Worst-case analysis

Jon Danielsson
London School of Economics

Lerby M. Ergun*
Bank of Canada
London School of Economics

Casper G. de Vries
Erasmus University Rotterdam

February 17, 2020

Abstract

Worst-case analysis studies the worst expected outcome over a predetermined time length. We find that if the distribution has a heavy tail, the non-parametric approach is always overly conservative, i.e downwards biased. Relying on semi-parametric extreme value theory (EVT) reduces the bias considerably in the case of the more heavy-tailed distributions. But for the less-heavy tails this relationship is reversed. Estimates for a large sample of US stock returns indicate that this pattern in the bias is indeed present in financial data. With respect to risk management, this induces an overly conservative capital allocation if the worst case is estimated incorrectly.

Keywords: Worst-case Analysis, EVT, Quantile Estimator, Risk Management
JEL codes: C01, C14, C58

*Corresponding author: lerbyergun@gmail.com, 234 Wellington Street, Ottawa, ON, K1A 0H9. Phone Number: +1-(613)-782-8904. We are grateful for comments from Jason Allen, Laurens de Haan, Maarten van Oordt and Chen Zhou. This work was supported by the Economic and Social Research Council (ESRC) in funding the Systemic Risk Centre [grant number ES/K002309/1] and the Netherlands Organisation for Scientific Research Mozaiek grant [grant number: 017.005.108].

Introduction

Worst-case analysis studies the most extreme outcome over a predetermined time length, with a typical question: what is the worst daily outcome in the next 10 years or 2,500 days? This type of analysis is increasingly common due to extreme weather patterns and the recent financial crises. Much of the bank stress testing scenarios is based on the worst observed historical event (BIS, 2017; EIOPA, 2014). Others use it as a value-at-risk (VaR) metric. In spite of its increasing importance, little is known about worst-case analysis and the properties of its estimators.

In this paper we compare the non-parametric and semi-parametric approaches for heavy-tailed distributions. Both worst-case estimators are biased. The method that produces the smallest bias depends on the heaviness of the tail. The semi-parametric approach produces a smaller bias for very heavy-tailed distributions. Given that the semi-parametric approach has a strictly smaller variance, under the mean square error (mse) criterion there is a strict preference for the semi-parametric estimator for very heavy-tailed distributed variables. We confirm this is the relevant case for the individual securities traded on the US stock exchanges.

There are generally three main approaches to worst-case analysis. The simplest is to directly read the worst case from the empirical distribution, in our case the historical maximum. This is the non-parametric approach (*NP*). One can also assume a model only for the tail of the distribution; this constitutes the semi-parametric approach (*SP*). The third approach is based on specifying a fully parametric distribution for all outcomes and estimating its parameters.

Of these three alternatives, the last is the only one that is not recommended.[1] The estimates are dominated by the more frequent center observations, so that the fit is optimal for a typical observation, but not the highest. Therefore, we focus on the *NP* and *SP* estimators and provide guidance on the appropriate use of either.

The *NP* quantile estimator is the maximum sample ordered observation, i.e. the most extreme order statistic. Through the use of extreme value theory (EVT) and under the assumption that the underlying process sat-

[1] Duffie and Pan (1997) give a comprehensive overview of the different methodologies and issues regarding VaR estimation.

isfies a first-order Hall expansion, we derive the asymptotic distribution of the most extreme order statistic. Using the asymptotic distribution we show that the *NP* estimator is upward biased. The bias is increasing in the heaviness of the tail. The variance of the *NP* approach is disproportionally large and does not exist for distributions with a tail index lower than or equal to 2.

The *SP* tail quantile estimator for the class of heavy-tailed distributions is the Weissman (1978) estimator. This *SP* estimator necessitates the estimation of the tail index. Given the tail index estimator by Hill (1975), Goldie and Smith (1987) show that the *SP* estimator is normally distributed with the bias and variance increasing in the heaviness of the tail. Additionally, the bias and variance are dependent on the number of order statistics, t, utilized to estimate the tail index.

The contribution of this paper lies in the comparison of the biases of the two worst-case estimators. We show that the choice of estimator with the lowest bias hinges on the tail index. For very heavy-tailed distributions, the *SP* estimator produces the smallest bias. This relationship reverses as the distribution becomes less heavy-tailed. For example, the Student-t distribution family the degrees of freedom and the tail index of the distribution are equal. For this distribution family the absolute bias of the *SP* estimator becomes larger than that of the *NP* approach for Student-t distribution with 5 degrees of freedom or more. For the unconditional distribution of the stationary solution to ARCH/GARCH type processes, which are also heavy tailed, the switching of the biases occurs around a tail index of 3.48. Given that the variance of the *NP* approach is strictly larger than that of the *SP* approach, for very heavy-tailed distributed variables the *SP* approach is strictly preferred based on the mse criteria. For the less heavy-tailed distributed variables, one needs to consider the bias-variance trade-off of the estimators.

The comparison of the biases puts forth two predictions, which we test for real world data. First, the difference between the *NP* and *SP* approach, i.e. *NP*-*SP*, is a decreasing function of the tail index (a higher tail index corresponds to a thinner tail). Second, this difference is increasing in t for $t > exp(2)$.[2] To investigate these predictions, we use the securities return data by the Center for Research in Security Price (CRSP) to estimate the two

[2]Here t corresponds to the t^{th}-order statistic used as a threshold for the tail index estimate. Given the other parameters, at $t = exp(2)$ the asymptotic bias of the *SP* approach attains its largest value.

worst-case estimators for each individual stock. We evaluate the relationship between the difference in the two estimates and the tail estimate and t.

The results from the empirical analysis reveal that for stocks with a very heavy-tailed distribution, the *SP* estimate is smaller than the *NP* estimate. The relative size of the biases switches for stocks with a tail index above 3. This suggests that the processes that generates individual stock returns induces an even larger bias than what a Student-t distributed variable would induce. The distribution is more likely similar to that of an ARCH/GARCH type process, as the large volume on volatility clustering would corroborate. In relation to the second prediction, we find that t is positively related to the difference *NP*-*SP*. Therefore, confirming the prediction put forth by the comparison of the variance and the bias of the two estimators.

In the next section we introduce the two quantile estimators and analyze their bias and variance. In the subsequent section we explore the extent of the bias in US securities data. The last section concludes.

1 Worst-case

To formally define the worst case consider a sample of size n,

$$\{X_1, X_2, ..., X_n\}$$

from distribution function, $F(x)$. The sorted sample, i.e. order statistics, can be represented as

$$M_n = X^{(1,n)} \geq X^{(2,n)} \geq ... \geq X^{(n,n)}.$$

Here we define the probability that the maximum is larger than some threshold x as $P(M_n > x)$. The worst-case in this setting is defined as

$$P^{\leftarrow}\left({}^1\!/\!{}_n\right) = x_{1/n}.$$

In many applications the challenge is to find the quantile at which there is a ${}^1\!/\!{}_n$ probability that the most severe outcome exceeds this quantile. Quantile estimation this deep into the tail is notoriously difficult. This paper, relies on EVT to contrast the bias and the variance of two approaches, namely: the non-parametric and semi-parametric estimator.

1.1 Non-parametric estimator

The non-parametric estimator is the maximum observation out of a sample. To characterize the bias and variance of this estimator, consider the class of distribution functions with regularly-varying tails, i.e.

$$\lim_{n\to\infty} \frac{F(-tx)}{F(-t)} = x^{-\alpha},$$

with $x > 0$ and $\alpha > 0$. In that case there exists a slowly varying function $L(x)$ such that we may write for large $x > 0$,

$$P(X \leq -x) \sim L(x)\, x^{-\alpha}$$

and

$$P(X > x) \sim L(x)\, x^{-\alpha}$$

By $x \sim y$ we mean that x is asymptotic to y. A function is slowly varying if,

$$\lim_{n\to\infty} \frac{L(-tx)}{L(-t)} = 1.$$

These distribution functions are characterized as being heavy-tailed.

To derive the bias of the worst observation as a worst-case estimator, we start with a relatively general approach. Consider the Hall expansion (Hall and Welsh, 1985) of a heavy-tailed distribution function,[3]

$$1 - F(x) = Ax^{-\alpha}\left[1 + Bx^{-\beta} + o\left(x^{-\beta}\right)\right] \tag{1}$$

as $x \to \infty$, where $\alpha > 0$, $A > 0$, $\beta > 0$ and B is a real number. Here A and B are the first and second-order scale parameters, where α and β are the first and second-order shape parameters. Furthermore, $o\left(x^{-\beta}\right)$ contains the higher-order terms of the Hall expansion. Assuming the X_i are i.i.d. and consider the first-order expansion, then

$$P(a_n M_n \leq s) = P\left(M_n \leq \frac{s}{a_n}\right) \approx \left[1 - \frac{s^{-\alpha}}{n}\right]^n,$$

[3]For the Pareto distribution we observe that the Hall expansion perfectly fits the first-order term. All of the standard heavy-tailed distributions, like the Student-t, Pareto, symmetric stable distribution or the unconditional distribution of the stationary solution to a GARCH(1,1) process, satisfy (1).

where $a_n = (An)^{-1/\alpha}$. For $n \to \infty$, on the basis of the classical extreme value theorem

$$\lim_{n\to\infty} \frac{P\left\{M_n \leq \frac{s}{a_n}\right\}}{e^{-s^{-\alpha}}} = 1.$$

Given a cumulative distribution function (cdf) that satisfies (1), we can derive the asymptotic expectation

$$A\,\mathrm{E}\left[M_n\right] = (An)^{\frac{1}{\alpha}}\,\Gamma\left(1-\frac{1}{\alpha}\right) \tag{2}$$

and the asymptotic variance

$$A\mathrm{Var}\left(M_n\right) = (An)^{\frac{2}{\alpha}}\left[\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma\left(1-\frac{1}{\alpha}\right)^2\right] \tag{3}$$

based on the first-order Hall expansion.[4] Here $\Gamma\left(\right)$ refers to the gamma function.

1.2 The semi-parametric approach

To contrast the bias and variance of the *NP* approach, we compare it to a semi-parametric estimator of the worst case. By inverting the first-order expansion in (1), using the empirical counterpart of $A = {}^{t}\!/_{n}\left[X^{(t,n)}\right]^{\alpha}$ measured at some threshold $X^{(t,n)}$ and $1 - F(y) = {}^{1}\!/_{n}$, one obtains the *SP* tail quantile estimator of Weissman (1978)

$$\hat{x}_{SP}\left(t\right) = X^{(t,n)} t^{1/\hat{\alpha}_t},$$

where $\widehat{\alpha}_t$ is estimated with the Hill estimator using the t largest order statistics.

For distributions where (1) applies, Goldie and Smith (1987) derive the distribution of the *SP* quantile estimator

$$\frac{\sqrt{t}}{\log\left(t\right)}\left(\frac{\hat{x}_{SP}\left(t\right)}{x^{(p)}} - 1\right) \sim N\left(-\frac{\mathrm{sign}\left(B\right)}{\sqrt{2\beta\alpha}}, \frac{1}{\alpha^2}\right), \tag{4}$$

where B and β are the second-order scale and shape parameters in (1). Here $x^{(p)}$ is the true quantile.

[4]See Appendix A for the derivation. For a derivation of the density for the lower order-statistics, see theorem 2.2.2 in Leadbetter et al. (1983). There EVT for the maximum is extended to lower order-statistics by means of their Poisson property.

1.3 Comparing the non-parametric and semi-parametric quantile estimators

The two approaches, the *NP* and *SP*, each have their own advantages and disadvantages. While the *NP* approach is much simpler to implement, the *SP* benefits from the use of multiple observations to further the accuracy of the estimation. However, the *SP* approach is dependent on correctly specifying the *SP* distribution and identifying a threshold $X^{(t,n)}$.

To shed more light on the use of these two estimators, we first compare their bias. From (2) and (4) the bias of the two approaches is as follows:

$$AE\left(\frac{\hat{x}_{SP}(t)}{x^{(p)}}-1\right)\approx-\frac{\operatorname{sign}(B)}{\sqrt{2\beta\alpha}}\frac{\log(t(n))}{\sqrt{t(n)}} \qquad SP \tag{5}$$

$$AE\left(\frac{\hat{x}_{NP}}{x^{(p)}}-1\right)\approx\Gamma\left(1-\frac{1}{\alpha}\right)-1. \qquad NP \tag{6}$$

As an approximation to $x^{(p)}$ for the non-parametric distribution function, we use the inverse of the first-order Hall expansion of the cdf.

A first observation from comparing the biases is that the asymptotic bias of the *NP* approach is independent of t, as opposed to the *SP* estimator. The asymptotic bias of the *SP* estimators goes to 0 as $t\to\infty$.[5] The *SP* approach benefits from a larger sample to more precisely estimate the parameters. This gives the *SP* approach a strict preference in large samples.

For intermediate values of t, the comparison between expressions (5) and (6) reveals that for particular parameter constellations the absolute bias of the *SP* approach is larger than the *NP* approach. For $\alpha=1$, $\Gamma\left(1-\frac{1}{\alpha}\right)=\infty$, which means for very heavy-tailed distributions the bias of the *NP* approach is large. For $\alpha\to\infty$ both biases tend to zero. However, $\Gamma'[1-{}^1\!/\!_\alpha]|_{\alpha=\infty}=-\gamma\alpha^{-2}$, where γ is the Euler-Mascheroni constant. Given β, this means that the absolute bias of the *NP* approach goes to zero at a faster rate than the *SP* approach. Therefore, it is possible that there exists a point $\alpha<\alpha^*$, where the absolute bias of the *SP* is larger than the *NP* approach.

In Figure 1 we portrays at which combination of α and β the bias of the *NP* estimator becomes smaller than the *SP* approach. One way of viewing the

[5] When studying the statistical properties of the tail, usually the conditions $t(n)\to\infty$ and ${}^{t(n)}\!/\!_n\to 0$ for $n\to\infty$ are imposed.

figure is by taking a fixed β, e.g. $\beta = 2$. For small values α, very heavy-tailed distributions, the bias of the *NP* estimator is larger than the *SP* estimator. As the α increases in value, we enter the gray area, i.e. the region where the bias of the *SP* approach is larger than the *NP* estimators. Therefore, the difference in the bias, *NP*-*SP*, is a decreasing function in α. The threshold t only plays a role in the bias of the *SP* estimator and has a negative relationship with the bias.[6] In case of the Student-t distribution family, $\beta = 2$ and α equals the degrees of freedom for the specific Student-t distribution.[7] We read that in the case of the family of Student-t distributions, the switching in the size of the biases occurs around $\alpha^* \approx 5$. Sun and de Vries (2018) show that for the ARCH/GARCH type processes $\beta = 1$, $B < 0$ and $\alpha > 0$. This implies that α^* is possibly smaller for a GARCH/ARCH process than that of the Student-t distribution family. Figure 1 shows that $\alpha^* \approx 3.48$ for the parameters of the ARCH/GARCH processes. For the family of symmetric stable and Fréchet distributions the bias is always smaller for the *SP* estimator. Given that for both distributions $\beta = \alpha$ and that for the symmetric stable distributions $\alpha < 2$ the bias of the *NP* approach is always larger.

Expressions (3) and (4) allow us to compare the variance of the estimators. They imply that the variance of the *NP* estimator goes to zero as $t \to \infty$. This is not the case for the variance of the *NP* estimators. The variance of the *SP* approach for intermediate $t(n)$ seems to be strictly smaller than that of the *NP* estimator. For $\alpha \leq 2$, the variance of the *NP* estimator does not exist and therefore for the very heavy-tailed distributions the variance of the *NP* is vastly larger than the *SP* approach. Figure 2 shows that for the more heavy-tailed distributions, the *NP* estimator has a disproportionally larger variance than the *SP* estimator. This result, combined with the comparison of the biases, indicates a strict preference for the use of the *SP* approach over the *NP* approach for very heavy-tailed distributed variables. Figure 3 in the Appendix depicts the ratio of the mse for the two worst-case estimators. The figure shows that under the mse criterion, the *SP* estimator is strictly preferred over the *NP* estimator for $\alpha < 13$, given $\beta = 2$.

The comparison of the biases leads to two empirical predictions. First, the difference between the *NP* and *SP* estimator is a decreasing function in α. Second, given that $t > exp(2)$, the difference is increasing in the number of order statistics used in the *SP* approach. This dictates a positive relationship

[6] Given α and β, the bias of *SP* approach reaches its maximum for $t = exp(2)$. Therefore, in the empirical exercise we analyze the cases where $t > exp(2)$.

[7] See Table 3 in the Appendix for the Hall expansion parameter values for the Student-t, symmetric stable and Fréchet distribution families.

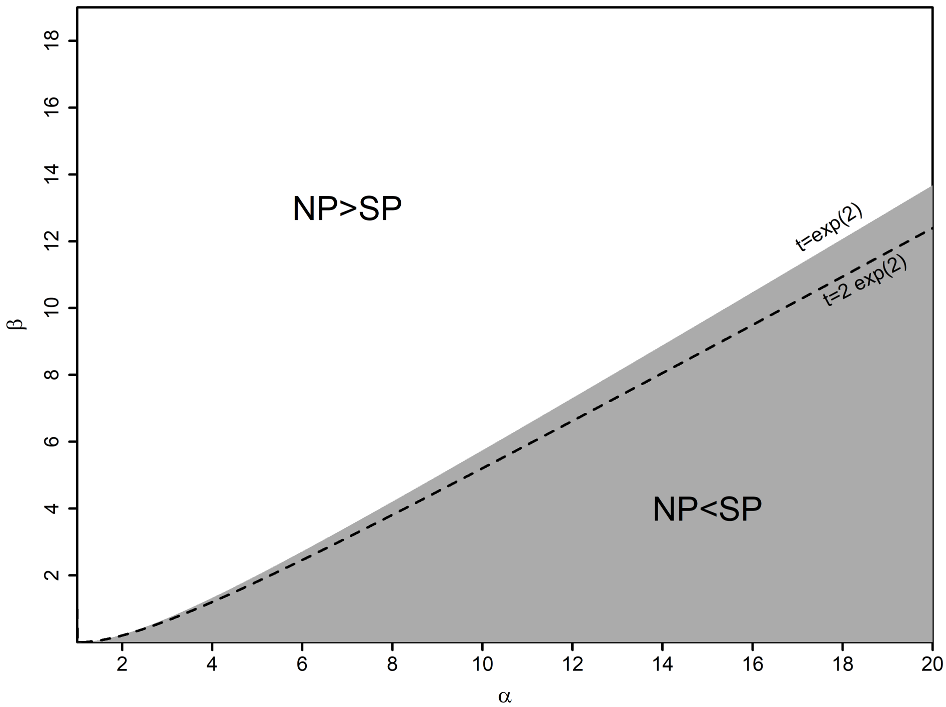

Figure 1: This figure depicts the comparison of the biases of the *SP* and *NP* estimator. On the vertical axis indicates the level of the second-order shape parameters in the Hall expansion. The horizontal axis indicates the level of the tail index. The white area shows the combinations for α and β where the absolute bias of the *NP* estimator is larger than the *SP* estimator. The gray area shows for which combination of α and β the absolute bias of the *SP* estimator is larger. The biases of the estimators are at $p = {}^1\!/_n$ as in Equation (5) and (6). For this figure we fix t at $exp(2)$. In case of the dotted line we fix $t = \mathbf{2}exp(2)$, shifting the gray area down and to the right.

between the difference in the biases and t.

2 Empirical Application

The CRSP dataset contains a large panel of daily stock prices for US stocks, the kind of assets financial institutions typically hold. The large cross-section of stocks allows us to compare a large number of worst-case estimates for the *NP* and *SP* estimators.

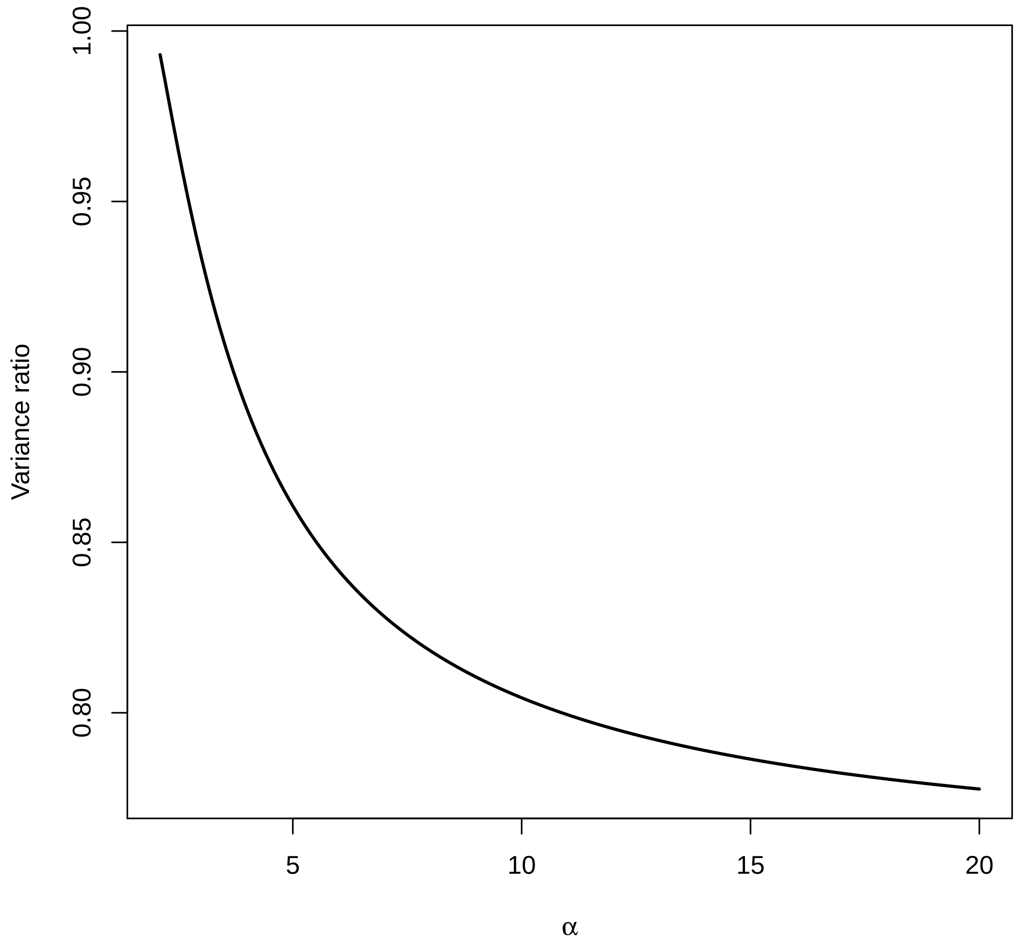

Figure 2: This figure displays the variance ratio of the semi-parametric and the non-parametric worst-case estimator as a function of α. The variance ratio is given by $\sigma^2_{NP}/(\sigma^2_{SP}+\sigma^2_{NP})$. For the variance of the semi-parametric estimator, we choose $t = exp(2)$.

2.1 Data

The CRSP database contains individual stock data from 31-12-1925 to 31-12-2015 for NYSE, AMEX, NASDAQ, and NYSE Arca. In the main analysis, $n = 888$ stocks are used. To be included, we require that stocks are traded on one of the four exchanges during the whole measurement period, which is between 01-01-1995 and 01-01-2011.[8] The choice of the specified fixed sample period is to ensure a large cross-sectional sample and a long enough time series for the EVT estimation.[9] Furthermore, this ensures that the empirical probability at the largest-order statistic is the same across securities.

2.2 Empirical analysis

The SP estimator requires the estimation of α and a choice of t. For this empirical application we use the Hill estimator to estimate the tail exponent α. This estimator depends on a selection of a high-order statistic as a threshold, i.e. $X^{(t,n)}$. This nuisance statistic is obtained by the KS-distance metric developed in Danielsson et al. (2016).[10] The KS-distance metric is focused on picking t to fit the quantile of the distribution. They show that alternative approaches, e.g. Danielsson et al. (2001) and Drees and Kaufmann (1998), underperform significantly, especially when it comes to the quantiles deep in the tail of the distribution.[11]

Given the estimate of the tail index and nuisance statistic, the quantile can be estimated semi-parametrically for each individual stock. The difference,

$$NP_i - SP_i = X_i^{(1,n)} - X_i^{(t_i,n)} t_i^{1/\hat{\alpha}_{t_i}}, \tag{7}$$

for stock i has an estimate of the tail index in the SP quantile estimator. To investigate the initial relationship between the bias in the two estimators, we sort the individual stocks by their estimated α_i. Based on $\hat{\alpha}_i$, the stocks are assigned to five different baskets with a range of $\{\hat{\alpha}_i < 2, 2 < \hat{\alpha}_i \leq 3, 3 < \hat{\alpha}_i \leq 4, 4 < \hat{\alpha}_i \leq 5, 5 \leq \hat{\alpha}_i\}$.

Table 1 portrays that the relative size of the bias changes as a function of $\hat{\alpha}_i$. For the individual stock returns the switch point is around $\hat{\alpha}^* = 3$. It is difficult to determine the exact switch point for real data, due to the unknown values of the second-order parameters, β and B, in the bias of the SP estimator. In addition, the Hill estimator is generally biased (Hall, 1982). The monotonic decrease in the average difference between the baskets is supportive of the result that the bias of the SP estimator overtakes the bias of

[8] We only use common stocks, which have share code 10 and 11. We also require the price of the stock to be above 5 dollars over the measurement period.

[9] The size of the time series for each individual firm is 4,030 days.

[10] The KS-distance metric chooses the threshold which minimizes the maximum quantile distance between the empirical and Pareto distribution. See Appendix A.1 for details.

[11] In Table 4 in the Appendix, we use a fixed sample fraction and obtain similar results.

Table 1: $NP_i - SP_i$ sorted by $\hat{\alpha}_i$

	All	$\hat{\alpha}_i < 2$	$2 \leq \hat{\alpha}_i < 3$	$3 \leq \hat{\alpha}_i < 4$	$4 \leq \hat{\alpha}_i < 5$	$\hat{\alpha}_i \geq 5$
Mean	-0.042	4.234	0.909	-0.371	-1.197	-1.861
Median	-0.701	5.749	0.754	-0.701	-1.337	-1.865
St. Dev.	2.235	5.664	2.926	1.300	1.154	0.604
Q0.01	-3.745	-3.567	-4.650	-2.520	-3.480	-2.921
Q0.99	7.708	9.186	10.472	3.087	1.768	-0.491
Rank Sum test	0.00002	0.250	0.397	0.000	0.000	0.000
N	888	4	329	392	144	19

(a) Left tail

	All	$\hat{\alpha}_i < 2$	$2 \leq \hat{\alpha}_i < 3$	$3 \leq \hat{\alpha}_i < 4$	$4 \leq \hat{\alpha}_i < 5$	$\hat{\alpha}_i \geq 5$
Mean	0.028	6.597	1.232	-0.649	-1.518	-2.203
Median	-0.832	4.382	1.137	-1.008	-1.464	-2.106
St. Dev.	2.797	6.893	2.903	1.565	1.458	0.995
Q0.01	-4.000	0.051	-4.314	-3.320	-4.257	-4.097
Q0.99	7.973	27.236	8.219	4.069	1.689	-0.334
Rank Sum test	0.001	0.000	0.113	0.000	0.000	0.000
N	884	22	314	384	151	13

(b) Right tail

This table reports summary statistics for the difference between the largest order statistic and the semi-parametric quantile estimator, $NP_i - SP_i$, for the left tail and right tail of US stocks returns. For the SP_i estimator, α_i is estimated with the Hill estimator. To determine the number of order statistics for the Hill estimator, we use the KS-distance metric described in Danielsson et al. (2016). Column 1 reports the summary statistics of $NP_i - SP_i$ for all stocks. The second column reports the summary statistics of the difference for the stock with $\hat{\alpha}_i \leq 2$. Columns 3 through 6 report the summary statistics for the stocks with the corresponding $\hat{\alpha}_i$. The first three rows report the mean, median and standard deviation for $NP_i - SP_i$ of the corresponding baskets. $Q_{0.01}$ and $Q_{0.99}$ report the 1% and 99% quantile for the distribution of the different basket of stocks. The next row reports the Wilcoxon signed-rank test p-value, testing non-parametrically for a difference in mean rank. N is the number of stocks in each basket. The individual stock data is from the CRSP dataset. The securities need to be traded on NYSE, AMEX, NASDAQ, and NYSE Arca exchanges over the period from 01-01-1995 to 01-01-2011. To be included, the stock price over the sample needs to be above 5 dollars.

the NP quantile estimator. The results for the difference in the median of each basket convey the same story.

The standard deviation, 1% and 99% quantiles of the buckets show that although the mean and median showcase a switch between the severity of the bias, this might be statistically insignificant. Therefore, we employ the Wilcoxon signed-rank sum test to test for the difference in size of NP_i and SP_i estimates. We find that for stocks with a modestly heavy-tailed return distribution, $\hat{\alpha}_i > 3$, the estimates are significantly different from one another. The SP_i quantile estimates tend to have larger values than the NP_i quantile estimates for these stocks. This is reversed and insignificant for $\hat{\alpha}_i \leq 3$. The same pattern emerges for the right tail of the distribution.

Table 2: Bias in stock returns

	NP-SP					
	Left tail			Right tail		
$\hat{\alpha}_i$	-1.260^{***}		-0.723^{***}	-1.809^{***}		-1.032^{***}
	(0.093)		(0.115)	(0.117)		(0.152)
$t_i/n * 100$		0.336^{***}	0.222^{***}		0.417^{***}	0.263^{***}
		(0.024)	(0.029)		(0.026)	(0.034)
Constant	4.132^{***}	-0.911^{***}	1.774^{***}	5.930^{***}	-1.156^{***}	2.652^{***}
	(0.316)	(0.093)	(0.436)	(0.393)	(0.115)	(0.571)
Observations	864	864	864	867	867	867
R^2	0.174	0.191	0.226	0.216	0.227	0.266

This table reports the regression results for the difference between the largest order statistic and the semi-parametric quantile estimator, $NP_i - SP_i$, for US stocks. For the SP_i estimator, α_i is estimated with the Hill estimator. To determine the number of order statistics for the Hill estimator we use the KS-distance metric described in Danielsson et al. (2016). Here $t_i/n * 100$ is the percentage of order statistics from the total sample used to estimate the Hill estimate. We include only stocks with $t_i > exp(2)$. The individual stock data is from the CRSP dataset. The securities need to be traded on NYSE, AMEX, NASDAQ, and NYSE Arca exchanges over the period from 01-01-1995 to 01-01-2011. To be included, the stock price over the sample needs to be above 5 dollars.

Table 2 reports the results of regressing $NP_i - SP_i$ on their respective tail index and nuisance parameter t_i. The signs of the coefficient estimates are as predicted by the comparison of the biases. The coefficient of $\hat{\alpha}_i$ in the first column shows that an increase of the tail index by 1 decreases the difference in the worst-case return estimates by 1.260 percentage points. The difference switches from positive to negative around a tail index of 3.3. Based on the comparison, in the previous section, between the parameter values of the Student-t and ARCH/GARCH processes, the value of $\alpha^* = 3.3$ indicates that the data possibly comes from an ARCH/GARCH type process. The vast literature on volatility clustering in financial data corroborates this (Engle, 1982).

When including t_i/n in the regression, the coefficient is as predicted. An increase in the number of order statistics, past $t > exp(2)$, decreases the bias in the SP approach and therefore increases the difference in the worst-case estimates. Both $\hat{\alpha}_i$ and t_i have a significant effect on the difference in estimates. This holds for both the left and right tail of the distribution.[12]

In the regressions presented in Table 2 we use $\hat{\alpha}_i$ instead of the true tail index. The measurement error in $\hat{\alpha}_i$ could be correlated with $NP_i - SP_i$. To address this issue we use an instrumental variable approach. In a two-stage least-square regression, we use kurtosis, skewness and the standard deviation of the empirical distribution as instruments for the tail index.[13] Table 5, in the Appendix, shows that the higher moments of the return distribution explain a large portion of the variation in $\hat{\alpha}_i$. The second-stage regression shows that the relationship between the bias and the tail index is not driven by the measurement error in $\hat{\alpha}_i$.

3 Conclusion

With worst-case analysis becoming increasingly common in both policymaking and practice, it is of interest to evaluate the qualities of common methods for such applications. The simplest and perhaps the most common way is to estimate the worst case by taking the most positive outcome in the historical sample. Alternatively, one could estimate the lower tail of the distribution by semi-parametric methods and use that to calculate the worst case.

Based on the comparison of the bias our overall conclusion is that either method is best, depending on how heavy the tails are and their specific shape. Generally, for the heaviest, the semi-parametric approach is best, and as it thins, the historical maxima eventually becomes less biased. The preference for the semi-parametric approach is further reinforced by the strictly higher variance of the non-parametric estimator. Individual stocks with a relatively heavy tail have on average a lower semi-parametric worst-case estimate. This relationship is reversed for stocks with a thinner heavy tail. Taking both the bias and variance of the estimators into account suggests that the semi-parametric approach is the more appropriate choice for very heavy-tailed distributed variables.

[12] To demonstrate robustness of the results, Figure 4 in the Appendix shows the estimates of the third and sixth model for a 10-year annual rolling window between 1975 to 2015.

[13] We have excluded the top and bottom 5% of the sample to prevent the tail observations from influencing the instruments. Results based on the full sample are quantitatively equivalent to the censored sample.

References

BIS, 2017. Supervisory and bank stress testing: Range of practices. Tech. rep., Bank for International Settlements.

Danielsson, J., Ergun, L. M., De Haan, L., De Vries, C., 2016. Tail index estimation: Quantile driven threshold selection. SRC Working Paper .

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

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

Duffie, D., Pan, J., 1997. An overview of value at risk. The Journal of Derivatives 4, 7–49.

EIOPA, 2014. Eiopa insurance stress test. Tech. rep., Consultative Document EIOPA.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society 50, 987–1007.

Fisher, R. A., Tippett, L., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190.

Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics 44, 423–453.

Goldie, C. M., Smith, R. L., 1987. Slow variation with remainder: Theory and applications. The Quarterly Journal of Mathematics 38, 45–71.

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

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

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

Leadbetter, M. R., Lindgren, G., Rootzén, H., 1983. Extremes and related properties of random sequences and processes. Springer Series in Statistics .

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

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

A Expectation and Variance

To formally derive the expectation and the variance of the maximum and the lower-order statistics, consider a sample of size n,

$$\{X_1, X_2, ..., X_n\}$$

from distribution function, $F(x)$. The sorted sample, i.e. order statistics, can be represented as

$$M_n = X^{(1,n)} \geq X^{(2,n)} \geq ... \geq X^{(n,n)}.$$

The order-statistics follow a binomial distribution:

$$G^{(k,n)}(x) = \sum_{r=0}^{k-1} \binom{n}{r} [1 - \mathrm{F}(x)]^r [\mathrm{F}(x)]^{n-r}. \tag{8}$$

Suppose one is interested in the distribution of the maximum realization:

$$\Pr(M_n < x) = [\mathrm{F}(x)]^n = G^{(1,n)}(x). \tag{9}$$

Similar to the standard central limit theorem for the asymptotic distribution of the arithmetic mean, Fisher and Tippett (1928) and Gnedenko (1943) provide a limit theorem for the asymptotic distribution of the maximum, i.e. EVT.

EVT gives the conditions under which there exist sequences b_n and a_n such that

$$\lim_{n\to\infty} [\mathrm{F}(a_n x + b_n)]^n \to \mathrm{G}^{(1,n)}(x).$$

Now suppose $X_1, ..., X_n$ have distribution function from the class with regularly varying tails F, i.e.

$$\lim_{s\to\infty} \frac{1 - F(sx)}{1 - F(s)} = x^{-\alpha}, \quad \alpha > 0. \tag{10}$$

Given this regular variation property and appropriate norming constants a_n and b_n, the heavy-tailed limit distribution, $G^{(1,n)}$, takes the form of the Fréchet distribution. Theorem 2.2.2 in Leadbetter et al. (1983) extends the EVT for the maximum to lower-order statistics by means of the Poisson property of the lower-order statistics. In particular, the asymptotic distribution of the k^{th} largest order statistic is

$$\mathrm{G}^{(k,n)}(x) \to \mathrm{G}^{(1,n)}(x) \sum_{s=0}^{k-1} \frac{\left(-\log\left[\mathrm{G}^{(1,n)}(x)\right]\right)^s}{s!}. \tag{11}$$

Therefore, for distributions functions with regularly varying tails we have,

$$\mathrm{G}^{(k,n)}\left(x\right)=e^{-a_n^\alpha x^{-\alpha}}\sum_{s=0}^{k-1}\frac{\left(a_n^\alpha x^{-\alpha}\right)^s}{s!}.$$

For the density we find

$$g^{(k,n)}\left(x\right)=\alpha a_n^\alpha x^{-\alpha-1}e^{-a_n^\alpha x^{-\alpha}}\left[\frac{\left(a_n^\alpha x^{-\alpha}\right)^{k-1}}{[k-1]!}\right].$$

Given the density, determining the expectation of the k^{th} order statistic is straightforward:

$$E\left[X^{(n-k+1,n)}\right]=\int_0^\infty x\alpha a_n^\alpha x^{-\alpha-1}e^{-a_n^\alpha x^{-\alpha}}\left[\frac{\left(a_n^\alpha x^{-\alpha}\right)^{k-1}}{[k-1]!}\right]dx.$$

Applying a change of variable $y=a_n^\alpha x^{-\alpha}$ we get

$$\begin{aligned}E\left[X^{(n-k+1,n)}\right]&=\frac{a_n}{k-1}\int_0^\infty y^{\frac{1}{\alpha}}y^{k-1}e^{-y}dy\\&=\frac{a_n}{[k-1]!}\Gamma\left[k-\frac{1}{\alpha}\right].\end{aligned}$$

Given the above expectation, determining the variance of the order statistics is a trivial matter:

$$\begin{aligned}var\left[X^{(n-k+1,n)}\right]&=E\left[\left(X^{(n-k+1,n)}\right)^2\right]-E\left[X^{(n-k+1,n)}\right]^2\\&=\frac{a_n^2}{[k-1]!}\Gamma\left[k-\frac{2}{\alpha}\right]-\left[\frac{a_n}{[k-1]!}\Gamma\left[k-\frac{1}{\alpha}\right]\right]^2\end{aligned}\tag{12}$$

A.1 KS-distance metric

The purpose of the KS-distance metric is to find the optimal number of order statistics to estimate the tail index with the Hill estimator. This method achieves this by minimizing the distance between the empirical distribution and Pareto distribution over the quantile dimension. The starting point for locating t^* is the first-order term of Hall's power expansion:

$$\mathrm{F}(x) = 1 - Ax^{-\alpha}[1 + o(1)]. \tag{13}$$

This function is identical to a Pareto distribution if the higher-order terms are ignored. By inverting (13), we get the quantile function

$$x = \left(\frac{1 - F(x)}{A}\right)^{\frac{1}{-\alpha}}. \tag{14}$$

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

$$q(t, i) = \left(\frac{i}{t}\right)^{-1/\widehat{\alpha}_t} X^{(t,n)}. \tag{15}$$

Here $X^{(t,n)}$ is the t^{th} order statistic such that t/n comes closest to the probability level $1 - F(x)$.

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

$$t^* = \arg\inf_t \left[\sup_i \left|X^{(i,n)} - q(t, i)\right|\right], \qquad \text{for } t = 1, ..., T, \tag{16}$$

where $T > t$ is the region over which the KS-distance metric is measured. Here $X^{(i,n)}$ is the order statistic and $q(t, i)$ is the estimated quantile from (15). This is done for different levels of t. The t, which produces the smallest maximum horizontal deviation along all the tail observations until T, is the t^* for the Hill estimator.

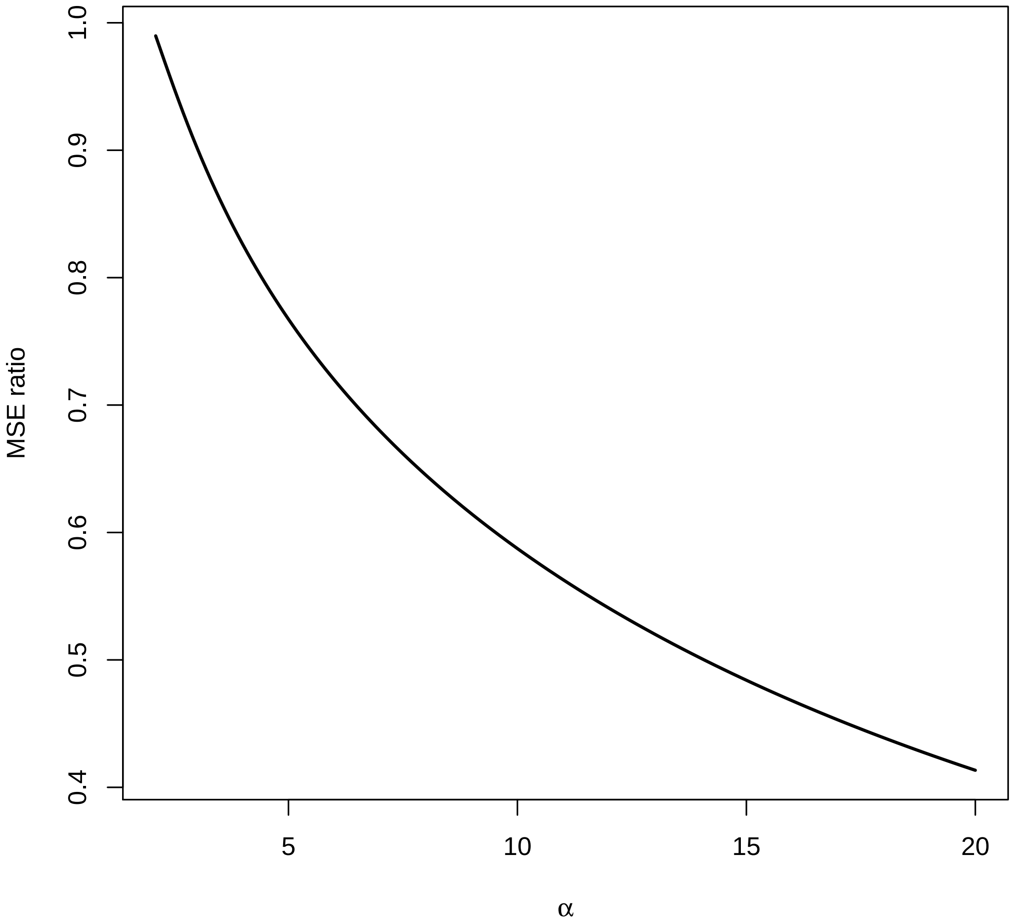

Figure 3: This figure displays the MSE ratio of the semi-parametric and the non-parametric worst-case estimator as a function of α. From (2), (3) and (4), we construct the $MSE = Variance + Bais^2$. The MSE ratio is given by $^{MSE_{NP}}/_{(MSE_{SP}+MSE_{NP})}$. For the MSE of the semi-parametric estimator we choose $t = exp(2)$ and $\beta = 2$.

Table 3: 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 4: Bias in stock returns: Fixed threshold

	NP-SP	
	Left tail	Right tail
$\hat{\alpha}_i$	-0.481^{***}	-0.852^{***}
	(0.092)	(0.138)
Constant	4.351^{***}	6.286^{***}
	(0.405)	(0.594)
Observations	889	889
R^2	0.030	0.041

This table reports the regression results for the difference between the largest order statistic and the semi-parametric quantile estimator, $NP_i - SP_i$, for US stocks. For the SP_i estimator, α_i is estimated with the Hill estimator. The number of order statistics is fixed at 0.25% of the total sample. The individual stock data is from the CRSP dataset. The securities need to be traded on NYSE, AMEX, NASDAQ, and NYSE Arca exchanges over the period from 01-01-1995 to 01-01-2011. To be included, the stock price over the sample needs to be above 5 dollars.

Table 5: Bias in stock returns: IV Regression

	NP-SP			
	Left tail		Right tail	
	Stage 2	Stage 1	Stage 2	Stage 1
$\hat{\alpha}_i^{fitted}$	-1.568^{***}		-0.823^{**}	
	(0.269)		(0.400)	
$t_i/n * 100$	0.090^{*}	-0.092^{***}	0.294^{***}	-0.102^{***}
	(0.048)	(0.008)	(0.065)	(0.007)
Kurtosis		-0.014^{***}		-0.004^{***}
		(0.001)		(0.001)
Skewness		0.291^{***}		-0.105^{***}
		(0.031)		(0.026)
StDev		11.497^{***}		25.255^{***}
		(2.640)		(2.823)
Constant	4.911^{***}	3.475^{***}	1.880	3.147^{***}
	(1.004)	(0.064)	(1.481)	(0.071)
Observations	864	864	867	867
R^2	0.221	0.492	0.231	0.526
F Statistic	122.285^{***}	208.334^{***}	129.598^{***}	239.066^{***}

This table reports the regression results of the two-stage least-square estimation. We instrument the estimated tail index. In the **first stage** we estimate $\hat{\alpha}_i = b_0 + b_1 * Kurtosis_i + b_2 * Skewness_i + b_3 * StDev_i + b_4 * (t_i/n * 100) + \varepsilon_i$. Here kurtosis, skewness and standard deviation are the moments of the return distribution of stock i. We exclude the top and bottom 5% of the observations in the measurement of the higher moments. In the **second stage** we estimate $NP_i - SP_i = c_0 + c_1 * \hat{\alpha}_i^{fitted} + c_2 * (t_i/n * 100) + \nu_i$. For the SP_i estimator, α_i is estimated with the Hill estimator. To determine the number of order statistics for the Hill estimator we use the KS-distance metric described in Danielsson et al. (2016). Here ${}^{t_i}/_{n} * 100$ is the percentage of order statistics from the total sample to estimate the Hill estimate. We include only stocks with $t_i > exp(2)$. The individual stock data is from the CRSP dataset. The securities need to be traded on NYSE, AMEX, NASDAQ, and NYSE Arca exchanges over the period from 01-01-1995 to 01-01-2011. To be included, the stock price over the sample needs to be above 5 dollars.

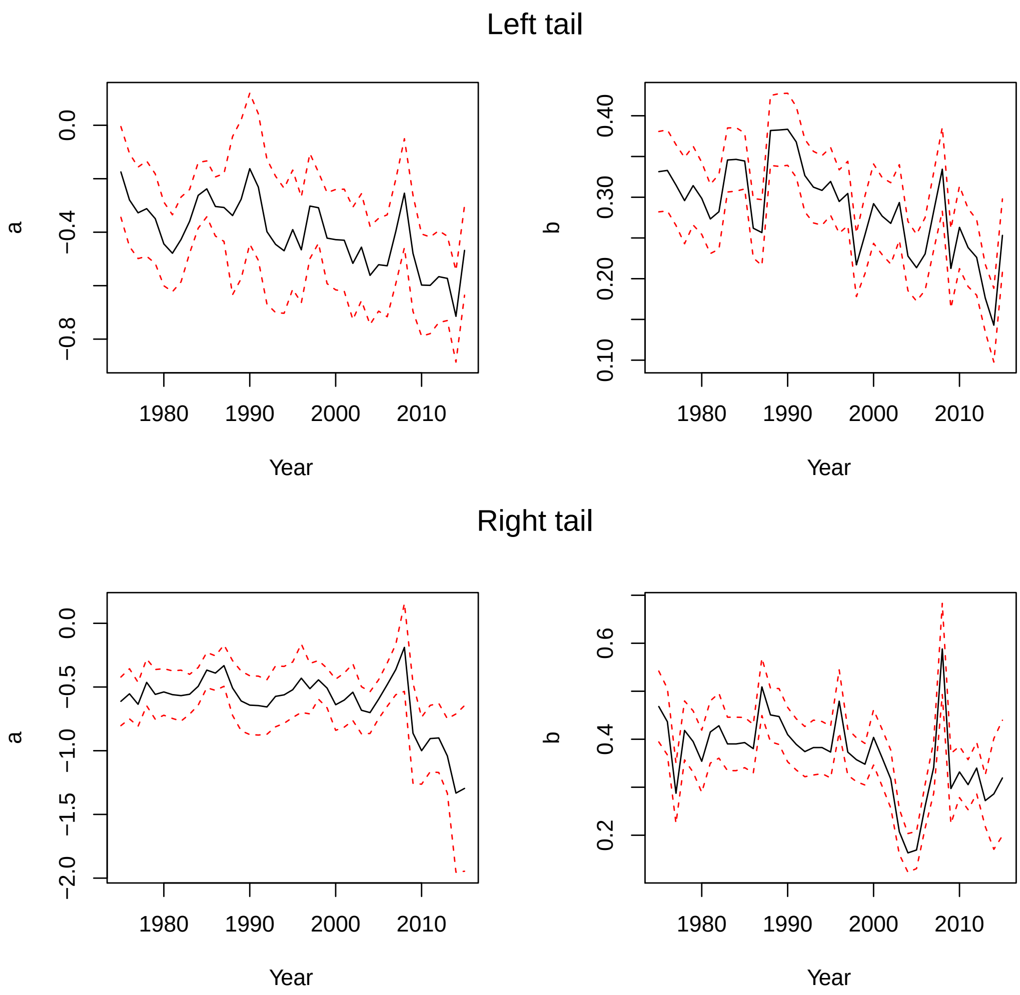

Figure 4: These figures depict the stability of the parameter estimates of Table 2. The solid lines are the parameter estimates over time and the dotted lines are their respective 95% error bounds. The two top and two bottom panels show the results for the left tail and right tail of the distribution, respectively. The left figures depict the results for the coefficient estimates $\hat{\alpha}_i$ and the right figures show the coefficient estimates ${}^{t_i}\!/_{n} * 100$. The regression equation, $NP_i - SP_i = c + a\ \hat{\alpha}_i + b\ {}^{t_i}\!/_{n} * 100 + e_i$, is estimated each year. In the estimation, the data from the preceding 10 years are used to estimate $NP_i - SP_i$, $\hat{\alpha}_i$, and ${}^{t_i}\!/_{n}$. We include only stocks with $t_i > exp(2)$. The individual stock data is from the CRSP dataset. The securities need to be traded on NYSE, AMEX, NASDAQ, and NYSE Arca exchanges over the period from 01-01-1965 to 01-01-2015. To be included, the stock price over the sample needs to be above 5 dollars.