Beyond the Sample: Extreme Quantile and Probability Estimation

By

Jon Danielsson and
Casper G. de Vries

DISCUSSION PAPER 298

July 1998

FINANCIAL MARKETS GROUP
AN ESRC RESEARCH CENTRE

LONDON SCHOOL OF ECONOMICS

ISSN 0956-8549-298

Beyond the Sample: Extreme Quantile and Probability Estimation*

Jon Danielsson
London School of Economics and Institute of Economic Studies at University of Iceland

Casper G. de Vries
Erasmus University Rotterdam and Tinbergen Institute

December 1997

Abstract

Economic problems such as large claims analysis in insurance and value-at-risk in finance, require assessment of the probability P of extreme realizations Q. This paper provides a semi-parametric method for estimation of extreme (P, Q) combinations for data with heavy tails. We solve the long standing problem of estimating the sample threshold of where the tail of the distribution starts. This is accomplished by the combination of a control variate type device and a subsample bootstrap technique. The subsample bootstrap attains convergence in probability, whereas the full sample bootstrap would only provide convergence in distribution. This permits a complete and comprehensive treatment of extreme (P, Q) estimation.

Keywords: Extreme value theory, tail estimation, risk analysis

* Corresponding author: C. G. de Vries, Tinbergen Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, e-mail *cdevries@few.eur.nl*. Danielsson's e-mail is *jond@hi.is*. This and related papers can be downloaded from *http://www.hag.hi.is/~jond/research*. Danielsson benefitted from an HCM fellowship of the EU and the Research Contribution of the Icelandic banks. We are grateful to Holger Drees, Laurens de Haan, Marc Henry, and Liang Peng for many useful suggestions. Some data studied in the paper was obtained from Olsen and Associates.

1 Introduction

Economic analysis often depends on assessment of the probability (P) of extreme quantiles (Q) . For example, insurance companies focus on the probability of ruin and commercial banks use the value-at-risk methodology to calculate the loss that can be incurred with a given low probability on their trading portfolio. In addition, value-at-risk is used as the basis for determination of the capital adequacy of financial institutions. Accurate estimation of the borderline in-sample and the out-of-sample (P, Q) combinations is essential for these problems. The tail characteristics are also important for econometric issues such as the convergence rate of regression estimators and the selection of appropriate test statistics.

In this paper we develop a semi-parametric estimator for the tails of the distribution. From statistical extreme value theory, see e.g. Leadbetter, Lindgren and Rootzen (1983), we know that 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 are declining exponentially fast or by a power. The distribution is said to be heavy tailed in the case of power decline so that not all moments are bounded; otherwise the distribution is said to be thin tailed. Hence, if one is only interested in the extreme (P, Q) combinations, one can rely on the asymptotic form of the tail of distribution instead of having to model the whole distribution. This gives the tail focused semi-parametric method an advantage over other methods in tail applications. This applies both to non-parametric methods in general, and parametric methods where the type of distribution is unknown. Estimating the parameters of the wrong distribution typically implies incorrect extreme (P, Q) estimates, both because of misspecification, and because the data in the center of the empirical distribution have too much influence over the parameter estimates of the wrong model; while if only the tails are modelled, this influence is absent. The semi–parametric method may also be superior to a non–parametric approach because the latter is difficult to use for constructing out-of-sample (P, Q) estimates.

If the data are generated by a heavy tailed distribution, then its distribution has, to a first order approximation, a Pareto type tail:

$$P\{X > x\} \approx ax^{-\alpha},\ a > 0,\ \alpha > 0,$$

as $x \to \infty$. By using the concept of regular variation, defined below, the Pareto nature of the tails of such distributions as the non-normal stable, Student-t, and the Fréchet, can easily be verified. The exponent α equals the number of bounded moments, and a is a scaling constant. Estimates of α are needed in order to construct extreme (P, Q) estimates. Suppose there exists a high threshold s above which $ax^{-\alpha}$ is a good approximation of $P\{X > x\}$, and let X_i denote the sample realizations such that $X_i > s$. Then the maximum likelihood estimator for $1/\alpha$ of the left truncated Pareto distribution is the average of the $\log(X_i/s)$. This estimator is known as the Hill (1975) estimator. It has been shown by Hall (1982) and Goldie and Smith (1987) that there exists a unique sequence of thresholds s_n as a function of the sample size n such that the bias squared and variance of the Hill estimator vanish at the same rate. Moreover, this sequence minimizes the asymptotic mean squared error (AMSE) of $1/\hat{\alpha}$. As we show later, given this sequence and the Hill estimate, the construction of extreme (P, Q) estimates is straightforward. The problem is therefore finding the optimal threshold s_n. Until now, it has not been known how to estimate s_n, except under very restrictive assumptions, see e.g. the recent survey by Embrechts, Kuppelberg and Mikosch (1997). This problem has hampered the practical implementation and adoption of extreme value methods, because a key part of the statistical procedure remained arbitrary. Most empirical papers proceed by plotting estimates of $1/\alpha$ against different choices for s_n. Subsequently, by eyeballing such a plot one tries to locate s_n where the bias squared and variance have the appearance of being in balance.

In this paper we solve the long standing problem of estimating s_n through a bootstrap of the MSE of $\widehat{1/\alpha}$, and by minimizing the bootstrap MSE through the choice of s_n. It is, however, not straightforward to construct such a bootstrap, because the theoretical benchmark value $1/\alpha$ is unknown. To solve this problem we use the idea behind control variates in Monte Carlo estimation, see e.g. Hendry (1984). We subtract from the Hill estimator an alternative estimator which converges in the MSE sense at the same rate, albeit with a different multiplicative constant. Hence, this difference statistic converges at the same rate and has a known theoretical benchmark which equals zero in the limit. The square of this difference statistic produces a viable estimate of the $\mathrm{MSE}[1/\alpha]$ that can be minimized with respect to the choice of the threshold s_n.

Unfortunately, it can be shown that the conventional bootstrap procedure of this difference statistic from the entire sample only generates s_n levels which relative to the optimal level converge in distribution. To attain the desired convergence in probability, we show that one needs to create resamples of smaller size than the original sample with a subsample bootstrap technique. The reason why the full sample bootstrap technique fails is due to the linearity of the estimators in $\log(X_i/s)$. By bootstrapping on the entire sample, one in essence recreates the full sample estimates, but the use of smaller resamples produces a weak law of large numbers effect.

We present the material in a comprehensive self contained manner, with known proven results either in the Appendix or referenced, with the benefit that statistical extreme value theory becomes accessible to economists. There is a host of interesting applications in economics and econometrics, some which have been underexploited. Loretan and Phillips (1994) use estimates of heaviness of the tails to determine whether the fourth moment is bounded or not, in order to decide upon the proper asymptotic distribution for the CUSUM statistic. Kearns and Pagan (1997) discuss the issue of tail index estimation with dependent financial returns data. Akgiray, Booth and Seifert (1988), Koedijk, Schafgans and de Vries (1990) and Longin (1996) use tail estimates to construct a nested test in order to discriminate between such non-nested models like the sum–stable and Student-t distributions. Information about the heaviness of the distribution is also useful for obtaining the convergence speed of OLS estimators in regression analysis, since the speed of convergence deteriorates if the innovations have finite variance but are heavy tailed instead of being normally distributed. Booth, Broussaard, Martikainen and Pattonen (1997) analyze the determination of margin calls in futures markets, and Jansen and de Vries (1991) studies the prediction of boom and crashes. Large claims analysis in insurance economics is studied by Beirlant, Teugels and Vynckier (1994), and Embrechts, Kuppelberg and Mikosch (1997).

We look at one application in some detail, Value-at-Risk. In addition, we provide an extensive amount of Monte Carlo experiments to test our estimator. We simulate from a number of i.i.d. and dependent heavy tailed distributions and stochastic processes, and estimate back known theoretical quantities.

2 Theory

We define the idea of heavy tails rigorously by means of the concept of regular variation in subsection 2.1 and then develop a whole class of tail estimators. The properties of these estimators and the optimal choice of s_n for a given distribution are discussed in subsection 2.3. Subsection 2.4 shows how s_n can be estimated from the data if the distribution is unknown. The last subsection provides the extreme probability–quantile (P, Q) estimators.

2.1 Regular Variation

A distribution function $F(x)$ is said to vary regularly at infinity with tail index α if

$$\lim_{t\to\infty} \frac{1 \Leftrightarrow F(tx)}{1 \Leftrightarrow F(t)} = x^{-\alpha}, \quad \alpha > 0, \quad x > 0. \tag{1}$$

The property of regular variation implies that the unconditional moments of X larger than α are unbounded. In this sense the class of regular varying distributions is heavy tailed. This assumption is essentially the only assumption which is needed for the analysis of the tail behavior of X. For expository reasons we only focus on the upper tail of the distribution; the analysis of the lower tail is analogous.

A more detailed parametric form for the upper tail of $F(x)$ can be obtained by taking a second order expansion of $F(x)$ as $x \to \infty$. While there are several expansions possible, de Haan and Stadtmuller (1996) show that there are only two non-trivial expansions. The first expansion is

$$F\left(x\right)=1\Leftrightarrow ax^{-\alpha}\left[1+bx^{-\beta}+o\left(x^{-\beta}\right)\right],\quad \beta>0,\quad \text{as } x\rightarrow\infty. \tag{2}$$

Under a mild extra condition, the expansion (2) implies the following expansion for the density

$$f\left(x\right)=a\alpha x^{-\alpha-1}+ab\left(\alpha+\beta\right)x^{-\alpha-\beta-1}+o\left(x^{-\alpha-\beta-1}\right). \tag{3}$$

Here, the theory will be developed on the basis of the expansion for the density (3). The density expansion facilitates the unified, comprehensive and streamlined treatment of statistical extreme value theory given below. The expansion (3) applies to the well known cases of non-normal sum–stable, Student-t, Fréchet, and other fat tailed distributions. The other non-trivial second order expansion is:

$$F\left(x\right)=1\Leftrightarrow ax^{-\alpha}\left[1+b\log x+o\left(\log x\right)\right].$$

The second order term in this expansion decays more slowly than the algebraic rate of the second order term in (2) and (3)[1].

2.2 k-Moment Ratio Tail Index Estimators

Consider the conditional $k\Leftrightarrow$th order log empirical moment from a sample $X_1,...,X_n$ of n i.i.d. draws from $F\left(x\right)$:

$$u_k\left(s_n\right)\equiv\frac{1}{M}\sum_{i=1}^{M}\left(\log\frac{X_i}{s_n}\right)^k\Bigg|\, X_i>s_n, \tag{4}$$

where s_n is a threshold that depends on n, and M is the random number of excesses. An alternative definition is

$$u_k\left(m_n\right)\equiv\frac{1}{m_n}\sum_{i=1}^{m_n}\left(\log\frac{X_{(i)}}{s_n}\right)^k,\qquad s_n=X_{(m_n+1)},$$

where $X_{(i)}$ are the descending order statistics, and s_n is a random threshold. Note that $u_k\left(s_n\right)$ and $u_k\left(m_n\right)$ are functions of the highest realizations of X. These two definitions yield identical results and are used interchangeably depending on the expediency of the proofs.

Danielsson, Jansen and de Vries (1996) introduced the following class of estimators for the inverse of the first order tail index, $1/\alpha$.

1. For this class the estimator we develop for $1/\alpha$ is consistent. But the slow decay of the second order term makes this class sufficiently different from the other class such that it does not easily fit within the streamlined presentation of the current paper, i.e. it would make the present paper overly long.

Definition 1 *The k-moment ratio estimator, denoted as $w_k(s_n)$, for the inverse tail index is*

$$w_k(s_n) \equiv \widehat{1/\alpha} = \frac{u_k(s_n)}{k u_{k-1}(s_n)}, \tag{5}$$

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

The specific case where $k = 1$, and $w_1(s_n) = u_1(s_n)$, is known as the Hill estimator proposed by Hill (1975). The theoretical properties of the Hill estimator are well documented by e.g. Hall (1982) and Goldie and Smith (1987), who develop the theory respectively on the basis of $u_k(m_n)$ and $u_k(s_n)$. The Hill estimator is considered here as part of the class of moment ratio estimators. There are several motives for a consideration of the whole class of k-moments ratio estimators. Below we show that the various members of the $k\Leftrightarrow$class have under certain conditions better bias and mean squared error properties than other members. Secondly, we show that at least two elements from this class are needed to pin down the optimal threshold s_n.

Hall (1982) and Goldie and Smith (1987) provide proofs of the moment properties of the Hill statistic w_1. We extend the proofs to the general case of w_k. The proofs are contained in Appendix A.

Theorem 1 *For the class of random variables that satisfy (3), the asymptotic bias of the k moment ratio estimator $w_k(s_n)$ is*

$$\mathrm{E}\left[w_k(s_n) \Leftrightarrow \frac{1}{\alpha}\right] = \Leftrightarrow \frac{b\beta\alpha^{k-2}}{(\alpha+\beta)^k} s_n^{-\beta} + o\left(s_n^{-\beta}\right). \tag{6}$$

Theorem 2 *Suppose the threshold s_n is chosen such that $Ms_n^{\alpha}/an \to 1$ in probability as $n \to \infty$. Then if (3) applies*

$$\mathrm{Var}\left[w_k(s_n) \Leftrightarrow \frac{1}{\alpha}\right] = \frac{\kappa(k)}{M\alpha^2} + o\left(\frac{1}{M}\right), \tag{7}$$

where

$$\kappa(k) = \frac{(2k)!}{(k!)^2} + \frac{(2k \Leftrightarrow 2)!}{((k \Leftrightarrow 1)!)^2} \Leftrightarrow 2\frac{(2k \Leftrightarrow 1)!}{k!\,(k \Leftrightarrow 1)!}. \tag{8}$$

The first few values of the $\kappa(k)$ function are given in Table 1. Note the rapid increase as k increases.

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

Table 1: Values of the $\kappa(k)$ function

Together these two theorems imply that for certain sequences of s_n the w_k are consistent estimators if the X_i are i.i.d. and satisfy the density restriction (3). The w_k estimators are also consistent under various forms of dependency. Leadbetter, Lindgren and Rootzen (1983) contains an extensive treatment of ARMA type dependence, which preservers the regular variation property, and de Haan, Resnick, Rootzen, and de Vries (1989) subsequently proved 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.

2.3 Optimal Choice of s_n

The asymptotic mean squared error (AMSE) of $w_k(s_n)$ follows from Theorems 1 and 2:

$$\mathrm{AMSE}(w_k(s_n)) \approx \frac{\kappa(k)}{a\alpha^2}\frac{s_n^\alpha}{n} + \frac{b^2\beta^2\alpha^{2k-4}}{(\alpha+\beta)^{2k}} s_n^{-2\beta}. \tag{9}$$

Which of the two terms on the right hand side of (9) asymptotically dominates the other, is determined by the rate by which $s_n \to \infty$ as $n \to \infty$. Moreover, there is a unique sequence s_n which asymptotically balances the two terms. We derive this sequence from the first order condition $\partial\,\mathrm{AMSE}/\partial s_n = 0$; it can be verified that the second order condition is satisfied as well.

Theorem 3 *Suppose that $Ms_n^\alpha/an \to 1$ in probability and that (3) applies. The unique AMSE minimizing asymptotic threshold $\overline{s}_n$ is*

$$\overline{s}_n(w_k) = \left[\frac{2ab^2\beta^3\alpha^{2k-3}}{(\alpha+\beta)^{2k}\kappa(k)}\right]^{\frac{1}{\alpha+2\beta}} n^{\frac{1}{\alpha+2\beta}}, \tag{10}$$

and the associated asymptotically minimal MSE of $w_k(s_n)$ equals

$$\overline{\mathrm{MSE}}\left[w_k(\bar{s}_n)\right] = \frac{\kappa(k)}{a\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2ab^2\beta^3\alpha^{2k-3}}{(\alpha+\beta)^{2k}\kappa(k)}\right]^{\frac{\alpha}{2\beta+\alpha}} n^{-\frac{2\beta}{2\beta+\alpha}}. \tag{11}$$

The asymptotic number of exceedances $\overline{m}_n$ where the bias and variance parts are balanced is computed by combining (10) with the supposition $Ms_n^{\alpha}/an \to 1$:

$$\overline{m}_n\left(w_k\right) = a\left[\frac{2ab^2\beta^3\alpha^{2k-3}}{\left(\alpha+\beta\right)^{2k}\kappa\left(k\right)}\right]^{-\frac{\alpha}{\alpha+2\beta}} n^{\frac{2\beta}{\alpha+2\beta}}. \tag{12}$$

From (9-11) it is straightforward to show that if s_n tends to infinity at a rate below $n^{1/(2\beta+\alpha)}$, the bias part in the MSE will dominate, while conversely the variance part dominates if s_n tends to infinity more rapidly than $n^{1/(2\beta+\alpha)}$. For the class $w_k\left(s_n\right)$ we show that on the basis of the AMSE criterion the only two elements of interest are w_1 and w_2.

Theorem 4 *The w_1 and w_2 statistics are the only two estimators in the class $w_k, k = 1, 2, 3...$, which are not dominated, in the sense of the AMSE criterion, for all $\beta/\alpha \in \mathbb{R}^+$ combinations.*

Proof. From (11) we have that, for a given n,

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

and where $c > 0$. Comparing

$$\overline{\mathrm{MSE}}\left(w_{k-1}\right) \gtreqless \overline{\mathrm{MSE}}\left(w_k\right),$$

we find that this is equivalent with

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

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

Because

$$1+\frac{\beta}{\alpha} \gtreqless 2^{\frac{\beta}{\alpha}} \quad \text{as} \quad \alpha \gtreqless \beta,$$

we find:

Corollary 1 *For $\beta > 0$, when the Hill and the w_2 statistics are each evaluated at their own asymptotic MSE minimizing thresholds:*

$$\overline{\mathrm{MSE}}\left(w_1\right) \gtreqless \overline{\mathrm{MSE}}\left(w_2\right) \quad \textit{as} \quad \alpha \gtreqless \beta.$$

From the previous formulas for the asymptotic bias and variance it can be seen that both the (optimal) variance and bias squared differ only with respect to the multiplicative constants $1/\alpha$ and $1/2\beta$. Hence

$$\mathrm{BIAS}^2(w_{k-1}) \gtreqless \mathrm{BIAS}^2(w_k) \Leftrightarrow 1+\beta/\alpha \gtreqless \left[\frac{\kappa(k)}{\kappa(k \Leftrightarrow 1)}\right]^{\beta/\alpha}.$$

This implies that if w_2 has a lower MSE than w_1, then w_1 is asymptotically more biased than the w_2. Therefore, w_2 dominates w_1 in both the MSE and bias sense for the cases where the first order term $x^{-\alpha}$ in the asymptotic expansion of $F(x)$ converges more rapidly than the second order term $x^{-\beta}$.

The asymptotic distribution of w_k is given in Theorem 5. The analysis follows Goldie and Smith (1987) who discuss the special case of the Hill statistic.

Theorem 5 *Suppose we choose s_n such that*

$$s_n n^{-\frac{1}{2\beta+\alpha}} \to \left[2ab^2\beta^3\alpha^{2k-3}(\alpha+\beta)^{-2k}\kappa(k)^{-1}\right]^{\frac{1}{2\beta+\alpha}}$$

in probability as $n\to\infty$. Then

$$\sqrt{M}\left(\alpha w_k(s_n) \Leftrightarrow 1\right)\sqrt{\kappa(k)} \to N\left(\Leftrightarrow\sqrt{\frac{\alpha}{2\beta}}\,\mathrm{sign}(b),1\right)$$

in distribution.

The usefulness of Theorem (5) depends on a number of factors. First, the point estimate $w_k(s_n)$ is conditional on the choice of s_n. If the asymptotically optimal threshold $\overline{s}_n$ can be estimated by $\hat{s}_n$ such that $\hat{s}_n/\overline{s}_n$ converges in probability to 1, then the asymptotic normality of $w_k(\overline{s}_n)$ also applies to the case where $\overline{s}_n$ is replaced by its estimated value. Hall and Welsh (1985) showed that this holds without restrictions on the convergence rate; it is an implication of the regular variation property of the MSE in (11). Second, the limiting normal distribution has a mean which depends on the unknown factor $\sqrt{\alpha/2\beta}\,\mathrm{sign}(b)$. If this latter factor can be estimated consistently, then it follows from Slutsky's theorem that the claim of the asymptotic normality of $w_k(\overline{s}_n)$ also applies to the case where the bias factor must be estimated. This issue will be taken up at the end of the next section.

2.4 Estimation of s_n

The estimation of s_n is a non-trivial problem, and up to now the published work on this problem is by Hall (1990). But Hall's paper only gives a very partial solution because it assumes that the first and second order tail indices are equal, i.e. $\alpha = \beta$. Nonetheless Hall's paper contains the important suggestion to employ a bootstrap procedure to estimate $\overline{s}_n$. Since $\overline{s}_n$ asymptotically minimizes the MSE, the idea is to solve this minimization problem by the bootstrap. Suppose that for $k = 1$ we obtain the standard bootstrap equivalent of the expectation

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

by

$$\frac{1}{R}\sum_{r=1}^{R}\left[\left(w_{1,r}\left(s_n\right) \Leftrightarrow \widetilde{w}_1\left(\tilde{s}_n\right)\right)^2\right], \tag{13}$$

where $w_{1,r}$ is calculated on a bootstrap resample of the original sample, R is the number of bootstrap resamples, and $\widetilde{w}_1\left(\tilde{s}_n\right)$ is some consistent initial estimate. Then one could minimize this statistic by choice of s_n. There are, unfortunately, three problems with this approach.

First, this procedure fails to pick up the bias. The log-linearity of the estimator implies that the bootstrap expectation E^R of $w_{1,r}$ given the empirical distribution function F_n equals

$$\mathrm{E}^R\left[w_{1,r}\left(s_n\right)\right| F_n] = w_1\left(s_n\right).$$

Hence, the bias is calculated to be zero. But, as was shown in the previous section, the optimal s_n must be such that the bias squared and variance are of the same order of magnitude. Hall (1990) nicely solved this problem by proposing a subsample bootstrap procedure. Suppose the bootstrap resamples are not of size n, but of smaller size $n_1 < n$, such that $n_1/n \to 0$ as $n_1, n \to \infty$. Replace $w_{1,r}\left(s_n\right)$ in (13) by $w_{1,r}\left(s_{n_1}\right)$. Then the average bias from the subsample estimates does not cancel against the bias of the full sample initial estimate $\widetilde{w}_1\left(s_n\right)$ because the latter is of smaller order, at the corresponding optimal values of s_n.

Second, the procedure only works when the first and second order tail indexes are restricted to be equal. The reason for this restriction can be understood as follows. Once the optimal threshold s_{n_1} has been estimated for the subsample size n_1, it has to be inflated to find the corresponding full sample equivalent. From (10) it follows that the relationship between the optimal threshold values for the different sample sizes is:

$$\bar{s}_n = \bar{s}_{n_1}\left(n/n_1\right)^{\frac{1}{2\beta+\alpha}}. \tag{14}$$

The corresponding formula for the number of excesses is from (12)

$$\bar{m}_n = \bar{m}_{n_1}\left(n/n_1\right)^{\frac{2\beta}{2\beta+\alpha}}. \tag{15}$$

Goldie and Smith (1987) show that (14) and (15) are equivalent solutions to the choice of a threshold. Hall assumes $\beta = \alpha$ and focuses on (15), so that the exponent equals $2/3$. This restriction applies to certain classes of fat tailed distributions like the type II extreme value distribution and the stable distributions. But for other distributions, like the Student-t class where $\beta = 2$ and α equals the degrees of freedom, the restriction does not apply. For a satisfactory general treatment of the class of heavy tailed densities the exponent in (14) or (15) has to be estimated.

The third problem is that the minimization of the subsample bootstrap MSE

$$\min_{s_{n_1}} \frac{1}{R} \sum_{r=1}^{R} \left[\left(w_{1,r} \left(s_{n_1} \right) \Leftrightarrow \widetilde{w}_1 \left(\tilde{s}_n \right) \right)^2 \right] \tag{16}$$

is still conditional on an initial full sample estimate $\widetilde{w}_1 \left(\tilde{s}_n \right)$. But for the entire procedure to work this initial estimate has to be such that $\tilde{s}_n / \overline{s}_n \to 1$ in probability, and hence requires an appropriate choice of $\widetilde{s}_n$. If $\alpha = \beta$ then one might use $w_1 \left(\tilde{m}_n \right)$ where $\tilde{m}_n = n^{2/3}$. This would still ignore the multiplicative constant in (12). But for the general case this is of no avail anyway, and hence $\tilde{m}_n$ or $\tilde{s}_n$ have to be estimated. However, we set out to find s_n in the first place. Hence, this problem undermines the entire procedure sketched thus far. Both the second and the third problem are solved below.

We propose replacing $w_{1,r} \left(s_{n_1} \right)$ by 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, albeit with a different multiplicative constant, as the AMSE of the w_k statistic. When the true value is known, the bootstrap AMSE is easily implemented, because we do not need an initial estimate like $\widetilde{w}_1 \left(\widetilde{s}_n \right)$ in (16), and the optimal s_n or m_n can be estimated.

The statistic we propose to use is:

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

We showed earlier the consistency of all w_k statistics as estimators of $1/\alpha$ for $s_n \approx cn^{\frac{1}{2\beta+\alpha}}$. The $z \left(\bar{s}_n \right)$ converges to 0 as $n \to \infty$. Hence, $\mathrm{MSE}[z] = \mathrm{E}\left[z^2 \right]$. We now show that the $\mathrm{AMSE}\left[z \right]$ has the same order of magnitude as the $\mathrm{AMSE}\left[w_k \right]$.

Theorem 6 *If s_n is such that $M s_n^{\alpha} / an \to 1$ in probability, then*

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

and

$$\mathrm{Var}\left[z \left(s_n \right) \right] = \frac{1}{\alpha^2} \frac{1}{M} + o\left(\frac{1}{M} \right).$$

Corollary 2 *Suppose that $Ms_n^\alpha/an \to 1$ in probability, then the* $\mathrm{AMSE}[z]$ *minimizing asymptotic threshold level $\bar{s}_n(z)$ reads*

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

By comparing $\bar{s}_n(w_k)$ from (10) with the $\bar{s}_n(z)$ from (18) we see that

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

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 2 we know that the asymptotic $\mathrm{MSE}[z]$ is minimized by $\bar{s}_n(z)$ from (18). However, the practical problem of finding this value remains. One possibility is a bootstrap of $z_n^2(s)$, i.e. calculate the bootstrap average $(1/R)\sum_r^R z_{n,r}^2(s)$, and minimize this average with respect to s. Unfortunately, for the following reason this does not produce an estimate which is asymptotic to $\bar{s}_n(z)$. In the proof to Theorem (5) we showed that $\sqrt{M}u_k$ is asymptotically normally distributed. By the Taylor expansion from the proof to Theorem (6) 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 M times the value of the $\mathrm{AMSE}[z]$ as stated in (40) in Appendix A. But because Mz^2 only converges in distribution, the average of the full sample bootstrap values $Mz_{n,r}^2$ 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 Taylor expansion of z as given in the proof to Theorem 6.

Instead of the convergence in distribution, for statistical purposes one would like to have convergence in probability. We will now show that the desired convergence in probability can be obtained through the subsample bootstrap procedure. Note that the argument for using the subsample bootstrap procedure is different from Hall's (1990) argument concerning the bias. Before we turn to the proof, we will provide an intuitive account of what the subsample bootstrap procedure achieves.

Consider the Hill estimator

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

where $Y_i(s_n) \equiv \log(X_i/s_n)$ given that $X_i > s_n$ and where X_i are descending order statistics. From Theorem 5 we know that for $s_n = \bar{s}_n$, $\left(1/\sqrt{M}\right)\sum_i^M Y_i(\bar{s}_n)$ is asymptotically normally distributed. Now suppose that s_n is not of the order $n^{\frac{1}{2\beta+\alpha}}$, cf. (10). Then it follows from (9) that either the bias dominates asymptotically, if $s_n = o\left(n^{\frac{1}{2\beta+\alpha}}\right)$, or that the variance dominates, if $n^{1/2\beta+\alpha} = o(s_n)$. This result for the Hill statistic first appeared in Hall (1982). Now 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. Because $n_1 \sim n^{1-\varepsilon}$, $\bar{s}_{n_1} = o\left(n^{\frac{1}{2\beta+\alpha}}\right)$. The subsample bootstrap average of the Hill statistic is:

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}\left(s_{n_1}\right) = \frac{1}{R}\sum_{r}^{R}w_{1,r}\left(s_{n_1}\right).$$

In the original sample, since $\bar{s}_n > \bar{s}_{n_1}$, the observations are ordered as follows:

$$X_{(1)} \geq ... \geq X_{(M)} > \bar{s}_n \geq X_{(M+1)} \geq ... \geq X_{(T)} > \bar{s}_{n_1} \geq X_{(T+1)} \geq ...$$

The bootstrapped statistic $w_{1,r}\left(\bar{s}_{n_1}\right)$ evaluated at the subsample optimal threshold value $\bar{s}_{n_1}$ is therefore an average from the set

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

It follows that as the number of subsamples R increases

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r}\sum_{i}^{M_r}Y_{i,r}\left(\bar{s}_{n_1}\right) \to \frac{1}{T}\sum_{i}^{T}Y_{(i)}\left(\bar{s}_{n_1}\right) \tag{19}$$

in probability. This result can be understood as follows. For a given n_1 the number of resamples R^*, say, for which M_r has the specific size m^*, $m^* \in \{1, 2, ..., n_1\}$, becomes more numerous as R increases. Hence, eventually the law of large numbers kicks in such that the sum of the averages $\sum_{R^*}\left(\sum^{m^*}Y_{i,r}\left(\overline{s}_{n_1}\right)/m^*\right)$ divided by the number of resamples R^* for which M_r equals m^* converges to $\sum^{T}Y_{(i)}/T$. The weighted average, with weights R^*/R, of these averages for specific m^* values then also converges to $\sum^{T}Y_{(i)}/T$. By the theorem from Hall (1982) we then have that as $n \to \infty$

$$\frac{\frac{1}{T_n}\sum\limits_{i}^{T_n}Y_{(i)}\left(\bar{s}_{n_1}\right) \Leftrightarrow \frac{1}{\alpha}}{\mathrm{E}\left[w_1\left(\bar{s}_{n_1}, n\right) \Leftrightarrow \frac{1}{\alpha}\right]} \to 1 \tag{20}$$

in probability. And hence this is also applies to the left hand side of (19). Note that $w_1\left(\bar{s}_{n_1}, n\right)$ stands for the Hill statistic calculated from the full sample but conditional on the smaller, subsample optimal, threshold value $\bar{s}_{n_1}$. The idea is 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 7 *Suppose model (3) applies. Let* $n_1 = O\left(n^{1-\varepsilon}\right)$ *for some* $0 < \varepsilon < 1$ *be the bootstrap resample size. For given* n *let* $R \to \infty$ *and determine* $\hat{s}_{n_1}$ *such that*

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

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

$$\hat{s}_{n_1}(z)\,/\bar{s}_{n_1}(z) \to 1$$

in probability. Note, $\bar{s}_{n_1}(z)$ *was given in (18).*

Proof. Use the above shorthand notation $Y_{i,r}^k(s) = \left(\log\frac{X_{i,r}}{s}\right)^k$, where $X_{i,r} \geq s$. By the Taylor expansion of z from the proof to Theorem (6) we can write (using the shorthand s_1 for s_{n1})

$$\begin{aligned}
&\frac{1}{R}\sum_{r}^{R}\left[z_r\left(s_1, n_1\right)\right]^2 \qquad (21)\\
= \;&\frac{1}{R}\sum_{r}^{R}\left\{\frac{1}{\alpha^2} + 4\left(\frac{1}{M_r}\sum_{i}^{M_r} Y_{i,r}\left(s_1\right)\right)^2 + \frac{\alpha^2}{4}\left(\frac{1}{M_r}\sum_{i}^{M_r} Y_{i,r}^2\left(s_1\right)\right)^2\right.\\
&\Leftrightarrow\frac{4}{\alpha}\frac{1}{M_r}\sum_{i}^{M_r} Y_{i,r}\left(s_1\right) + \frac{1}{M_r}\sum_{i}^{M_r} Y_{i,r}^2\left(s_1\right)\\
&\left.\Leftrightarrow 2\alpha\left(\frac{1}{M_r}\right)^2\sum_{i}^{M_r} Y_{i,r}\left(s_1\right)\sum_{i}^{M_r} Y_{i,r}^2\left(s_1\right)\right\} + o.
\end{aligned}$$

For each of the terms within the curled brackets we can use arguments similar to the ones that were used to find the asymptotic values in (19) and (20) by first driving $R \to \infty$ and subsequently taking $n \to \infty$. To this end suppose that $s_1 = o\left(\bar{s}_n\right)$. Hence, for the second term on the RHS of (21) asymptotically

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\left(\sum_{i}^{M_r} Y_{i,r}\right)^2$$

$$= \frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\sum_{i}^{M_r}(Y_{i,r})^2 + \frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\sum_{i\neq j}^{M_r}\sum_{j}^{M_r}Y_{i,r}Y_{j,r}$$

$$\approx \frac{1}{\alpha^2}\frac{1}{m(s_1)} + \frac{1}{\left(1+bs_1^{-\beta}\right)^2}\left(\frac{1}{\alpha}+\frac{bs_1^{-\beta}}{\alpha+\beta}\right)^2,$$

where $m(s_1) \approx an_1 (s_{n_1})^{-\alpha}$. The last step, when $n \to \infty$ while $ms_{n_1}^{\alpha}/n \to a$, follows from the arguments in the proof to Theorem 2. To see how the first step can be obtained consider e.g. the first term. By the reasoning that was applied to arrive at (19), we find that as $R \to \infty$

$$\frac{1}{R}\sum_{r}^{R}\frac{1}{M_r^2}\sum_{i}^{M_r}[Y_{i,r}(s_1)]^2 \to \frac{1}{m(s_1)}\frac{1}{T} \quad T$$

$$\approx \frac{1}{m(s_1)}\frac{4}{\alpha^3}+\frac{2}{\left(1+bs_1^{-\beta}\right)^2}\left(\frac{1}{\alpha^2}+\frac{bs_1^{-\beta}}{(\alpha+\beta)^2}\right)\left(\frac{1}{\alpha}+\frac{bs_1^{-\beta}}{\alpha+\beta}\right).$$

Substitute these expressions into the appropriate places within the curled brackets in (21). After some rearrangement, one arrives at

$$\frac{1}{R}\sum_r^R\left[z_r(s_1,n_1)\right]^2\approx\frac{1}{a\alpha^2}\frac{s_1^\alpha}{n_1}+\frac{b^2\beta^4}{\alpha^2(\alpha+\beta)^4}\frac{1}{s_1^{2\beta}}, \tag{22}$$

for any $s_1 = o(\bar{s}_n)$. By Corollary 2 this bootstrap $\mathrm{MSE}[z]$ value is minimized at $s_1 = \bar{s}_{n_1}$, where $\bar{s}_{n_1}$ is given in (18). Moreover, it is straightforward to show that on the one hand for $s_1 \in (0, \bar{s}_{n_1})$ the right hand side of (22) is monotonic and declining in s_1. On the other hand for $s_1 = o(\bar{s}_n)$ and $s_1 > \bar{s}_{n_1}$, the right hand side of (22) is monotonic and increasing. Thus $\bar{s}_{n_1}$ can be located asymptotically by searching for the minimum to

$$\frac{1}{R}\sum_r^R\left[z_r(s_1,n_1)\right]^2$$

as s_1 is increased from 'zero'. ■

Remark 1 *An analogous procedure, and proof applies to the interpretation of the Hill statistic with a fixed number of excesses and a random threshold. In that case one decreases the number of excesses from the maximal number to the value where*

$$\frac{1}{R}\sum_r^R\left[z_r(s_1,n_1)\right]^2$$

bottoms out.

Remark 2 *A proof of this claim by means of bounds on the upper class sequences for the empirical distribution function of the uniform distribution is available from Danielsson, de Haan, Peng and de Vries (1997). This proof applies to the more general class (2), but is also more involved. An altogether different approach is taken in a recent manuscript by Drees and Kaufmann (1997). They establish a law of iterated logarithm for* $\sqrt{m}w_1$. *The result is then used to construct a sequence* $\hat{m}_n$ *which is asymptotic to* $\bar{m}_n$.

The above yields an estimate $\hat{s}_{n_1}(z)$ of the optimal threshold $\bar{s}_{n_1}(z)$ such that $\hat{s}_{n_1}(z)/\bar{s}_{n_1}(z) \to 1$ in probability. A similar statement applies to the estimate for

$\hat{m}_{n_1}(z)$ for the optimal number of highest order statistics $\bar{m}_{n_1}(z)$. Note that $\bar{m}_n(z)$ can be calculated in the same way as that $\bar{m}_n(w_k)$ in (12) was obtained:

$$\bar{m}_n(z) = a\left(\frac{2ab^2\beta^5}{\alpha(\alpha+\beta)^4}\right)^{-\frac{\alpha}{2\beta+\alpha}} n^{\frac{2\beta}{2\beta+\alpha}}. \tag{23}$$

While it is more expedient to present the theoretical derivations in terms of the threshold interpretation, however, in practice the minimization of the bootstrapped $\mathrm{MSE}[z]$ is done in terms of the index m.

In the end we are not interested in the optimal $\bar{m}_n(z)$ from (23), but rather we need the optimal $\bar{m}_n(w_k)$ as in (12). These two quantities are related as follows:

$$\frac{\bar{m}_n(z)}{\bar{m}_n(w_k)} = \left[\left(\frac{\beta}{\alpha}\right)^2\left(1+\frac{\beta}{\alpha}\right)^{2k-4}\kappa(k)\right]^{-\frac{1}{1+\frac{2\beta}{\alpha}}}. \tag{24}$$

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 $\overline{m}_{n_1}(z)$ varies regularly, cf.(1) and (23).

Theorem 8 *A consistent estimator for β/α is*

$$\widehat{\beta/\alpha} = \frac{\log \hat{m}_{n_1}(z)}{2\log n_1 \Leftrightarrow 2\log \hat{m}_{n_1}(z)}. \tag{25}$$

Theorem (8) in combination with (24) implies that

$$\hat{m}_{n_1}(w_2) = \hat{m}_{n_1}(z)\left[\sqrt{2}\frac{\log \hat{m}_{n_1}(z)}{2\log n_1 \Leftrightarrow 2\log \hat{m}_{n_1}(z)}\right]^{\frac{2\log n_1 - 2\log \hat{m}_{n_1}(z)}{\log n_1}} \tag{26}$$

is a consistent estimator for $\bar{m}_{n_1}(w_2)$. 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 $\hat{m}_n(w_2)$.

Corollary 3 *Under the conditions of Theorems (3) and (7)*

$$\frac{\hat{m}_{n_1}(w_k)}{\bar{m}_n(w_k)}\left(\frac{n}{n_1}\right)^{\frac{2}{2+\alpha/\beta}} \to 1 \tag{27}$$

in probability.

Proof. Combine the results from both propositions. ■

One might contemplate using relation (27) as an equality and to replace α/β in the exponent by $\widehat{\alpha/\beta}$ from (25). Unfortunately, even though the $\widehat{\beta/\alpha}$ estimates in (25) is consistent, its rate of convergence is unknown. This frustrates using $\widehat{\beta/\alpha}$ in (27) 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 handily such that the multiplicative factor in (27) can be replaced by a known value.

Theorem 9 *Let $n_1 = O\left(n^{1-\varepsilon}\right)$ 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 $\bar{m}_{n_2}(z)$ from the subsample bootstrap procedure on subsample resamples of size n_2. Then*

$$\frac{\left(\hat{m}_{n_1}(z)\right)^2}{\bar{m}_n(z)\,\hat{m}_{n_2}(z)} \to 1 \tag{28}$$

in probability.

Proof. Similar to Corollary 3 we have that

$$\frac{\hat{m}_{n_1}(z)}{\bar{m}_n(z)}\left(\frac{n}{n_1}\right)^{\frac{2}{2+\alpha/\beta}} \overset{P}{\Leftrightarrow\!\!\rightarrow} 1$$

and

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

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

Combine result (28) with (26) to arrive at the 'consistent' estimator

$$\hat{m}_n(w_2) = \frac{\left(\hat{m}_{n_1}(z)\right)^2}{\hat{m}_{n_2}(z)}\left[\sqrt{2}\frac{\log m_{n_1}(z)}{2\log n_1 \Leftrightarrow 2\log \hat{m}_{n_1}(z)}\right]^{\frac{2\log n_1 - 2\log \hat{m}_{n_1}(z)}{\log n_1}}. \tag{29}$$

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

We have shown how s_n, or m_n, and β/α can be estimated on the basis of a double subsample bootstrap procedure. This procedure 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 their AMSE. The subsample bootstrap yields estimates of the $\mathrm{AMSE}\left(z_{n_1}\right)$ and $\mathrm{AMSE}\left(z_{n_2}\right)$. By the same arguments as were used in the proof to Theorem 9, one can show that

$$\left[\widehat{\mathrm{AMSE}\left(z_{n_1}\right)}\right]^2 / \widehat{\mathrm{AMSE}\left(z_{n_2}\right)} \tag{30}$$

is asymptotic to $\mathrm{AMSE}\left(z\left(\bar{m}_n\right)\right)$. The idea is then to choose n_1 by

$$\underset{n_1}{\arg\min} \left[\widehat{\mathrm{AMSE}\left(z_{n_1}\right)}\right]^2 / \widehat{\mathrm{AMSE}\left(z_{n_2(n_1)}\right)}. \tag{31}$$

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

Remark 3 *Note that in contrast to the previous literature, no arbitrary choice of parameters, in particular m_n or s_n, has to be made in our procedure. Only the tuning parameters concerning the grid size over which n_1 is varied and the number of bootstrap resamples has to be chosen. These are dictated by the available computing time.*

Finally, we need to address how $\mathrm{sign}\left(b\right)$ can be estimated consistently. Estimation of $\mathrm{sign}\left(b\right)$ is needed in Theorem 5 for the purpose of hypothesis testing. This can be achieved as follows. Recall the mean of $z\left(s_n\right)$ from Theorem 6:

$$\mathrm{E}\left[z\left(s_n\right)\right] = c s_n^{-\beta}\, \mathrm{sign}\left(b\right) + o\left(s_n^{-\beta}\right),$$

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

$$\widehat{\mathrm{sign}\left(b\right)} = \mathrm{sign}\left(z\left(s_n\right)\right), \text{ with } s_n < \overline{s}_n. \tag{32}$$

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

$$\mathrm{sign}\left(\left[w_2 \Leftrightarrow w_1\right] \Leftrightarrow \left[w_4 \Leftrightarrow w_3\right]\right).$$

It is straightforward to check that this estimator is also consistent.

2.5 Prediction of Extremes

The primary objective of the paper is to develop better estimators for borderline in–sample and out–of–sample quantile and probability (P, Q) combinations. The properties of the quantile and tail probability estimators follow from the properties of $\widehat{1/\alpha}$. Out–of–sample (P, Q) estimates are related in the same fashion as the in sample (P, Q) estimates, i.e. we establish an out–of–sample Bahadur-Kiefer result.

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 $x_p : 1 \Leftrightarrow F(x_p) = p$, and $x_t : 1 \Leftrightarrow F(x_t) = t$. 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 (2) with $\beta > 0$ we have

$$\frac{t}{p} = \left(\frac{x_p}{x_t}\right)^{\alpha} \frac{1 + bx_t^{-\beta} + o\left(x_t^{-\beta}\right)}{1 + bx_p^{-\beta} + o\left(x_p^{-\beta}\right)},$$

so that

$$x_p = x_t \left(\frac{t}{p}\right)^{1/\alpha} \left(\frac{1 + bx_p^{-\beta} + o\left(x_p^{-\beta}\right)}{1 + bx_t^{-\beta} + o\left(x_t^{-\beta}\right)}\right)^{1/\alpha}. \tag{33}$$

This suggests the following estimator. Ignore the higher order terms in the expansion, replace t by m/n and x_t by the $(m+1)$-th descending order statistic, and substitute for $1/\alpha$ an w_k estimator. This yields:

$$\hat{x}_p = X_{(m+1)} \left(\frac{m}{np}\right)^{w_k}. \tag{34}$$

Alternatively, employ the threshold interpretation of the w_k, i.e. the probability t is replaced by the random variable M/n with x_t fixed at s_n. This gives

$$\hat{x}_p = x_t \left(\frac{M}{np}\right)^{w_k}. \tag{35}$$

Theorem 10 *Suppose that the conditions of Theorem 5 do hold. In addition take* $x_t = \hat{s}_n$, $m = [tn]$, *t decreasing but* $tn \to \infty$. *Suppose that* np_n *converges to a constant* τ *which may be zero. Then the quantile estimator* $\hat{x}_p$ *is asymptotically normally distributed:*

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

The proof for the fixed number of order statistics is similar and omitted.

An estimator for the reverse problem can be developed as well. Rewrite (33)

$$p = t\left(\frac{x_t}{x_p}\right)^{\alpha} \frac{1 + bx_p^{-\beta} + o\left(x_p^{-\beta}\right)}{1 + bx_t^{-\beta} + o\left(x_t^{-\beta}\right)}, \tag{36}$$

and use

$$\hat{p} = \frac{M}{n}\left(\frac{x_t}{x_p}\right)^{\hat{\alpha}}. \tag{37}$$

Theorem 11 *Under the same conditions as in Theorem10, the excess probability estimator $\hat{p}$ is asymptotically normally distributed, that is*

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

in distribution.

Note that the asymptotic distributions of the normed quantiles and probabilities differ by a multiplicative factor of $\Leftrightarrow\alpha^2$. This is a Bahadur-Kiefer type result for out of sample (P, Q)combinations, cf. Serfling (1980). In 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 $(p, \hat{x})$ or $(\hat{p}, x_p)$ curves.

Remark 4 *The algorithm for computing $(p, \hat{x})$ or $(\hat{p}, x_p)$ and $w_2(m_n)$ is as follows. For a given choice of $n_1 < n$ draw R bootstrap resamples of size n_1. Calculate $\frac{1}{R}\sum_r^R \left[z_r\left(m_{n_1}, n_1\right)\right]^2$, i.e. the bootstrap MSE of the difference statistic z at each m_{n_1}; and find the $\hat{m}_{n_1}$ which minimizes this bootstrap MSE. Repeat this procedure for an even smaller resample size n_2, where $n_2 = \left(n_1\right)^2/n$. This yields $\hat{m}_{n_2}$. Subsequently calculate $\hat{m}_n$ from (29). Finally, estimate $1/\alpha$ by $w_2\left(\hat{m}_n\right)$. The choice for n_1 is made from (31). By using this procedure two tuning parameters have to be chosen, the number of bootstrap resamples and the search grid size. Last, one estimates the desired (P, Q) combinations from either (34) or (37).*

3 Estimation and Simulation

We investigate the performance of our estimators, both with Monte Carlo experiments and application to real world problems. In the first experiment, we evaluate the tail index and quantile estimators for a number of heavy tailed distributions and stochastic processes while in the second experiment, we generate data from one model and estimate back another fully parametric model. In the applications we first evaluate the tail shape of several financial returns, and then investigate the determination of capital requirements by Value-at-Risk methods.

3.1 Monte Carlo Experiments

We generate pseudo random numbers from several known distributions and stochastic processes. These are listed in Table 2. The tail index estimator w_2 and the quantile estimator $\hat{x}_p$ are applied to simulated data, and the results compared with their theoretical values. For the Student-t the tail index equals the degrees of freedom; for the extreme value of the Fréchet type and the log Pareto distributions the hyperbolic coefficient equals the tail index; and for the non–normal symmetric stable the characteristic exponent equals the tail index. The log Pareto has a distribution for which the second order term decays slower than the power decay of (2), cf. Footnote (1):

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

It can be obtained as the distribution of the product of two i.i.d. Pareto random variates. The Student(3) $\mathrm{SV}_{(0.1,0.9)}$ model denotes the following process

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

This process generates volatility clusters but Y_t still reflects the fair game property of financial returns, and it is therefore related to the ARCH class of models. It was designed in this specific way because it follows that the marginal distribution function of Y_t has a Student-t(3) distribution, for which we know all theoretical parameters of the expansion at infinity. The MA(1,1) Student t(3) refers to the MA1 process, $Y_t = X_t + X_{t+1}$ where the X_i are i.i.d. Student-t(3) distributed. Therefore Y_t has a convoluted Student-t marginal distribution, for which we can compute all the relevant theoretical values.

The last process to be simulated is the GARCH(1,1) process with normal innovations. We use two processes, $\mathrm{GARCH(2.0)}_{(0.05,0.8,0.2)}$ and $\mathrm{GARCH(4.0)}_{(0.05,0.6,0.2)}$ where the number in the brackets is the theoretical tail index, and the subscripted values are the mean, MA, and AR parameters of the volatility process. Kearns and Pagan (1997) suggest that tail index estimation may not be straightforward for financial returns data

Table 2: 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	—	—
	$m/\overline{m}$	—	—	—	—

due to the presence of volatility clusters. By including the GARCH and SV processes in the simulation, where the parameters have been chosen with an eye towards the Kearns and Pagan (1997) paper, we can investigate this issue. We conclude that our method still performs well in these cases.

The simulations consist of 250 replications with sample size 5,000. The minimum sample size for prudent application of extreme value methods lies around 1,500. Given the currently available sizes of financial data sets, 5000 is a reasonable number.

The simulation estimation proceeds as outlined in Remark 4**.** Estimation was performed by searching over the minimum $\mathrm{MSE}\,(n)$ by varying n_1 in steps of 300 from 800 up to $4,200,$ as suggested in Remark 1. We note that in an application to a particular data set, a much finer grid can easily be implemented. For each choice of n_1 and n_2 we drew 500 subsamples in the bootstrap procedure.

In Table 2 we report the estimate w_2 of $1/\alpha$, the estimates of the ratio β/α, and the ratio of the optimal number of highest order statistics m_n to the theoretical value $\overline{m}_n$ where the latter value is known. 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 3 reports results from the quantile estimation, with probabilities $1/n$ and $1/3n.$ We show the theoretical values, the mean, the coefficient of variation over the simulations, and the average of the sample maxima.

From Table 2 we see that the tail index estimator works reasonably well, w_2 is mostly within two standard errors of the true $1/\alpha,$ and it is often within one standard error of $1/\alpha.$ This holds also for the dependent data. The stable distribution with a characteristic exponent close to 2 is, however, more heavily biased. This is due to the fact that when the characteristic exponent 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, see e.g. Gielens, Straetmans and de Vries (1996) for further details.

The estimate of β/α is less precise than the estimate of $1/\alpha.$ The reason is that the measurements of second order tail parameters is more difficult, because these are second order parameters of the Taylor expansion of the distribution at infinity. The extreme realizations are less informative about this second order behavior than the first order behavior. Nevertheless, the β/α estimates are often within two standard errors of the true values. 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 $\beta/\alpha.$

For most economic purposes the usefulness of our procedure resides in the estimation of out-of-sample (P,Q) combinations, rather than in the precise value of the tail index $1/\alpha.$ Table 3 reports quantile estimates $\hat{x}_p$ on the basis of the estimator in (34). We report the quantile estimate for the borderline in sample probability $1/n$ and the out-

Table 3: Simulation Results: Quantile Estimation

Distribution			Predicted		Sample	
	n	True	mean	c.v.	mean	c.v.
Student t(1)	5000	1591.6	653.6	.36	14180	5.12
	15000	4774.7	5320	.47	—	—
Student t(4)	5000	10.915	11.54	.18	13.68	.37
	15000	14.450	15.97	.23	—	—
Stable(1.40)	5000	153.18	133.4	.47	435.2	1.93
	15000	335.57	282.8	.32	—	—
Stable(1.80)	5000	30.398	21.01	.21	60.68	1.20
	15000	56.028	32.66	.26	—	—
Type II Extreme (1)	5000	5000	5562	.33	20370	2.39
	15000	15000	17560	.39	—	—
Type II Extreme (4)	5000	8.409	8.547	.08	9.875	.35
	15000	11.067	11.35	.10	—	—
Log Pareto (4)	5000	15.65	17.02	.11	19.35	.37
	15000	21.09	23.76	.13	—	—
Student t(3) SV$_{(0.1,0.9)}$	5000	17.598	18.63	.21	24.9	.57
	15000	25.432	28.07	.26	—	—
MA Student t(3)	5000	22.452	22.3	.26	25.52	.87
	15000	32.243	32.17	.34	—	—
GARCH(2.0)$_{(.05,0.8,0.2)}$	5000	—	17.06	.57	18.46	1.01
	15000	—	31.01	.71	—	—
GARCH(4.0)$_{(.05,.6,.2)}$	5000	—	4.737	.30	5.548	.72
	15000	—	6.941	.37	—	

Table 4: Performance of GARCH and Worst Case Analysis

	Worst Case	GARCH	Extreme
True $x_{1/5000}$	17.60	17.60	17.60
mean of min	⇔24.32	⇔19.06	⇔17.54
s.e. of min	20.99	18.60	4.18
max of min	⇔10.11	⇔6.48	⇔9.60
min of min	⇔311.07	⇔158.48	⇔33.77

of-sample probability $1/3n$ $(n = 5,000)$. To estimate $x_{1/n}$, the financial industry often uses either the so called worst case analysis or historical simulation. In the former case, one uses the maximum or minimum sample realizations, and in the latter case one uses the average of the extreme realizations in bootstrapped replications of the sample data. Table 3 reports the average of the maximum in the 250 simulations, i.e. the average worst case analysis. As can be seen from Table 3, this procedure invariably carries with it more uncertainty and bias than the semi-parametric method. Our semi-parametric technique is in essence a method which extrapolates the tail shape of the empirical distribution function. Hence, it relies on the contribution of more than a single order statistic, and thereby reduces variance and bias. In fact, one can show that the average worst case analysis overpredicts the quantiles, and that if our semi–parametric method overpredicts as well, the former method is necessarily more upward biased.

From Table (3) we see the quantile estimator $\hat{x}_p$ performs decently when judged by the mean and the coefficient of variation. The coefficient of variation (c.v.), i.e. s.e./mean, is more convenient than the standard error for the purpose of comparison, and hence we only report the c.v. Since the true mean is unknown in the GARCH case, we use the sample mean. The c.v. does not change much when moving from $p = 1/n$ to $p = 1/3n$ in the extreme value technique case. Moreover, the performance of the quantile estimator is fairly homogeneous across distributions and stochastic processes in terms of the c.v. The worst case analysis shows c.v.'s which are consistently larger, at least twice as large, than c.v.'s for the semi–parametric method.

A comparison between our semi-parametric and fully parametric approaches is also of interest. We set up an experiment where the researcher is given the knowledge that the data is fat tailed, but not the correct class of parametric models. We perform one experiment with sample sizes 5,000 and 250 replications where data is generated from the Student-t $\mathrm{SV}_{(0.3,0.68)}$ model, while the GARCH(1,1) model is estimated back. The estimated GARCH(1,1) model is subsequently used to generate 500 series of 5,000 observations where each simulated series is used to evaluate one simulated estimate of the $\hat{x}_{1/n}$ quantile. Again this approach is compared with the semi-parametric and the worst case analysis.

From Table 4 we see that GARCH outperforms the worst case analysis, but the semi-parametric procedure is still better and has much lower variance. The reason is that the tail procedure is less prone to model misspecification because it does not rely on specific distributional assumptions**.** It only uses the limit expansion for heavy tailed distributions. The semi-parametric tail estimates therefore do not have to serve two masters by matching the parameters to satisfy both tail and center characteristics of the model.

3.2 Asset Returns and Value-at-Risk

The literature already contains tail index and quantile estimates for financial series. Typically those estimates are obtained from applying a graphical procedure for locating the start of the tail, see e.g. Embrechts, Kuppelberg and Mikosch (1997). We used a set of the highest frequency data from the Olsen company on foreign exchange rate quotes. The data set contains 1.4 million quotes on the USD-DM spot contract from October 1992 to September 1993. These quotes are aggregated into 52558 equally spaced 10 minute returns, and we use the first and last 5,000 observations. The data are as in Danielsson and de Vries (1997), and a complete description of the data can be found there. Similarly, we use the first and last 5,000 daily returns from the daily S&P 500 index over the period 1928 to 1997. For these four datasets we compute a number of standard statistics. The mean and standard error are annualized by using a factor of 250 and 52558 for the stock index and FOREX 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 $1/\alpha$, β/α, $\hat{x}_{1/n}$, and $\hat{x}_{1/3n}$. Between brackets we give the 95% confidence band.

From Table 5 we note that the FOREX data are fairly symmetric, while the S&P data clearly reflects the abysmal period of the 1930's, and the fat years of the present decade. Nevertheless, the black Monday of 1987 is present in the last 5,000 S&P returns sample as can be seen from the minimum. The tail index estimates hover around 3. The second order parameter β appears to be larger than α. From an economic point of view the interesting estimates are the quantile estimates. Table 5 reveals that the risk of investing in stocks has come down over time, and that the risk in the FOREX market is fairly stable over the period of one year.

The other application concerns the Value-at-Risk (VaR) on a portfolio of assets. Financial institutions have to regularly assess their capital adequacy to cover adverse market movements, and have to report a number which reflects the (minimum) loss of their portfolio if a negative return in the lowest quantile materializes; this loss is called VaR. For example, the lowest quantile may be the 0.5% quantile, and the VaR is the loss that materializes if exactly this return materializes. Several methods are used in practice, e.g. historical simulation (HS), the J. P. Morgan RiskMetrics, and normal moving

Table 5: Daily S&P 500 and Olsen DM/US 10 Minute forex. Lower Tail

Dataset Observations	S&P 500 First 5,000	S&P 500 Last 5,000	10 minute DM/US First 5,000	10 minute DM/US Last 5,000
Annualized				
mean	⇔. 75%	10.0%	112.9%	⇔23.0%
s.e.	26.4%	14.7%	19.2%	12.8%
skewness	0.089	⇔3.126	0.464	⇔0.0814
Kurtosis	8.143	77.96	7.886	17.190
Minimum	⇔13.166	⇔22.8	⇔0.693	⇔0.655
$1/\alpha$	0.276	0.346	0.301	0.374
$(\cdot, \cdot)$	(0.24, 0.31)	(.31, .36)	(0.27, 0.32)	(0.33, 0.40)
β/α	1.97	1.105	1.60	1.760
$\hat{x}_{1/n}$	⇔13.3	⇔8.618	⇔0.690	⇔0.668
$(\cdot, \cdot)$	(⇔11.8, ⇔17.1)	(⇔7.82, ⇔10.9)	(⇔0.62, ⇔0.85)	(⇔0.58, ⇔0.91)
$\hat{x}_{1/3n}$	⇔18.0	⇔12.601	⇔0.958	⇔1.01
$(\cdot, \cdot)$	(⇔16.0, ⇔23. 1)	(⇔11. 4, ⇔15. 9)	(⇔. 86, ⇔1. 17)	(⇔. 88, ⇔1. 4)

Table 6: Estimation Results: Average Number of Realized Portfolios that were Larger than VaR Predictions

Tail Percentage	5%	0.5%	0.1%	0.005%
Expected Number of Violations	50	5	1	0.05
Expected Frequency of Violations	20 days	200 days	3.8 years	77 years
RiskMetrics	52.45(7.39)	10.65(2.73)	4.85(2.06)	1.58(1.29)
Historical Simulation	43.24(10.75)	3.69(2.39)	0.95(1.03)	—
Tail Estimator	43.14(11.10)	4.23(2.55)	1.06(1.13)	0.06(0.23)

Daily observations in testing = 1000 over period 930115 to 961230. Window size in HS and TK = 1500, initial staring date for window 870210. Random portfolios = 500. Standard errors in parenthesis. Probabilities expressed in precentages with sum=100%

average. HS is non-parametric and uses the lowest quantiles of a historical sample. RiskMetrics combines the normal innovations with an IGARCH model. For 500 randomly weighted portfolios of seven US stocks, i.e. J.P. Morgan, 3M, McDonalds, Intel, IBM, Xerox, and Exxon, we calculate the VaR on the basis of the first two industry standard methods, and our semi-parametric approach. To implement our procedure on the portfolio, we created vectors of portfolio returns by simulating from the original return per individual stock and then combining these on the basis of initial portfolio weights to arrive at a vector of portfolio returns (this is the same procedure as followed by historical simulation.) The length of the vector was set at 1,500. To this vector we applied our tail estimator procedure. We then reserve 1,000 days at the end of the sample, do sequential one day VaR prediction, and count the number of days where the realized return exceeded the VaR. The results of these different techniques towards the VaR problem are in Table (6).

We can see that the IGARCH normal based RiskMetrics performs well at the 5% level, but is unable to cope with lower probabilities. Using the empirical distribution function (historical simulation), gives results which are similar to the semi–parametric approach in sample. It can not provide estimates for the 0.005% level, or an event which happens once every 77 years, since we only have windows with 1500 days. The semi–parametric tail estimator preforms well at all probability levels, except those which are far inside the sample.

4 Conclusion

In this paper we develop a complete semi–parametric method for estimating large, i.e. borderline in–sample and out-of-sample, probability-quantile (P, Q) combinations for heavy tailed distributions. The key parameter to be estimated is the tail index which determines the tail shape and dominates the large (P, Q) estimators. For both tail index estimation and large (P, Q) estimation, it is essential to know the number of extreme order statistics that have to be taken into account. In this paper we solve the hitherto unknown determination of the optimal number of extreme order statistics by means of a two step subsample procedure in combination with a control variate type method. The subsample bootstrap procedure is essential for convergence in probability; a full sample bootstrap would deliver convergence in distribution. The theory is presented in a comprehensive and self contained manner.

In addition to establishing the theoretical properties of our estimators, we subject them to a number of Monte Carlo experiments, where we simulate from a variety of heavy tailed distributions and stochastic processes. These experiments are designed so that the finite sample properties of the estimators can be established for most DGP's. For the most frequently used fat tailed distributions, and stochastic processes such as GARCH, we demonstrate that the estimators have good performance.

There are a number of problems in economics where extreme value analysis plays a key role. We focus on two applications in finance. First we investigate the change in risk over time with of the daily SP-500 index, and 10 minute foreign exchange quotes, then we apply the method to the problem of determining capital requirements in trading portfolios with the Value-at-Risk method. We demonstrate that there are large gains to be made in precision if one uses the semi–parametric method advocated here, rather than a fully parametric or a non–parametric approach.

We are currently working on several other applications of our tail estimator, e.g. the problem of capital determination for financial institutions, optimal hedging, options pricing, and intra–day risk management. Further theoretical work is focused on the rate of convergence problems in regression estimators.

A Mathematical Derivations

Before we can proof the first two theorems we need the following calculus result. The argument is used repeatedly.

Lemma 1 *Given the model (3) and conditional on* $M \geq 1$

$$\mathrm{E}\left[u_k\left(s_n\right)\right] = k!\left(\frac{1}{\alpha^k} + \frac{b s_n^{-\beta}}{(\alpha+\beta)^k}\right) + o\left(s_n^{-\beta}\right), \tag{38}$$

for $s_n \to \infty$ *as* $n \to \infty$.

Proof. From calculus we have the following result:

$$\begin{aligned}
\alpha \int_s^\infty \left(\log \frac{x}{s}\right)^k x^{-\alpha-1} dx &= \alpha s^{-\alpha} \int_1^\infty (\log y)^k \, y^{-\alpha-1} dy \\
&= k s^{-\alpha} \int_1^\infty (\log y)^{k-1} \, y^{-\alpha-1} dy \\
&= \frac{k!}{\alpha^{k-1}} s^{-\alpha} \int_1^\infty y^{-\alpha-1} dy \\
&= \frac{k!}{\alpha^k} s^{-\alpha}.
\end{aligned}$$

Now apply this result twice to compute the conditional expected value of $(\log x/s)^k$ when the density adheres to (3). Hence, the conditional expectation in (38) follows from

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

as $s \to \infty$. ■

To obtain the mean of w_k one first takes a first order expansion of w_k in u_k and u_{k-1}. Subsequently, using the fixed threshold interpretation of $u_k(s_n)$, one needs to compute the conditional expected value of $(\log x/s)^k$. This is facilitated by means of the calculus result from Lemma (1) above.

Proof of Theorem 1. Develop $w_k(s_n)$ into a first order Taylor expansion of the ratio of the two arguments $u_k/k!$ and $u_{k-1}/(k \Leftrightarrow 1)!$ around the point $\left(1/\alpha^k, 1/\alpha^{k-1}\right)$ and with remainder o:

$$w_k = \frac{1}{\alpha} + \alpha^{k-1}\left(\frac{u_k}{k!} \Leftrightarrow \frac{1}{\alpha^k}\right) \Leftrightarrow \alpha^{k-2}\left(\frac{u_{k-1}}{(k \Leftrightarrow 1)!} \Leftrightarrow \frac{1}{\alpha^{k-1}}\right) + o. \tag{39}$$

By application of Lemma (1) we get asymptotically

$$\mathrm{E}\left[w_k(s_n) \Leftrightarrow \frac{1}{\alpha}\right] = \alpha^{k-1} \frac{bs^{-\beta}}{(\alpha+\beta)^k} \Leftrightarrow \alpha^{k-2} \frac{bs^{-\beta}}{(\alpha+\beta)^{k-1}} + o\left(s^{-\beta}\right).$$

From this result the claim in (6) is easily established. ■

We proceed by investigating the variance of the estimators. Before we can do so, we need to say more about the relation between the size of the threshold and the number of excesses. Note that the Bernoulli character of the empirical distribution function implies for M from (4):

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

Hence, by the weak law of large numbers we know that for the proportion of excesses over a fixed but large threshold s

$$\frac{M}{n} \to as^{-\alpha}\left[1 + bs^{-\beta}\right] + o\left(s^{-\alpha-\beta}\right), \quad \text{as } n \to \infty$$

in probability. Now let $s_n \to \infty$ as $n \to \infty$, and choose the sequence s_n such that

$$\frac{Ms_n^{\alpha}}{an} \to 1$$

in probability. This ensures that the contribution of the second order term becomes negligibly small in large samples. We can now state the result for the variance. The proof is similar to the proof for the mean but involves more terms.

Proof of Theorem 2. By using the Taylor expansion in (39) we have that

$$\begin{aligned}\mathrm{Var}\left[w_k\left(s_n\right) \Leftrightarrow \frac{1}{\alpha}\right] &= \alpha^{2k-2}\,\mathrm{Var}\left[\frac{u_k}{k!}\right] + \alpha^{2k-4}\,\mathrm{Var}\left[\frac{u_{k-1}}{(k \Leftrightarrow 1)!}\right] \\ &\quad \Leftrightarrow 2\alpha^{2k-3}\,\mathrm{Cov}\left[\frac{u_k}{k!}, \frac{u_{k-1}}{(k \Leftrightarrow 1)!}\right] + o.\end{aligned}$$

We calculate the various parts by using the definition of u_k in (4), the independence of X_i and X_j , and Lemma (1). For the variance part we find:

$$\begin{aligned}\mathrm{Var}\left[\frac{u_k}{k!}\right] &= \frac{1}{(k!)^2}\left\{\mathrm{E}\left[u_k^2\right] \Leftrightarrow \left(\mathrm{E}\left[u_k\right]\right)^2\right\} \\ &= \frac{1}{(k!)^2}\left\{\frac{1}{M}\,\mathrm{E}\left[\left(\log\frac{X_1}{s}\right)^{2k}\right] \Leftrightarrow \frac{1}{M}\left(\mathrm{E}\left[\left(\log\frac{X_1}{s}\right)^{k}\right]\right)^2\right\} \\ &= \frac{1}{M}\left\{\frac{(2k)!}{(k!)^2}\left[\frac{1}{\alpha^{2k}} + \frac{bs^{-\beta}}{(\alpha+\beta)^{2k}}\right] \Leftrightarrow \right. \\ &\quad \left.\left[\frac{1}{\alpha^{2k}} + \frac{2bs^{-\beta}}{\alpha^k(\alpha+\beta)^k} + \frac{b^2s^{-2\beta}}{(\alpha+\beta)^{2k}}\right]\right\} + o\left(\frac{1}{M}\right) \\ &= \frac{1}{M}\frac{1}{\alpha^{2k}}\left(\frac{(2k)!}{(k!)^2} \Leftrightarrow 1\right) + o\left(\frac{1}{M}\right).\end{aligned}$$

Where the last step follows from our assumption regarding s_n and M. Similarly,

$$
\begin{aligned}
& \operatorname{Cov}\left[\frac{u_k}{k!}, \frac{u_{k-1}}{(k \Leftrightarrow 1)!}\right] \\
= \; & \frac{1}{k!\,(k \Leftrightarrow 1)!}\frac{1}{M}\left\{ \mathrm{E}\left[\left(\log\frac{X_1}{s}\right)^{2k-1}\right] \Leftrightarrow \right. \\
& \left. \mathrm{E}\left[\left(\log\frac{X_1}{s}\right)^{k}\right] \mathrm{E}\left[\left(\log\frac{X_1}{s}\right)^{k-1}\right]\right\} \\
= \; & \frac{1}{k!\,(k \Leftrightarrow 1)!}\frac{1}{M}\left\{ (2k \Leftrightarrow 1)!\left[\frac{1}{\alpha^{2k-1}} + \frac{bs^{-\beta}}{(\alpha+\beta)^{2k-1}}\right]\right. \\
& \Leftrightarrow k!\,(k \Leftrightarrow 1)!\left[\frac{1}{\alpha^{2k-1}} + \frac{1}{\alpha^{k-1}}\frac{bs^{-\beta}}{(\alpha+\beta)^{k}} + \right. \\
& \left.\left. \frac{1}{\alpha^{k}}\frac{bs^{-\beta}}{(\alpha+\beta)^{k-1}} + \frac{\left(bs^{-\beta}\right)^2}{(\alpha+\beta)^{2k-1}}\right]\right\} + o\left(\frac{1}{M}\right) \\
= \; & \frac{1}{M}\frac{1}{\alpha^{2k-1}}\left(\frac{(2k \Leftrightarrow 1)!}{k!\,(k \Leftrightarrow 1)!} \Leftrightarrow 1\right) + o\left(\frac{1}{M}\right).
\end{aligned}
$$

Putting the various parts into place then yields the claim. ■

Proof of Theorem 5. Let $G_s(y)$ be the conditional distribution of $Y = \log X/s$, given that $X > s$. For a given threshold s the dominant terms of the mean $\mu(s)$, the variance $\sigma^2(s)$ and the third moment of Y can be easily obtained by using the Lemma (1). Note that the second moment is bounded away from zero, and that the third moment is bounded above as $s \to \infty$. By assumption and conditional on M, the random variables $\log X_1/s_n, ..., \log X_m/s_n$ are i.i.d. with distribution function $G_s(y)$. Therefore by Liapounov's double array central limit theorem with independence within rows, see (Serfling 1980, sect. 1.9.3), we have that

$$\frac{\sqrt{M}\,(u_k \Leftrightarrow \mu)}{\sigma}$$

converges in distribution to a standard normal distribution as $n \to \infty$. Note that as $n \to \infty$, the number of extreme order statistics $M \to \infty$ with probability 1. By (39) and Cramér's theorem (1974, sect. 28.4), it follows that $\sqrt{M}\,(\alpha w_k \Leftrightarrow 1)$ converges in distribution to a normal distribution. The mean and variance of this normal distribution readily follow from Theorems 1 and 2, and the fact that $\sqrt{M}/s_n^{\beta}$ converges in probability to $\overline{m}_n^{1/2}/\overline{s}_n^{\beta}$, where $\overline{m}_n$ and $\overline{s}_n$ are given in (12) and (10) respectively ■

Proof of Theorem 6. Use the Taylor expansion of (39) for w_2 to show that

$$z = w_2 \Leftrightarrow w_1 = \frac{1}{\alpha} \Leftrightarrow 2u_1 + \frac{\alpha}{2}u_2 + o.$$

By Lemma (1) we then find

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

And similar to the proof of Theorem (2), we derive

$$\begin{aligned} \mathrm{Var}\left[z\right] &= 4\,\mathrm{Var}\left[u_1\right] \Leftrightarrow 2\alpha\,\mathrm{Cov}\left[u_1,u_2\right] + \frac{\alpha^2}{4}\mathrm{Var}\left[u_2\right] + o\left(\frac{1}{M}\right) \\ &= \frac{4}{\alpha^2 M} \Leftrightarrow 4\alpha\frac{1}{\alpha^3 M}\left(\frac{6}{2} \Leftrightarrow 1\right) + \frac{\alpha^2}{4}\frac{4}{M\alpha^4}\left(\frac{24}{4} \Leftrightarrow 1\right) + o\left(\frac{1}{M}\right) \\ &= \frac{1}{\alpha^2}\frac{1}{M} + o\left(\frac{1}{M}\right). \end{aligned}$$

■

Proof of Corollary 2. From Theorem (6) we calculate the $\mathrm{AMSE}\left[z\right]$ as

$$\frac{1}{\alpha^2 a}\frac{s^{\alpha}}{n} + \frac{b^2\beta^4}{\alpha^2\left(\alpha+\beta\right)^4}s^{-2\beta}. \tag{40}$$

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

Proof of Theorem 8. Note that $\bar{m}_n\left(z\right)$ from (23) is itself a regularly varying function with tail index $2\beta/\left(\alpha+2\beta\right)$. By Proposition 1.7 from Geluk and de Haan (1987) on the properties of regularly varying functions we have that

$$\frac{\log \bar{m}_{n_1}\left(z\right)}{\log n_1} \to \frac{2\beta/\alpha}{1+2\beta/\alpha}$$

in probability as $n \to \infty$. Then use the fact that $\hat{m}_{n_1}\left(z\right)/\overline{m}_{n_1}\left(z\right) \to 1$ in probability. ■

Proof of Theorem 10. Use the threshold interpretation and write

$$\hat{x}_p = \hat{x}_p\left(w_k, M\right).$$

Expand $\hat{x}_p\left(.,.\right)$ into a first order Taylor series around the point

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

This gives

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

Where in the last result $x_p^{-\beta} \to 0$ and $x_t^{-\beta} \to 0$ is used. Thus the first term in the sequence is asymptotic to $\sqrt{m}\,(w_k \Leftrightarrow 1/\alpha)$. We already know that $\sqrt{m}\,(\alpha w_k \Leftrightarrow 1)$ converges in distribution to the normal distribution function by Theorem (5). The second term can be split into two terms:

$$\frac{1}{\alpha}\frac{\sqrt{m}}{\log(m/np)}\left\{ b\frac{\frac{M}{nt}x_t^{-\beta} \Leftrightarrow x_p^{-\beta}}{1+bx_p^{-\beta}} + \frac{\frac{M/n}{t} \Leftrightarrow 1}{1+bx_p^{-\beta}} \right\} np =$$

$$\frac{b}{\alpha}\left(x_t^{-\beta}\sqrt{m}\right)\frac{\frac{M/n}{t} \Leftrightarrow \left(\frac{x_p}{x_t}\right)^{-\beta}}{\log(m/np)} np +$$

$$\frac{1}{\alpha\left(1+bx_p^{-\beta}\right)}\frac{\sqrt{m}}{\log(m/np)}\left[\frac{M/n}{ax_t^{-\alpha}\left(1+bx_t^{-\beta}+o\left(x_t^{-\beta}\right)\right)} \Leftrightarrow 1\right] np,$$

and where we use the fact that $t = ax_t^{-\alpha}\left(1+bx_t^{-\beta}+o\left(x_t^{-\beta}\right)\right)$. Consider the first term. By assumption x_t is asymptotic to $cn^{1/(2\beta+\alpha)}$, while $\sqrt{m}$ is asymptotic to $\sqrt{a}c^{-\alpha/2}n^{\beta/(2\beta+\alpha)}$. Hence $\lim_{n\to\infty} x_t^{-\beta}\sqrt{m} = \sqrt{a}\; c^{-\beta-\alpha/2}$, constant. Evidently $\mathrm{plim}\; M/nt_n = 1$, so that $\mathrm{plim}\, M/\left[nt\log(m/np)\right] = 0$. Moreover

$$\lim_{n\to\infty}\left(\frac{x_p}{x_t}\right)^{-\beta} = \lim_{n\to\infty}\left(\frac{t}{p}\right)^{-\beta/\alpha} = \lim_{n\to\infty}\left(\frac{np_n}{m_n}\right)^{\beta/\alpha} = 0.$$

Thus the first term equals 0 in probability as $n \to \infty$ (recall $np \to \tau$ by assumption). With respect to the second term, recall that by assumption

$$M/n \to ac^{-\alpha}n^{-\alpha/(2\beta+\alpha)}$$

in probability. On the other hand by construction

$$ax_t^{-\alpha} \to ac^{-\alpha}n^{-\alpha/(2\beta+\alpha)}.$$

Hence

$$\underset{n\to\infty}{\mathrm{plim}}\left[\frac{M/n}{ax_t^{-\alpha}\left(1+bx_t^{-\beta}\right)} \Leftrightarrow 1\right] x_t^{\beta} = \Leftrightarrow b.$$

We already showed that $\lim_{n\to\infty} x_t^{-\beta}\sqrt{m}$ is constant and by assumption $\lim_{n\to\infty} np = \tau$. The factor in the denominator $\log\left(m/np\right)$, though, diverges. Therefore the second term converges to 0 as well. ■

Proof of Theorem 11. Develop $\hat{p}\left(M, 1/w_k\right)$ into a Taylor series around the point

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

Thus

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

From this point onwards, the proof is similar to the proof of the previous proposition. ■

References

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

Beirlant, Jan, Jozef L. Teugels and Petra Vynckier. 1994. Extremes in non-life Insurance. In *Extremal value theory and applications*, ed. J. Galambos, J. Lechner and Emil Simiu. Dordrecht: Kluwer pp. 489–510.

Booth, G. Geoffrey, John Paul Broussaard, Teppo Martikainen and Vesa Pattonen. 1997. “Prudent margin levels in the finnish stock index futures market.” *Management Science* , forthcoming.

Cramér, H. 1974. *Mathematical methods of statistics.* Princeton: Princeton University Press.

Danielsson, Jon and Casper G. de Vries. 1997. “Tail index and quantile estimation with very high frequency data.” *Journal of Empirical Finance* 4:241–257.

Danielsson, Jon, Dennis W. Jansen and Casper G. de Vries. 1996. “The method of moments ratio estimator for the tail shape parameter.” *Communication in Statistics, Theory and Methods* 25-4:711–720.

Danielsson, Jon, L. de Haan, L. Peng and C.G. de Vries. 1997. “Using a bootstrap method to choose the sample fraction in tail index estimation.”. Tinbergen Institute Rotterdam, TI 97-016/4.

de Haan, L., S. I. Resnick, H. Rootzen, and C. G. de Vries. 1989. “Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes.” *Stochastic Process and their Applications* 32:213–224.

de Haan, L. and U. Stadtmuller. 1996. “Generalized regular variation of second order.” *Journal of the Australian Mathematical Society, series A* 61:381–395.

Drees, H. and E. Kaufmann. 1997. “Selecting the optimal sample fraction in univariate extreme value estiamtion.”. Mimeo, University of Cologne and University of Siegen.

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

Geluk, J. and L. de Haan. 1987. “Regular variation, extensions and Tauberian theorems.”. Amsterdam: CWI Tract 40.

Gielens, Geert, Stefan Straetmans and Casper G. de Vries. 1996. “Fat tail distributions and local thin tail alternatives.” *Communication in Statistics, Theory and Methods* 25-4:705–710.

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

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, Peter. 1982. "On some simple estimates of an exponent of regular variation." *J. R. Statist. Soc. B* 42:37–42.

Hall, Peter and A.H. Welsh. 1985. "Adaptive estimates of parameters of regular variation." *Annals of Statistics* 13:331–341.

Hendry, David F. 1984. Monte Carlo experimentation in econometrics. In *Handbook of Econometrics vol. II,* ed. Z. Griliches and M. W. Intriligator. Amsterdam: North Holland pp. 939–976.

Hill, B. M. 1975. "A simple general approach to inference about the tail of a distribution." *Annals of Statistics* 35:1163–1173.

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

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

Kearns, P. and Adrian Pagan. 1997. "Estimating the density tail index for financial time series." *Review of Economics and Statistics* 79:171–175.

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

Leadbetter, M. R., G. Lindgren and H. Rootzen. 1983. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics New York: Springer-Verlag.

Longin, F. M. 1996. "The aymptotic distribution of extreme stock market returns." *Journal of Business* 69:383–408.

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

Resnick, Sidney and Cătălin Stărică. 1996. "Tail index estimation for dependent data.". mimeo, School of ORIE, Cornell University.

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