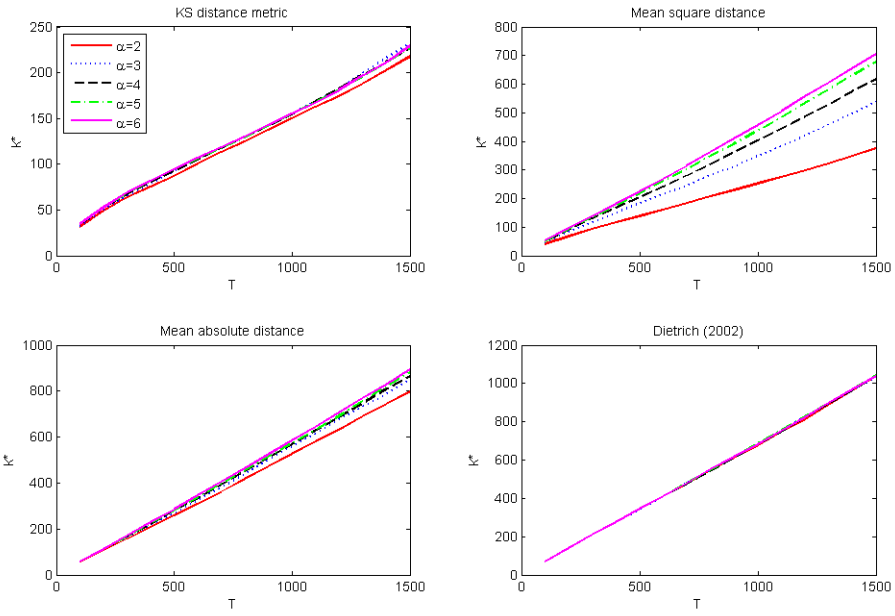
This Figure depicts simulation results of the average optimally chosen $\alpha(k)$ for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Fréchet distributions with degrees 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 10: Optimal k for quantile metrics (Symmetric Stable distribution)



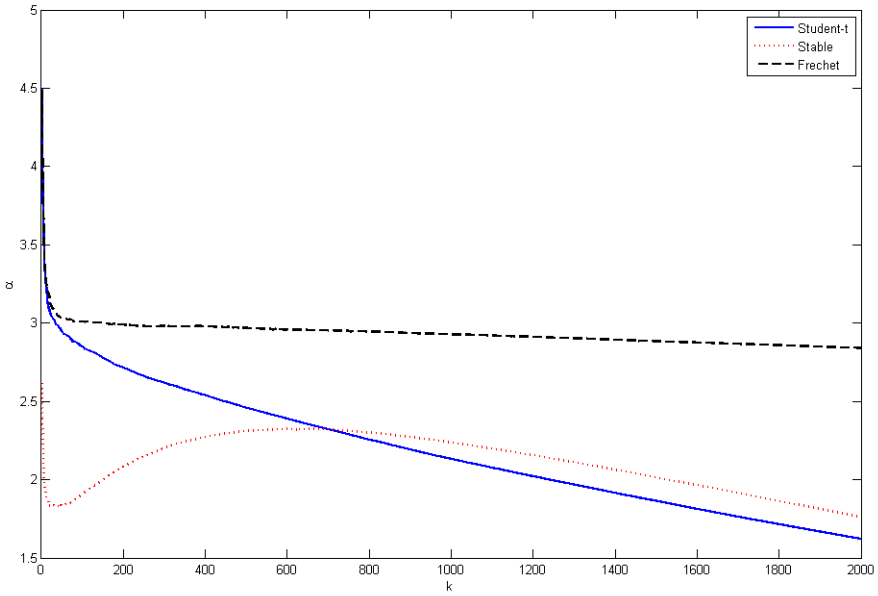
This Figure depicts simulation results of the average optimally chosen k for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Symmetric Stable distributions with degrees 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 11: Optimal k for quantile metrics (Fréchet distribution)



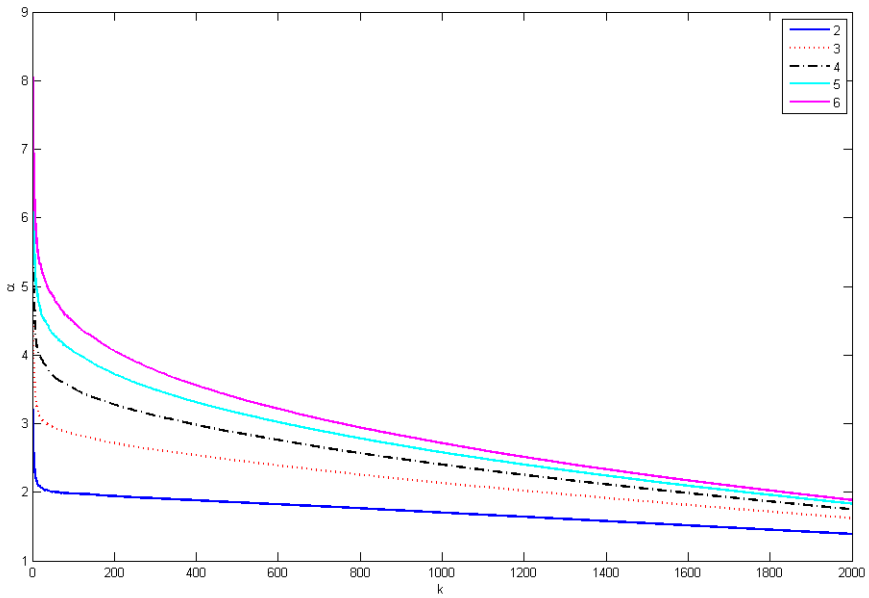
This Figure depicts simulation results of the average optimally chosen k for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Fréchet distributions with degrees of freedom 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 12: Shape of the Hill Plot for different distributions



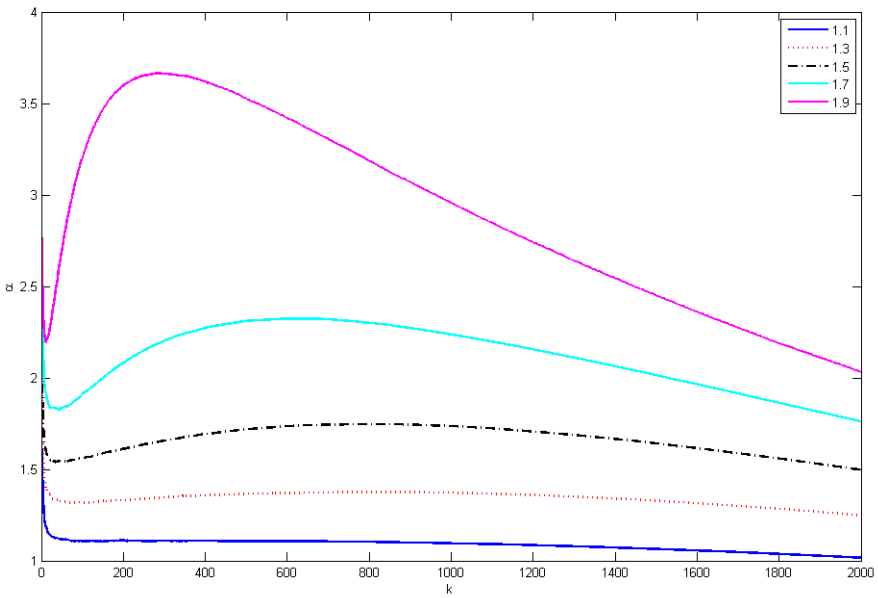
The Figure depicts the Hill estimates for the Student-t(3), Symmetric Stable(1.7), and Fréchet(3) distribution against the number of order statistics used in the estimation. These graphs are constructed by taking the average Hill estimates over 500 simulations.

Figure 13: The shape of the Hill plot for the Student-t distribution



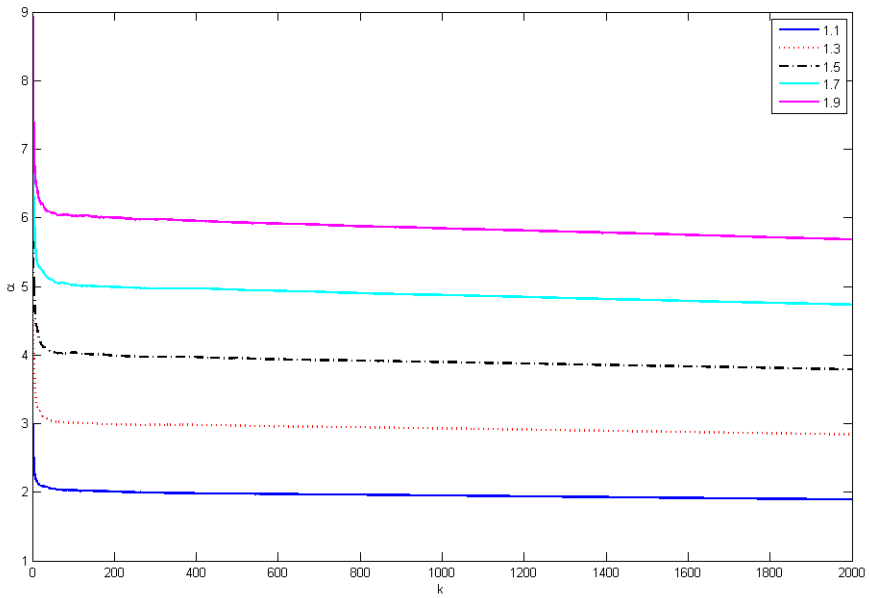
The Figure depicts the Hill estimates for the Student-t distribution family against the number of order statistics used in the estimation. This is done for different corresponding tail indices. These graphs are constructed by taking the average Hill estimates over 500 simulations.

Figure 14: The shape of the Hill plot for the Symmetric Stable distribution



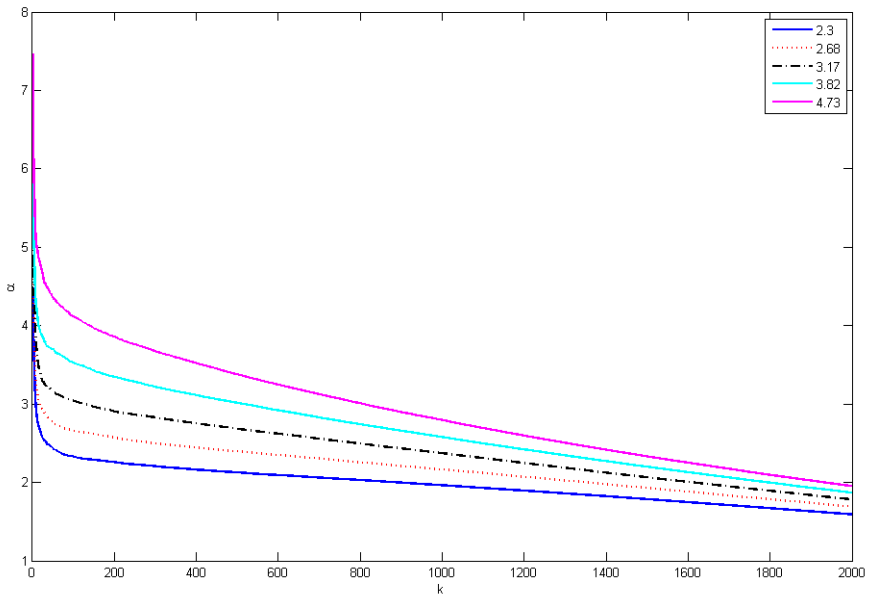
The Figure depicts the Hill estimates for the Symmetric Stable distribution family against the number of order statistics used in the estimation. This is done for different corresponding tail indices. These graphs are constructed by taking the average Hill estimates over 500 simulations.

Figure 15: The shape of the Hill plot for the Fréchet distribution



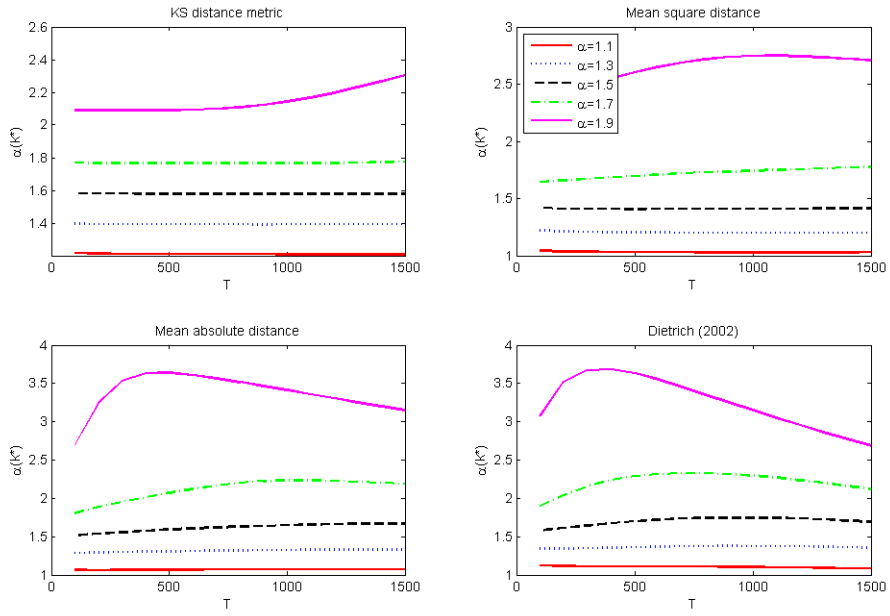
The Figure depicts the Hill estimates for the Fréchet distribution family against the number of order statistics used in the estimation. This is done for different corresponding tail indices. These graphs are constructed by taking the average Hill estimates over 500 simulations.

Figure 16: The shape of the Hill plot for the ARCH stochastic process



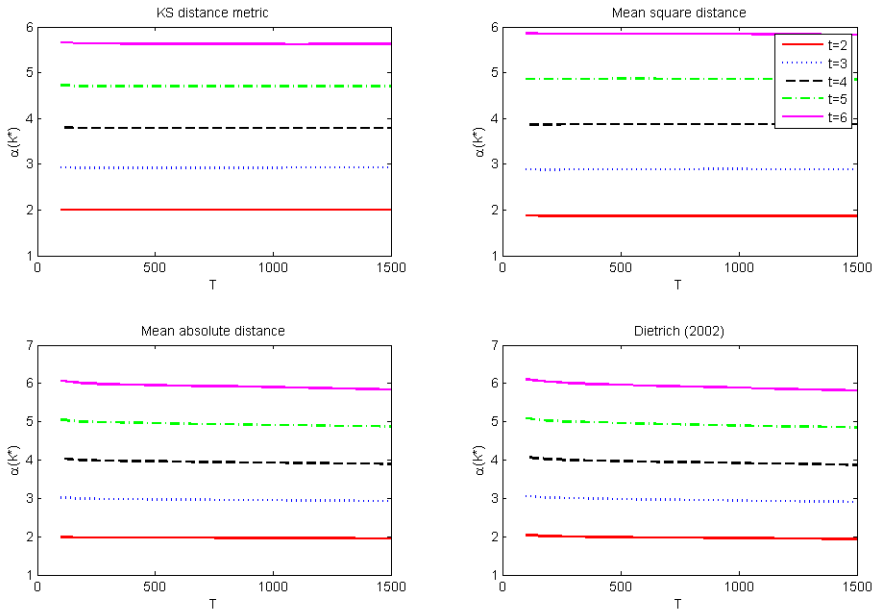
The Figure depicts the Hill estimates for the ARCH(1) process against the number of order statistics used in the estimation. This is done for different corresponding tail indices. These graphs are constructed by taking the average Hill estimates over 500 simulations.

Figure 17: Optimal $\hat{\alpha}$ for quantile metrics (Symmetric Stable distribution)



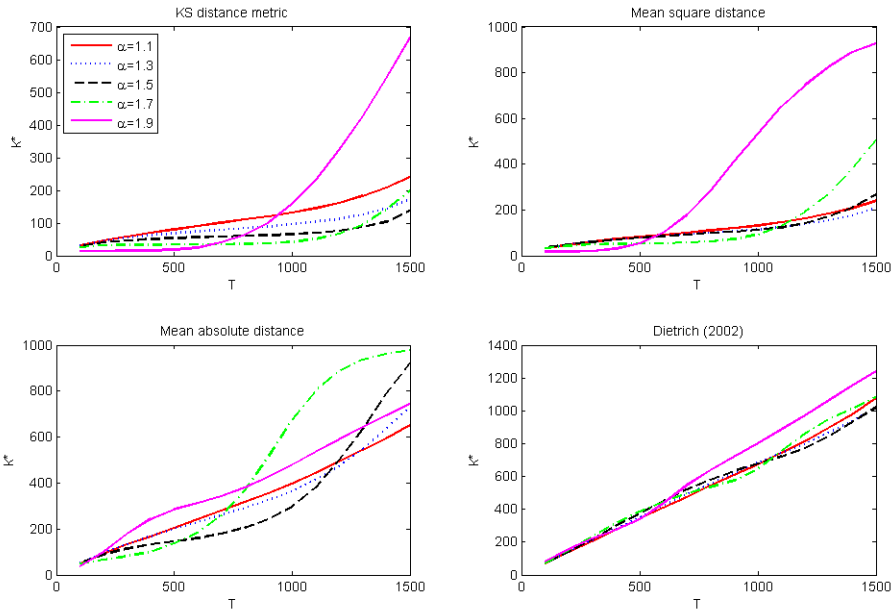
This Figure depicts simulation results of the average optimally chosen $\alpha(k)$ for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Symmetric Stable distributions 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 18: Optimal $\hat{\alpha}$ for quantile metrics (Fréchet distribution)



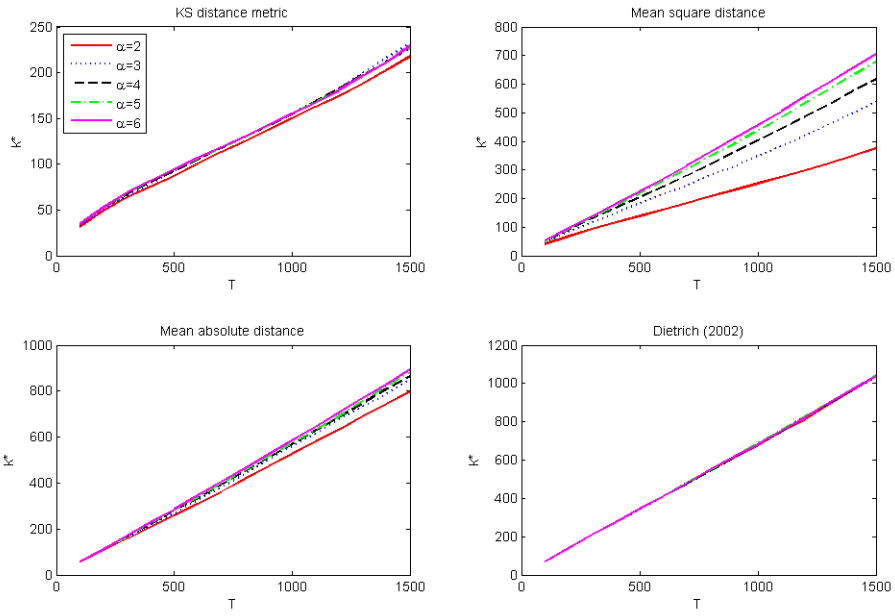
This Figure depicts simulation results of the average optimally chosen $\alpha(k)$ for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Fréchet distributions with degrees 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 19: Optimal k for quantile metrics (Symmetric Stable distribution)



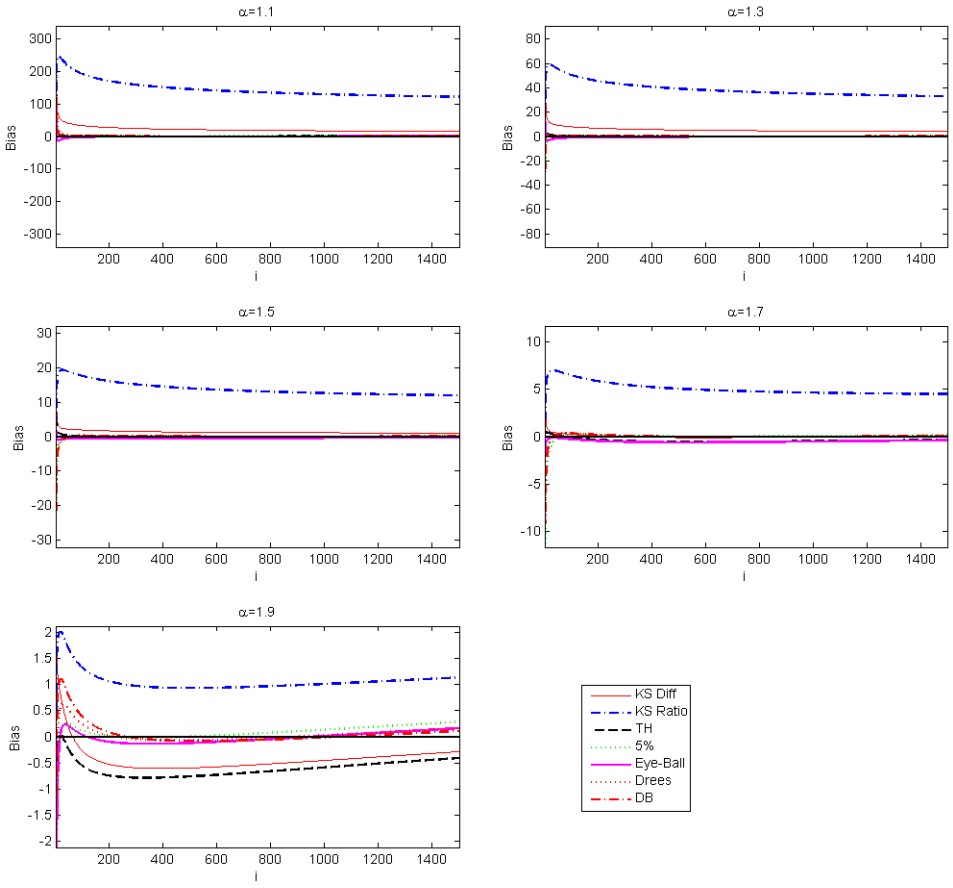
This Figure depicts simulation results of the average optimally chosen k for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Symmetric Stable distributions with degrees 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 20: Optimal k for quantile metrics (Fréchet distribution)



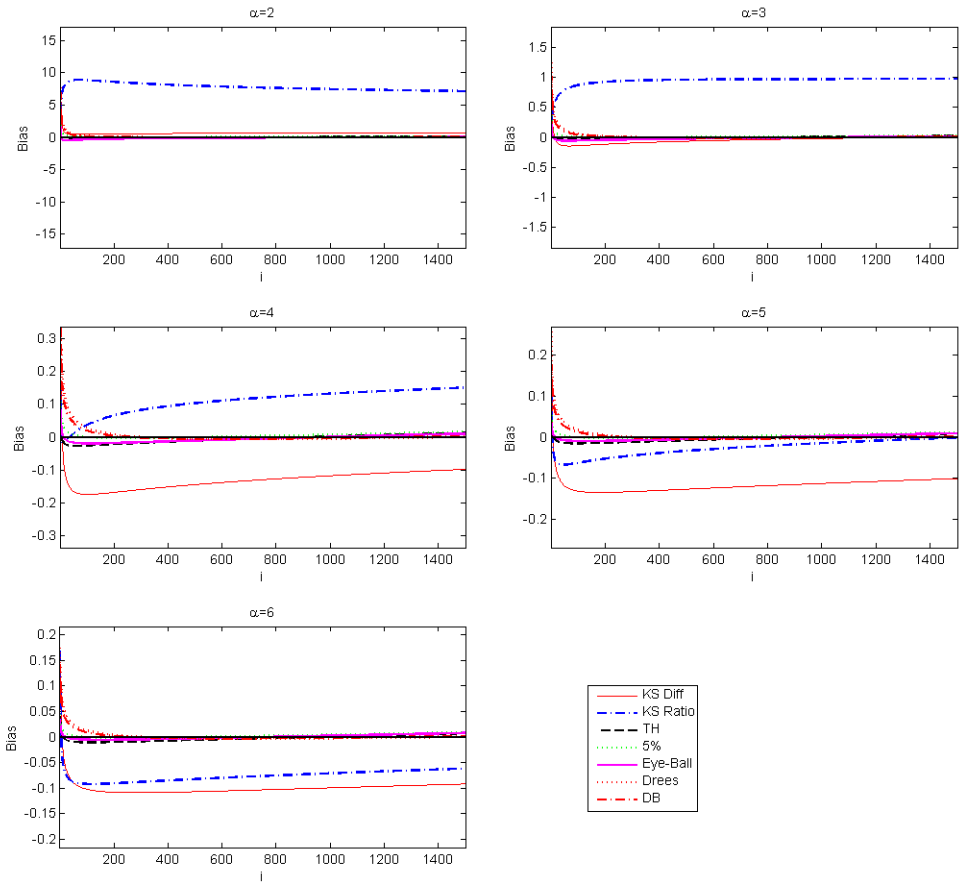
This Figure depicts simulation results of the average optimally chosen k for a given level of T. Here T is the number of extreme order statistics over which the metric is optimized. In the upper left graph this is done for the KS distance metric for different Fréchet distributions with degrees of freedom 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 21: Quantile estimation bias different methodologies (Symmetric Stable distribution)



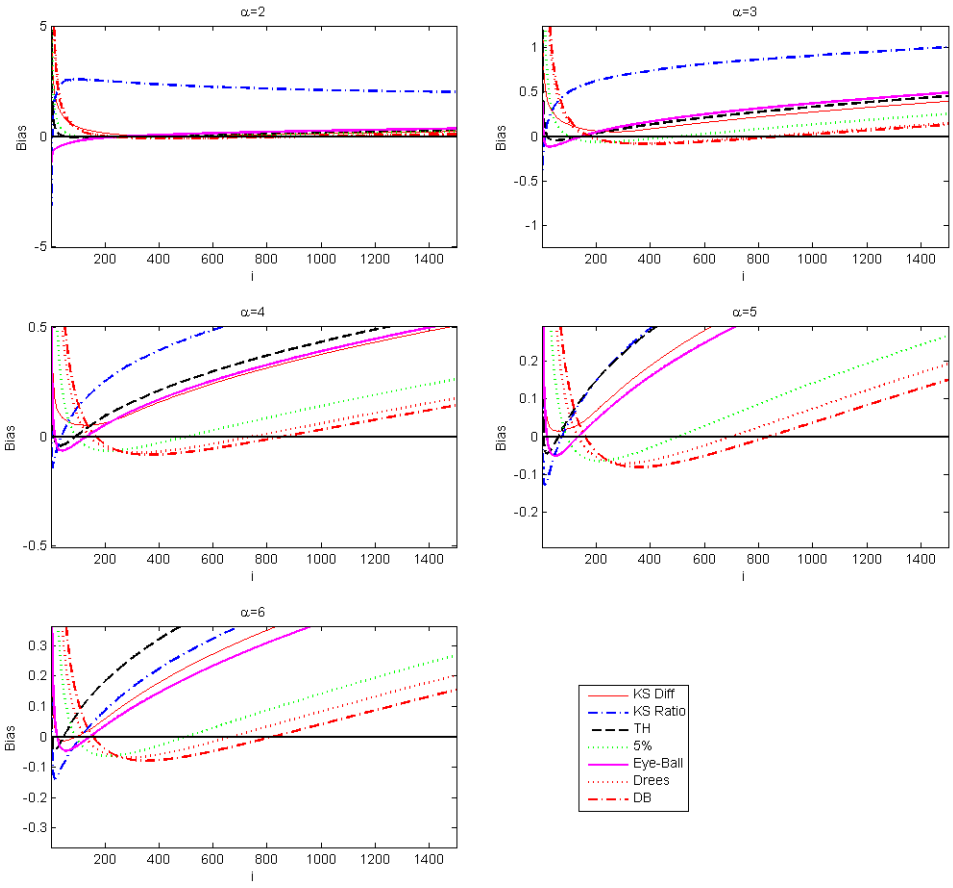
This Figure show the bias induced by using the quantile estimator presented in Equation (9). We use the k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the Symmetric Stable distribution family with the shape parameter indicated at the top of the picture. The i on the horizontal axis gives the probability level i/n at which the quantile is estimated. Moving rightwards along the x-axis represents a move towards the center of the distribution.

Figure 22: Quantile estimation bias different methodologies (Fréchet Distribution)



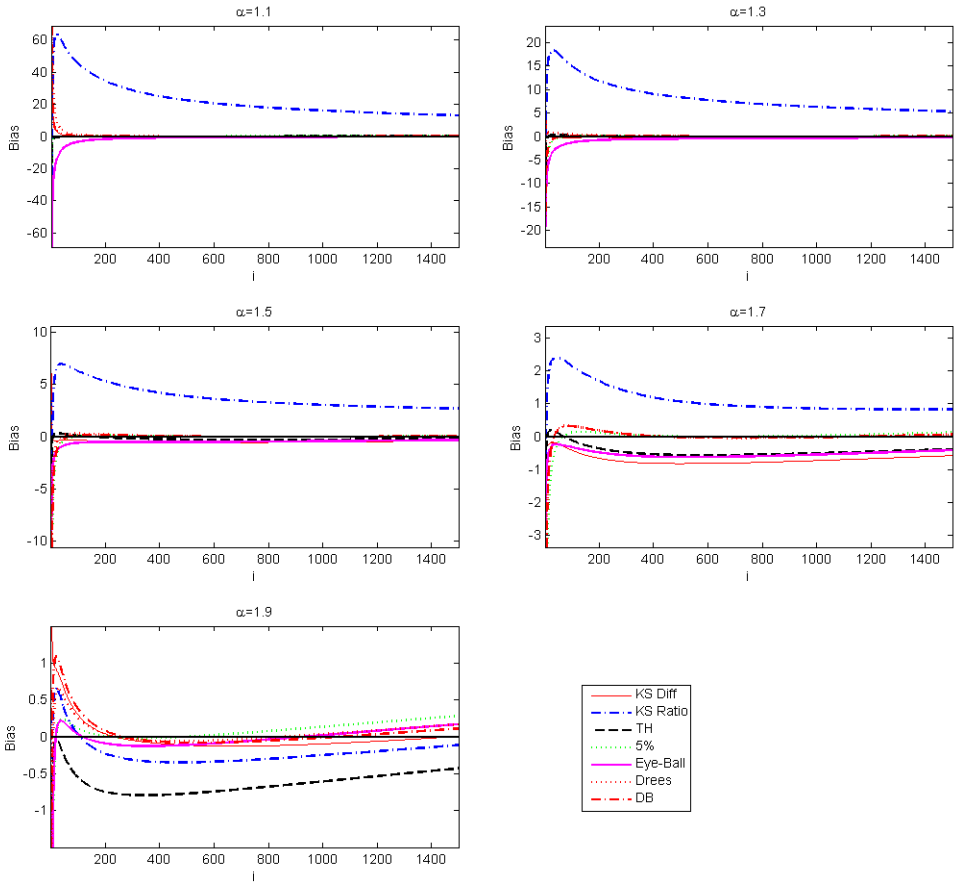
This Figure show the bias induced by using the quantile estimator presented in Equation (9). We use the k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the Fréchet distribution family with the shape parameter indicated at the top of the picture. The i on the horizontal axis gives the probability level i/n at which the quantile is estimated. Moving rightwards along the x-axis represents a move towards the center of the distribution.

Figure 23: Quantile estimation different methodologies median difference (Student-t distribution)



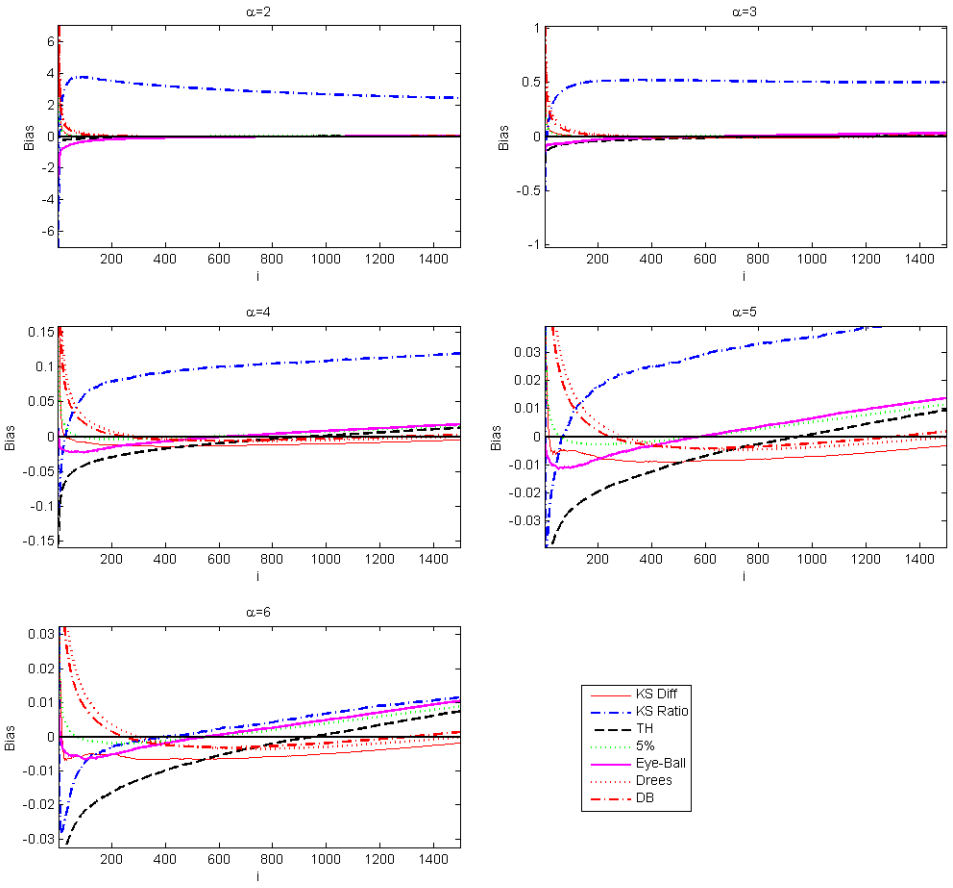
This Figure show the median difference induced by using the quantile estimator presented in Equation (9). We use the k^* from the different methodologies to estimate $\alpha(k^*)$ and the scale parameter $A(k^*)$ for the quantile estimator. The 10,000 samples of size $n = 10,000$ are drawn from the Student-t 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 24: Quantile estimation different methodologies median difference (Symmetric Stable distribution)



This Figure show the median difference induced by using the quantile estimator presented in Equation (9). 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 25: Quantile estimation different methodologies median difference (Fréchet distribution)



This Figure show the median difference induced by using the quantile estimator presented in Equation (9). 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.