TI 2004-090/2
Tinbergen Institute Discussion Paper

Optimal Confidence Intervals for the Tail Index and High Quantiles

Ana Ferreira[1]
Casper G. de Vries[2]

[1] *ISA, Technical University of Lisbon, Portugal,*
[2] *Faculty of Economics, Erasmus Universiteit Rotterdam, and Tinbergen Institute.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Optimal confidence intervals for the tail index and high quantiles

Ana Ferreira and Casper G. de Vries*

August 17, 2004

Abstract

The aim of the paper is to obtain confidence intervals for the tail index and high quantiles taking into account the optimal rate of convergence of the estimators. The common approach to obtaining confidence intervals as presented in the literature is to use the normal distribution approximation at a non-optimal rate. Instead, we propose to use the optimal rate, but then a bias term with unknown sign has to be estimated. We provide an estimator for this sign and the full programme to obtain the optimal confidence intervals. Moreover, we demonstrate the gain in coverage, and show the relevance of these confidence intervals by calculating the reduction in capital requirements in a financial Value at Risk exercise. Simulation results are also presented.

It is well known that extreme value parameter estimators which balance the asymptotic bias squared and variance yield the lower asymptotic mean square error. Here we demonstrate the relevance of using the confidence bands for the quantiles using the optimal number of order statistics on simulated and actual data. It appears that if one does not correct for the sign factor the confidence bands are considerably larger. In the financial application for the determination of appropriate capital buffers usage of the optimal confidence band implies a considerable reduction in capital provisioning. The band without the correction term sometimes requires about 10% more capital vis a vis the optimal band. Since investment banks nowadays have to provision against such losses by holding capital, a reduction in capital requirements in the order of 10% gives quite a significant reduction in operating costs.

Key Words: tail index, bias sign, optimal rate, confidence intervals

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from some unknown distribution function F, and denote the order statistics by $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$.

*Ana Ferreira, ISA, Technical University of Lisbon, Portugal (email: anafh@isa.utl.pt); and Casper G. de Vries, Erasmus University Rotterdam, Tinbergen Institute, H14-25, PO Box 1738, 3000 DR Rotterdam, Netherlands (email: cdevries@few.eur.nl). De Vries is corresponding author.

Suppose F satisfies the maximum domain of attraction condition (Fisher and Tippett, 1928; Gnedenko, 1943), with positive extreme value index. In terms of regularly varying functions, this is equivalent to: For some $\gamma > 0$

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad \text{for all } x > 0. \tag{1}$$

Then F is said to have a regularly varying tail with index $-1/\gamma$ (i.e. $1-F \in \mathrm{RV}_{-1/\gamma}$).

For any non-decreasing function f, let $f^{\leftarrow}$ denote its left-continuous inverse, that is $f^{\leftarrow}(y) = \inf\{x : f(x) \geq y\}$. Let $U = (1/(1-F))^{\leftarrow}$. Consider the following refinement of condition (1) (e.g. de Haan, 1994; de Haan and Stadtmüller, 1996). Suppose there exists a function a, with constant sign near infinity and $\lim_{t\to\infty} a(t) = 0$, such that

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{a(t)} = \frac{x^\rho - 1}{\rho}, \quad \text{for all } x > 0,\ \rho < 0. \tag{2}$$

The following estimator of γ was proposed by Hill (1975),

$$\hat{\gamma}(k) = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}. \tag{3}$$

It is well known that Hill's estimator has, in general, large variance for small values of k and large bias for large values of k. Hence, when estimating γ, one usually looks for a k value which balances between these two vices.

Let k_n be an intermediate sequence, i.e. $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$, and

$$H_{n,k_n} = \sqrt{k_n} \left(\frac{\hat{\gamma}(k_n)}{\gamma} - 1 \right). \tag{4}$$

Under condition (2) and if $a(n/k_n)\sqrt{k_n} \to \lambda$, $\lambda \in (-\infty, \infty)$, H_{n,k_n} converges in distribution to a normal random variable with mean $\lambda/(\gamma(1-\rho))$ and variance 1 (e.g. Hall, 1982; Dekkers et al., 1989). The best rate of convergence is attained when $a(n/k_n)\sqrt{k_n} \to \lambda \neq 0$, and in this case the limiting distribution has non-zero mean.

Often when using Hill's estimator in applied problems, for simplicity one uses (4) with $\lambda = 0$ in order to construct a confidence interval for γ (e.g. Cheng and Peng, 2001 and Caers et al. 1998; though a host of other papers could be cited as well). The reason for this shortcut is perhaps that the construction of the confidence interval at the optimal rate involves other parameters, which have to be estimated. In section 2 we construct a confidence interval for gamma, using (4) but with the sequence k_n following the optimal rate of convergence, in the sense of minimizing the asymptotic mean square error. Moreover, we demonstrate the gain in coverage. In order to implement this confidence interval, several related problems have to be solved: One needs an adaptive way to obtain the optimal sequence k_n, moreover one needs to estimate two new parameters consistently, the second order parameter ρ and the sign of the asymptotic bias. For the adaptive choice of the optimal sequence k_n we follow Danielsson et al. (2001). For the estimation of the parameter ρ we follow Danielsson et al. (2001) and Fraga Alves et al. (2001). For the sign of the asymptotic bias, in Section 3 we introduce a new estimator and show its consistency. In Section 4 we obtain optimal confidence intervals for high quantiles. From the results it follows that the reverse problem of tail probability estimation and confidence interval construction can be obtained in a similar fashion, left to the reader. Section 5 gives simulation results. In section 6 a financial application is studied. In a Value-at-Risk exercise it is shown that the capital requirements of the financial intermediary can be reduced due to using the 'optimal' confidence band vis a vis the simple confidence band.

Related papers on confidence interval estimation are Caers et al. (1998) and Cheng and Peng (2001). In the first paper the authors also use the bootstrap methodology to obtain the optimal k_n, though rather differently from the methodology considered here, but then they end up considering $\lambda = 0$ to obtain the confidence intervals. In the Cheng and Peng paper the authors try to find k_n optimising the confidence interval but their criterion is quite different from ours. They look for the optimal sequence in the non-optimal range of values satisfying $a(n/k_n)\sqrt{k_n} \to 0$. Cheng and Peng (2001) also point out the importance of the sign of the asymptotic bias but they do not discuss explicitly its estimation.

We restrict ourselves to the case $\rho < 0$, since optimality results for the choice of k_n are well established for this case.

2 Optimal confidence intervals for the tail index

Let $a(n/k_n)\sqrt{k_n} \to \lambda \in \mathsf{R}$. Denote by Φ the standard normal distribution function and $z_\alpha = \Phi^{\leftarrow}(1-\alpha)$. As a first approach towards constructing a confidence interval with significance level α for γ, based on (4) and its limiting distribution, solve $-z_\alpha < H_{n,k_n} - \lambda/(\gamma(1-\rho)) < z_\alpha$ for γ to get

$$\frac{\hat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{z_{\alpha/2} + \sqrt{k_n}} < \gamma < \frac{\hat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{-z_{\alpha/2} + \sqrt{k_n}}, \tag{5}$$

provided $\sqrt{k_n} - z_{\alpha/2} > 0$.

Let k_n^0 denote the 'optimal' sequence, in the sense of minimizing the mean square error of the limiting distribution (Hall and Welsh, 1985; Dekkers and de Haan, 1993). Under our conditions this sequence is easily calculated if one assumes, moreover, that the regularly varying function a behaves, asymptotically, as a power function,

$$a(t) \sim \; ct^{\rho}, \quad c \neq 0, \tag{6}$$

as $t \to \infty$. Then, our assumptions are equivalent to assuming Hall's model

$$1 - F(x) = \mathrm{C}x^{-1/\gamma} \left(1 + \mathrm{D}x^{\rho/\gamma} + o(x^{\rho/\gamma})\right), \quad \mathrm{C} > 0, \mathrm{D} \neq 0, \quad x \to \infty$$

where, from (2) and (6) we have $D = c\gamma^{-1}\rho^{-1}\mathrm{C}^{\rho}$. Therefore (Hall and Welsh (1985))

$$k_n^0 \sim \left(\frac{\gamma^2(1-\rho)^2}{-2\rho\, c^2}\right)^{1/(1-2\rho)} \quad n^{-2\rho/(1-2\rho)}. \tag{7}$$

Then it is easy to see that the value λ for this sequence k_n^0, is asymptotic to $\mathrm{sign}(c)\gamma(1-\rho)/\sqrt{-2\rho}$. In this case (5) simplifies to

$$\frac{\hat{\gamma}(k_n^0)\sqrt{k_n^0}}{z_{\alpha/2} + \mathrm{sign}(c)/\sqrt{-2\rho} + \sqrt{k_n^0}} < \gamma < \frac{\hat{\gamma}(k_n^0)\sqrt{k_n^0}}{-z_{\alpha/2} + \mathrm{sign}(c)/\sqrt{-2\rho} + \sqrt{k_n^0}}.$$

Now, in order to obtain a confidence interval from this inequality, we need to approximate adaptively k_n^0, and estimate ρ and $\mathrm{sign}(c)$. From Hall and Welsh (1985, Th. 4.1), an adaptive choice $\hat{k}_n^0$ can be used for which

$$\frac{\hat{k}_n^0}{k_n^0} \xrightarrow{P} 1. \tag{8}$$

For ρ we need a consistent estimator and for $\mathrm{sign}(c)$ an estimator $\widehat{\mathrm{sign}}$ satisfying

$$\mathrm{P}\{\widehat{\mathrm{sign}} = \mathrm{sign}(c)\} \to 1 \qquad (n \to \infty). \tag{9}$$

Then the following theorem holds.

Theorem 2.1. *Suppose* (2) *and* (6)*. Let* $\hat{k}_n^0$ *satisfy* (8)*,* $\widehat{sign}$ *satisfies* (9) *and let* $\hat{\rho}$ *be a consistent estimator for* ρ*. Then, as* $n \to \infty$*,*

$$\sqrt{\hat{k}_n^0}\left(\frac{\hat{\gamma}(\hat{k}_n^0)}{\gamma} - 1\right) - \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}$$

converges, in distribution, to a standard normal random variable. Therefore, as $n \to \infty$*,*

$$P\left(\frac{\hat{\gamma}(\hat{k}_n^0)\sqrt{\hat{k}_n^0}}{z_{\alpha/2} + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}} + \sqrt{\hat{k}_n^0}} < \gamma < \frac{\hat{\gamma}(\hat{k}_n^0)\sqrt{\hat{k}_n^0}}{-z_{\alpha/2} + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}} + \sqrt{\hat{k}_n^0}}\right) \to 1 - \alpha \qquad (10)$$

which gives an asymptotic confidence interval for γ*, with confidence coefficient* $1 - \alpha$*.*

Note that for the cases where the true γ is near zero a one-sided confidence interval can alternatively be considered. The extension of our results to this case is obvious.

2.1 Accuracy of the confidence interval

Denote the confidence interval based on (5), where k_n is such that $\lambda = 0$ and ρ is replaced by a consistent estimator, by $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$. In the following we shall see that the confidence interval (10) is more accurate than $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$.

Fix

$$\lim_{n\to\infty} P(\gamma \in (\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})) = 1 - \alpha,$$

for each $\gamma > 0$. Then, the probabilities of covering the wrong value γ',

$$P(\gamma' \in [\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]) \qquad (11)$$

should be as small as possible, for each $\gamma' > 0$. In fact for $\gamma' \neq \gamma$ and all sequences k_n with $a(n/k_n)\sqrt{k_n} \to \lambda$ this probability converges to zero, since the lower and upper limits of the confidence interval converge to γ in probability. Therefore next we compare the probabilities of wrong coverage as $\gamma'_n/\gamma \to 1$ $(n \to \infty)$.

For the confidence interval (10) the probability of wrong coverage equals

$$\begin{aligned} P\Big(& - z_{\alpha/2}\frac{\gamma'_n}{\gamma} \\ & \le \sqrt{\hat{k}_n^0}\left(\frac{\hat{\gamma}(\hat{k}_n^0)}{\gamma} - 1\right) - \frac{\widehat{\text{sign}}}{\sqrt{-2\hat{\rho}}} + \left(1 - \frac{\gamma'_n}{\gamma}\right)\left(\sqrt{\hat{k}_n^0} - \frac{\widehat{\text{sign}}}{\sqrt{-2\hat{\rho}}}\right) \\ & \le z_{\alpha/2}\frac{\gamma'_n}{\gamma}\Big). \end{aligned}$$

Hence for sequences γ'_n with $\sqrt{k_n^0}(1 - \gamma'_n/\gamma) \to \nu \neq 0, \pm\infty$, this probability converges to $\Phi(z_{\alpha/2} - \nu) - \Phi(-z_{\alpha/2} - \nu)$. Now take k_n with $a(n/k_n)\sqrt{k_n} \to 0$.

Then for sequences γ_n^* with $\sqrt{k_n}(1-\gamma_n^*/\gamma) \to \nu^* \neq 0, \pm\infty$, probability (11) equals

$$P\left(-z_{\alpha/2}\frac{\gamma_n^*}{\gamma} + \sqrt{k_n}\left(\frac{\gamma_n^*}{\gamma}-1\right) \right.$$
$$\left. \leq \sqrt{\hat{k}_n}\left(\frac{\hat{\gamma}(k_n)}{\gamma}-1\right) \leq z_{\alpha/2}\frac{\gamma_n^*}{\gamma} + \sqrt{k_n}\left(\frac{\gamma_n^*}{\gamma}-1\right)\right)$$

which converges to $\Phi(z_{\alpha/2}-\nu^*) - \Phi(-z_{\alpha/2}-\nu^*)$. Therefore in order to compare the two probabilities take a common sequence, e.g. γ_n^*. Then in the first case the probability of covering the wrong values converges to zero, whilst in the second case it is equal to $\Phi(z_{\alpha/2}-\nu^*) - \Phi(-z_{\alpha/2}-\nu^*) > 0$.

3 Estimation of the sign of the bias of Hill's estimator

For convenience, in this section we shall use the Hill process parameterised continuously. The following result was taken from Drees et al. (2000), Corollary 1.

Lemma 3.1. *Let k_n denote an arbitrary intermediate sequence. Under condition (2), there exists a probability space carrying $X_1, X_2, \ldots$ and a sequence of Brownian motions W_n, such that*

$$\sup_{t_n \leq t \leq 1} t^{1/2}\left|\hat{\gamma}([k_n t]) - \left(\gamma + \frac{\gamma}{\sqrt{k_n}}\frac{W_n(t)}{t} + a(\frac{n}{k_n})\frac{t^{-\rho}}{1-\rho}\right)\right|$$
$$= o_p\left(k_n^{-1/2} + a(\frac{n}{k_n})\right)$$

for all $t_n \to 0$, satisfying $k_n t_n \to \infty$.

In this expansion for the Hill process the term

$$t^{-\rho}/(1-\rho)a(n/k_n)\sqrt{k_n},$$

is the bias of the Hill estimator. Note that if $a(n/k_n)\sqrt{k_n} \to \lambda \neq 0$, its asymptotic sign is determined by the sign of the function $a(n/k_n)$, which equals $\mathrm{sign}(c)$ provided (6) holds. For instance if $t = 1$, the sign of the expected value of the limiting variable of $\sqrt{k_n}(\hat{\gamma}-\gamma)$ equals $\mathrm{sign}(\lambda/(1-\rho))$ $= \mathrm{sign}(c)$.

Let a_n, b_n and c_n be intermediate sequences such that

$$a_n < b_n \leq c_n \text{ for all } n, \text{ and } a_n/b_n \to \nu \in [0, 1). \tag{12}$$

We suggest the following estimator for the sign of the bias,

$$\widehat{\text{sign}} = \text{sign}\left(\hat{\gamma}(c_n) - \frac{1}{b_n - a_n + 1}\sum_{i=a_n}^{b_n}\hat{\gamma}(i)\right). \tag{13}$$

Theorem 3.1. *Assume* (2), (12) *and moreover that* b_n *satisfies* $a(\frac{n}{b_n})\sqrt{b_n} \to \infty$. *Then*

$$P\{\widehat{sign} = sign(c)\} \to 1, \qquad n \to \infty.$$

Proof. Lemma 3.1 implies that

$$\int_{a_n/b_n}^{1} \left|\hat{\gamma}([b_n t]) - \left(\gamma + \frac{\gamma}{\sqrt{b_n}}\frac{W_n(t)}{t} + a(\frac{n}{b_n})\frac{t^{-\rho}}{1-\rho}\right)\right| dt = o_p(b_n^{-1/2} + a(\frac{n}{b_n})).$$

Therefore,

$$\lim_{n\to\infty}\frac{\hat{\gamma}(c_n) - \frac{1}{b_n-a_n+1}\sum_{i=a_n}^{b_n}\hat{\gamma}(i)}{a(\frac{n}{b_n})} = \lim_{n\to\infty}\frac{\hat{\gamma}(c_n) - \frac{b_n}{b_n-a_n}\int_{a_n/b_n}^{1}\hat{\gamma}_n([b_n t])dt}{a(\frac{n}{b_n})}$$

$$= \lim_{n\to\infty}\left\{\frac{\gamma}{a(\frac{n}{b_n})}\left(\frac{W_n(1)}{\sqrt{c_n}} - \frac{b_n}{\sqrt{b_n}(b_n - a_n)}\int_{a_n/b_n}^{1}\frac{W_n(t)}{t}dt\right)\right.$$
$$\left. + \frac{1}{1-\rho}\left(\frac{a(\frac{n}{c_n})}{a(\frac{n}{b_n})} - \frac{b_n}{b_n - a_n}\int_{a_n/b_n}^{1} t^{-\rho}dt\right) + o_p(\frac{1}{a(\frac{n}{b_n})\sqrt{b_n}})\right\}.$$

Since $a_n/b_n \to \nu \in [0, 1)$, we have that $(b_n/(b_n - a_n))\int_{a_n/b_n}^{1} W_n(t)/t\,dt = O_p(1)$. Hence taking $b_n \leq c_n$ such that $a(\frac{n}{b_n})\sqrt{b_n} \to \infty$, the first and last factors in the last equality go to zero, in probability. For the second factor, just note that under the given conditions $a(n/c_n)/a(n/b_n) \geq 1$ and that $(1/(1-\nu))\int_{\nu}^{1} t^{-\rho}dt < 1$, for all $\nu \in [0, 1)$ and $\rho < 0$. Hence we have that the second factor converges in probability to some positive constant. □

Remark 3.1. Although we have excluded the case $\rho = 0$ in condition (2), the results of this section are still valid for $\rho = 0$, provided $b_n < c_n$ in (12).

Remark 3.2. Other proposals to estimate the sign could be to use two consistent estimators of γ, for instance $\hat{\gamma}_1 = \hat{\gamma}$ and $\hat{\gamma}_2 = (2k_n)^{-1/2} \sqrt{\sum_{i=0}^{k_n-1}(\log X_{n-i,n} - \log X_{n-k,n})^2}$. Both admit expansions of the type $\hat{\gamma}_i =$

$\gamma + c_i k_n^{-1/2} P_i + d_i a(n/k_n) + o_p(k_n^{-1/2}) + o_p(a(n/k_n))$, where c_i, d_i are some known constants and P_i are normal (0,1) random variables. Hence, for large k_n (i.e. $a(n/k_n)k_n^{-1/2} \to \infty$) and with $a(n/k_n) \sim c\,(n/k_n)^{\rho}$ we have

$$\text{sign}\left(\frac{\hat{\gamma}_1}{\hat{\gamma}_2} - 1\right) = c\,\frac{d_1 - d_2}{\gamma}.$$

But typically one encounters two kind of problems with this sort of estimators. First of all, they are very sensitive to the choice of k. Commonly a plot of $\text{sign}(\hat{\gamma}_1\hat{\gamma}_2^{-1} - 1)$ versus k frequently alters sign for small and moderate values of k. Secondly, since these estimators have similar behaviour, e.g. they have the same sign of the bias and predominance of one bias over the other, in a plot of $\text{sign}(\hat{\gamma}_1\hat{\gamma}_2^{-1} - 1)$ for k large these features turn out to be the most predominant.

4 Optimal confidence intervals for high quantiles

Suppose one is given a small probability p and one wants to estimate the quantile $x : P(X > x) = p$. We are interested in studying the situations where p is indeed very small, for instance where this small probability corresponds to an event that has never been observed. More specifically, $p = p_n$ must depend on n (size of the sample), since we use asymptotic theory, and we shall assume $np_n \to$ constant (≥ 0).

As before let k_n be an intermediate sequence. To estimate a high quantile $x_n = F^{\leftarrow}(1 - p_n)$ we suggest the following estimator (Dekkers and de Haan, 1989; Ferreira et al., 2002)

$$\hat{x}(k_n) = X_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}(k_n)}. \tag{14}$$

The following result is from Ferreira et al. (2002).

Lemma 4.1. *Assume* (2), (6) *and* $np_n \to$*constant* (≥ 0). *Let* k_n *be an intermediate sequence such that*

$$a(n/k_n)\sqrt{k_n} \to \lambda \in (-\infty, \infty)$$

and

$$\log(k_n/np_n)/\sqrt{k_n} \to 0.$$

Then

$$\frac{\sqrt{k_n}}{\gamma \log(\frac{k_n}{np_n})}\left(\frac{\hat{x}(k_n)}{x_n} - 1\right) \tag{15}$$

converges in distribution to a normal random variable, with mean $\lambda/(\gamma(1-\rho))$ *and variance 1.*

Let ${}^{q}k_n^0$ denote the sequence k_n minimizing the mean square error of the limiting distribution of (15). Then from Ferreira et al. (2002) it follows that (when $\gamma > 0$) ${}^{q}k_n^0 \sim k_n^0$ where k_n^0 is from (7). Hence a consistent estimator of ${}^{q}k_n^0$ is $\hat{k}_n^0$ from Section 2, that is,

$$\frac{\hat{k}_n^0}{{}^{q}k_n^0} \xrightarrow{P} 1. \tag{16}$$

Moreover, also from this paper it follows that Lemma 4.1 still holds if k_n is replaced by $\hat{k}_n^0$.

Therefore, in analogy with Section 2, the following theorem holds. For more details on the proof we refer to Ferreira et al. (2002). Motivated by an application on VAR estimation, we consider here a one-sided confidence interval. The changes to obtain a similar result for a two-sided confidence interval are obvious.

Theorem 4.1. *Suppose* (2), (6), *and that* $np_n \to$ *constant* (≥ 0) *and* $\log(p_n) = o(n^{-\rho/(1-2\rho)})$, *as* $n \to \infty$. *Let* $\hat{k}_n^0$ *satisfy* (16), $\widehat{sign}$ *satisfy* (9) *and let* $\hat{\rho}$ *be a consistent estimator for* ρ. *Then, as* $n \to \infty$,

$$\frac{\sqrt{\hat{k}_n^0}}{\hat{\gamma}(\hat{k}_n^0)\log(\frac{\hat{k}_n^0}{np_n})}\left(\frac{\hat{x}(\hat{k}_n^0)}{x_n} - 1\right) - \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}$$

converges, in distribution, to a standard normal random variable. Therefore, as $n \to \infty$,

$$P\left(x_n < \hat{x}(\hat{k}_n^0)\left(1 + \frac{\hat{\gamma}(\hat{k}_n^0)\log(\frac{\hat{k}_n^0}{np_n})}{\sqrt{\hat{k}_n^0}}\left(-z_\alpha + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}\right)\right)^{-1}\right) \to 1-\alpha \tag{17}$$

which gives a one-sided asymptotic confidence interval for x_n, *with confidence coefficient* $1-\alpha$.

The analysis of the accuracy of the confidence intervals for the quantiles is analogous to the discussion in Section 2 regarding the confidence interval of the tail index, and hence is skipped.

5 Simulations

We evaluate the performance of the confidence intervals for the tail index and high quantiles. For the estimation of the optimal sequence k_n we follow the bootstrap algorithm proposed by Danielsson et al. (2001). This bootstrap procedure also provides an estimator of the second order parameter, say $\hat{\rho}_1$. In Appendix A we describe the main ideas behind this bootstrap procedure.

To estimate ρ we also investigate the following estimator proposed by Fraga Alves et al. (2001). Let $M_n^{(\alpha)}(k_n) = (1/k_n)\sum_{i=0}^{k_n-1}(\log X_{n-i,n} - \log X_{n-k,n})^{\alpha}$ and

$$T_n^{(1,2,3,0)}(k_n) = \frac{\log M_n^{(1)}(k_n) - \log\left(\frac{M_n^{(2)}(k_n)}{2}\right)/2}{\log\left(\frac{M_n^{(2)}(k_n)}{2}\right)/2 - \log\left(\frac{M_n^{(3)}(k_n)}{6}\right)/3}.$$

The estimator is given by

$$\hat{\rho}_2 = 3\frac{T_n^{(1,2,3,0)}(k_n) - 1}{T_n^{(1,2,3,0)}(k_n) - 3},$$

provided $1 \leq T_n^{(1,2,3,0)} < 3$ (otherwise we shall say it is not defined). Fraga Alves et al. (2001) proved consistency under condition (2) and for k_n satisfying $a(\frac{n}{k_n})\sqrt{k_n} \to \infty$.

To estimate the sign of the asymptotic bias we use (13) with $a_n = \log n$ and $b_n = c_n = n/\log\log n$ (see section 5.3 for more details).

We considered i.i.d. pseudo random numbers from the following distributions:

1. Student-t distribution with $\nu = 1$ and 4 degrees of freedom, for which $\gamma = 1/\nu$, $\rho = -2/\nu$, $a(t) \sim (2/3)\Pi^2 t^{-2}$ if $\nu = 1$ and $a(t) \sim (20/24\sqrt{3})\, t^{-1/2}$ if $\nu = 4$ (for the general formulas to obtain the scale constant of the function a we refer to Martins, 2000). Hence the sign of the bias is positive.

2. Fréchet distribution, $F_{\mu,\sigma}(x) = \exp\{-((x-\mu)/\sigma)^{-1/\gamma}\}$, for which we have, if $\mu \neq 0$ and $0 < \gamma < 1$ then $\rho = -\gamma$ and $a(t) \sim -(\mu\gamma/\sigma)t^{-\gamma}$; if $\gamma = 1$ then $\rho = -1$ and $a(t) \sim (1/2 - \mu/\sigma)t^{-1}/\sigma$; if $\mu = 0$ or $\gamma > 1$ then $\rho = -1$ and $a(t) \sim (\gamma/2)t^{-1}$. We shall consider (μ, σ, γ) equal to (0,1,1) and (1,1,1). Then note that when $\gamma = 1$ the sign of the bias equals the sign$(1/2 - \mu/\sigma)$.

5.1 Simulations for the tail index

Tables 1 and 2 present the simulation results based on 500 samples of size $n = 2\,000$. Table 1 gives the boostrap results, reporting the mean and standard deviation of $\hat{k}_n^0$, the mean and root mean square error of $\widehat{\gamma}$ and $\widehat{\rho}$, and the percentage of simulations in which the estimator of the sign of the bias yielded the correct sign.

Table 2 gives the results for the confidence intervals of size 98%, 96% and 90%. These where obtained by using the following alternative inputs in the computation of the confidence intervals:

a) $\hat{k}_n^0 + \hat{\rho}_1$ - (10) where the estimates of ρ are from the bootstrap procedure,

b) $\hat{k}_n^0 + \hat{\rho}_2$ - (10) where the estimates of ρ are from Fraga Alves et al. (2001),

c) $\hat{k}_n^0 + \rho$ - (10) with true ρ and correct sign(c),

d) $\hat{k}_n^0 + (\lambda = 0)$ - from (5) with $\lambda = 0$,

e) $(\hat{k}_n^0)^{.8} + (\lambda = 0)$ - from (5) with $k_n = (\hat{k}_n^0)^{.8}$ and $\lambda = 0$. Note that $(\hat{k}_n^0)^{.8}$ is a rather arbitrary choice, the only requirement that it should be smaller (of smaller order) than $\hat{k}_n^0$.

For each confidence interval $[\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]$, its coverage probability $P(\gamma \in [\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]) \sim 1 - \alpha$ was checked. Furthermore it is checked if coverage is equally weighted in each tail, where for the left-hand side it is desirable that $P(\gamma < \underline{\gamma}_{n,k_n}) \sim \alpha/2$ and for the right-hand side that $P(\gamma < \overline{\gamma}_{n,k_n}) \sim 1 - \alpha/2$. In Table 2 these are shown in the order: total coverage, left-hand side coverage and right-hand side coverage.

From Table 1 we see that for both distributions the tail index estimator performs reasonably well. There is more variation and bias in the second order parameter ρ, moreover one cannot rank the two alternative estimators as each performs better for one of the two distributions. For these two distributions, the sign estimator performs very well. The first distribution is the Cauchy model and sometimes produces extremely large realizations, which affect the quantile estimates, see fn.1 to the table.

$\hat{k}_n^0$		$\hat{\gamma}(\hat{k}_n^0)$		$\hat{\rho}_1$		$\hat{\rho}_2$		$\widehat{\text{sign}}$	$\hat{x}(\hat{k}_n^0)$	
mean	st.dev.	mean	rmse	mean	rmse	mean	rmse	% true	mean	rmse
$t_1 : \gamma = 1, \rho = -2, \text{sign} = +, x_n = 636.6$										
200.	117.	1.00	.16	-1.31	.81	-1.01	1.00	100.0	706.1[1]	353.5[1]
$t_4 : \gamma = .25, \rho = -.5, \text{sign} = +, x_n = 8.6$										
33.	37.	.29	.08	- .55	.21	-.69	.19	100.0	9.1	2.9
$\text{F}_{0,1} : \gamma = 1, \rho = -1, \text{sign} = +, x_n = 1999.5$										
414.	231.	1.03	.11	-2.13	1.44	-1.26	.31	99.6	2470.6	1161.5
$\text{F}_{1,1} : \gamma = 1, \rho = -1, \text{sign} = +, x_n = 2000.5$										
708.	247.	.94	.08	-3.58	2.79	-[2]	-[2]	91.0	1634.8	714.6

[1] without extreme quantile estimate of 378782281

[2] not defined in most of the samples

Table 1: Bootstrap estimates and percentage $\widehat{\text{sign}}$ equals the true sign, 500 samples of size 2 000 (see text for details).

Comparing the confidence intervals $(\hat{k}_n^0 + \hat{\rho}_1)$ and $(\hat{k}_n^0 + \rho)$, with $(\hat{k}_n^0 + (\lambda = 0))$, in general we consider the first two better (particularly for α small), since for these the mean lengths are smaller (except for Fréchet (1,1)) and the coverage probabilities are usually much closer to those expected. The fact that the mean lengths are larger in the cases $(\hat{k}_n^0 + \hat{\rho}_1)$ and $(\hat{k}_n^0 + \rho)$ for Fréchet (1,1), is due to the negative sign of the bias.

Sometimes the confidence intervals with $\lambda = 0$ seem to give close coverage probabilities in the right tail, but note that a coverage probability of 100% is totally non-informative, a situation so often obtained for these cases. Indeed what we get are biased confidence intervals. For all distributions associated with positive bias the upper limit of the confidence intervals are so large that they are too often larger than the true value. On the other hand, the lower limits are so large that again, they are too often larger than the true value. Systematically we see that the contribution to the coverage probability, considering both sides, for being less than 100% always comes from the wrong coverage of the lower limit. Similar considerations can be made for the distribution associated with negative bias.

From these simulations we learn that when the bias of Hill's estimator is positive, in general the right-hand side confidence intervals are quite precise; the same observation applies when the bias is negative for left-hand side confidence intervals.

Our simulations indicate that the inclusion of second order information in the construction of the confidence intervals gives significant improvement.

Table 2: Means of the center points, the confidence interval lengths and coverage probabilities, 500 samples of size 2 000 (see text for details).

CDF	Method	α = 2% mean center	mean length	Cov. Prob. 98%	1%	99%
t_1	$\hat{k}_n^0 + \hat{\rho}_1$	.98	.36	91.	3.	95.
	$\hat{k}_n^0 + \hat{\rho}_2$	.99	.37	92.	3.	95.
	$\hat{k}_n^0 + \rho$	1.01	.40	94.	4.	98.
	$\hat{k}_n^0 + (\lambda = 0)$	1.07	.48	92.	8.	100.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.13	.77	99.	1.	100.
t_4	$\hat{k}_n^0 + \hat{\rho}_1$	.29	.23	80.	14.	94.
	$\hat{k}_n^0 + \hat{\rho}_2$	.30	.26	85.	14.	99.
	$\hat{k}_n^0 + \rho$	.29	.24	85.	13.	98.
	$\hat{k}_n^0 + (\lambda = 0)$	.43	.48	79.	21.	100.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	.48	.62	90.	10.	100.
$F_{0,1}$	$\hat{k}_n^0 + \hat{\rho}_1$	1.02	.27	80.	17.	97.
	$\hat{k}_n^0 + \hat{\rho}_2$	1.01	.27	82.	15.	97.
	$\hat{k}_n^0 + \rho$	1.01	.27	83.	14.	97.
	$\hat{k}_n^0 + (\lambda = 0)$	1.06	.32	79.	21.	100.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.10	.56	98.	2.	100.
$F_{1,1}$	$\hat{k}_n^0 + \hat{\rho}_1$	.97	.19	69.	0.	70.
	$\hat{k}_n^0 + \rho$	.98	.20	80.	1.	81.
	$\hat{k}_n^0 + (\lambda = 0)$	.95	.18	61.	0.	61.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.02	.37	98.	0.	98.

CDF	Method	α = 4% mean center	mean length	Cov. Prob. 96%	2%	98%	α = 10% mean center	mean length	Cov. Prob. 90%	5%	95%
t_1	$\hat{k}_n^0 + \hat{\rho}_1$	.97	.31	85.	6.	91.	.96	.25	72.	10.	82.
	$\hat{k}_n^0 + \hat{\rho}_2$	.97	.32	87.	4.	91.	.96	.25	75.	9.	83.
	$\hat{k}_n^0 + \rho$	1.00	.34	89.	6.	95.	.98	.27	79.	10.	89.
	$\hat{k}_n^0 + (\lambda = 0)$	1.05	.40	88.	11.	99.	1.03	.30	81.	15.	96.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.08	.64	97.	2.	99.	1.04	.48	91.	5.	97.
t_4	$\hat{k}_n^0 + \hat{\rho}_1$	.27	.19	74.	15.	89.	.26	.14	63.	20.	83.
	$\hat{k}_n^0 + \hat{\rho}_2$	.29	.21	82.	15.	97.	.27	.16	70.	20.	90.
	$\hat{k}_n^0 + \rho$	.28	.20	81.	14.	95.	.26	.15	68.	18.	86.
	$\hat{k}_n^0 + (\lambda = 0)$	.38	.36	75.	25.	100.	.34	.25	67.	33.	100.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	.41	.46	87.	13.	100.	.35	.31	82.	17.	100.
$F_{0,1}$	$\hat{k}_n^0 + \hat{\rho}_1$	1.01	.23	75.	20.	95.	1.00	.18	67.	23.	90.
	$\hat{k}_n^0 + \hat{\rho}_2$	1.01	.24	78.	18.	95.	1.00	.19	68.	22.	90.
	$\hat{k}_n^0 + \rho$	1.00	.23	78.	17.	95.	1.00	.19	69.	21.	90.
	$\hat{k}_n^0 + (\lambda = 0)$	1.05	.27	76.	24.	100.	1.04	.21	67.	31.	97.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.07	.47	96.	3.	99.	1.05	.36	90.	9.	99.
$F_{1,1}$	$\hat{k}_n^0 + \hat{\rho}_1$	.96	.17	59.	1.	60.	.96	.13	46.	3.	49.
	$\hat{k}_n^0 + \rho$	.98	.17	73.	2.	75.	.97	.14	57.	4.	61.
	$\hat{k}_n^0 + (\lambda = 0)$	.95	.16	52.	0.	52.	.95	.13	43.	1.	44.
	$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.01	.32	96.	1.	96.	1.00	.26	90.	2.	92.

Distribution	α = 2% mean upp lim	Cov. Prob. 98%	α = 4% mean upp lim	Cov. Prob. 96%	α = 10% mean upp lim	Cov. Prob. 90%
t_1						
$\hat{k}_n^0 + \hat{\rho}_1$	1631.7	93.	1268.5	85.	951.6	70.
$\hat{k}_n^0 + \hat{\rho}_2$	1584.8	93.	1215.9	86.	916.4	69.
$\hat{k}_n^0 + \rho$	2225.6	96.	1434.0	92.	1022.6	77.
$\hat{k}_n^0 + (\lambda = 0)$	6693.6	98.	2731.7	98.	1477.6	93.
t_4						
$\hat{k}_n^0 + \hat{\rho}_1$	11.4	80.	10.6	73.	9.6	62.
$\hat{k}_n^0 + \hat{\rho}_2$	11.7	85.	10.9	80.	9.9	67.
$\hat{k}_n^0 + \rho$	11.3	82.	10.6	77.	9.6	64.
$\hat{k}_n^0 + (\lambda = 0)$	14.8	95.	13.5	93.	11.9	86.
$F_{0,1}$						
$\hat{k}_n^0 + \hat{\rho}_1$	5012.8	94.	4141.0	91.	3299.2	84.
$\hat{k}_n^0 + \hat{\rho}_2$	4769.2	95.	3919.2	91.	3143.9	82.
$\hat{k}_n^0 + \rho$	4478.5	95.	3739.8	91.	3028.7	81.
$\hat{k}_n^0 + (\lambda = 0)$	9482.0	99.	6566.3	99.	4283.6	95.
$F_{1,1}$						
$\hat{k}_n^0 + \hat{\rho}_1$	7045.5	87.	5099.8	81.	3274.5	66.
$\hat{k}_n^0 + \rho$	11181.1	97.	5779.6	93.	3753.8	80.
$\hat{k}_n^0 + (\lambda = 0)$	4209.4	83.	3801.4	75.	2697.5	53.

Table 3: Means of upper limits of the quantile confidence intervals and estimated coverage probabilities, 500 samples of size 2 000 (see text for details).

5.2 Simulations for high quantiles

Table 3 gives the simulation results for the one-sided confidence intervals of sizes 98%, 96% and 90%. These are based on the same samples used in tail index estimation. Note the insensitivity of the confidence intervals under the zero bias assumption for t_1 and $F_{0,1}$, where the coverages remain the same whether $\alpha = 2\%$ or 4%. Note the very large upper confidence limits when compared with the others considering the bias information.

5.3 Additional considerations on the sign estimation

The sign estimator also depends on the chosen values of the tuning parameters a_n, b_n and c_n. We found that for many common distributions the choice of $a_n = \log n$ and $b_n = c_n = n/\log\log n$ is quite reasonable. Mainly we have just chosen simple sequences verifying the conditions of Theorem 3.1.

While these choices may appear reasonable and the Table 1 results regarding the sign estimator appear comforting, the following example shows that there are cases in which the estimator requires a large data set and a juidicous choice of the number of observations that have to be taken into account. Suppose the logarithmic daily stock returns net of trend growth follow a symmetric Student distribution. Empirically the Student-t with about three or four degrees of freedom yields a decent fit. Although the i.i.d. assumption neglects the observed clustering of volatilities, the Student law describes the unconditional stock returns distribution pretty well. The density of the Student distribution with 3 degrees of freedom reads

$$f(x) = 2\pi^{-1}3^{-1/2}\left[1 + x^2/3\right]^{-2}.$$

By direct integration, one finds

$$F(x) = \frac{1}{2}\frac{2x\sqrt{3} + 6\arctan\frac{1}{3}x\sqrt{3} + 2\left(\arctan\frac{1}{3}x\sqrt{3}\right)x^2 + 3\pi + \pi x^2}{\pi\left(3 + x^2\right)}.$$

To understand the behavior of the sign estimator, we obtain the expansion of the distribution at large quantiles. Recall Hall's (1982) second order expansion,

$$1 - F(x) = Cx^{-\alpha}[1 + Dx^{-\beta} + Sx^{-\phi} + o(x^{-\phi})], \quad x \to \infty \tag{18}$$

where $C > 0, D \neq 0, S \neq 0$ and $\phi > \beta$. Here $\alpha = 1/\gamma$ and $\beta = -\rho/\gamma$. By the monotone density theorem (18) implies

$$f(x) = \alpha Cx^{-\alpha-1}[1 + \frac{\alpha+\beta}{\alpha}Dx^{-\beta} + \frac{\alpha+\phi}{\alpha}Sx^{-\phi} + o(x^{-\phi})]. \tag{19}$$

By the transformation $y = x^{-2}$ and a Taylor expansion around $y = 0$ of the Student-3 density, one obtains

$$f(x) = 3\frac{2\sqrt{3}}{\pi}x^{-4}[1 - 6x^{-2} + 27x^{-4} + o(x^{-4})].$$

Thus $\alpha = 3$, $\beta = 2$, $\phi = 4$, $C = 2\sqrt{3}/\pi$, $D = -18/5$, $S = 81/7$.

Since logarithmic returns are additive, one can find the distribution of the two-day return from the convolution of the Student distribution. The density of the 2-convoluted Student-3 is obtained by integration:

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x-s)f(s)ds \\ &= 12\sqrt{3}\pi^{-1}\left(x^2 + 60\right)\left(12 + x^2\right)^{-3}. \end{aligned}$$

An approximation similar to the one for the Student-3 gives

$$g(x) = 3\frac{4\sqrt{3}}{\pi}x^{-4}[1 + 24x^{-2} - 1296x^{-4} + o(x^{-4})]$$

Hence for the convoluted Student law, we find $\alpha = 3$, $\beta = 2$, $\phi = 4$, $C = 4\sqrt{3}/\pi$, $D = 72/5$, $S = -3888/7$.

The interesting point to note is that the signs of the second and third order scaling constants D and S differ between the student law and its convolution. Thus the sign estimator should register a switch in sign. But whether one will be successful in registering such a sign switch depends on the bias properties of the Hill estimator. The sheer size of the scale of the coefficient S for the convoluted law hampers registering this sign switch in practice. To see this note that, cf. De Haan (1990),

$$\begin{aligned} E[\widehat{1/\alpha}|s] &= \frac{1}{1-F(s)} \int_s^\infty \log(\frac{x}{s}) f(x) dx \qquad (20) \\ &= \frac{1}{\alpha} - D\frac{\beta}{\alpha(\alpha+\beta)} s^{-\beta} + \{D^2 \frac{\beta}{\alpha(\alpha+\beta)} - S\frac{\phi}{\alpha(\alpha+\phi)}\} s^{-2\beta} \\ &\quad + o(s^{-2\beta}),\ s \to \infty. \end{aligned}$$

By using the coefficients derived above, we obtain for the Student law

$$E[\widehat{\gamma}|s] \approx 0.333 + 0.480 s^{-2} - 0.476 s^{-4},$$

while for the convoluted model

$$E[\widehat{\gamma}|s] \approx 0.333 - 1.920 s^{-2} + 133.443 s^{-4}. \qquad (21)$$

The Student law has the first order term in $E[\widehat{\gamma}|s]$ positive and the second negative, while for the convoluted law the first term is negative and the second term is positive. The size of the second order term in the expansion of the convoluted law is also relatively large.

If we plot the expressions for $E[\widehat{1/\alpha}|s]$ at different values for the threshold values this implies quite a distinct pattern for the two models. The plot in Figure 1 gives the values for $E[\widehat{1/\alpha}|s]$ for the laws and the approximate expression $E[\widehat{\gamma}|s]$ from (21) for the convoluted distribution. Along the Y-axis we plot γ. To make the X-axis readable, we have transformed the quantiles s as follows

$$(1 - F(s)) * 10^5.$$

Thus the number along the X-axis can be interpreted as the number of highest order statistics which are needed in a sample of ten thousand observations to yield an estimate of the tail index if there were no variance. The plot is therefore comparable to a so called Hill plot that would obtain if one used actual data. Note that large values of s correspond to a low

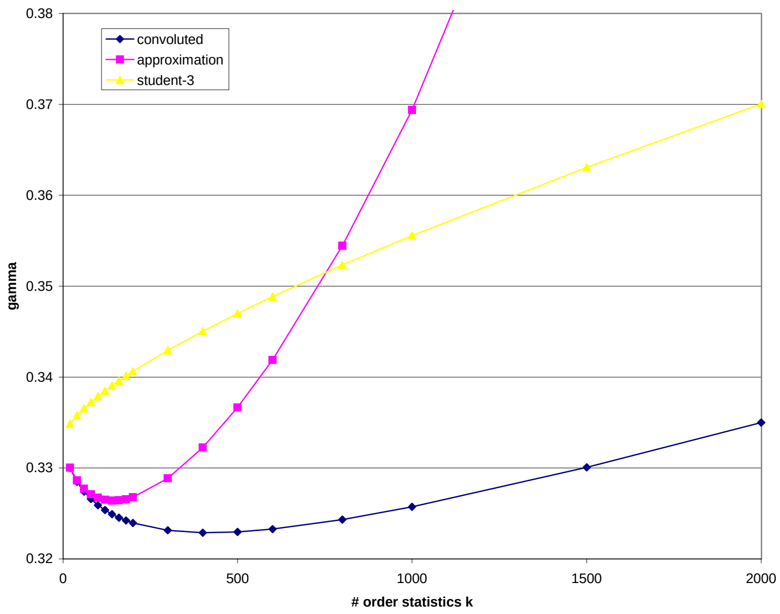

Figure 1: Theoretical Hill plot; sample size 100000.

number of 'order statistics'. In Figure 1 the two plots $E[\widehat{1/\alpha}|s]$ are based on (20) using the integrals directly, while $E[\widehat{\gamma}|s]$ is based on the second order approximation from (21) for the convoluted model. The Student model has a one way bias, but the bias implied by the convoluted distribution is clearly U-shaped.[1] The second order approximation picks up the U-shaped form but quickly deteriorates if one goes too much inside the sample by using too many of the highest order statistics.

In practice it may thus be difficult to assess the correct sign if the Hill plot is U-shaped. Note that this example is not far fetched, since e.g. all symmetric stable distributions with characteristic coefficients between one and two also imply alternating signs in the expansion $E[\widehat{\gamma}|s]$. The Student example is nice since it has such different behavior for the convoluted and unconvoluted model, and because of the additivity of logarithmic returns. In simulations for the Student model and the convoluted data, we also observed

[1] To get some intuition for the fact that the Student model has a one way bias, note the similarity of the absolute values of the coefficients in the approximation $E[\widehat{\gamma}|s]$ for the student law.

the U-shape pattern for the convoluted law, but we noted that in some runs the U-shape is buried under the uncertainty in the data.

6 Data analysis

We started the paper by noting that it is well known that estimators which balance the asymptotic bias squared and variance yield the lower asymptotic mean square error. Nevertheless, in practice confidence bands are commonly based on the estimators evaluated at the asymptotically suboptimal number of order statistics (taking $\lambda = 0$), such that the signed bias factor is omitted. Here we demonstrate the relevance of using the confidence bands for the quantiles using the optimal number of order statistics on actual data. It is shown that these can yield a considerable reduction in capital loss estimates.

We used daily price quotes over the period 1-1-1980 to 14-5-2002 on four quite different financial series, each of them comprising 5835 observations; all available from datastream. The first contract is the US dollar per UK pound spot foreign exchange rate contract, abbreviated as the forex contract. The forex contract is also of interest since forex risk is an important risk driver in international portfolios of pension funds. The second series is the S&P500 total return index, and the third contract is the Dutch Nedlloyd share price. The latter contract is known to be very volatile due to the cyclical business of sea transport, while the US index is naturally better diversified and hence less volatile, compare e.g. the S&P and Nedlloyd quantile estimates at $p = 1/n$ given in Table 4. The fourth contract is the French based Alcatel stock. The daily price quotes p_t are used to compute daily continuously compounded returns r_t by taking the logarithmic first differences of the price series, i.e. $r_t = \ln(p_t/p_{t-1})$. Since forex data for currencies from countries with similar monetary policies are known to be symmetrically distributed, we used the absolute returns for the forex series (except for the few zero quotes). Stock returns generally exhibit a positive mean due to positive growth of the economy. Therefore for the stock return data we focussed on the loss returns only. Still, the loss returns comprised approximately 50% of the data.

In Table 4 the tail parameter estimates are displayed. The gamma point estimates indicate that the number of bounded moments are between 3 and 5. We record the bootstrap based rho estimate from Danielsson et al. (2001) as $\hat{\rho}_1$, and the one based on Fraga Alves et al. (2001) recorded as $\hat{\rho}_2$. It can be seen that these differ quite considerably, but as we will see later, this difference is not so important for the construction of the confidence

Series	$\hat{k}_n^0$	$\hat{\gamma}(\hat{k}_n^0)$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\widehat{sign}$	$\hat{x}_{n-\hat{k}_n^0}$	$\hat{x}(\hat{k}_n^0)$ at $p = 1/n$
forex	8	0.201	-0.310	-0.639	+	0.028	0.043
S&P	225	0.308	-1.263	-0.711	+	-0.016	-0.090
Nedlloyd	12	0.255	-0.430	-0.725	+	-0.106	-0.201
Alcatel	34	0.281	-0.641	-0.739	+	-0.077	-0.207

Table 4: Parameter estimates

bands as is the inclusion of the sign correction factor. Nonetheless it is worth mentioning the proximity of all the estimates obtained from $\hat{\rho}_2$. The subsample bootstrap estimates of the optimal number of order statistics $\hat{k}_n^0$ is on the low side for the first and third series. The procedure sometimes runs into boundary problems due to insufficient data. In case of the forex contract, the plot of the bootstrap constructed mean square error reveals the surface is very flat over the range between $k = 8$ and 20 approximately, so that the global minimum is difficult to locate. The mean square error plot for the S&P and Alcatel series reveals unique and clearly identifiable minima, while the forex and Nedlloyd mean square error plots display multiple local minima for small values of k.

In Figures 2 and 3 we have plotted the Hill estimator against the number of order statistics k for the two individual stock series. The patterns of the Hill plots should be compared with the theoretical plots given in Figure (1). The Nedlloyd stock seems to have a one way upward bias, like the case of the Student law from the previous section. But the behavior of the Alcatel stock is quite different and displays the U-type behavior we noticed for the convoluted Student law. For this reason the positive sign we estimate for the Alcatel stock is perhaps not entirely convincing. In analogy with the plot for the convoluted Student law, the possible negative sign may be hard to notice due to a large positive third order coefficient.

The confidence bands for the tail index gamma are displayed in Table 5. We give three different bands at three different confidence levels (at the 2%, 4% and 10% level respectively). The first band is the sign factor corrected (optimal asymptotic mean square error) band, the second is the zero λ based band used in most studies.[2] The third band is also sign factor corrected, but where $\hat{\rho}_2$ is used instead of $\hat{\rho}_1$. There are some differences between the

[2] Some studies may on purpose prefer the estimates evaluated such that $\lambda = 0$, since the criterion function gives more (negative) weight to the asymptotic bias term. For these cases it is difficult to pick a specific number of order statistics, since such studies usually do not provide an automatic procedure for selecting the number of order statistics. Hence, even if the sign factor is ignored in the construction of the confidence band, we still use the same number of order statistics as for the case when the sign factor is included.

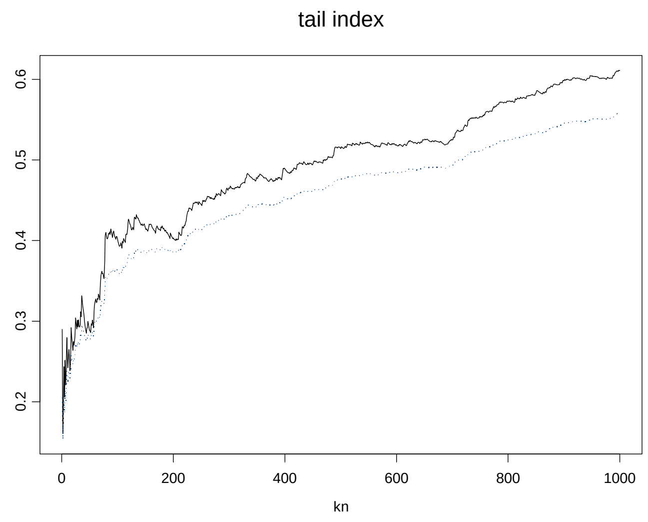

Figure 2: Tail estimates for Nedloyd data. The dotted line represents the alternative estimator in the bootstrap.

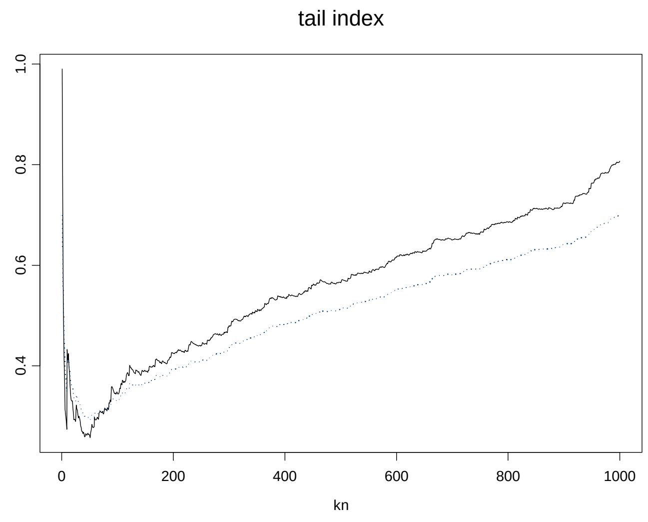

Figure 3: Tail estimates for Alcatel data. The dotted line represents the alternative estimator in the bootstrap.

series	α = 2%		α = 4%		α = 10%	
	LL	UL	LL	UL	LL	UL
forex						
signcorr.	.09	.32	.09	.28	.10	.23
nocorr.	.11	1.14	.12	.74	.13	.48
signcorr.(FA)	.09	.41	.10	.34	.11	.28
S&P						
signcorr.	.26	.35	.26	.34	.26	.33
nocorr.	.27	.36	.27	.36	.28	.35
signcorr.(FA)	.25	.34	.26	.34	.27	.33
Nedlloyd						
signcorr.	.13	.40	.13	.36	.14	.31
nocorr.	.15	.78	.16	.63	.17	.49
signcorr.(FA)	.13	.45	.14	.40	.15	.33
Alcatel						
signcorr.	.18	.37	.19	.35	.20	.32
nocorr.	.20	.47	.21	.43	.22	.39
signcorr.(FA)	.18	.38	.19	.36	.20	.33

Table 5: Tail index confidence bands

first and the last band, but the most glaring differences are in comparison with the second band. It appears that if one does not correct for the sign factor the confidence bands are considerably larger. This is basically due to a larger upper limit UL, the lower limits more or less all coincide. But the exception is the S&P series, where all three are quite close. The latter is due to the larger ρ values, see (17) for the influence of the second order parameter ρ. On the other hand, the larger is $\hat{k}_n^0$ the lower is the influence of the second order components.

A confidence band for the quantile estimates hinges on the choice of the quantile. We decided to report the quantiles located at the border of the sample, i.e. we took $p = 1/n$. Results are in Table 6. As in the previous table we report three different type of bands. Since these are about the possible loss, we report the left one-sided confidence interval. To indicate that we worked with the absolute returns in case of the forex series, the loss quantiles are reported positively in this case. Again the band based on the zero λ presumption yields much higher loss limits at the desired confidence level. What does this imply economically speaking? Consider e.g. the case of Nedlloyd, and suppose that an investment bank has taken a stake of 10 million in the company. From the first column labeled "$\widehat{q}(\hat{k}_n^0)$ at $p = 1/n$" in Table 6 one sees that once per 22 years there is a day on which this investment bank loses two or more million of its ten million investment. But taking into account the uncertainty pertaining to this estimate, one has to add another half million at the 2% level if one uses the bias corrected

series	$\hat{x}(\hat{k}_n^0)$ at $p = 1/n$	$\hat{x}_{\alpha=2\%}$	$\hat{x}_{\alpha=4\%}$	$\hat{x}_{\alpha=10\%}$
forex	.043			
signcorr.		.048	.046	.043
nocorr.		.062	.053	.053
signcorr.(FA)		.052	.049	.045
S&P	-.090			
signcorr.		-.107	-.103	-.097
nocorr.		-.116	-.112	-.105
signcorr.(FA)		-.104	-.100	-.094
Nedlloyd	-.201			
signcorr.		-.245	-.229	-.209
nocorr.		-.323	-.297	-.263
signcorr.(FA)		-.259	-.242	-.219
Alcatel	-0.207			
signcorr.		-0.259	-0.243	-0.223
nocorr.		-0.319	-0.296	-0.265
signcorr.(FA)		-0.263	-0.246	-0.225

Table 6: Quantile confidence bands

band. The band without the correction term requires quite a bit more, i.e. at least 1.2 million extra! Since investment banks nowadays have to provision against such losses by holding capital, a reduction in capital requirements by 0.7 million on a investment of 10 million gives quite a significant reduction in costs. Compare this case with the case of an investment in the S&P composite. For the case of an index investor with 10 million invested in the S&P composite, the extra loss stemming from the use of the confidence band without correction factor is more moderate (an extra hundred thousand).

7 Conclusion

The paper obtains confidence intervals for the tail index and high quantiles taking into account the optimal rate of convergence of the estimators. The common approach to obtaining confidence intervals in the applied literature is to ignore the bias term with unkown sign and use the zero bias approximation at a sub-optimal rate. We provide an estimator for the sign in the bias part and present the full programme for obtaining the optimal confidence intervals. Simulations demonstrate quite a considerable gain regarding the width of the confidence interval and regarding the coverage. In practice the tighter confidence intervals imply a considerable reduction in capital requirements for financial firms who must provision against high losses. Future research should perhaps be focussed on investigating the sensitivity of the sign estimator to higher order terms.

Appendix A. Tail index and quantile bootstrap estimation

The adaptive bootstrap method proposed by Danielsson et al. (2001) is a two-step sub-sample bootstrap method. From a sample of size n, in a first step take r independent bootstrap sub-samples of size n_1, where n_1 must be of the order $n^{1-\varepsilon}$, $0 < \varepsilon < 1/2$. For the simulations we took ε equal to .05 in all cases. For r we used 500. Let ${}^1\hat{\gamma}(k_n)$ and ${}^2\hat{\gamma}(k_n)$ be two consistent estimators of γ. Then, let

$$k_1^* = \ \text{argmin}_k \ \frac{1}{r} \sum_{i=1}^{r} \left({}^1\hat{\gamma}_i^*(k) - {}^2 \hat{\gamma}_i^*(k) \right)^2 \tag{A.1}$$

where ${}^1\hat{\gamma}_i^*$ and ${}^2\hat{\gamma}_i^*$ are the estimates based on the i-th bootstrap sub-sample of size n_1. In a second step, repeat step 1 but with n_1 replaced by $n_2 = n_1^2/n$, to get k_2^* say. Then, it can be shown that

$$\hat{k_n^0} = \frac{(k_1^*)^2}{k_2^*} C(\hat{\gamma}, \hat{\rho})$$

is a consistent estimator of k_n^0, where $C(\gamma, \rho)$ is some known constant depending on γ and ρ, and $\hat{\gamma}$ and $\hat{\rho}$ are consistent estimators of γ and ρ respectively. To estimate ρ one shows that

$$\hat{\rho} = \frac{\log k_1^*}{-2 \log n_1 + 2 \log k_1^*}$$

is a consistent estimator.

When estimating quantiles a similar algorithm can be used, where in (A.1) ${}^1\hat{\gamma}(k_n)$ and ${}^2\hat{\gamma}(k_n)$ are replaced by quantile estimators. Still, since $k_n^0/{}^q k_n^0 \sim 1$ both procedures with gamma or quantile estimators provide a consistent estimator of ${}^q k_n^0$.

Table A.1 gives the bootstrap results from minimizing the bootstrap mse based on two quantile estimators. Compare these results with the ones in Table 1. The quantile estimates in Table A.1 are more accurate, but the ρ (and γ) estimates are less accurate. It turns out that in terms of confidence intervals, the bootstrap mse based on the quantile estimators yields results comparable to the confidence intervals based on the bootstrap mse of the tail index estimators.

$^q\hat{k}_n^0$		$\hat{\gamma}(^q\hat{k}_n^0)$		$\hat{\rho}_1$		$\hat{x}(^q\hat{k}_n^0)$	
mean	st.dev.	mean	rmse	mean	rmse	mean	rmse
156.	130.	.94	.19	-1.08	1.08	621.0	332.4

Table A.1: Bootstrap estimates, $t_1 : \gamma = 1, \rho = -2, \text{sign} = +, x_n = 636.6$.

References

[1] Caers, J., Beirlant, J. and Vynckier, P. (1998). Bootstrap confidence intervals for tail indices. *Comput. Stat. and Data Analysis* **26**, 259-277.

[2] Cheng, S. and Peng, L. (2001). Confidence intervals for the tail index. *Bernoulli* **7**, 751-760.

[3] Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Analysis* **76**, 226-248.

[4] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833-1855.

[5] Dekkers, A.L.M. and de Haan, L. (1989). On the estimation of the extreme value index and large quantile estimation. *Ann. Statist.* **17**, 1795-1832.

[6] Dekkers, A.L.M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Analysis* **47**, 173-195.

[7] Drees, H. de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254-274.

[8] Ferreira, A. de Haan, L. and Peng, L. (2002). On optimizing the estimation of high quantiles of a probability distribution. To appear in Statistics. EURANDOM Technical Report 1999-042, Eindhoven.

[9] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24**, 180-190.

[10] Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2002). A new class of semi-parametric estimators of the second order parameter. *Accepted for publication in Portugaliæ Mathematica.*

[11] Gnedenko, B.V. (1943). Sur la distribution limite du terme du maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

[12] de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica* **44**, 45-68.

[13] de Haan, L. (1994). Extreme value statistics. In:*Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simiu, eds.), 93-122. Kluwer Academic, Dordrecht.

[14] de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Australian Math. Soc. Ser. A* **61**, 381-395.

[15] Hall, P. (1982). On some simple estimators of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37-42.

[16] Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

[17] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

[18] Martins, M.J. (2000). *Heavy tails estimation - variants to the Hill estimator*. PhD thesis (in Portuguese), University of Lisbon. Portugal