Where do Extremes Matter?*

Jon Danielsson
London School of Economics

Casper G. de Vries
Erasmus University Rotterdam

February 2002

Abstract

Extreme value theory has been applied to many areas of economics where the data is heavy tailed, e.g. in the analysis of market structure and risk forecasting. Accurate inference has, however, been hindered by the lack of consistent procedures for determining the start of the tail. A double subsample bootstrap procedure is proposed to solve this problem. The accuracy of the procedure is accessed with Monte Carlo experiments. Subsequently it is applied to Gibrat's and Zipf's laws, as well as the estimation of financial risk.

JEL: C0, C4, D4, G0

*Corresponding author: Jon Danielsson `j.danielsson@lse.ac.uk` . Danielsson benefitted from an HCM fellowship of the EU. We are grateful to Holger Drees, Laurens de Haan, Marc Henry, Oliver Linton, Namwom Hyung, and Liang Peng. Some of the data studied in the paper was obtained from Olsen and Associates, and Charles van Marrewÿk provided us with the city size data. Our papers can be downloaded from `www.RiskResearch.org`.

1 Introduction

Economic analysis is often dependent on the accurate estimation of the tails of distributions, in applications such as market structure and risk analysis. While the probability theory of extremes (extreme value theory, EVT) is well developed, the determination of the threshold where the tails begin remains a fundamental statistical problem. This is important because most analysis of extremes depends on using the highest or lowest observations to estimate the distribution of the tail area. Estimates of tail shapes are in general biased and inefficient unless the threshold is accurately determined. The reason is that the linear relationship between log size and log rank, which most estimators depend on, typically only holds asymptotically. As a result, finding the tail threshold is a nontrivial undertaking. Our main objective is the development of a procedure for optimally estimating the tail threshold. We subsequently apply our method to a range of problems in market structure and risk analysis.

Extreme value theory (EVT) is the study of the probability theory governing tail shapes of distributions for extreme observations. The limit law for the extreme order statistics is one of three types which are determined by whether the distribution has a finite endpoint or not, and by whether the tails of the densities fall exponentially fast or by a power. When the tails decline exponentially, all moments are bounded, and the distribution is said to be thin tailed. If however, the tails decline by a power, not all moments are defined, and the distribution is said to be heavy tailed. Typically for the applications considered here, the tails are heavy, and the discussion below focuses on that case.

A long–standing issue in economics concerns the relation between size and rank. The earliest published examples date back to Pareto's (1898) work on income distributions. Current examples are Chung and Cox (1994) who analyze the income distribution of superstars with Pareto (1896) laws, Sutton (1997) who employs Gibrat's (1931) law to consider the size distribution of the largest firms, e.g. to predict mergers, and Gabaix (1999) who uses the law of Zipf (1949) to analyze the population sizes of large US metropolitan areas. The common theme in all these applications is the assumption that the the largest observations of the quantity of interest, be it city size, market capitalization, or income distribution adheres to the hyperbolic Pareto distribution. This implies that the log size–log rank relationship will be linear. If, however, the Pareto distribution only holds asymptotically in the tails, the relationship between log size and log rank will be nonlinear except in the extreme tails. As a result, parameter estimates are sensitive to the number of observations in the tail, a point noted by e.g. Mansfield (1962) and Ijiri and Simon (1977).

Similarly, financial risk analysis is dependent on accurate forecasts of the probability of tail events. This problem has been addressed with a variety of methods.

However, if one needs to forecast very low probability events, extreme value theory (EVT) has to be used. EVT has seen widespread applications in risk analysis and especially stress testing in both insurance (see e.g. Embrechts et al., 1997, ch. 2), and finance (see e.g. Jansen and de Vries, 1991; Daníelsson and de Vries, 1997; Quintos et al., 2001; Longin, 2001; Chernozhukov and Umantsev, 2001). In these applications, a key question concerns the determination of the tails to ensure robustness of the estimated tail of the distribution.

Extreme value theory has also found useful applications in econometrics, e.g. in non–nested testing (e.g. Akgiray et al., 1988; Koedijk et al., 1990), and convergence rates of regression estimators (e.g. Loretan and Phillips, 1994). In these applications the determination of the tail threshold is also critically important. Furthermore, a related problem appears in optimal bandwidth selection, and similar methods as those proposed below have been applied to that problem, (see e.g. Henry and Robinson, 1999).

A key feature of EVT is that the distribution of the tails is only estimated with observations from the tail area. This implies accurate inference since parameter estimates are not affected by assumptions about what happens in the center of the distribution. In EVT, the rate of decline of the tails is determined by the tail index, α. The tail index is estimated by using the most extreme observations above a threshold s_n, where n is the sample size. The most common estimator of the tail index is the Hill estimator, which is generally considered to have more desirable properties than other estimators, see e.g. Hall and Welsh (1984). The efficient determination of the tail threshold, s_n, requires an optimal assessment of the trade–off between bias and variance.

In special cases, i.e., when relationship of the log rank to log size is linear, the Hill estimator is unbiased, and one simply uses all available data. This corresponds with the case where the data is exactly Pareto distributed. In general however, the log rank to log size relationship is nonlinear. This corresponds with the Pareto distribution only being a first order approximation to the true distribution far away from the center. In this case, the determination of the optimal threshold necessitates considering higher order approximations of the true distribution.

Unfortunately, finding the first order parameters and the tail threshold s_n when second order factors cause bias is not a trivial exercise. A key objective of this paper is the estimation of s_n which determines the location where the first order Pareto parameter approximation of the cumulative distribution function is appropriate. We first estimate the relevant parameters of the generalized Pareto, and subsequently calculate the tail threshold.

Our estimator of s_n is based on first bootstrapping the mean squared error (MSE) of $1/\hat{\alpha}$, and the exponent of the Pareto distribution function and subsequently minimizing the MSE with respect to s_n. Unfortunately, since α is unknown, one does not have a theoretical benchmark for the calculation of the MSE. As a replacement for the true α, we propose using an alternative estimator which converges in the MSE sense at the same rate, but with a different constant. Hence, this difference statistic converges at the same rate and has a known theoretical benchmark which equals zero in the limit. The idea for extracting the optimal sequence s_n, i.e. finding the best MSE sequence, is similar to the use of control variates to improve efficiency in Monte Carlo simulations. We therefore obtain an estimate of $\text{MSE}(1/\hat{\alpha})$, which can be minimized with respect to the choice of the threshold s_n. A full sample bootstrap of $\text{MSE}(1/\hat{\alpha})$, only generates s_n levels which converge in distribution. The weak law of large numbers does not operate on averages of the most extreme order statistics. To attain convergence in probability, we show that one needs to create resamples of smaller size than the original sample with a subsample bootstrap technique. By bootstrapping on the entire sample, one in essence recreates the full sample estimates, where the use of smaller resamples induces a weak law of large numbers effect. In some resamples of reduced sample sizes, intermediate order statistics become extreme observations, enabling the law of large numbers to take effect.

We first introduce power laws in Section 2, and then present the theoretical statistical analysis in Section 3, and a method for tail forecasting in Section 4. The empirical properties of our estimator are accessed in Section 5 by simulating from a number of heavy tailed distributions and stochastic processes where we are able to compare the Monte Carlo results with known analytic solutions. We then apply our technique in Section 6 to ranks–size relationships and estimate the risk in some financial return data sets. All proofs are contained in the Appendix.

2 Power Laws

Suppose we consider applications in market structure, such as the size ranking of corporations, personal wealth, or the population sizes of the world's largest cities. Alternatively, consider the estimation of the probability of large losses in financial markets.

The unifying element in each of these applications is that they depend on the accurate estimation of rank–order relationships. Furthermore, since in each case the tails of the underlying data is fat, extreme value theory is ideally suited the analysis of the largest outcomes. Unfortunately, the usual assumption of the log rank–order relationship being linear only holds asymptotically. Therefore, common estimators such as the least squares and Hill estimators discussed below are biased. The main contribution of this paper is the development of estimators for optimally estimating the tail shape of the rank–order distribution.

2.1 Unconditional Models: Pareto and Zipf

The earliest example of a formal analysis of rank–order relationships, and especially extreme realizations is by Pareto (1896) who considers income distributions. His approach is to use a static model, where he proposes a continuous distribution, now known as the *Pareto distribution*. Suppose that x is a random variable, then the Pareto distribution takes the form:

$$F(x) = 1 - ax^{-\alpha}.$$

Since, at the time, only the wealthiest individuals paid income tax, this is only a model of the upper tail and the highest incomes. The Pareto distribution has seen widespread applications in the analysis of incomes and other areas, e.g. by Champernowne (1953) and more recently by Chung and Cox (1994).

A discrete form of the Pareto distribution was proposed by Zipf (1949) who is concerned with the size of cities. For a city of population size x, the probability it is larger than some population size X is given by:

$$\Pr[x > X] = 1 - aX^{-\alpha}. \tag{1}$$

In this model, α can be estimated by the regression:

$$\log(y) = \mu + \alpha \log(x). \tag{2}$$

where x is size, and y rank.

In both the Pareto distribution and Zipf's law, the key coefficient is α, which determines the rate of decline of the tails. In the field of extreme value theory, α is known as the *tail index*.

2.2 Conditional Models: Gibrat

In an attempt to explain rank–order relationships, primarily in firm size, Gibrat (1931) proposes a stochastic growth model which now is called Gibrat's law. Suppose x_t is the size of a firm x at a given time, where size may be any measure of firm size, e.g. revenue or market capitalization. The evolution of x over time is then governed by:

$$X_t = c + \delta_t X_{t-1}, \tag{3}$$

where δ_t is a random growth factor, and c is a constant. In general, δ_t is not conditional on firm size, but may be dependent on other economic variables, including the size of competing firms. As a result, the rank order relationship among a group of competing firms is conditional on their relative sizes. While these stochastic growth models do not incorporate any economic optimization, subsequent literature has proposed optimizing models of market structure, see Sutton (1997) for a survey.

Several authors have noted that the implied distribution of random variables generated by an equation like (3) is the Pareto distribution. Champernowne (1953) obtains this result for a specific case, while Ijiri and Simon (1977) exploit this further, and Gabaix (1999) relates Gibrat's Law to Zipf's law. Kesten (1973) demonstrates in full generality, under which conditions the stochastic growth model (4) has a stationary distribution with Pareto tails as a solution. He does this by showing that that power law declines, as in the Pareto's distribution, stem from the stochastic nature of the coefficients in a linear difference stochastic equation, now known as the Kesten equation:

$$X_t = \epsilon_t + \gamma_t X_{t-1}, \quad \epsilon_t, \gamma_t \geq 0, \tag{4}$$

where $\{\gamma_t, \epsilon_t\}$ are IID, and $\exists \kappa > 0$ such that $E[\gamma_t^\kappa] = 1$. Then $X_t \overset{D}{\rightarrow} X_\infty$ and $\exists c > 0$ such that as $s \to \infty$ the $\Pr[X_\infty > s] \sim cs^{-\kappa}$.[1]

There are two common and important assumptions embedded in the Gibrat–Zipf models: First, the constancy of α. This is a common assumption in the economic growth literature, and has been applied to city sizes by e.g. Glaeser et al. (1995). The second assumption, i.e. that α equals one, is less obvious. Ijiri and Simon (1977) argue that this is natural in special cases, and Gabaix (1999) argues that under some conditions this follows naturally from Gibrat's law in the limit.

An important side effect of α equaling one, is that it implies that all moments of the underlying distribution are undefined. Only moments lower than α are defined. Hence research assuming that $\alpha = 1$ necessarily cannot use the mean and variance in its analysis.

2.3 Extreme Value Theory

Consider a set of random variables, $X_1, \ldots, X_n$, where we are only interested in the largest outcomes. In a fundamental result, reminiscent of the Central Limit Theorem, it can be established that the distribution of appropriately scaled extreme outcomes converges to one of three particular distributions, irrespective of the actual distribution of X. Of these three distributions, one, the Weibull, has a finite endpoint, while the second, the Gumbel, has exponentially declining tails and includes the normal in its domain of attraction as its most prominent member. The last limit distribution is the Fréchet, which takes the form $\exp[-(-x)^\alpha]$, where α is known as the *tail index*. The Pareto distribution (1) is the most prominent member in the domain of attraction of the Fréchet class. If the Pareto holds exactly, a maximum likelihood estimator of the tail index is the Hill estimator which is equivalent to the linear regression (2). As a result, the tail index is the same statistic as the power in the power law in Gibrat's and Zipf's laws.

[1] There are some other conditions, see e.g. de Haan et al. (1989) or Embrechts et al. (1997).

2.4 Financial Risk Analysis

A different application of power laws is the estimation of financial risk. Because financial returns are well–known to be heavy tailed, the relevant limit distribution for large outcomes of financial return processes is the Fréchet, implying that financial returns have Pareto type tails. As a result, EVT has become a common technique for forecasting extreme financial risk. As such, financial risk analysis and the Gibrat–Zipf laws share important traits, especially in statistical analysis, and many of the estimation issues are the same.

2.5 What about Dependence?

There is no reason to believe that any of the data types considered here, be they city sizes, firm sizes, or financial returns, are independent. Fortunately, the analysis does not rely on the assumption of independence.

The fundamental result from which the power laws are derived, i.e. the Kesten equation, (4) explicitly assumes that the underlying data is not independent. It is well known that the power laws hold under common forms of dependence, i.e., ARMA. In the specific case of financial returns, de Haan et al. (1989) demonstrate that an ARCH process is one form of the Kesten equation, implying Pareto tails. Several authors, e.g. Hsing (1991), and Resnick and Stărică (1996) show that the Hill estimator is a consistent estimator under ARMA and ARCH type dependent processes.

2.6 Which observations to Include?

If the Pareto distribution holds exactly, then one should use all available observations in order to estimate α efficiently. Unfortunately, this will only happen in special circumstances. In general, the Pareto approximation only holds asymptotically and both the least squares (LS) estimator (2) and the Hill estimator will be biased.

Several authors, (e.g. Mansfield, 1962; Gabaix, 1999; Ijiri and Simon, 1977) note that power laws do not hold exactly for all observations, and that some minimum threshold for firm or city sizes needs to be established. Ijiri and Simon (1977) further note the specific form of the bias in the LS estimator of (2), and attempt to correct for the bias by including non–linear terms in the regression. This is precisely where EVT makes a valuable contribution in understanding the power laws. The reason is that by using EVT, one can quantify the deviation from the Pareto distribution and derive optimal estimators of the tail index and hence the Gibrat and Zipf coefficients. This is possible because one can measure the increase in bias as the sample includes more observations away from the maxima, and the increased variance as the sample size decreases. A key objective of this paper is the development of technique to measure the magnitudes of the of the variance and bias in order to offset them optimally and hence obtain an optimal estimate of the tail index.

3 Estimation Theory

A key difficulty for the statistical implementation of extreme value theory (EVT) is the determination of how many observations to use in the estimation of the tail index, α. In the special case when the tails follow the Pareto distribution, the Hill estimator can be used with all available observations. In that case the Hill estimates attain the Cramer–Rao lower bound. In contrast, the focus on a larger class of distribution functions which only the tails have a Pareto shape. To this end consider the class of distribution functions which satisfy:

$$F(x) = 1 - ax^{-\alpha}\left[1 + bx^{-\beta} + o\left(x^{-\beta}\right)\right], \quad \beta > 0, \quad \text{as } x \to \infty. \tag{5}$$

This class covers such diverse distributions as the Student-t, sum–stable, F, Burr, and the Fréchet. Due to the presence of the higher order terms, however, the Hill estimator will be biased, where the bias increases with the number of observations. As a result, the tail threshold must be estimated in such a way that the included higher order statistics balance the variance against the bias, which in turn necessitates estimating the higher order parameter β. We develop below an estimator for the optimal threshold s_n by using a double subsample bootstrap technique. The results are presented as a series of Theorems with all proofs relegated to the Appendix.

3.1 k-Moment Ratio Tail Index Estimators

Let $\{X_1, ..., X_n\}$ be a sample of size n of IID random values with common distribution $F(x)$ which satisfies (5). Consider the descending order statistics from this sample around a given threshold $s : X_{(1)} \geq \cdots \geq X_{(m)} > s \geq X_{(m+1)} \geq \cdots \geq X_{(n)}$. The set $\mathcal{A}$ is the set of order statistics which strictly exceed s :

$$\mathcal{A}(n, s, m) = \left\{ X_{(1)}, ..., X_{(m)} \middle| X_{(i)} > s, \; i = 1, ..., n \right\}.$$

The number of elements in $\mathcal{A}$ will be random if the threshold s is fixed before sampling; the number of elements is then denoted by capital M or $M(s)$ to reveal its dependence on the threshold choice. Also, we will use the indicator function $\mathcal{I}(X_i \in \mathcal{A})$ which takes the value of 1 when $X_{(i)}$ is in $\mathcal{A}$ and 0 otherwise.

Define the conditional $k-$th order log empirical conditional moment from a sample $X_1, ..., X_n$ of n IID draws from $F(x)$ as follows:

$$u_k(s) \equiv \frac{1}{M} \sum_{i=1}^{n} \mathcal{I}(X_i \in \mathcal{A}) \left(\log \frac{X_i}{s} \right)^k, \tag{6}$$

and where $u_k(s) = 0$ if $\mathcal{A}$ is empty. Our interest in this statistic is motivated by the following obvious result, which is stated without proof, and used repeatedly in the subsequent analysis:

Lemma 1

$$\alpha \int_s^\infty \left(\log \frac{x}{s} \right)^k x^{-\alpha-1} dx = k! \alpha^{-k} s^{-\alpha}$$

From Lemma 1 it is immediate that the log empirical conditional moments are bounded in mean if $F(x)$ adheres to (5), even though the k-th mean of X_i is unbounded. Lemma 1 also motivates a class of estimators for the tail index α. Daníelsson et al. (1996) introduced the following class of estimators for the inverse of the first order tail index, $1/\alpha$:

Definition 2 *The k-moment ratio estimator, denoted as $w_k(s)$, for the inverse tail index is*

$$w_k(s) \equiv \widehat{1/\alpha} = \frac{u_k(s)}{k u_{k-1}(s)}, \tag{7}$$

where $k = 1, 2, ...$ are integer valued, and $u_0(s) = 1$.

The specific case where $k = 1$ is the Hill estimator whose properties are documented by e.g. Hall (1982) or Goldie and Smith (1987). The Hill estimator is a special case off the class of moment ratio estimators.

Hall (1982) and Goldie and Smith (1987) obtain the moment properties of the Hill statistic w_1. We extend the proofs to the general case of w_k. Define the statistic $v_k(s)$:

$$v_k(s) \equiv \frac{1}{\alpha} + \alpha^{k-1} \left(\frac{u_k}{k!} - \frac{1}{\alpha^k} \right) - \alpha^{k-2} \left(\frac{u_{k-1}}{(k-1)!} - \frac{1}{\alpha^{k-1}} \right). \tag{8}$$

We first obtain the mean and variance for this linearized statistic. In Theorem 7 we subsequently show that the limiting distribution of $w_k(s)$ as $n \to \infty$ has the same mean and variance. Note that the choice of the threshold s is a function of the sample size n. This is indicated by writing s_n, though the notation is sometimes suppressed for brevity.

Theorem 3 *For the class of random variables with density that satisfies (5), and letting $s_n^\alpha/n \to 0$, $s_n \to \infty$ as $n \to \infty$ the bias of the linearized statistic v_k is*

$$\mathrm{E}_{\mathcal{A}}\left[v_k(s_n) - \frac{1}{\alpha}\right] = -\frac{b\beta\alpha^{k-2}}{(\alpha+\beta)^k} s_n^{-\beta} + o\left(s_n^{-\beta}\right). \tag{9}$$

Theorem 4 *For the class of random variables with density that satisfies (5); and letting $s_n^\alpha/n \to 0$, $s_n \to \infty$ the variance of the linearized statistic v_k is*

$$\mathrm{Var}_{\mathcal{A}}\left[v_k(s_n) - \frac{1}{\alpha}\right] = \frac{s_n^\alpha}{an}\frac{\kappa(k)}{\alpha^2} + o\left(\frac{s_n^\alpha}{n}\right), \tag{10}$$

where

$$\kappa(k) = \frac{(2k)!}{(k!)^2} + \frac{(2k-2)!}{((k-1)!)^2} - 2\frac{(2k-1)!}{k!(k-1)!}. \tag{11}$$

The first few values of the $\kappa(k)$ function are given below. Note the rapid increase as k increases. [2]

k	1	2	3	4	5	6	7
$\kappa(k)$	1	2	6	20	70	252	924

3.2 Optimal Theoretical Choice of s

By combining the variance (10) and the bias squared (9), we obtain the mean squared error (MSE) of the linearized statistic $v_k(s)$:

$$\mathrm{MSE}(v_k(s_n)) = \frac{\kappa(k)}{a\alpha^2}\frac{s_n^\alpha}{n} + \frac{b^2\beta^2\alpha^{2k-4}}{(\alpha+\beta)^{2k}} s_n^{-2\beta} + o\left(\frac{s_n^\alpha}{n}\right) + o\left(s_n^{-2\beta}\right). \tag{12}$$

From (12) we see that there is a delicate balance between bias and variance depending on how fast $s_n \to \infty$ as $n \to \infty$. Either the bias part or variance part dominates, or they just balance, (see Hall, 1982). If the objective is to minimize the asymptotic mean squared error (AMSE), then it is optimal to let the threshold s depend on the sample size n in such a way that both leading terms in (12) vanish at the same rate. The AMSE is given by the leading terms from (12):

$$\mathrm{AMSE}(v_k(s_n)) = \frac{\kappa(k)}{a\alpha^2}\frac{s_n^\alpha}{n} + \frac{b^2\beta^2\alpha^{2k-4}}{(\alpha+\beta)^{2k}} s_n^{-2\beta}.$$

From the first order condition $\partial\,\mathrm{AMSE}/\partial s_n = 0$, one determines the optimal rate by which $s_n \to \infty$ as $n \to \infty$ such that the two terms in the AMSE are balanced and the AMSE vanishes at the best possible rate.

[2] The results above are easily extended to non–integer log moments, but this is not needed for our analysis.

Theorem 5 *For n sufficiently large, the unique AMSE minimizing asymptotic threshold $\overline{s}_n$ is*

$$\overline{s}_n(v_k) = \left[\frac{2ab^2\beta^3\alpha^{2k-3}}{(\alpha+\beta)^{2k}\kappa(k)}\right]^{\frac{1}{\alpha+2\beta}} n^{\frac{1}{\alpha+2\beta}}, \tag{13}$$

and the associated minimal MSE of $v_k(s_n)$ equals

$$\overline{\mathrm{MSE}}(v_k(\bar{s}_n)) = \frac{\kappa(k)}{a\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2ab^2\beta^3\alpha^{2k-3}}{(\alpha+\beta)^{2k}\kappa(k)}\right]^{\frac{\alpha}{2\beta+\alpha}} n^{-\frac{2\beta}{2\beta+\alpha}} + o\left(n^{-\frac{2\beta}{2\beta+\alpha}}\right). \tag{14}$$

The use of bars in the notation in (13–14) reflect that the chosen s_n sequence minimizes the AMSE. Hall and Welsh (1984) show that this rate cannot be improved upon by other estimators. Since $1 - F(s) = as^{-\alpha}\left[1 + O\left(s^{-\beta}\right)\right]$, the following result for the number of upper order statistics is immediate:

$$n^{-\frac{2\beta}{2\beta+\alpha}} M(\bar{s}_n) \to a\left[\frac{2ab^2\beta^3\alpha^{2k-3}}{(\alpha+\beta)^{2k}\kappa(k)}\right]^{-\frac{\alpha}{\alpha+2\beta}} \tag{15}$$

with probability 1 as $n \to \infty$.

From (12–14) it follows that if s_n tends to infinity at a rate below $n^{1/(2\beta+\alpha)}$, the bias part in the MSE dominates, while conversely the variance part dominates if s_n tends to infinity more rapidly than $n^{1/(2\beta+\alpha)}$. For the class $v_k(s_n)$ we show that on the basis of the AMSE criterion the only two elements of interest are $v_1(s)$ and $v_2(s)$.

Theorem 6 *The $v_1(s)$ and $v_2(s)$ statistics are the only two estimators in the class $v_k(s), k = 1, 2, 3...$, which are not dominated, in the sense of the AMSE criterion, for all $\beta/\alpha \in \mathbb{R}^+$ combinations.*

Because $1+\frac{\beta}{\alpha} \gtreqless 2^{\frac{\beta}{\alpha}}$ as $\alpha \gtreqless \beta$, we find that for $\beta > 0$, when the $v_1(s)$ and the $v_2(s)$ statistics are each evaluated at their own asymptotic MSE minimizing thresholds denoted as $\bar{s}_1$ and $\bar{s}_2$ respectively that:

$$\overline{\mathrm{AMSE}}(v_1(\bar{s}_1)) \gtreqless \overline{\mathrm{AMSE}}(v_2(\bar{s}_2)) \quad \text{as} \quad \alpha \gtreqless \beta.$$

This implies that if v_2 has a lower AMSE than v_1, then v_1 is asymptotically more biased than the v_2.

The asymptotic distribution of w_k from 7 is given by:

Theorem 7 *Suppose we choose $s_n = \bar{s}_n(v_k)$ from (13). Then as $n \to \infty$*

$$\sqrt{M}\frac{(\alpha w_k(\bar{s}_n) - 1)}{\sqrt{\kappa(k)}} \xrightarrow{D} N\left(-\sqrt{\frac{\alpha}{2\beta}}\,\mathrm{sign}(b), 1\right)$$

The estimator $w_k(s_n)$ is conditional on the choice of s_n. If the asymptotically optimal threshold $\bar{s}_n$ can be estimated by $\widehat{s}_n$ such that $\widehat{s}_n/\bar{s}_n$ converges in probability to 1, then the asymptotic normality of $w_n(\bar{s}_n)$ also applies to the case where $\bar{s}_n$ is replaced by its estimated value.

The limiting normal distribution has a mean which depends on the unknown nuisance factor $\sqrt{\alpha/2\beta}\,\mathrm{sign}(b)$. If this latter factor can be estimated consistently, however, then it follows from Slutsky's theorem that the claim of the asymptotic normality of $w_k(\bar{s}_n)$ also applies to the case where the bias factor is estimated.

We can extend upon the convergence in distribution in Theorem 7 by considering convergence in moments. Convergence in moments of $w_1(\bar{s}_n)$ is straightforward. For $w_2(\bar{s}_n)$ convergence in moments is given by:

Lemma 8 *Suppose $F(x)$ in (5) is exactly Pareto. Then as $s_n \to \infty$, $s_n^\alpha/n \to 0$ when $n \to \infty$*

$$\lim_{n\to\infty} E\left[w_2(\bar{s}_n)\right] = \alpha^{-1}$$

Lemma 8 extends to the general case in (5), e.g. if $b < 0$ the distribution function $\exp(-\alpha t)$ bounds the conditional probability of $\log(X/s_n)$. Moreover, a similar proof can be used to obtain convergence in the second moment.

The $w_k(s)$ estimators are also consistent under various forms of dependency. It can be shown that regular variation is preserved under ARMA type dependence. de Haan et al. (1989) prove that the unconditional distribution of ARCH processes satisfy the regular variation property. Hsing (1991) and Resnick and Stărică (1996) show that the Hill estimator is a consistent estimator under respectively ARMA and ARCH type dependent processes by using point process techniques.

An alternative definition of the statistic $w_k(s_n)$ is to make the number of highest order statistics in set $\mathcal{A}$ a deterministic sequence m_n, and to identify the threshold as the m_n+1-th highest order statistic X_{m_n+1}. This switches the randomness from M to the thresholds. Goldie and Smith (1987) argue that the two formulations are essentially equivalent. Hall (1982) is based in the alternative interpretation and derives a theorem analogous to Theorem (7) for the $w_1\left(X_{(\bar{m}+1)}\right)$ statistic, where $\bar{m}_n = O\left(M_n(\bar{s}_n)\right)$.

3.3 Estimation of the Tail Threshold s_n

Hall (1990) proposes a bootstrap procedure for estimating the tail threshold in the special case where $\alpha = \beta$. His method has two important limitations. First, it only yields convergence in distribution, implying that statistics depending on s_n do not converge in probability. Second, it assumes that $\alpha = \beta$ which rules out relevant classes of distributions, such as the Student-t. We propose a technique addressing both issues.

Before we can state our main results, we need to provide some preliminary results. In addition, we establish some intuition behind our result.

For the case $k = 1$ the standard bootstrap equivalent of the expectation

$$\mathrm{MSE}(w_1) = \mathrm{E}\left[\left(w_1\left(s_n\right) - \frac{1}{\alpha}\right)^2\right]$$

is

$$\frac{1}{R}\sum_{r=1}^{R}\left[\left(w_{1,r}\left(s_n\right) - \left(\frac{1}{\alpha}\right)\right)^2\right], \tag{16}$$

where $w_{1,r}$ is calculated on a bootstrap resample of the original sample and R is the number of bootstrap resamples. The problem is, however, that the benchmark $1/\alpha$ for the MSE is unknown. We suggest replacing $1/\alpha$ by one of the other available estimators as the benchmark

$$\frac{1}{R}\sum_{r=1}^{R}\left[\left(w_{1,r}\left(s_{n_1}\right) - \widetilde{w}_1\left(\tilde{s}_n\right)\right)^2\right],$$

and where $\widetilde{w}_1\left(\tilde{s}_n\right)$ is some consistent alternative estimator like $w_2(s_n)$, which has the same convergence speed as $w_{1,r}\left(s_{n_1}\right)$, but a different multiplicative constant. In effect, $w_{1,r}\left(s_{n_1}\right) - \widetilde{w}_1\left(\tilde{s}_n\right)$ is an estimator of zero. We therefore replace $w_{1,r}\left(s_{n_1}\right)$ with a statistic for which the true value is known, i.e., is independent of α, but with an AMSE that has the same convergence rate. It does, however, have a different multiplicative constant, as the AMSE of the w_1 statistic, and we correct later in the procedure for the multiplicative constant. This is reminiscent of the use of control variates for variance reduction in Monte Carlo estimation.

In particular, we focus on the statistic:

$$z\left(s_n\right) \equiv w_2\left(s_n\right) - w_1\left(s_n\right). \tag{17}$$

We showed earlier the consistency of all w_k statistics as estimators of $1/\alpha$ for $s_n = \bar{s}_n = c_k n^{\frac{1}{2\beta+\alpha}}$. It follows that the $z\left(\bar{s}_n\right)$ statistic converges to 0 as $n \to \infty$. We now show that the AMSE$\left(z\right)$ has the same order of magnitude as the AMSE$\left(w_k\right)$. To this end define the linearized counterpart to $z\left(s_n\right)$, using (8):

$$q\left(s_n\right) \equiv \frac{1}{\alpha} - 2u_1\left(s_n\right) + \frac{\alpha}{2}u_2\left(s_n\right).$$

We can now establish:

Theorem 9 *For the class of random variables where the density satisfies (5), then* $s_n \to \infty$, $s_n^{\alpha}/n \to 0$ *as* $n \to \infty$

$$\mathrm{E}_{\mathcal{A}}\left[q\left(s_n\right)\right] = \frac{b\beta^2}{\alpha\left(\alpha+\beta\right)^2}s_n^{-\beta} + o\left(s_n^{-\beta}\right),$$

and

$$\mathrm{Var}_{\mathcal{A}}\left[q\left(s_n\right)\right]=\frac{1}{\alpha^2}\frac{s_n^\alpha}{an}+o\left(\frac{s_n^\alpha}{n}\right).$$

Corollary 10 *For n sufficiently large the* $\mathrm{AMSE}\left(q\right)$ *minimizing asymptotic threshold level* $\bar{\bar{s}}_n\left(q\right)$ *is*

$$\bar{\bar{s}}_n\left(q\right)=\left(\frac{2ab^2\beta^5}{\alpha\left(\alpha+\beta\right)^4}\right)^{\frac{1}{2\beta+\alpha}}n^{\frac{1}{2\beta+\alpha}}. \tag{18}$$

By comparing $\bar{s}_n\left(v_k\right)$ from (13) with the $\bar{\bar{s}}_n\left(q\right)$ from (18) we see that

$$\frac{\bar{\bar{s}}_n\left(q\right)}{\bar{s}_n\left(v_k\right)}=\left(\frac{\beta^2\left(\alpha+\beta\right)^{2k-4}}{\alpha^{2k-2}}\kappa\left(k\right)\right)^{\frac{1}{2\beta+\alpha}}.$$

Hence the two threshold values only differ with respect to their multiplicative constants, but increase at the same rate with respect to the sample size n. From Corollary 10 we have that the asymptotic MSE(q) is minimized by $\bar{\bar{s}}_n\left(q\right)$. By the analogue to Theorem 7 for the $z\left(s_n\right)$ statistic in (18), $\bar{\bar{s}}\left(q\right)$ then also asymptotically minimizes the MSE of the $z_n\left(s_n\right)$ statistic.

However, the practical problem of finding this value remains. One possibility is a bootstrap of $z^2\left(s_n\right)$, i.e. calculate the bootstrap average $\left(1/R\right)\sum_r^R z_r^2\left(s_n\right)$, and minimize this average with respect to s_n. Before we do this, we have to worry about convergence in the correct mode. The full sample bootstrap does not produce an estimate which is asymptotic to $\bar{\bar{s}}_n\left(q\right)$. In the proof to Theorem 7 we showed that $\sqrt{M}u_k$ is asymptotically normally distributed. By the Taylor expansion from the proof to Theorem 9 it then readily follows that $\sqrt{M}z$ is also asymptotically normally distributed. Hence Mz^2 is asymptotic to a $\chi^2_{(1)}$ distributed random variable. The mean of Mz^2 is easily shown to be asymptotic to $\left(\bar{\bar{s}}_n\right)^\alpha/n$ times the value of the $\mathrm{AMSE}\left(z\right)$ as given in (36). But because Mz^2 only converges in distribution, the average of the full sample bootstrap values Mz_r^2 has the same distributional properties as Mz^2. To show this, use the log–linearity of the u_1 and u_2 in the data and consider the linearized statistic $q\left(s\right)$.

To be able to back out the optimal rate, we need convergence in probability. Moreover, for the practical implementation of bootstrap procedure we need in addition convergence in moments. We now show that the desired convergence can be obtained through the subsample bootstrap procedure. Before we turn to the proof, we discuss the main intuition.

Consider the Hill estimator

$$w_1\left(s_n\right)=\frac{1}{M}\sum_1^M Y_i\left(s_n\right),$$

where as in Lemma 8 $Y_i(s_n) \equiv \log\left(X_{(i)}/s_n\right)$, $X_{(i)} > s_n$. Suppose we bootstrap this statistic. The bootstrap average of the Hill statistic is:

$$\frac{1}{R}\sum_r^R w_{1,r}(s_n) = \frac{1}{R}\sum_r^R \frac{1}{M_r}\sum_i^{M_r} Y_{i,r}(s_n).$$

The bootstrapped statistic $w_{1,r}(\bar{s}_n)$ evaluated at the optimal threshold value $\bar{s}_n$ is therefore an average from the set

$$\left\{Y_{(1)}(\bar{s}_n), ..., Y_{(M)}(\bar{s}_n)\right\}.$$

It follows by the weak law of large numbers that as the number of samples R increases

$$\frac{1}{R}\sum_r^R \frac{1}{M_r}\sum_i^{M_r} Y_{i,r}(\bar{s}_n) \xrightarrow{\text{P}} \frac{1}{T}\sum_i^T Y_{(i)}(\bar{s}_n) \text{ as } R \to \infty.$$

By Theorem 7 we know that for $s_n = \bar{s}_n$ from (13), $\left(1/\sqrt{M}\right)\sum_i^M Y_i(\bar{s}_n)$ is asymptotically normally distributed. Now suppose that s_n is not of the order $n^{\frac{1}{2\beta+\alpha}}$, cf. (13). It follows from (12) that either the bias dominates asymptotically, if $s_n = o\left(n^{\frac{1}{2\beta+\alpha}}\right)$, or that the variance dominates, if $n^{1/2\beta+\alpha} = o(s_n)$, see Hall (1982). Consider taking subsample resamples of size n_1 such that $n_1 = O\left(n^{1-\varepsilon}\right)$, where $0 < \varepsilon < 1$. Let $\bar{s}_{n_1} be$ the AMSE minimizing threshold level for the sample size n_1 so that $n_1 \sim n^{1-\varepsilon}$, $\bar{s}_{n_1} = o\left(n^{\frac{1}{2\beta+\alpha}}\right)$. Since $\bar{s}_n > \bar{s}_{n_1}$, the bootstrapped statistic $w_{1,r}(\bar{s}_{n_1})$ evaluated at the subsample optimal threshold value $\bar{s}_{n_1}$ is therefore an average from the larger set

$$\left\{Y_{(1)}(\bar{s}_{n_1}), ..., Y_{(M)}(\bar{s}_{n_1}), Y_{(M+1)}(\bar{s}_{n_1}), ..., Y_{(T)}(\bar{s}_{n_1})\right\}.$$

As before, it follows that as the number of subsamples R increases

$$\frac{1}{R}\sum_r^R \frac{1}{M_r}\sum_i^{M_r} Y_{i,r}(\bar{s}_{n_1}) \xrightarrow{\text{P}} \frac{1}{T}\sum_i^T Y_{(i)}(\bar{s}_{n_1}) \text{ as } R \to \infty \tag{19}$$

Define the bootstrap statistic

$$Q(\bar{s}_{n_1}) = \sqrt{m(\bar{s}_{n_1})}[\frac{1}{T_n}\sum_i^{T_n} Y_{(i)}(\bar{s}_{n_1}) - \frac{1}{\alpha}], \tag{20}$$

where $m(\bar{s}_{n_1}) = O(n^{(1-\varepsilon)\frac{2\beta}{2\beta+\alpha}})$ cf. (15). As was shown by Hall (1982, th. 2, part 3), for $\varepsilon > 0$, so that $n_1 < n$, $\bar{s}_{n_1} < \bar{s}_n$ then $Q(\bar{s}_{n_1}) \to 0$ in probability. Hence this is also applies to the left hand side of (19) after subtracting $1/\alpha$ and premultiplication with $\sqrt{m(\bar{s}_{n_1})}$. The idea is thus that subsample bootstrap averages conditional on the subsample optimal threshold value are comparable to the corresponding full sample statistic evaluated at a smaller threshold than $\bar{s}_n$. But conditional on this smaller threshold value, the full sample statistic converges in probability rather than in distribution. This embedding idea is the essence of the proof to our main result.

Theorem 11 *Suppose $F(x)$ satisfies (5). Let $n_1 = O(n^{1-\varepsilon})$ for some $0 < \varepsilon < 1$ be the bootstrap resample size. For given n let $R \to \infty$ and determine $\widehat{s}_{n_1}$ such that*

$$\frac{1}{R}\sum_{r}^{R}\left[z_r\left(\widehat{s}_{n_1}\right)\right]^2$$

is minimal. Then, as $n \to \infty$

$$\widehat{s}_{n_1}(z)/\overline{\overline{s}}_{n_1}(q) \xrightarrow{P} 1$$

Note, $\overline{\overline{s}}_{n_1}(q)$ was defined in (18).

It is straightforward to show that for given sample size for $s_1 \in \left(0, \overline{\overline{s}}_{n_1}(q)\right)$ the AMSE from (38) is monotonic and declining in s_1. Also, for $s_1 = o\left(\overline{\overline{s}}_{n_1}(q)\right)$ and $s_1 > \overline{\overline{s}}_{n_1}(q)$, the right hand side of (38) is monotonic and increasing. The monotonicity result implies that $\overline{\overline{s}}_{n_1}(q)$ can be located by searching for the minimum to

$$\frac{1}{R}\sum_{r}^{R}\left[z_r(s_1)\right]^2$$

as s_1 is increased from zero. For this procedure to work we need in fact convergence in the 2^{nd} moment of z . Thus in practice we minimize the second empirical moment of the z-statistic, rather than minimizing the q-statistic. Therefore to ensure that the convergence in probability from Theorem 11 can be used, convergence in the mean of $(m(s_1)/R)\sum_r (z_r(\widehat{s}_{n1}))^2$ is needed. By the argument that was used to prove Lemma 8, this convergence immediately follows; we omit an explicit proof for the sake of brevity.

An analogous procedure, and proof applies to the interpretation of the Hill statistic with a fixed number of excesses and a random threshold that case we also have that $\hat{m}_{n_1}(z)/\overline{\overline{m}}_{n_1}(z) \to^{\mathrm{P}} 1$, and

$$\overline{\overline{m}}_n(z) = a\left(\frac{2ab^2\beta^5}{\alpha(\alpha+\beta)^4}\right)^{-\frac{\alpha}{2\beta+\alpha}} n^{\frac{2\beta}{2\beta+\alpha}}. \tag{21}$$

While it is more expedient to present the theoretical derivations in terms of the threshold interpretation, however, in practice the minimization of the bootstrapped

MSE (z) is done in terms of the index m. We minimize MSE (z) by stepwise increases in the lowest extreme order statistic that is used as the stochastic threshold in the calculation of the estimators.

In the end we are interested in the optimal $\bar{m}_n(w_k)$ instead of $\overline{\overline{m}}_n(z)$ from (21). These two quantities are related as follows, recall (15):

$$\frac{\overline{\overline{m}}_n(z)}{\bar{m}_n(w_k)} = \left[\left(\frac{\beta}{\alpha}\right)^2\left(1+\frac{\beta}{\alpha}\right)^{2k-4}\kappa(k)\right]^{-\frac{1}{1+\frac{2\beta}{\alpha}}}. \tag{22}$$

Hence, a conversion from $\hat{m}_n(z)$ to $\hat{m}_n(w_k)$ requires a consistent estimate of the ratio of the first and second order tail parameters β/α. The following result exploits the fact that $\overline{\overline{m}}_{n_1}(z)$ varies regularly.

Theorem 12 *A consistent estimator for β/α is*

$$\widehat{\beta/\alpha} = \frac{\log \hat{m}_{n_1}(z)}{2\log n_1 - 2\log \hat{m}_{n_1}(z)}. \tag{23}$$

Theorem 12 in combination with (22) implies that

$$\hat{m}_{n_1}(w_2) = \hat{m}_{n_1}(z)\left[\sqrt{2}\frac{\log \hat{m}_{n_1}(z)}{2\log n_1 - 2\log \hat{m}_{n_1}(z)}\right]^{\frac{2\log n_1 - 2\log \hat{m}_{n_1}(z)}{\log n_1}} \tag{24}$$

is a consistent estimator for $\overline{\overline{m}}_{n_1}(w_2)$. Similar expressions can be obtained for $\hat{m}_{n_1}(w_1)$. But these estimators do not exploit all the information which is available in the full sample, because these are restricted to the subsample size n_1.

The second conversion we need is to go from $\hat{m}_{n_1}(w_2)$ to $\hat{m}_n(w_2)$. Theorem 11, (13), and (15) give:

Corollary 13 *Suppose Theorem (11) applies, then*

$$\frac{\hat{m}_{n_1}(w_k)}{\overline{\overline{m}}_n(w_k)}\left(\frac{n}{n_1}\right)^{\frac{2}{2+\alpha/\beta}} \xrightarrow{P} 1 \tag{25}$$

One might consider using relation (25) as an equality and to replace α/β in the exponent by $\widehat{\alpha/\beta}$ from (23). Unfortunately, even though the $\widehat{\beta/\alpha}$ estimates in (23) is consistent, its rate of convergence is unknown. This frustrates using $\widehat{\beta/\alpha}$ in (25) because α/β appears in the exponent (and hence its convergence rate may be too slow, i.e. less than $\varepsilon \log n$). A solution is to do a second bootstrap on an even further reduced subsample size n_2, and to choose n_2 such that the multiplicative factor in (25) can be replaced by a known value.

Theorem 14 *Let* $n_1 = O(n^{1-\varepsilon})$ *for some* $0 < \varepsilon < 1/2$ *and choose* $n_2 = n_1^2/n$. *Suppose* $\hat{m}_{n_2}(z)$ *is the consistent estimator of* $\bar{m}_{n_2}(z)$ *(21), from the subsample bootstrap procedure on subsample resamples of size* n_2. *Then*

$$\frac{(\hat{m}_{n_1}(z))^2}{\overline{\overline{m}}_n(z)\,\hat{m}_{n_2}(z)} \xrightarrow{P} 1. \tag{26}$$

Combine result (26) with (24) to arrive at the *consistent estimator*

$$\hat{m}_n(w_2) = \frac{(\hat{m}_{n_1}(z))^2}{\hat{m}_{n_2}(z)}\left[\sqrt{2}\frac{\log m_{n_1}(z)}{2\log n_1 - 2\log \hat{m}_{n_1}(z)}\right]^{\frac{2\log n_1 - 2\log \hat{m}_{n_1}(z)}{\log n_1}}. \tag{27}$$

The other variants like $\hat{m}_n(w_1)$ follow easily.

The s_n, or m_n, and β/α can therefore be estimated by a double subsample bootstrap procedure which rests on a choice for the subsample sizes $n_1 = n^{1-\varepsilon}$, where $\frac{1}{2} > \varepsilon > 0$, and $n_2 = n_1^2/n$. Asymptotically any n_1 such that $\frac{1}{2} > \varepsilon > 0$ yields a consistent estimate of α. Hence, asymptotic arguments provide little guidance in choosing between any of the n_1, which is desired for practical purposes. We propose the following criterion.

The basis for our estimator of α is the minimization of its AMSE. The subsample bootstrap yields estimates of the $\text{AMSE}(z_{n_1})$ and $\text{AMSE}(z_{n_2})$. By the same arguments as were used in the proof to Theorem 14, one can show that

$$\left[\widehat{\text{AMSE}(z_{n_1})}\right]^2/\widehat{\text{AMSE}(z_{n_2})} \tag{28}$$

is asymptotic to $\text{AMSE}(z_n)$. We then choose n_1 by

$$\arg\min_{n_1}\left[\widehat{\text{AMSE}(z_{n_1})}\right]^2/\widehat{\text{AMSE}\left(z_{n_2(n_1)}\right)}. \tag{29}$$

Choosing n_1 in this way keeps the estimated MSE to a minimum.

Finally, we need to consistent estimator of $\text{sign}(b)$ for the purpose of diagnosis. Recall the mean of $q(s_n)$ from Theorem 9:

$$\text{E}_{\mathcal{A}}[q(s_n)] = cs_n^{-\beta}\,\text{sign}(b) + o\left(s_n^{-\beta}\right),$$

where $c > 0$. This suggests the following consistent estimator

$$\widehat{\text{sign}(b)} = \text{sign}(z(s_n)), \text{ with } s_n < \overline{\overline{s}}_n. \tag{30}$$

Note that we choose $s_n < \overline{\overline{s}}_n$, or alternatively $m_n > \overline{\overline{m}}_n$, to guarantee that the bias asymptotically dominates the variance. We also experimented with the following estimator for $\text{sign}(b)$

$$\text{sign}([w_2 - w_1] - [w_4 - w_3]).$$

It is straightforward to check that this estimator is consistent as well.

4 Prediction of Extremes

A major application of extreme value theory is the estimation of borderline in–sample and out–of–sample probability–quantile (P, Q) combinations. We show that the properties of the quantile and tail probability estimators follow from the properties of $\widehat{1/\alpha}$.

Consider two excess probabilities p and t with $p < 1/n < t$, where n is the sample size. Associated with p and t are large quantiles x_p and x_t, where $x_p : 1 - F(x_p) = p$, and $x_t : 1 - F(x_t) = t$. Since $p < 1/n$, it is likely that $x_p > \max\{X_1, \ldots, X_n\}$. The quantile x_p can be estimated by extrapolating the empirical distribution function $F_n(x)$ by means of its regular variation properties. Using the expansion of $F(x)$ in (5) we use $F(x_p)/F(x_t)$ to obtain the following estimator. Ignore the higher order terms in the expansion, replace the probability t by the random variable M/n with x_t fixed at s_n. Substitute for $1/\alpha$ any $w_k(s_n)$ estimator. This gives

$$\hat{x}_p = x_t \left(\frac{M}{np} \right)^{w_k}. \tag{31}$$

We can now prove:

Theorem 15 *Suppose that the conditions of Theorem 7 and (11) do hold. In addition take $x_t = \widehat{s}_n$. Suppose that np_n converges to a constant τ which may be zero. Then the quantile estimator $\hat{x}_p$ is asymptotically normally distributed:*

$$\frac{\sqrt{M}}{\log(M/np)} \left(\frac{\hat{x}_p}{x_p} - 1 \right) / \sqrt{\kappa(k)} \sim N\left(-\frac{\operatorname{sign}(b)}{\sqrt{2\beta\alpha}}, \frac{1}{\alpha^2} \right).$$

An estimator for the reverse problem can be developed as well. By solving for p in (31) we get:

$$\hat{p} = \frac{M}{n} \left(\frac{x_t}{x_p} \right)^{\hat{\alpha}}. \tag{32}$$

Without proof, since it essentially duplicates the arguments used to prove Theorem 15, we state:

Theorem 16 *Under the same conditions as in Theorem 15, the excess probability estimator $\hat{p}$ is asymptotically normally distributed, that is*

$$\frac{\sqrt{M}}{\log(x_t/x_p)} \left(\frac{\hat{p}}{p} - 1 \right) / \sqrt{\kappa(k)} \xrightarrow{D} N\left(\alpha^2 \frac{\operatorname{sign}(b)}{\sqrt{2\alpha\beta}}, \alpha^2 \right)$$

Note that the asymptotic distributions of the normed quantiles and probabilities differ by a multiplicative factor of $-\alpha^2$. This is a Bahadur-Kiefer type result for out of sample (P, Q) combinations, (see Serfling, 1980). In other words, it does not matter from which axis one looks at the distance between the empirical distribution function and the distribution function, even if out–of–sample the empirical distribution function is replaced by the $(p, \hat{x}_p)$ or $(\hat{p}, x_p)$ curves.

5 Monte Carlo Experiments

In order to access the performance of our estimator for the tail threshold, we generate pseudo random numbers from several known distributions and stochastic processes. The criteria for choosing these particular distributions was that they should resemble observed heavy tailed economic variables such as the market capitalization, city size, and return variables used below. In all cases we know the true α, and in most cases the true threshold value and quantiles.

The distributions chosen are Student–t, the sum–stable–Levy, Extreme Type II, and the log Pareto which is defined as:

$$F(x) = 1 - x^{-\alpha}[1 + \alpha \log x].$$

The stochastic processes are the normal GARCH, Student-t MA which is an MA process with Student–$t_{(3)}$ innovations, and a Student-t stochastic volatility (SV) model. The Student(3) $\text{SV}_{(\beta,\gamma)}$ is defined as:

$$\begin{aligned} Y_t &= U_t W_t H_t^2, \quad , \quad \Pr[U_t = -1] = 0.5, \ \Pr[U_t = 1] = 0.5 \\ H_t &= \beta Q_t + \gamma H_{t-1}, \quad Q_t \sim N(0,1), \ \ \beta = 0.1, \ \gamma = 0.9 \\ W_t &= \sqrt{\frac{1-\gamma^2}{\beta^2}} \frac{\sqrt{3}}{\sqrt{Z_t}}, \quad Z_t \ \sim \chi_{(3)}. \end{aligned}$$

This process generates volatility clusters, and was designed in this specific way because it follows that the stationery marginal distribution function of Y_t is Student–$t_{(3)}$ distributed, and hence we can determine the theoretical second order parameters, which for the GARCH model are unknown. The sample size was 5,000, and the simulations were replicated 250 times, where the w_2 statistic was used in the estimation.

The estimates of the model parameters, i.e. $1/\alpha$, β/α, and the ratio of the estimated threshold to the true threshold $m_n/\overline{m}_n$, are reported in Table 1. The estimates of α are quite accurate, the lowest root mean squared error (RMSE) is 1.7% for the Type II Extreme(4) value distribution function, while highest RMSE values are 17% for the sum–stable (1.8). This is due to the fact that when the characteristic exponent of the Stable distribution equals 2, the stable law switches from being fat tailed to the normal distribution which has thin tails. Thus while $1/\alpha$ jumps at the left end from the open interval $(0.5, \infty)$ to 0, the estimator smoothly interpolates between 0.5 and 0. For the stochastic processes, the RMSE ranges from 6.6% for the MA process to 11% for one of the GARCH models. The estimates of β/α are not as precise as for α, since β is a second order parameter, and the extreme realizations are less informative about the second order behavior than the first order behavior. The optimal number of order statistics m_n is a function of the second order parameters, and hence it is not surprising to observe a similar behavior as for β/α.

The results in Table 2 are more interesting from an economic point of view since they are concerned with the estimation of out–of–sample probability—quantile (P, Q) combinations. We report the quantile estimate for the borderline in sample probability $p = 1/n$ and the out–of–sample probability $p = 1/3n$. For comparison, we also report what is known in the financial industry as "worst case analysis" where the maximum value is used as an estimator of the expected maximum. The main criteria for analysis is the average quantile estimates across simulations and the scale free coefficient of variation (CV). The quantile estimator $\hat{x}_p$ performs well when judged by the mean and the CV. The true mean is unknown in the GARCH case. The CV does not change much when moving from the in–sample $p = 1/n$ to the out–of–sample $p = 1/3n$. Moreover, the performance of the quantile estimator is fairly consistent across distributions and stochastic processes in terms of the CV. In contrast, the worst case analysis has considerably worse performance, e.g., it's CV's are consistently more that twice as large than their tail estimation counterparts.

5.1 Tail Determination and Rank Order Relationships

Typical estimates of α in the market structure literature are in the range of 1–2. In order to gauge the impact of choosing different threshold levels, we simulate 1,000 realizations from a Student–$t_{(1.5)}$ distribution, which has similar tail properties as some of our data sets. We plot the log rank–order relationship in Figure 1, and use two arbitrary threshold levels to estimate α with least squares using (2). In this case we know that the true value of α is 1.5, however, the two estimates give us 1.3 and 1.9, for tail sample sizes 50 and 250, respectively. In effect it is the nonlinearity of the tail which biases the estimation, illustrating the importance of optimally determining the tails.

6 Estimation

We apply our estimator of the tail distribution to two distinct problems, size–rank relationships and financial risk analysis.

6.1 Market Structure

We employ two different datasets to estimate the Zipf–Gibrat coefficients:

- Firm size: the 1000 largest U.S. corporations in 1999, ranked by revenue. The data was obtained from Fortune.com.
- City size: the largest cities in Russia, UK, India, Japan, Brazil, US, and China.

The results are shown in Table 3 and graphically in Figures 2 and 4.

For firm size we see that the optimal number of observations in the tail is 32, implying a tail index of 2.1. By using all observations in the regression we get an estimate of the tail index of 1.2.

The optimal number of observations in the tail, m_n for the city size data is 59, with a tail index of 2.0. However, when using all observations, the tail index is estimated as 1.0. Figure 3 shows what happens when we vary m_n from almost no observations (15) in the tail, to using almost all observations (1652). The estimates of α vary considerably, from 0.98 to 2.24, with the estimates generally declining as the tail sample size increases. This result is common in EVT estimation, giving rise to the term *Hill horror plot* for plots like Figure 3. If the data was exactly Pareto distributed, the Hill horror plot should be roughly linear and horizontal.

Gabaix (1999) estimates the Zipf coefficient (α) for 135 largest U.S. metropolitan areas in 1991, and finds a value of 1. Figure 1 in his paper which plots this dataset shows clear signs of non–linearity of the type discussed above. Our city size data (Figure 2) has the same features.

We present results from using two tail sample sizes, the first from our procedure, and the second with all observations. This results in a wide range of estimates of the Gibrat–Zipf coefficients. The reason is that the log rank–size is not linear but convex, as can be seen in Figures 2 and 4. The deviation from linearity is very clear, and it is obvious from the Figures that the choice of threshold strongly influences the estimate of α. Therefore, any choice of a tail threshold, except the one obtained here, will be less accurate than our estimates.

6.2 Forecasting Financial Risk

We estimate the risk in financial return series with four datasets; two with high frequency foreign exchange datasets and two with daily returns on the SP–500 stock index[3]. For each dataset, we compute a number of standard statistics. The

[3]The FX data was obtained from Olsen and Associates and is quotes on the USD–DM spot contract from October 1992 to September 1993. We also use the first and last 5,000 daily returns from the daily S&P 500 index over the period 1929 to 2000.

mean and standard error are annualized by using a factor of 250 and 52,558 for the stock index and FX datasets respectively. In addition the skewness, kurtosis, and the minimum log return are reported. Subsequently, we applied our estimation procedure to the lower tail of the data, and report estimates of parameters $1/\alpha$, β/α, and the quantiles $x_{1/n}$, and $x_{1/3n}$. Between brackets we give the 95% confidence band.

The results are reported in Table 4. We note that the FX tails are fairly symmetric, while the S&P tails are more asymmetric, reflecting the great depression and the 1987 crash. The tail index estimates are close to 3. From an economic point of view the interesting estimates are the quantile estimates. Table 4 reveals that the risk of investing in stocks has come down over time, and that the risk in the FX market is fairly stable over the period of one year. It is especially interesting to note the difference in the risk forecasts in the 1929–1946 period compared to 1981–2000. Even if the realized maximum loss is lower in the first period, the forecasts give a different picture. In the first period we expect larger losses than predicted by the realized minima, while in the second period this result reverses. In other words, our forecast indicates that the 1987 crash had a probability much lower than 1/5000, the probability forecast that obtains if it is estimated with only the probability of the realized sample minima.

7 Conclusion

Accurate estimation of the tail threshold is essential in many applications of market structure analysis, and financial risk forecasting. The estimation of the tail shape in those applications usually depends on the data being Pareto distributed, an assumption that in usually incorrect in practice. Extreme value theory (EVT) provides unique insight into these problems, not only in providing a formal environment in which to analyze these problems, but more importantly in enabling optimal estimation of tail shapes.

Unfortunately, determining the tail threshold proves to be very challenging in most cases. We propose a bootstrap procedure for estimating the optimal tail threshold in extreme value analysis, where we use a two step subsample bootstrap procedure for estimating the threshold such that the bias and variance of the estimated tail index decline at the same rate and achieve convergence in probability to normal distribution function of the estimated tail index.

We evaluate our estimator with a series of Monte Carlo experiments using a variety of heavy tailed distributions and stochastic process chosen for the resemblance to economic data. Subsequently, we apply our estimator to market structure problems, and financial risk forecasting.

A Mathematical Derivations

Before we can proof the first two theorems, we need the following preliminary results. If s is fixed sufficiently high, then by (5) and the binomial distribution we find that the expected number of elements in the set $\mathcal{A}$ equals

$$\mathrm{E}\left(M\right)=nas^{-\alpha}\left[1+bs^{-\beta}+o\left(s^{-\beta}\right)\right].$$

We also need the expectation of M given that there is at least one element in $\mathcal{A}$, which we write as

$$\mathrm{E}_1\left(M\right)\equiv\mathrm{E}\left(M|M\geq1\right).$$

For large n

$$\begin{aligned}\mathrm{E}_1\left(M\right)&=\mathrm{E}\left(M\right)+\frac{F\left(s\right)^n}{1-F\left(s\right)^n}\mathrm{E}\left(M\right)\\&=ans^{-\alpha}+o\left(ns^{-\alpha}\right).\end{aligned}$$

From the Taylor series of $1/n$ around $1/\mathrm{E}\left(M\right)$, one shows that

$$\mathrm{E}_1\left(\frac{1}{M}\right)=\frac{s^{\alpha}}{an}+o\left(\frac{s^{\alpha}}{n}\right).$$

Let $\mathrm{E}_{\mathcal{A}}\left(\cdot\right)$ denote the conditional expectations operator where the expectation is computed with respect to the measure $dF\left(x\right)$ conditional on $X_{(i)}$ being an element in the set $\mathcal{A}$.

Lemma 17 *If the density $f\left(x\right)$ adheres to (5), then the conditional expectation of the $k-th$ order log empirical moment is*

$$\mathrm{E}_{\mathcal{A}}\left(u_k\left(s_n\right)\right)=k!\left(\frac{1}{\alpha^k}+\frac{bs_n^{-\beta}}{\left(\alpha+\beta\right)^k}\right)+o\left(s_n^{-\beta}\right), \quad (33)$$

for $s_n\to\infty$, $s_n^{\alpha}/n\to0$ as $n\to\infty$.

Proof. Note that conditionally on s, the upper order statistics in $\mathcal{A}$ are independently distributed. Hence, the expected value of $u_k\left(s\right)$ equals the expected value of a single element from the sum in (6). Apply Lemma 1 to each of the three terms in the density expression (5). Hence, the conditional expression (33) follows from

$$\begin{aligned}\mathrm{E}_{\mathcal{A}}\left(u_k\left(s\right)\right)&=\frac{1}{1-F\left(s\right)^n}\frac{1}{1-F\left(s\right)}\int_s^{\infty}\left(\log\frac{x}{s}\right)^k f\left(x\right)dx\\&=\frac{k!}{1+bs^{-\beta}}\left[\frac{1}{\alpha^k}+\frac{bs^{-\beta}}{\left(\alpha+\beta\right)^k}\right]+o\left(s^{-\beta}\right).\end{aligned}$$

■

Proof of Theorem 3. By application of Lemma 17 we get that

$$\mathrm{E}_{\mathcal{A}}\left[v_k(s)\right] = \frac{1}{\alpha} + \alpha^{k-1}\frac{bs^{-\beta}}{(\alpha+\beta)^k} - \alpha^{k-2}\frac{bs^{-\beta}}{(\alpha+\beta)^{k-1}} + o\left(s^{-\beta}\right) \tag{34}$$

$$= \frac{1}{\alpha} - \frac{b\beta\alpha^{k-2}}{(\alpha+\beta)^k}s^{-\beta} + o\left(s^{-\beta}\right) \tag{35}$$

■

Proof of Theorem 4. By definition of the linearized statistic $v_k(s)$ in (8) and the variance operator

$$\mathrm{Var}_{\mathcal{A}}\left[v_k(s) - \frac{1}{\alpha}\right] = \alpha^{2k-2}\,\mathrm{Var}_{\mathcal{A}}\left[\frac{u_k(s)}{k!}\right] + \alpha^{2k-4}\,\mathrm{Var}_{\mathcal{A}}\left[\frac{u_{k-1}(s)}{(k-1)!}\right] - 2\alpha^{2k-3}\,\mathrm{Cov}_{\mathcal{A}}\left[\frac{u_k(s)}{k!}, \frac{u_{k-1}(s)}{(k-1)!}\right].$$

We calculate the various parts by using the definition of $u_k(s)$ in (6), the independence of X_i and X_j, Lemma 17 and the above preliminary result for $E_1(1/M)$. For the variance part we find when $s_n \to \infty$, $s_n^\alpha/n \to 0$ as $n \to \infty$:

$$\begin{aligned}
\mathrm{Var}_{\mathcal{A}}\left[\frac{u_k(s)}{k!}\right] =& \frac{1}{(k!)^2}\left\{\mathrm{E}_{\mathcal{A}}\left[u_k^2(s)\right] - \left(\mathrm{E}_{\mathcal{A}}\left[u_k(s)\right]\right)^2\right\} \\
=& \frac{1}{(k!)^2}\frac{1}{P\{M \geq 1\}}\left\{\sum_{m=1}^{n}\frac{P[M=m]}{m^2}\mathrm{E}_{\mathcal{A}}\left[\left(\sum_{i=1}^{m}\left(\log\frac{X_i}{s}\right)^k\right)^2\right]\right. \\
&\left. - \left(\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^k\right]\right)^2\right\} \\
=& \frac{1}{(k!)^2}\frac{1}{1-F(s)^n}\left\{\sum_{m=1}^{n}\frac{P\{M=m\}}{m}\left(\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^{2k}\right]\right)\right. \\
&\left. - \left(\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^k\right]\right)^2\right\} \\
=& \frac{s^\alpha}{an}\left\{\frac{(2k)!}{(k!)^2}\left[\frac{1}{\alpha^{2k}} + \frac{bs^{-\beta}}{(\alpha+\beta)^{2k}}\right] - \right. \\
&\left.\left[\frac{1}{\alpha^{2k}} + \frac{2bs^{-\beta}}{\alpha^k(\alpha+\beta)^k} + \frac{b^2s^{-2\beta}}{(\alpha+\beta)^{2k}}\right]\right\} + o\left(\frac{s^\alpha}{n}\right) \\
=& \frac{s^\alpha}{an}\frac{1}{\alpha^{2k}}\left(\frac{(2k)!}{(k!)^2} - 1\right) + o\left(\frac{s^\alpha}{n}\right).
\end{aligned}$$

For the covariance, a similar procedure yields

$$\mathrm{Cov}_{\mathcal{A}}\left[\frac{u_k(s)}{k!},\frac{u_{k-1}(s)}{(k-1)!}\right]=\frac{1}{k!\,(k-1)!}\frac{1}{1-F(s)^n}\left\{\sum_{m=1}^{n}\frac{P\{M=m\}}{m}\right\}\times$$
$$\left\{\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^{2k-1}\right]-\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^{k}\right]\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(1)}}{s}\right)^{k-1}\right]\right\}$$
$$=\frac{1}{k!\,(k-1)!}\frac{s^{\alpha}}{an}\left\{(2k-1)!\left[\frac{1}{\alpha^{2k-1}}+\frac{bs^{-\beta}}{(\alpha+\beta)^{2k-1}}\right]\right.$$
$$-k!\,(k-1)!\left[\frac{1}{\alpha^{2k-1}}+\frac{1}{\alpha^{k-1}}\frac{bs^{-\beta}}{(\alpha+\beta)^{k}}+\right.$$
$$\left.\left.\frac{1}{\alpha^{k}}\frac{bs^{-\beta}}{(\alpha+\beta)^{k-1}}+\frac{\left(bs^{-\beta}\right)^2}{(\alpha+\beta)^{2k-1}}\right]\right\}+o\left(\frac{s^{\alpha}}{n}\right)$$
$$=\frac{s^{\alpha}}{an}\frac{1}{\alpha^{2k-1}}\left(\frac{(2k-1)!}{k!\,(k-1)!}-1\right)+o\left(\frac{s^{\alpha}}{n}\right).$$

■

Proof of Theorem 6. From (14) we have that, for a given n,

$$\overline{\mathrm{AMSE}}\left(v_k\left(\overline{s}\right)\right)=c\kappa\left(k\right)^{2\beta/(2\beta+\alpha)}\left[\frac{\alpha}{\alpha+\beta}\right]^{\frac{2\alpha k}{2\beta+\alpha}},$$

and where $c>0$. Comparing

$$\overline{\mathrm{AMSE}}\left(v_{k-1}\left(\overline{s}\right)\right)\gtreqless\overline{\mathrm{AMSE}}\left(v_k\left(\overline{s}\right)\right),$$

we find that this is equivalent with

$$1+\frac{\beta}{\alpha}\gtreqless\left[\frac{\kappa\left(k\right)}{\kappa\left(k-1\right)}\right]^{\frac{\beta}{\alpha}}.$$

Now note that $\left[\kappa\left(k\right)/\kappa\left(k-1\right)\right]^{\beta/\alpha}$ dominates $1+\beta/\alpha$ for all values of $\beta/\alpha>0$ if $\kappa\left(k\right)/\kappa\left(k-1\right)>e\approx 2.71$. This holds for $k=3,4,...$ ■

Proof of Theorem 7. For $s_n\to\infty$, $s_n^{\alpha}/n\to 0$ as $n\to\infty$ it is a direct consequence of the proof to Lemma 17 that the statistics $\log\left(X_{(i)}/s_n\right)$ for which $X_{(i)}\in\mathcal{A}$ do have a nonzero second moment and bounded third moment, since

$$\mathrm{E}_{\mathcal{A}}\left[\left(\log\frac{X_{(i)}}{s_n}\right)^{k}\right]=\frac{k!}{\alpha^k}+O\left(s_n^{-\beta}\right).$$

This statement carries over to the case where $s_n=\bar{s}_n\left(v_k\right)\to\infty$ as n$\to\infty$. Conditionally on M, the order statistics in set $\mathcal{A}$ are independently distributed.

Therefore by Liapounov's double array central limit theorem with independence within rows, see (Serfling 1980, sect. 1.9.3), we have that

$$\sqrt{M}u_k\left(s_n\right)$$

converges to a normal distribution for any sequence $M \to \infty$, $s_n \to \infty$. In particular for $s_n = \bar{s}_n\left(v_k\right)$, the number of elements M in the set $\mathcal{A}$ is such that $M_n\left(\bar{s}\right) \to \infty$ with probability 1, recall (15). This result carries over to the linearized statistic $v_k\left(s_n\right)$, i.e. $\sqrt{M}v_k\left(\bar{s}_n\right)$ converges to a normal distribution. By (8) and Cramér's delta method, it follows that $\sqrt{M}\left(\alpha w_k\left(\bar{s}_n\right)-1\right)$ also converges in distribution to a normal distribution. The mean and variance of this normal distribution readily follow from Theorems 3 and 4, and the fact that $\sqrt{M}/s_n^{\beta}$ converges in probability to $\left(\alpha+\beta\right)k\sqrt{\kappa\left(k\right)}/\sqrt{2b^2\beta^3\alpha^{2k-3}}$, which follows directly from combining (15) with (13). ■

Proof of Lemma 8. By the Pareto assumption

$$P\left\{\log\left(\frac{X}{\bar{s}_n}\right) > t \,\middle|\, X > \bar{s}_n\right\} = e^{-\alpha t}.$$

Denote $Y_i = \log\left(X_{(i)}/\bar{s}_n\right)$, where $X_{(i)}$ is the $i-th$ order statistic that exceeds $\bar{s}_n$. We have the chain

$$\begin{aligned}
2\,\mathrm{E}\left[\left(w_2\left(\bar{s}_n\right)\right)^2\right] &\leq \mathrm{E}\left[\left(\frac{Y_1^2}{Y_2+\cdots+Y_m}+\cdots+\frac{Y_m^2}{Y_1+\cdots+Y_{m-1}}\right)^2\right] \\
&= m\,\mathrm{E}\left[\left(\frac{Y_1^2}{Y_2+\cdots+Y_m}\right)^2\right] \\
&\quad + m\left(m-1\right)\mathrm{E}\left[\frac{Y_1^2Y_m^2}{\left(Y_2+\cdots+Y_m\right)\left(Y_1+\cdots+Y_{m-1}\right)}\right].
\end{aligned}$$

Since from the convolution of exponentials

$$P\left\{\sum_{i=1}^{m} Y_i > x \mid Y_i > 0\right\} = \int_x^{\infty}\frac{\alpha^m}{\Gamma\left(m\right)}t^{m-1}e^{-\alpha t}dt,$$

a transformation of variables argument gives

$$\mathrm{E}\left[\left(\frac{1}{\sum_{i=1}^{m}Y_i}\right)^{\ell+1}\right] = \frac{\alpha^{\ell+1}}{\left(m-1\right)\left(m-2\right)...\left(m-\ell-1\right)}.$$

Use this to rewrite the previous inequality

$$2\,\mathrm{E}\left[\left(w_2\left(\bar{s}_n\right)\right)^2\right] \leq m\frac{\alpha^4\,\mathrm{E}\left[Y_1^4\right]}{\left(m-2\right)\left(m-3\right)} + m\left(m-1\right)\frac{\alpha^2\,\mathrm{E}\left[Y_1^2\right]}{\left(m-2\right)\left(m-3\right)},$$

where we used Holder's in equality to obtain the second part (for $X, Y \geq 0$: $E[XY] \leq \sqrt{E[X^2]\, E[Y^2]}$). Using the moments of the exponential d.f. we get

$$E\left[\left(w_2\left(\overline{s}_n\right)\right)^2\right] \leq 30, \quad \forall m \geq 4.$$

This bound is sufficient for uniform integrability of the mean and hence the convergence in the mean. ■

Proof of Theorem 9. By Lemma 17 we have

$$\begin{aligned} \mathrm{E}_{\mathcal{A}}\left[q\left(s\right)\right] &= \frac{1}{\alpha} - \frac{2}{1+bs^{-\beta}}\left[\frac{1}{\alpha} + \frac{bs^{-\beta}}{\alpha+\beta}\right] + \frac{\alpha}{1+bs^{-\beta}}\left[\frac{1}{\alpha^2} + \frac{bs^{-\beta}}{\left(\alpha+\beta\right)^2}\right] + o\left(s^{-\beta}\right) \\ &= \frac{bs^{-\beta}}{1+bs^{-\beta}}\frac{\beta^2}{\alpha\left(\alpha+\beta\right)^2} + o\left(s^{-\beta}\right). \end{aligned}$$

And similar to the proof of Theorem 4, we derive

$$\begin{aligned} \mathrm{Var}_{\mathcal{A}}\left[q\left(s\right)\right] &= 4\,\mathrm{Var}\left[u_1\left(s\right)\right] - 2\alpha\,\mathrm{Cov}\left[u_1\left(s\right), u_2\left(s\right)\right] + \frac{\alpha^2}{4}\,\mathrm{Var}\left[u_2\left(s\right)\right] \\ &= \left\{\frac{4}{\alpha^2} - 4\alpha\frac{1}{\alpha^3}\left(\frac{6}{2} - 1\right) + \frac{\alpha^2}{4}\frac{4}{\alpha^4}\left(\frac{24}{4} - 1\right)\right\}\frac{s^\alpha}{an} + o\left(\frac{s^\alpha}{n}\right) \\ &= \frac{1}{\alpha^2}\frac{s^\alpha}{an} + o\left(\frac{s^\alpha}{a}\right). \end{aligned}$$

■

Proof of Corollary 10. From Theorem 9 we calculate the AMSE $[q]$ as

$$\frac{1}{\alpha^2}\frac{s^\alpha}{an} + \frac{b^2\beta^4}{\alpha^2\left(\alpha+\beta\right)^4}s^{-2\beta}. \tag{36}$$

Minimizing the AMSE with respect to s then yields the claim. ■

Proof of Theorem 11. Use again the shorthand notation $Y_{i,r}^k\left(s\right) = \left(\log\frac{X_{i,r}}{s}\right)^k$, where $X_{i,r} \geq s$. The bootstrap MSE version of the linearized statistic $q\left(s_n\right)$ is (we use s_1 as shorthand for s_{n_1}):

$$\frac{1}{R}\sum_{r}^{R}\left[q_r\left(s_1\right)\right]^2 = \tag{37}$$

$$\begin{aligned} &\frac{1}{R}\sum_{r}^{R}\left\{\frac{1}{\alpha^2} + 4\left(\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}\left(s_1\right)\right)^2 + \frac{\alpha^2}{4}\left(\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}^2\left(s_1\right)\right)^2\right. \\ &-\frac{4}{\alpha}\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}\left(s_1\right) + \frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}^2\left(s_1\right) \\ &\left.-2\alpha\left(\frac{1}{M_r}\right)^2\sum_{i}^{M_r}Y_{i,r}\left(s_1\right)\sum_{i}^{M_r}Y_{i,r}^2\left(s_1\right)\right\}. \end{aligned}$$

For each of the terms within the curled brackets first drive $R \to \infty$ and subsequently take $n \to \infty$. To this end suppose that $s_1 = o(\bar{s}_n)$. Hence, for the second term on the right hand side of (37) we can write

$$\frac{1}{R}\sum_r^R \frac{1}{M_r^2}\left(\sum_i^{M_r} Y_{i,r}(s_1)\right)^2 = \frac{1}{R}\sum_r^R \frac{1}{M_r^2}\sum_i^{M_r}(Y_{i,r})^2 + \frac{1}{R}\sum_r^R \frac{1}{M_r^2}\sum_{i\neq j}^{M_r}\sum_j^{M_r} Y_{i,r}Y_{j,r}.$$

Consider the first term of this expression. We find that as $R \to \infty$

$$\frac{1}{R}\sum_r^R \frac{1}{M_r^2}\sum_i^{M_r}[Y_{i,r}(s_1)]^2 \xrightarrow{\text{P}} \frac{s_1^\alpha}{an_1}\frac{1}{T}\sum_i^T Y_{(i)}^2(s_1)$$

Where we use

$$\text{plim}_{R\to\infty}\sum_{i=1}^{n_1}\frac{R_i^*}{R}\frac{1}{i} = \frac{s_1^\alpha}{an_1}.$$

Similarly, the second term tends in probability to

$$\left(1 - \frac{s_1^\alpha}{an_1}\right)\frac{1}{T(T-1)}\sum_{i\neq j}^T\sum_j^T Y_{(i)}(s_1)\,Y_{(j)}(s_1).$$

Recombine the two terms. Recall the way in which we calculated the variance part in the proof of Theorem (4). Subsequently drive $n \to \infty$ to show that when we choose $m(\bar{s}_1) = O\left(n^{1-\varepsilon\frac{2\beta}{2\beta+\alpha}}\right)$

$$m(\bar{s}_1)\left\{\frac{1}{R}\sum_r^R \frac{1}{M_r^2}\left(\sum_i^{M_r} Y_{i,r}\right)^2 - \left[\frac{1}{\alpha^2}\frac{s_1^\alpha}{an_1} + \frac{1}{\left(1+bs_1^{-\beta}\right)^2}\left(\frac{1}{\alpha} + \frac{bs_1^{-\beta}}{\alpha+\beta}\right)^2\right]\right\} \xrightarrow{\text{P}} 0$$

By similar reasoning, one finds the other terms on the right hand side of (37) are asymptotic to respectively:

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\left(\sum_{i}^{M_r}Y_{i,r}^2(s_1)\right)^2 \approx \frac{s_1^{\alpha}}{an_1}\frac{20}{\alpha^4}+\frac{4}{\left(1+bs_1^{-\beta}\right)^2}\left(\frac{1}{\alpha^2}+\frac{bs_1^{-\beta}}{(\alpha+\beta)^2}\right)^2;$$

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}(s_1) \approx \frac{1}{1+bs_1^{-\beta}}\left(\frac{1}{\alpha}+\frac{bs_1^{-\beta}}{\alpha+\beta}\right);$$

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}^2(s_1) \approx \frac{2}{1+bs_1^{-\beta}}\left(\frac{1}{\alpha^2}+\frac{bs_1^{-\beta}}{(\alpha+\beta)^2}\right);$$

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\sum_{i}^{M_r}Y_{i,r}(s_1)\sum_{i}^{M_r}Y_{i,r}^2(s_1) \approx \frac{s_1^{\alpha}}{an_1}\frac{4}{\alpha^3}+\frac{2}{\left(1+bs_1^{-\beta}\right)^2}\left(\frac{1}{\alpha^2}+\frac{bs_1^{-\beta}}{(\alpha+\beta)^2}\right)\left(\frac{1}{\alpha}+\frac{bs_1^{-\beta}}{\alpha+\beta}\right).$$

Substitute these expressions into the appropriate places within the curled brackets in (37). After rearrangement, one finds that

$$m(s_1)\left\{\frac{1}{R}\sum_{r}^{R}\left[q_r(s_1)\right]^2-\left[\frac{1}{a\alpha^2}\frac{s_1^{\alpha}}{n_1}+\frac{b^2\beta^4}{\alpha^2(\alpha+\beta)^4}\frac{1}{s_1^{2\beta}}\right]\right\}\xrightarrow{\text{P}}0 \tag{38}$$

for any $s_1 = o(\bar{s}_n)$. By Corollary 10 and its proof, the asymptotic value of this bootstrap MSE $[q]$ is minimized at $s_1 = \overline{\overline{s}}_{n_1}(q)$, where $\overline{\overline{s}}_{n_1}(q)$ is given in (18). Since by previous arguments $z(s_1) = O_p(q(s_1))$, $\overline{\overline{s}}_{n_1}(q)$ also minimizes the asymptotic MSE of the z–statistic. ■

Proof of Theorem 12. Note that $\overline{\overline{m}}_n(z)$ from (21) is itself a regularly varying function with tail index $2\beta/(\alpha+2\beta)$. By the properties of regularly varying functions we have that

$$\frac{\log\overline{\overline{m}}_{n_1}(z)}{\log n_1}\xrightarrow{\text{P}}\frac{2\beta/\alpha}{1+2\beta/\alpha}$$

as $n_1\to\infty$. Then use the fact that $\hat{m}_{n_1}(z)/\overline{\overline{m}}_{n_1}(z)\xrightarrow{\text{P}}1$ ■

Proof of Theorem 14. Similar to Corollary 13 we have that

$$\frac{\hat{m}_{n_1}(z)}{\overline{\overline{m}}_n(z)}\left(\frac{n}{n_1}\right)^{\frac{2}{2+\alpha/\beta}}\xrightarrow{\text{P}}1$$

and

$$\frac{\hat{m}_{n_2}(z)}{\hat{m}_{n_1}(z)}\left(\frac{n_1}{n_2}\right)^{\frac{2}{2+\alpha/\beta}}\xrightarrow{\text{P}}1.$$

Division combined with the fact that we choose $nn_2/n_1^2 = 1$ yields the claim. ■

Proof of Theorem 15. Consider the quantile estimator (31) based on the deterministic threshold interpretation of the tail index estimator

$$\hat{x}_p = x_t \left(\frac{M/n}{p}\right)^{w_k(s)}, \; x_t = s,$$

and write $\hat{x}_p = \hat{x}_p\left(w_k(s), \frac{M}{n}\right)$. Expand $\hat{x}_p\left(w_k(s), \frac{M}{n}\right)$ into a first order Taylor series around the point

$$\left(1/\alpha, t\left(1+bx_p^{-\beta}\right)/\left(1+bx_t^{-\beta}\right)\right).$$

This gives

$$\begin{aligned}
\hat{x}_p\left(w_k(s), \frac{M}{n}\right) = {} & x_t\left(\frac{t}{p}\right)^{1/\alpha}\left(\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)^{1/\alpha} + \\
& x_t\left(\frac{t}{p}\right)^{1/\alpha}\left[\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right]^{1/\alpha}\log\left(\frac{t}{p}\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)\left[w_k(s)-\frac{1}{\alpha}\right] + \\
& \frac{1}{\alpha}x_t\left(\frac{t}{p}\right)^{1/\alpha}\left[\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right]^{1/\alpha}\left(\frac{t}{p}\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)^{-1}\left[\frac{M}{n}-t\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right] + \\
& O_p\left(\left[w_k(s)-\frac{1}{\alpha}\right]^2+\left[\frac{M}{n}-t\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right]^2\right).
\end{aligned}$$

Since $x_p > x_t$, and due to the property of slowly varying functions, we can write

$$\begin{aligned}
x_t\left(\frac{t}{p}\right)^{\frac{1}{\alpha}}\left(\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)^{\frac{1}{\alpha}} &= x_p\left(1+o\left(x_t^{-\beta}\right)\right)^{\frac{1}{\alpha}} \\
&= x_p\left(1+o\left(x_t^{-\beta}\right)\right).
\end{aligned}$$

Hence, we can write

$$\begin{aligned}
\hat{x}_p\left(w_k(s), \frac{M}{n}\right) = {} & x_p\left(1+o\left(x_t^{-\beta}\right)\right) \\
& \left\{1+\log\left(\frac{t}{p}\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)\left[w_k(s)-\frac{1}{\alpha}\right]+\frac{1}{\alpha}\left[\frac{1}{t}\frac{M}{n}\frac{1+bx_t^{-\beta}}{1+bx_p^{-\beta}}-1\right]\right\} \\
& + O_p\left(\left[w_k(s)-\frac{1}{\alpha}\right]^2+\left[\frac{M}{n}-t\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right]^2\right).
\end{aligned}$$

Rearrangement and premultiplication gives

$$\frac{\frac{\sqrt{M}}{\log\left(\frac{M}{np}\right)}\left(\frac{\hat{x}_p}{x_p}-1\right)}{\sqrt{\kappa(k)}} = \frac{\sqrt{\frac{M}{\kappa(k)}}}{\log\left(\frac{M}{np}\right)}\left\{\left(\log\left(\frac{t}{p}\right)+\log\left(\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)\right)\right.$$
$$\left(w_k(s)-\frac{1}{\alpha}\right)+\frac{1}{\alpha}\left[\frac{M}{nt}\frac{1+bx_t^{-\beta}}{1+bx_p^{-\beta}}-1\right]+o\left(x_t^{-\beta}\right)$$
$$\left.+O_p\left(\left[w_k(s)-\frac{1}{\alpha}\right]^2+\left[\frac{M}{n}-t\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right]^2\right)\right\}.$$

Take $x_t=\bar{s}_n$ and let $n\to\infty$. Moreover, by assumption $np_n\to\tau$, $\tau\geq 0$ a constant. Recall that $M\bar{s}_n^{\alpha}/an\to 1$ in p, and by definition $t=1-F(x_t)=1-F(\bar{s})\approx a(\bar{s}_n)^{-\alpha}$. Thus

$$\text{plim}_{n\to\infty}\frac{\log\left(\frac{t}{p}\right)}{\log\left(\frac{\frac{M}{n}}{p}\right)}=1,$$

and

$$\text{plim}_{n\to\infty}\frac{\log\left(\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}\right)}{\log\left(\frac{M}{np}\right)}=0.$$

It follows from Theorem 7 and Slutsky's Theorem that

$$\frac{\log\frac{t}{p}+\log\frac{1+bx_p^{-\beta}}{1+bx_t^{-\beta}}}{\log\left(\frac{M}{np}\right)}\frac{\sqrt{M}\left(w_k(\bar{s}_n)-\frac{1}{\alpha}\right)}{\sqrt{\kappa(k)}}\xrightarrow{\text{D}}N\left(-\frac{\text{sign}(b)}{\sqrt{2\alpha\beta}},\frac{1}{\alpha^2}\right)$$

in distribution. The statement of the theorem now follows if the other terms can be shown to converge to zero in probability. Also note that this part of the proof shows that the properties of the quantile estimators are driven solely by the properties of the tail estimator.

From the above assumptions and choice for x_t it readily follows that

$$\frac{1}{\alpha}\left[\frac{M}{nt}\frac{1+bx_t^{-\beta}}{1+bx_p^{-\beta}}-1\right]=O_p\left(n^{\frac{-\beta}{\alpha+2\beta}}\right).$$

From (15) we know that

$$\sqrt{M}=O_p\left(n^{\frac{\beta}{\alpha+2\beta}}\right).$$

By assumption $np_n \to \tau$, so that $\log\left(\frac{M}{np_n}\right)$ diverges. Putting everything together we find

$$\text{plim}_{n\to\infty} \frac{\frac{\sqrt{M}}{\log\left(\frac{M}{np}\right)} \frac{1}{\alpha} \left[\frac{M}{nt} \frac{1+bx_t^{-\beta}}{1+bx_p^{-\beta}} - 1\right]}{\sqrt{\kappa\left(k\right)}} = 0.$$

By using the foregoing arguments the last two terms are readily shown to vanish in probability. Finally, note that we may replace $\bar{s}_n$ by $\hat{s}_n$ on the basis of Theorem 11.

■

Table 1: Simulation Results: Parameters

Distribution	Parameter	Mean	s.e.	RMSE	True
Student t(1)	$1/\alpha$	1.012	0.075	0.075	1.000
	β/α	1.398	0.305	0.675	2.000
	$m/\overline{m}$	1.702	0.691	0.984	1.000
Student t(4)	$1/\alpha$	0.286	0.054	0.064	0.250
	β/α	0.600	0.165	0.193	0.500
	$m/\overline{m}$	2.702	1.982	2.610	1.000
Stable(1.4)	$1/\alpha$	0.670	0.047	0.065	0.714
	β/α	1.497	0.268	0.565	1.000
	$m/\overline{m}$	7.425	1.993	6.726	1.000
Stable(1.8)	$1/\alpha$	0.392	0.040	0.168	0.556
	β/α	1.226	0.204	0.304	1.000
	$m/\overline{m}$	21.708	6.492	21.698	1.000
Type II Extreme (1)	$1/\alpha$	1.026	0.062	0.067	1.000
	β/α	2.042	0.623	1.213	1.000
	$m/\overline{m}$	2.461	1.127	1.844	1.000
Type II Extreme (4)	$1/\alpha$	0.257	0.016	0.017	0.250
	β/α	2.043	0.623	1.214	1.000
	$m/\overline{m}$	2.462	1.127	1.844	1.000
Log Pareto (4)	$1/\alpha$	0.302	0.020	0.055	0.250
	β/α	2.068	0.702	2.183	0.000
	$m/\overline{m}$	—	—	—	—
Student t(3) $SV_{(0.1,0.9)}$	$1/\alpha$	0.360	0.060	0.066	0.333
	β/α	0.705	0.181	0.185	0.667
	$m/\overline{m}$	2.484	1.627	2.200	1.000
MA(1,1) Student t(3)	$1/\alpha$	0.313	0.077	0.079	0.333
	β/α	0.664	0.239	0.239	0.667
	$m/\overline{m}$	5.994	5.061	7.103	1.000
$GARCH(2.0)_{(0.05,0.8,0.2)}$	$1/\alpha$	0.485	0.112	0.113	0.500
	β/α	0.905	0.317	—	—
	$m/\overline{m}$	—	—	—	—
$GARCH(4.0)_{(0.05,0.6,0.2)}$	$1/\alpha$	0.326	0.073	0.105	0.250
	β/α	0.695	0.232	—	—
	$m/\overline{m}$	—	—	—	—

The simulations consist of 250 replications with sample size 5,000. Estimation was performed by searching over the minimum MSE (n) by varying n_1 in steps of 300 from 800 up to $4,200$. For each choice of n_1 and n_2 we drew 500 subsamples in the bootstrap procedure. For each value we report the mean, standard error (s.e.), root mean squared error (RMSE), and the theoretical value, where these values are known.

Table 2: Simulation Results: Quantile Estimation

Distribution			Predicted		Sample	
	$1/p$	True	mean	CV	mean	CV
Student t(1)	$5,000$	1591.6	653.6	.36	14180	5.12
	$15,000$	4774.7	5320	.47	—	—
Student t(4)	$5,000$	10.915	11.54	.18	13.68	.37
	$15,000$	14.450	15.97	.23	—	—
Stable(1.40)	$5,000$	153.18	133.4	.47	435.2	1.93
	$15,000$	335.57	282.8	.32	—	—
Stable(1.80)	$5,000$	30.398	21.01	.21	60.68	1.20
	$15,000$	56.028	32.66	.26	—	—
Type II Extreme (1)	$5,000$	$5,000$	5562	.33	20370	2.39
	$15,000$	$15,000$	17560	.39	—	—
Type II Extreme (4)	$5,000$	8.409	8.547	.08	9.875	.35
	$15,000$	11.067	11.35	.10	—	—
Log Pareto (4)	$5,000$	15.65	17.02	.11	19.35	.37
	$15,000$	21.09	23.76	.13	—	—
Student t(3) $SV_{(0.1,0.9)}$	$5,000$	17.598	18.63	.21	24.9	.57
	$15,000$	25.432	28.07	.26	—	—
MA Student t(3)	$5,000$	22.452	22.3	.26	25.52	.87
	$15,000$	32.243	32.17	.34	—	—
$GARCH(2.0)_{(.05,0.8,0.2)}$	$5,000$	—	17.06	.57	18.46	1.01
	$15,000$	—	31.01	.71	—	—
$GARCH(4.0)_{(.05,.6,.2)}$	$5,000$	—	4.737	.30	5.548	.72
	$15,000$	—	6.941	.37	—	

These are results from the same simulations as in Table 1. The probability level $1/5,000$ corresponds to the expected maximum, and $1/15,000$ is an out–of–sample forecast. The true quantiles for a GARCH process are not known. We show the EVT forecast of the quantiles from (31) along with the coefficient of variation (CV), i.e. s.e./mean. Finally, we show the average sample maxima and its CV.

Table 3: City and Firm Size

	Firm size	City size
number of observations	1,000	1,653
mean	72,633	0.47
standard error	14,177	0.83
max	189,058	9.92
min	1,162	0.100
EVT results		
$\hat{\alpha}$	2.1	2.0
$\hat{\beta}$	3.4	2.5
$\hat{M}$	32	59
LS results		
$\hat{\alpha}(n)$	1.0	1.2

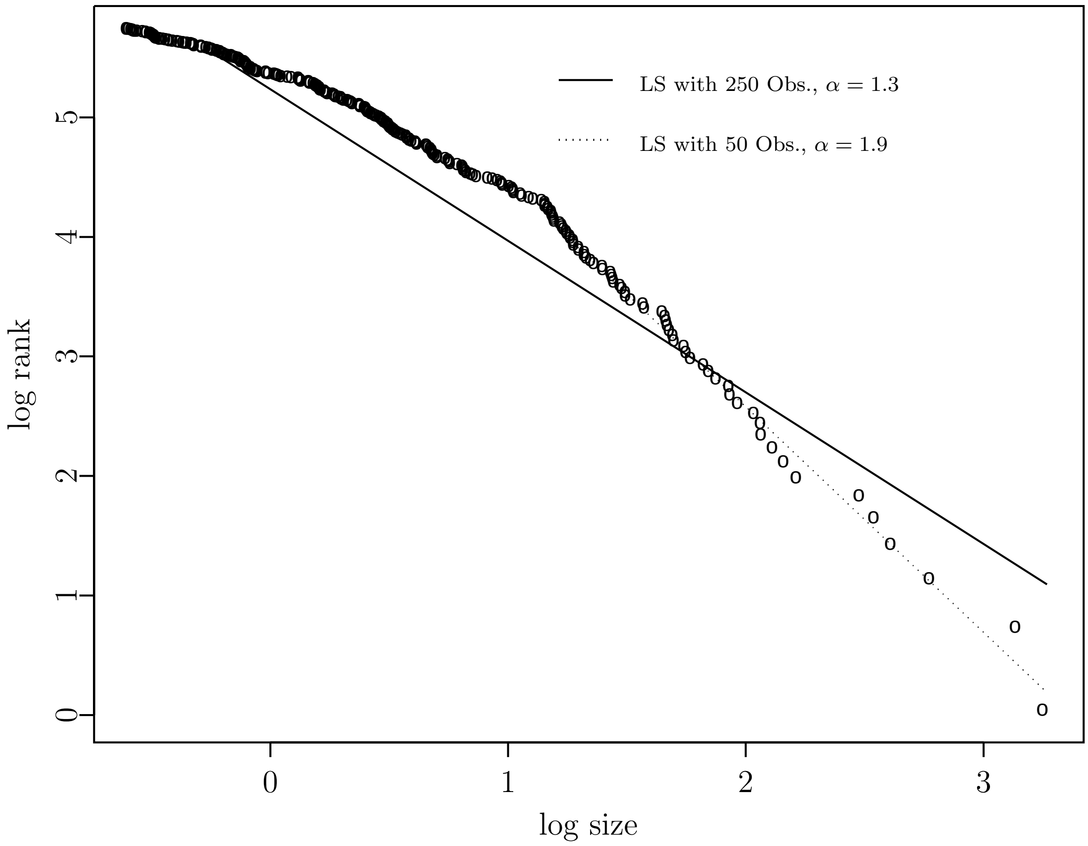
The city size data is all cities in Russia, UK, India, Japan, Brazil, US, China, with more that 100,000 inhabitants, in millions. The firm size data is the revenue of the Fortune 1000 firms from the year 1999 in the US, in USD Billions. The least squares (LS) results are obtained by using all observations in the sample.

Table 4: Daily S&P 500 and Olsen DM/US 10 Minute forex. Lower Tail

	S&P 500		10 minute DM/US	
Observations	First 5,000	Last 5,000	First 5,000	Last 5,000
Annualized				
mean	−1.6%	11.4%	112.9%	− 23.0%
s.e.	26.4%	16.2%	19.2%	12.8%
Skewness	0.09	−2.5	0.464	−0.0814
Kurtosis	8.143	55.0	7.886	17.190
Minimum	−0.13	−0.23	−0.693	−0.655
$1/\alpha$	0.24	0.35	0.301	0.374
$(\cdot,\cdot)$	$(-.17,-.43)$	$(-.29,-.45)$	$(0.27,0.32)$	$(0.33,0.40$
β/α	1.7	2.3	1.60	1.760
$\hat{x}_{1/n}$	−.124	−.099	−0.690	−0.668
$(\cdot,\cdot)$	$(-.069,-.18)$	$(-.058,-.14)$	$(-0.62,-0.85)$	$(-0.58,-0.91)$
$\hat{x}_{1/3n}$	−.161	−.143	−0.958	−1.01
$(\cdot,\cdot)$	$(-.073,-.25)$	$(-.72,-.21)$	$(-.86,-1.17)$	$(-.88,-1.4$

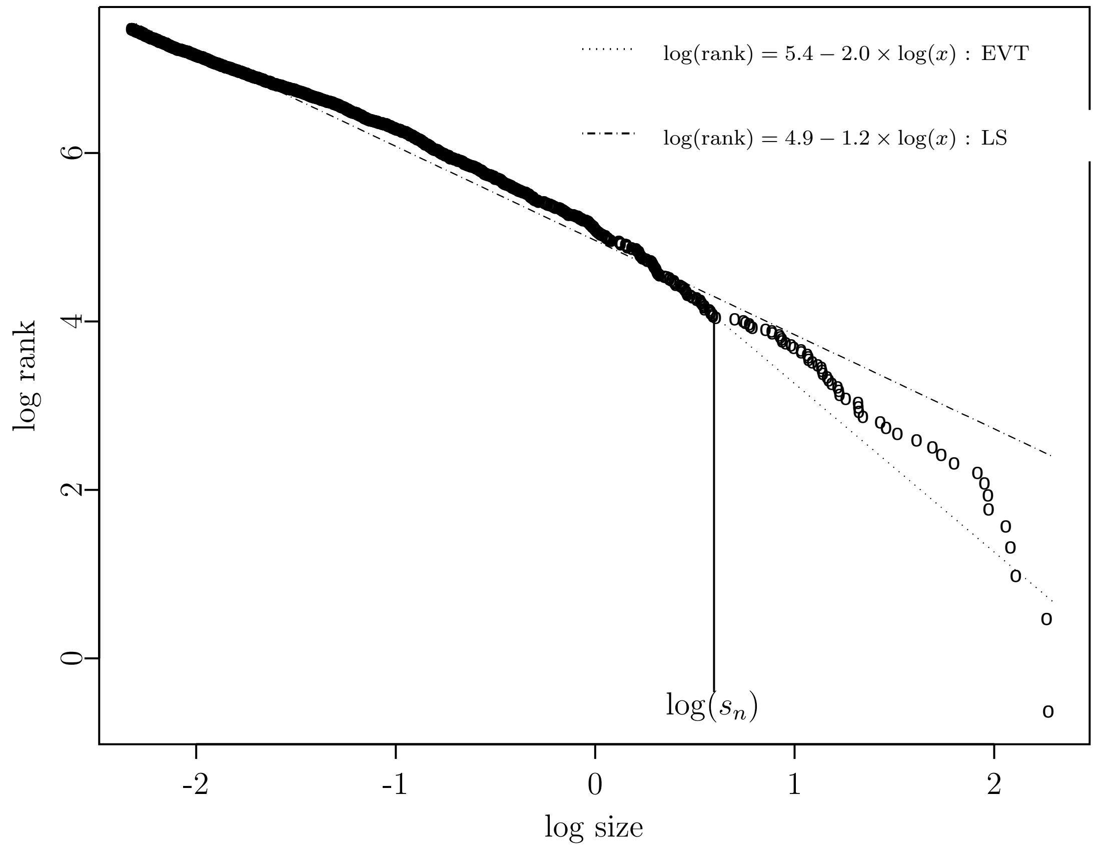
The SP–500 dataset is daily returns from 1929/01/02 to 1946/07/03 and 1981/03/18 to 2000/12/30. The FX data was obtained from Olsen and Associates and contains 1.4 million quotes on the USD–DM spot contract from October 1992 to September 1993. These quotes are aggregated into 52,558 equally spaced 10 minute returns, and we use the first and last 5,000 observations. In addition to the various sample statistics, the table shows estimates of $1/\alpha$ along with the 95% confidence interval. The quantile forecasts, $\hat{x}$, are for the expected maxima, probability equal to $1/5,000$ and for an output sample probability $1/15,000$, along with a 95% confidence interval.

Figure 1: Log Size – Log Rank, 1000 Realizations from a $t_{(1.5)}$



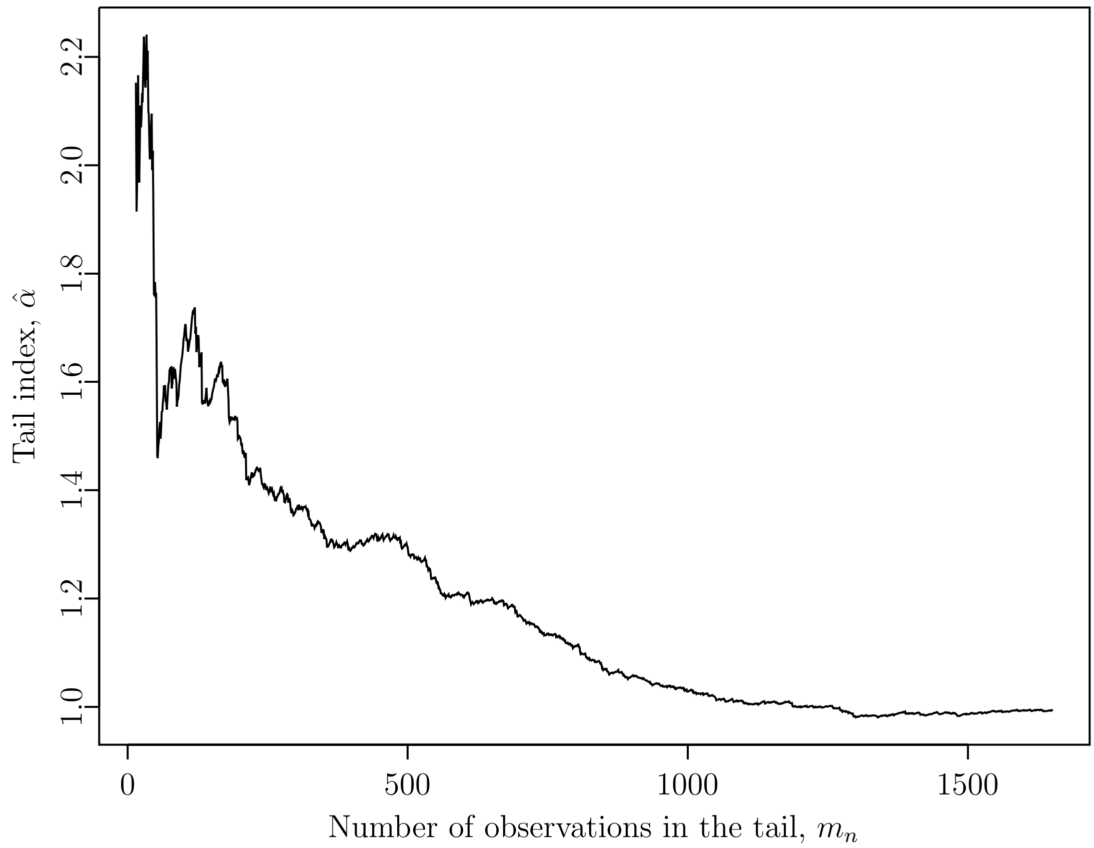
This is a plot of log size vs. log rank for 1000 relations of a Student $t_{(1.5)}$ distribution. The two regression lines are estimated by using 50 and 250 observations.

Figure 2: Log Size – Log Rank Cities



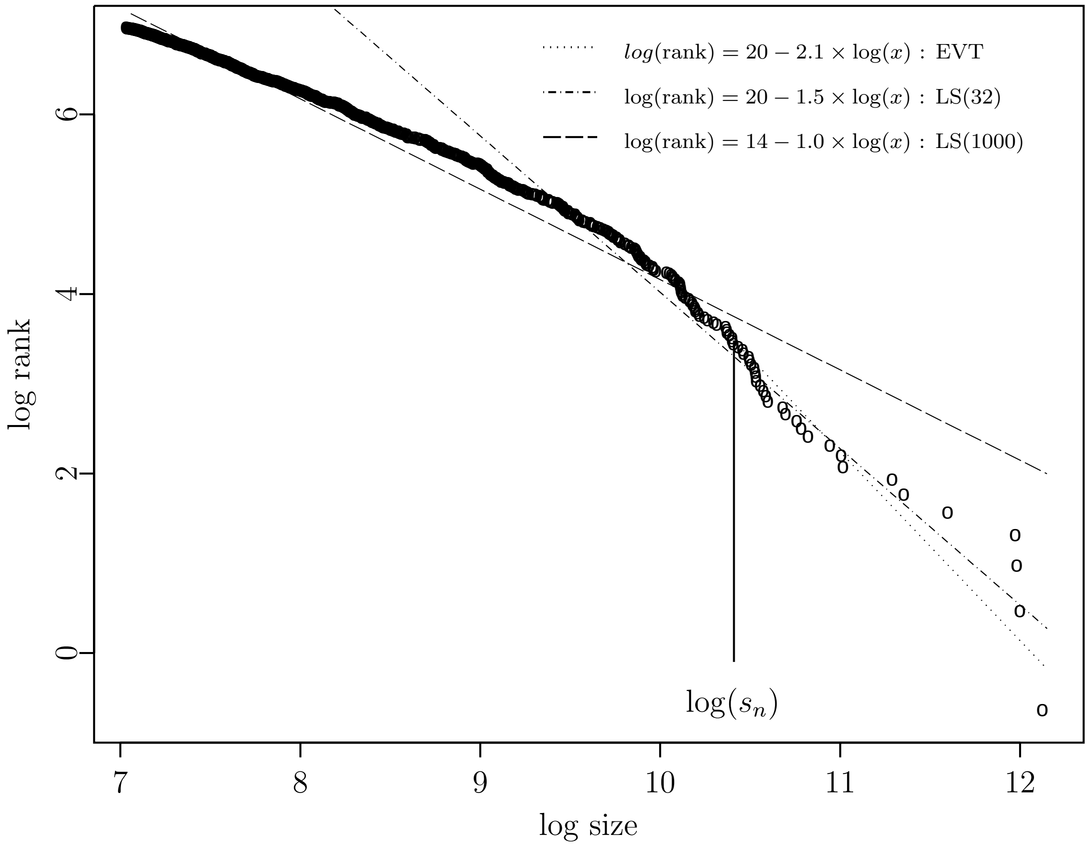
This is a plot of log size vs. log rank for the cities. The two regression lines are estimated with the tail observations only as well as using all observations.

Figure 3: Impact of Tail Threshold on the Tail Index for City Sizes



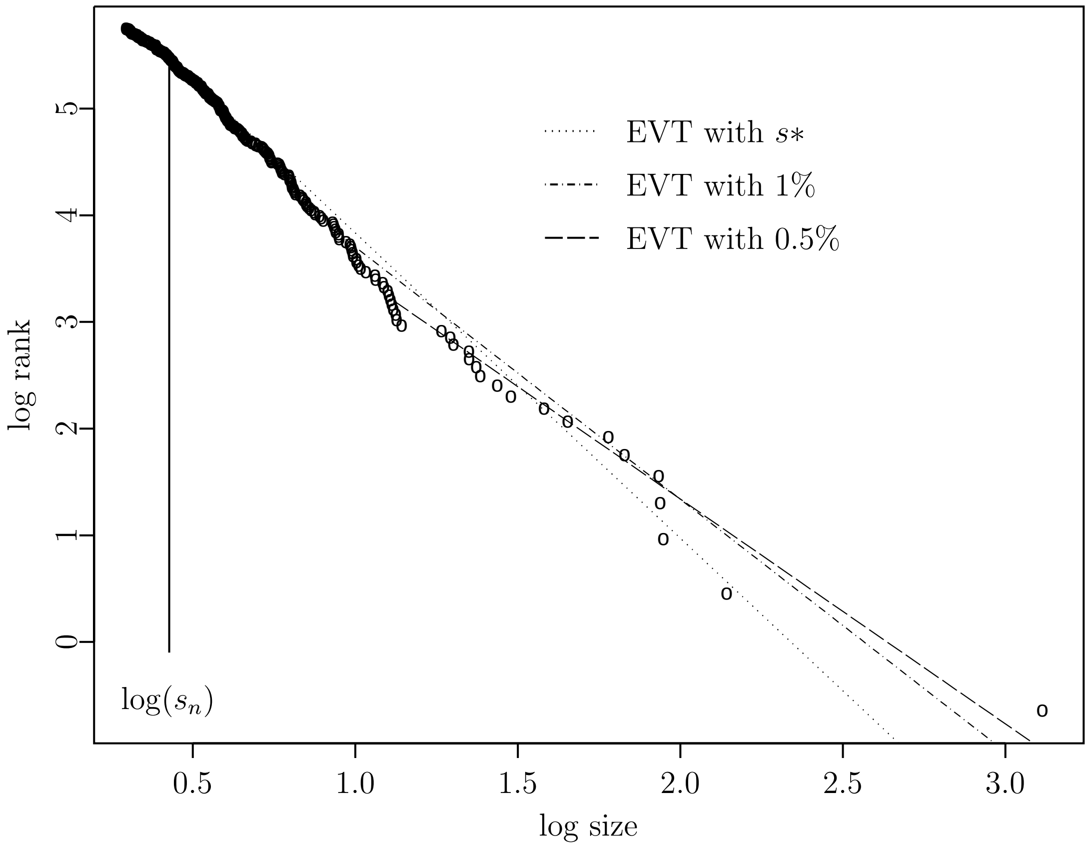
The x axis show the number of observations in the tail, m_n, which ranges from 15 to 1652, i.e. very few observations in the tail to almost all observations in the tail. The y axis show the estimated tail index for the city size data. In the absence of a procedure for determining m_n, we have estimates of α which range from 0.98 to 2.24.

Figure 4: Log size – Log Rank Fortune 1000



This is a plot of log size vs. log rank for the US Fortune 1000 largest firms, ranked by revenue. The three regression lines are estimated with the tail observations only as well as using all observations.

Figure 5: Log size – Log Rank SP–500 Index



This is a plot of log size vs. log rank for 5000 observations of the SP-500 from 1981-03-18 to 2000-12-30. The optimal estimate of the tail index $\alpha(s^*) = 2.86$ along with the $\alpha(1\%) = 2.29$ and $\alpha(0.5\%) = 2.05$.

References

Akgiray, V., Booth, C., and Seifert, B. (1988). Distribution properties of Latin American black market exchange rates. *Journal of International Money and Finance*, 71:37–48.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, 63, 250:318–351.

Chernozhukov, V. and Umantsev, L. (2001). Conditional value-at-risk: Aspects of modeling and estimation. MIT Dept. of Economics Working Paper No. 01-19, `www.mit.edu/people/vchern`.

Chung, K. and Cox, A. (1994). A stochastic model of superstardom: An application of the Yule distribution. *Review of Economic Studies*, pages 771–775.

Daníelsson, J. and de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257.

Daníelsson, J., Jansen, D. W., and de Vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communication in Statistics, Theory and Methods*, 25–4:711–720.

de Haan, L., Resnick, S., Rootzén, H., and de Vries, C. G. (1989). Extremal behavior of solutions to a stochastic difference equation with application to ARCH processes. *Stochastic Processes and their Applications*, 32:213–224.

Embrechts, P., Kuppelberg, C., and Mikosch, T. (1997). *Modelling extremal events.* Berlin: Springer Verlag.

Gabaix, X. (1999). Zipfss law for cities: an explanaiton. *The Quarterly Journal of Economics*, August:739–767.

Gibrat, R. (1931). *Les inégalités économiques; applications : aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel.* Librairie du Recueil Sirey, Paris.

Glaeser, E. L., Scheinkman, J., and Shleifer, A. (1995). Economic growth in a cross–section of cities. *Journal of Monetary Economics*, 36:117–143.

Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics*, Oxford 2nd series:38, 45–71.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc. B*, 42:37–42.

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

Hall, P. and Welsh, A. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Annals of Statistics*, 12:1072–1084.

Henry, M. and Robinson, P. (1999). Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. *Econometric Theory*, 15:299–336.

Hsing (1991). On tail index estimation using dependent data. *Annals of Statistics*, 19:1547–1569.

Ijiri, Y. and Simon, H. A. (1977). *Skew Distributions and the Sizes of Business Firms*, volume 24 of *Studies in Mathematical and Managerial Economics, ed. H. Theil*. North–Holland.

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

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, CXXXI:207–248.

Koedijk, K. G., Schafgans, M., and de Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of International Economics*, 29:93–108.

Longin, F. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56 no 2.

Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1:211–248.

Mansfield, E. (1962). Entry, Gibrat's law, and the growth of firms. *American Economic Review*, 52, 5:1023 – 1051.

Pareto, V. (1896). *Cours d'Economie Politique*. Geneva, Switzerland: Droz.

Quintos, C., Fan, Z., and Phillips, P. C. (2001). Structural change in tail behavior and the Asian financial crisis. *Review of Economic Studies*, 68(3):633–664.

Resnick, S. and Stărică, C. (1996). Tail index estimation for dependent data. MISC, School of ORIE, Cornell University.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.

Sutton, J. (1997). Gibrats legacy. *Journal of Economic Literature*, pages 40–59.

Zipf, G. (1949). *Human Behavior and the Principle of Last Effort*. Cambridge, MA: Addison-Wesley.