Journal of Multivariate Analysis **76**, 226–248 (2001)
doi:10.1006/jmva.2000.1903, available online at http://www.idealibrary.com on IDEAL®

Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation[1]

J. Danielsson[2]

London School of Economics, London, United Kingdom, and University of Iceland, Iceland

and

L. de Haan,[3] L. Peng, and C. G. de Vries

Erasmus University, Rotterdam, The Netherlands, and Tinbergen Institute, Rotterdam, The Netherlands

Received March 19, 1998

Tail index estimation depends for its accuracy on a precise choice of the sample fraction, i.e., the number of extreme order statistics on which the estimation is based. A complete solution to the sample fraction selection is given by means of a two-step subsample bootstrap method. This method adaptively determines the sample fraction that minimizes the asymptotic mean-squared error. Unlike previous methods, prior knowledge of the second-order parameter is not required. In addition, we are able to dispense with the need for a prior estimate of the tail index which already converges roughly at the optimal rate. The only arbitrary choice of parameters is the number of Monte Carlo replications. © 2001 Academic Press

AMS 1991 subject classifications: 62G32, 62F40.

Key words and phrases: tail index, bootstrap, bias, mean squared error, optimal extreme sample fraction.

1. INTRODUCTION

Let $X_1, X_2, \dots$ be independent random variables with a common distribution function F which has a regularly varying tail

$$1 - F(x) = x^{-1/\gamma} L(x) \qquad x \to \infty, \quad \gamma > 0, \tag{1.1}$$

[1] Some data studied in the paper was obtained from Olsen and Associates.

[2] Danielsson benefitted from an HCM fellowship of the EU and the Research Contribution of the Icelandic banks.

[3] To whom correspondence should be addressed: L. de Haan, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: ldehaan@cs.few.eur.nl.

0047-259X/01 $35.00

where L is a slowly varying function and $1/\gamma$ is the index of regular variation, or the tail index. This is the case if F is in the domain of attraction of an extreme-value distribution with positive index or if F is in the domain of attraction of a stable distribution with index $0<\alpha<2$. Various estimators for estimating γ have been proposed (see Hill, 1975; Pickands, 1975; de Haan and Resnick, 1980; Hall, 1982; Mason, 1982; Davis and Resnick, 1984; Csörgő, *et al.* 1985, Hall and Welsh, 1985). We concentrate on the best known estimator, Hill's estimator,

$$\gamma_n(k) := \frac{1}{k} \sum_{i=1}^{k} \log X_{n,\,n-i+1} - \log X_{n,\,n-k},$$

where $X_{n,1} \leqslant \cdots \leqslant X_{n,n}$ are the order statistics of $X_1, \ldots, X_n$.

It is well known that if $k = k(n) \to \infty$ and $k(n)/n \to 0$, then

$$\gamma_n(k) \to \gamma, \qquad n \to \infty,$$

in probability (Mason, 1982). This follows since $k(n) \to \infty$ implies that eventually infinitely many order statistics are involved, allowing for the use of the law of large numbers, while the condition $k(n)/n \to 0$ means that the tail and nothing else is estimated. An asymptotic normality result for $\gamma_n(k)$ is needed for the construction of a confidence interval. Hall (1982) showed that if one chooses $k(n)$ by

$$k_0(n) := \arg\min_k \text{Asy}\, E(\gamma_n(k) - \gamma)^2,$$

where Asy E denotes the expectation with respect to the limit distribution, then

$$\sqrt{k_0(n)}(\gamma_n(k_0(n)) - \gamma) \xrightarrow{d} N(b, \gamma^2),$$

so that the optimal sequence $k_0(n)$ results in an asymptotic bias b. One can evaluate $k_0(n)$ asymptotically when the first- and second-order regular variation properties of the underlying distribution are known. A version of that result is our Theorem 1. In fact, $k_0(n)$ is the value which just balances the asymptotic variance and bias components of $E(\gamma_n(k) - \gamma)^2$.

Our framework is a second-order condition connected with (1.1). There exists a function A^*, not changing sign near infinity, such that

$$\lim_{t \to \infty} \left(\left(\frac{1 - F(x)}{1 - F(t)} - x^{-1/\gamma} \right) \middle/ A^*(t) \right) = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho/\gamma}$$

for $x>0$ and where $\rho \leqslant 0$ is the second-order parameter. A reformulated version of this condition with the inverse function U of $1/(1-F)$ is needed: There exists a function A, not changing sign near infinity, such that

$$\lim_{t\to\infty} \frac{U(tx)/U(t)-x^{\gamma}}{A(t)} = x^{\gamma}\,\frac{x^{\rho}-1}{\rho}. \tag{1.2}$$

The function $|A|$ is regular varying at infinity with index ρ. We write $|A| \in \mathrm{RV}_{\rho}$. We solve the optimality issue when ρ is strictly negative. Under this condition $k_0(n)$ can be expressed in terms of γ, ρ and the second-order rate function A.

Our aim is to determine the optimal sequence $k_0(n)$ solely on the basis of the sample, i.e., to determine an estimator $\hat{k}_0(n)$ such that

$$\sqrt{\hat{k}_0(n)}(\gamma_n(\hat{k}_0(n))-\gamma) \xrightarrow{d} N(b, \gamma^2). \tag{1.3}$$

For this it is sufficient to prove

$$\frac{\hat{k}_0(n)}{k_0(n)} \to 1, \tag{1.4}$$

in probability (Hall and Welsh, 1985). To find such $\hat{k}_0(n)$ we need two steps. We apply two subsample bootstrap procedures. This solves the problem under the extra assumption that $A(t)=ct^{\rho}$ with $\rho<0$ and $c\neq 0$, but otherwise ρ and c unknown.

The published literature at the time of writing did not contain a solution for the estimation of $k_0(n)$ except for very special cases. The most advanced is Hall (1990), who obtained an estimator $\hat{k}_0(n)$ which satisfies (1.4) under two extra assumptions, that ρ is known and that a prior estimate of γ is available such that this estimator already converges roughly at the optimal rate.[4] We are able to dispense with these assumptions. Nevertheless, Hall's (1990) suggestion to use a bootstrap method was very instrumental for the development of our automatic and general procedure.

As a byproduct of our approach we obtain a consistent estimator for the second-order parameter ρ; cf. Eq. (3.9) below. We believe this result to be new to the literature as well.

A completely different approach to the problem is taken in a recent paper by Drees and Kaufmann (1998). The Drees and Kaufmann method requires the choice of a tuning parameter. In our case the equivalence of this tuning parameter is the choice of the bootstrap resample size n_1. Below we present a fully automatic procedure for obtaining n_1 in the sense that

[4] Hall (1990) also uses the same idea to select the bandwidth in kernel estimation procedures. There, however, the second assumption is rather innocuous, but this is not the case for the problem at hand.

a heuristic algorithm is used to determine the bootstrap sample size (see Section 4). An explicit procedure for the choice of the resample size appears to be new to the literature as well.

2. MAIN RESULTS

Let $X_{n,1} \leqslant \cdots \leqslant X_{n,n}$ be the order statistics of $X_1, ..., X_n$. Hill's estimator is defined by

$$\gamma_n(k) := \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}.$$

Various authors have considered the asymptotic normality of γ_n; see Hall (1982). We can minimize the mean squared error of γ_n to get the asymptotically optimal choice of k, but it depends on the unknown parameter γ and the function $A(t)$ (see Dekkers and de Haan, 1993). We apply the powerful bootstrap tool to find the optimal number of order statistics adaptively.

The asymptotic mean squared error of γ_n is defined as

$$\text{AMSE}(n, k) := \text{Asy } E(\gamma_n(k) - \gamma)^2.$$

The AMSE will be estimated by a bootstrap procedure. Subsequently, we minimize the estimated AMSE to find the optimal k-value adaptively. For this to work two problems need to be solved. Even if one were given γ, then the regular bootstrap is not ensured to yield an AMSE estimate which is asymptotic to AMSE (n, k). Moreover, one does not know γ in the first place. The first problem can be solved by using a bootstrap resample size n_1 which is of smaller order than n. Therefore resamples $\mathscr{X}^*_{n_1} = \{X^*_1, ..., X^*_{n_1}\}$ are drawn from $\mathscr{X}_n = \{X_1, ..., X_n\}$ with replacement. Let $n_1 < n$ and $X^*_{n_1,1} \leqslant \cdots \leqslant X^*_{n_1,n_1}$ denote the order statistics of $\mathscr{X}^*_{n_1}$ and define

$$\gamma^*_{n_1}(k_1) := \frac{1}{k_1} \sum_{i=1}^{k_1} \log X^*_{n_1,n_1-i+1} - \log X^*_{n_1,n_1-k_1}.$$

Hall (1990) proposes the bootstrap estimate

$$\widehat{\text{AMSE}}(n_1, k_1) = E((\gamma^*_{n_1}(k_1) - \gamma_n(k))^2 \mid \mathscr{X}_n).$$

In this setup k has to be chosen such that $\gamma_n(k)$ is consistent. Then an estimate of k_1 for sample size n_1 is obtained. The problem is, however, that

k is unknown. Therefore we replace $\gamma_n(k)$ in the above expression with a more suitable statistic. This can be achieved by using a control variate.

Define

$$M_n(k) = \frac{1}{k} \sum_{i=1}^{k} (\log X_{n,n-i+1} - \log X_{n,n-k})^2.$$

Note that $M_n(k)/(2\gamma_n(k))$ is another consistent estimator of γ, which also balances the bias squared and variance if k tends to infinity with the rate of $k_0(n)$. Only the multiplicative constant differs. Therefore, we propose to use the bootstrap estimate for the mean squared error,

$$Q(n_1, k_1) := E((M^*_{n_1}(k_1) - 2(\gamma^*_{n_1}(k_1))^2)^2 \mid \mathscr{X}_n),$$

where $M^*_{n_1}(k_1) = \frac{1}{k_1} \sum_{i=1}^{k_1} (\log X^*_{n_1, n_1-i+1} - \log X^*_{n_1, n_1-k_1})^2$.

It can be shown that the statistics $M_n(k)/(2\gamma_n(k)) - \gamma_n(k)$ and $\gamma_n(k) - \gamma$ have similar asymptotic behavior in particular, both have asymptotic mean zero. Accordingly, as is shown in the following two theorems, the k-value that minimizes $\text{AMSE}(n, k)$ and the k-value that minimizes Asy $E(M_n(k) - 2(\gamma_n(k))^2)^2$ are of the same general order (with respect to n), under some conditions.

THEOREM 1. *Suppose* (1.2) *holds and* $k \to \infty$, $k/n \to 0$. *Determine* $k_0(n)$ *such that* $\text{AMSE}(n, k)$ *is minimal. Then*

$$k_0(n) = \frac{n}{s^-(\gamma^2(1-\rho)^2/n)}(1 + o(1)) \in RV_{-2\rho/(1-2\rho)}, \qquad as \quad n \to \infty,$$

where s^- *is the inverse function of* s, *with* s *given by*

$$A^2(t) = \int_t^\infty s(u)\, du (1 + o(1)) \qquad as \quad t \to \infty.$$

For the existence of such a monotone function see Lemma 2.9 *of Dekkers and de Haan* (1993). *Moreover, for fixed* $\delta > 0$ *and* $n \to \infty$,

$$k_0(n) \sim \arg\min_k E(\gamma_n(k) - \gamma)^2\, 1_{\{|\gamma_n(k) - \gamma| < k^{\delta - 1/2}\}}.$$

THEOREM 2. *Suppose* (1.2) *holds and* $k \to \infty$, $k/n \to 0$. *Determine* $\bar{k}_0(n)$ *such that* Asy $E(M_n(k) - 2(\gamma_n(k))^2)^2$ *is minimal. Then*

$$\bar{k}_0(n) = \frac{n}{s^-(\gamma^2(1-\rho)^4/(n\rho^2))}(1 + o(1)), \qquad as \quad n \to \infty.$$

Moreover, for fixed $\delta > 0$ *and* $n \to \infty$,

$$\bar{k}_0(n) \sim E(M_n(k) - 2(\gamma_n(k))^2)^2 \, 1_{\{|M_n(k) - 2(\gamma_n(k))^2| < k^{\delta - 1/2}\}}.$$

COROLLARY 3.

$$\frac{\bar{k}_0(n)}{k_0(n)} \to \left(1 - \frac{1}{\rho}\right)^{2/(1-2\rho)} \qquad (n \to \infty).$$

The next theorem is our main result and shows that the optimal k_1 for a subsample of size n_1 can be estimated consistently. The method used in proving this result is more involved but similar to the method that is used in proving Theorem 1.

THEOREM 4. *Suppose* (1.2) *holds and* $k_1 \to \infty$, $k_1/n_1 \to 0$, $n_1 = O(n^{1-\varepsilon})$ *for some* $0 < \varepsilon < 1$. *Determine* $k^*_{1,0}(n_1)$ *such that*

$$Q(n_1, k_1) = E((M^*_{n_1}(k_1) - 2(\gamma^*_{n_1}(k_1))^2)^2 \mid \mathcal{X}_n)$$

is minimal. Then

$$\frac{k^*_{1,0}(n_1) \, s^-(\gamma^2(1-\rho)^4/(n_1\rho^2))}{n_1} \xrightarrow{P} 1, \qquad as \quad n \to \infty.$$

Theorem 4 gives the optimal k_1 for sample size n_1, but we need the optimal value for the sample size n. This can be achieved modulo a conversion factor.

COROLLARY 5. *Suppose* (1.2) *holds for* $A(t) = ct^\rho$, $t \to \infty$, *and* $k_1 \to \infty$, $k_1/n_1 \to 0$, $n_1 = O(n^{1-\varepsilon})$ *for some* $0 < \varepsilon < 1$. *Then*

$$\left(\frac{n_1}{n}\right)^{-2\rho/(2\rho-1)} \frac{k^*_{1,0}(n_1)}{\bar{k}_0(n)} \xrightarrow{P} 1, \qquad as \quad n \to \infty.$$

The conversion factor can be calculated consistently as follows.

THEOREM 6. *Let* $n_1 = O(n^{1-\varepsilon})$ *for some* $0 < \varepsilon < 1/2$ *and* $n_2 = (n_1)^2/n$. *Suppose* (1.2) *holds for* $A(t) = ct^\rho$, $t \to \infty$, *and* $k_i \to \infty$, $k_i/n_i \to 0$ $(i = 1, 2)$. *Determine* $k^*_{i,0}$ *such that*

$$E((M^*_{n_i}(k_i) - 2(\gamma^*_{n_i}(k_i))^2)^2 \mid \mathcal{X}_n)$$

is minimal $(i=1,2)$. *Then*

$$\frac{\dfrac{(k^*_{1,0}(n_1))^2}{k^*_{2,0}(n_2)}\left(\dfrac{(\log k^*_{1,0}(n_1))^2}{(2\log n_1-\log k^*_{1,0}(n_1))^2}\right)^{(\log n_1-\log k^*_{1,0}(n_1))/\log n_1}}{k_0(n)} \xrightarrow{p} 1 \quad (2.1)$$

as $n\to\infty$.

Remark 1. From Theorem 6 we can achieve the optimal choice of k asymptotically. Therefore, by using the asymptotically optimal choice of k, Hill's estimator will also be asymptotically optimal.

COROLLARY 7. *Suppose the conditions of Theorem* 6 *hold. Define*

$$\hat{k}_0(n):=\frac{(k^*_{1,0}(n_1))^2}{k^*_{2,0}(n_2)}\left(\frac{(\log k^*_{1,0}(n_1))^2}{(2\log n_1-\log k^*_{1,0}(n_1))^2}\right)^{(\log n_1-\log k^*_{1,0}(n_1))/\log n_1}$$

Then $\gamma_n(\hat{k}_0)$ *has the same asymptotic efficiency as* $\gamma_n(k_0)$.

To summarize, the algorithm for computing $\gamma_n(\hat{k}_0)$ is as follows. For a given choice of n_1 draw bootstrap resamples of size n_1. Calculate $Q(n_1,k_1)$, i.e., the bootstrap AMSE, at each k_1; and find the $k^*_{1,0}(n_1)$ which minimizes this bootstrap AMSE. Repeat this procedure for an even smaller resample size n_2, where $n_2=(n_1)^2/n$. This yields $k^*_{2,0}(n_2)$. Subsequently, calculate $\hat{k}_0(n)$ from the formula in Corollary 7. Finally, estimate γ by $\gamma_n(\hat{k}_o)$. By using this procedure two tuning parameters have to be chosen, the number of bootstrap resamples and n_1. The number of bootstrap resamples is determined by the computational facilities and can be chosen on the basis of a stopping criterion where either the resampling is stopped once the fluctuations in the bootstrap MSE's fall below a certain level or once a bound on run time is hit. The choice of n_1 is made as follows.

From Theorem 6 we know that for any ε such that $0<\varepsilon<1/2$ the $n_1=n^{1-\varepsilon}$ is an appropriate choice. Hence, asymptotic arguments provide little guidance in choosing between any of the possible n_1. We use the following heuristic procedure. In the proof to Theorem 6 we will show that

$$\frac{\bar{k}_o k^*_{2,0}}{(k^*_{1,0})^2}\to 1,$$

in probability. By very similar arguments one can show that

$$\text{Asy } E(M_n(\bar{k}_0)-2(\gamma_n(\bar{k}_0))^2)^2\,\frac{Q(n_2,k^*_{2,0})}{(Q(n_1,k^*_{1,0}))^2}\to 1,$$

in probability, as well. Thus an estimator for Asy $E(M_n(\bar{k}_0) - 2(\gamma_n(\bar{k}_0))^2)^2$ is the ratio

$$R(n_1) := \frac{(Q(n_1, k^*_{1,0}))^2}{Q(n_2, k^*_{2,0})}.$$

The finite sample n_1 is now chosen such that $R(n_1)$ is minimal. Note that this criterion is the finite sample analogue of the asymptotic criterion that is used for locating $\bar{k}_0(n)$. In practice this criterion is implemented by working with a grid of n_1 values over which $R(n_1)$ is minimized. The grid size is again determined by the available computing time.

3. PROOFS

Let $Y_1, ..., Y_n$ be independent random variables with common distribution function $G(y) = 1 - y^{-1}$ $(y \geqslant 1)$. Let $Y_{n,1} \leqslant \cdots \leqslant Y_{n,n}$ be the order statistics of $Y_1, ..., Y_n$. Note that $\{X_{n,n-i+1}\}_{i=1}^n \stackrel{d}{=} \{U(Y_{n,n-i+1})\}_{i=1}^n$ with the function U defined in the Introduction.

LEMMA 8. *Let $0 < k < n$ and $k \to \infty$. We have:*

(1) *For $n \to \infty$, $Y_{n,n-k}/(n/k) \to 1$ in probability.*

(2) *For $n \to \infty$, (P_n, Q_n) is asymptotically normal with means zero, variances 1 and 20 respectively, and covariance 4, where*

$$P_n := \sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{k} \log Y_{n,n-i+1} - \log Y_{n,n-k} - 1\right\}$$

and

$$Q_n := \sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{k} (\log Y_{n,n-i+1} - \log Y_{n,n-k})^2 - 2\right\}.$$

Proof. The proof is similar to the proof of Lemma 3.1 of Dekkers and de Haan (1993). ∎

Proof of Theorem 1. We use the method of Dekkers and de Haan (1993), which we outline, since a similar reasoning is used in the proofs of Theorems 2 and 4.

Relation (1.2) is equivalent to the regular variation of the function

$$|\log U(t) - \gamma \log t - c_0|$$

with index ρ for some constant c_0 (see Geluk and de Haan, 1987, II.1). Then (1.2) holds with

$$A(t) = \rho(\log U(t) - \gamma \log t - c_0).$$

Applying extended Potter's inequalities to the function A, we get that for each $0 < \varepsilon < 1$ there exists $t_0 > 0$ such that for $t \geqslant t_0$ and $tx \geqslant t_0$,

$$(1-\varepsilon)\, x^{\rho} e^{-\varepsilon\, |\log x|} - 1 \leqslant \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)/\rho}$$
$$\leqslant (1+\varepsilon)\, x^{\rho} e^{\varepsilon\, |\log x|} - 1. \tag{3.1}$$

Applying this relation with t replaced by $Y_{n,n-k}$ and x replaced by $Y_{n,n-i+1}/Y_{n,n-k}$, adding the inequalities for $i = 1, 2, ..., k$, and dividing by k, we get

$$\gamma_n \approx \gamma + \frac{\gamma P_n}{\sqrt{k}} + \rho^{-1} A(Y_{n,n-k})(1 \pm \varepsilon) \left\{ \frac{1}{k} \sum_{i=1}^{k} \left(\frac{Y_{n,n-i+1}}{Y_{n,n-k}} \right)^{\rho \pm \varepsilon} - 1 \right\}.$$

Now

$$\sum_{i=1}^{k} \left(\frac{Y_{n,n-i+1}}{Y_{n,n-k}} \right)^{\rho \pm \varepsilon} \stackrel{d}{=} \sum_{i=1}^{k} Y_i$$

with $Y_1, ..., Y_k$ i.i.d. with common distribution function $1 - 1/x$. Hence by the weak law of large numbers,

$$\gamma_n \approx \gamma + \frac{\gamma P_n}{\sqrt{k}} + \rho^{-1}(1 \pm \varepsilon) \left(\frac{1}{1 - \rho \mp \varepsilon} - 1 \right) A(Y_{n,n-k}),$$

i.e.,

$$\gamma_n = \gamma + \frac{\gamma P_n}{\sqrt{k}} + (1-\rho)^{-1} A\left(\frac{n}{k}\right) + o_p\left(A\left(\frac{n}{k}\right)\right).$$

(Note that in the latter term we have replaced $Y_{n,n-k}$ by n/k, which can be done since $|A|$ is regularly varying.) Hence

$$\text{Asy } E(\gamma_n - \gamma)^2 \approx \frac{\gamma^2}{k} + \frac{A^2(n/k)}{(1-\rho)^2}.$$

We can assume (see Lemma 2.9 of Dekkers and de Haan, 1993) that A^2 has a monotone derivative s which is then regularly varying with index $2\rho - 1$. Consequently, $s^{-}(1/t)$ (s^{-} denoting the inverse of s) is regularly varying with index $1/(1-2\rho)$. The first result of the theorem is then obtained by minimizing the right-hand side of the equation above. For the proof of the second statement of Theorem 1 we are going to replace the o_p-terms by o-terms on part of the sample space. Define for some $0 < \delta_0 < 1/2$ the set

$$E_n := \left\{ \omega\colon |P_n|, |D_n^{\pm}|, \left| \frac{k}{n} Y_{n,n-k} - 1 \right| < k^{\delta_0 - 1/2} \right\}$$

with

$$D_n^{\pm} := \frac{1}{k} \sum_{i=1}^{k} (Y_{n,n-i+1}/Y_{n,n-k})^{\rho \pm \varepsilon} - (1 - \rho \mp \varepsilon)^{-1}.$$

Now take ε and t_0 as in (3.1). Then, provided $(n/k)(1 - k^{\delta_0 - 1/2}) \geqslant t_0$, we have $Y_{n,n-k} \geqslant t_0$ on E_n. Also, since A is regularly varying we have

$$\left| A(Y_{n,n-k}) - A\left(\frac{n}{k}\right) \right| < 2\varepsilon A\left(\frac{n}{k}\right)$$

on E_n. Using these two facts and the inequalities (3.1) we find that

$$\left| \gamma_n(k) - \gamma - \frac{\gamma P_n}{\sqrt{k}} + \frac{A(n/k)}{(1-\rho)} \right| < \varepsilon A\left(\frac{n}{k}\right)$$

on the set E_n (so we have $o(A)$ instead of $o_p(A)$). Hence for $n \to \infty$ and any intermediate sequence $k(n)$,

$$\frac{E(\gamma_n(k) - \gamma)^2 \, 1_{\{|\gamma_n(k) - \gamma| < k^{\delta - 1/2}\}} 1_{E_n}}{\dfrac{\gamma^2}{k} + \dfrac{A^2(n/k)}{(1-\rho)^2}} \to 1.$$

Next, we show that the contribution of the set E_n^c to the expectation can be neglected. For example,

$$E(\gamma_n(k)-\gamma)^2\, 1_{\{|\gamma_n(k)-\gamma|<k^{\delta-1/2}\}}\, 1_{\{|P_n|>k^{\delta_0-1/2}\}} \leqslant k^{2\delta-1} P\{|P_n|>k^{\delta_0-1/2}\},$$

and by Bennett's inequality (cf. Petrov, 1975, Chap. III.5) we can show that

$$P\{|P_n|>k^{\delta_0-1/2}\}\leqslant k^{-\beta}$$

for any $\beta>0$, eventually. Hence

$$\lim_{n\to\infty}\frac{E(\gamma_n(k)-\gamma)^2\, 1_{\{|\gamma_n(k)-\gamma|^2\leqslant k^{\delta-1/2}\}}\, 1_{\{|P_n|>k^{\delta_0-1/2}\}}}{\frac{\gamma^2}{k}+\frac{A^2(n/k)}{(1-\rho)^2}}=0.$$

The reasoning in the case any of the other conditions of the set E_n is violated is exactly the same (but for $(k/n)\, Y_{n,n-k}$ we first have to transform the inequality into an inequality for its inverse, $(1/k)\sum_{i=1}^n 1_{\{Y_i>(n/k)\,x\}}$, and apply Bennett's inequality). Hence

$$E(\gamma_n(k)-\gamma)^2\, 1_{\{|\gamma_n(k)-\gamma|\leqslant k^{\delta-1/2}\}}\approx\frac{\gamma^2}{k}+\frac{A^2(n/k)}{(1-\rho)^2}$$

The rest of the proof is the same as before.

Proof of Theorem 2. From the proof of Theorem 1 we get

$$\gamma_n \stackrel{d}{=} \gamma+\frac{\gamma P_n}{C}+d_1 A(Y_{n,n-k})+o_p(A(n/k)) \tag{3.2}$$

with $d_1=1/(1-\rho)$ and hence

$$\gamma_n^2 \stackrel{d}{=} \gamma^2+\frac{2\gamma^2 P_n}{\sqrt{k}}+2\gamma d_1 A(Y_{n,n-k})+o_p(A(n/k)). \tag{3.3}$$

Similarly,

$$M_n \stackrel{d}{=} 2\gamma^2+\frac{\gamma^2 Q_n}{\sqrt{k}}+d_2 A(Y_{n,n-k})+o_p(A(n/k)) \tag{3.4}$$

where $d_2=2\gamma(2-\rho)/(1-\rho)^2$. The rest of the proof is similar to that of Theorem 1. ∎

Proof of Theorem 4. Let G_n denote the empirical distribution function of n independent, uniformly distributed random variables. As n is large enough and $n_1 = O(n^{1-\varepsilon})$, we have

$$1/2 \leqslant \sup_{0<t\leqslant n_1(\log n_1)^2} tG_n^-\left(\frac{1}{t}\right) \leqslant 2 \qquad \text{a.s.} \tag{3.5}$$

and

$$\sup_{t\geqslant 2}\left|\sqrt{t}\left(G_n\left(\frac{1}{t}\right)-\frac{1}{t}\right)\right| \leqslant \frac{\log n}{\sqrt{n}} \qquad \text{a.s.}$$

(see Shorack and Wellner, 1986, Eqs. (10) and (17), Chap. 10.5). Hence

$$\sup_{4\leqslant t\leqslant n_1(\log n_1)^2}\left|\sqrt{\frac{1}{G_n^-\left(\frac{1}{t}\right)}}\left[G_n\left(G_n^-\left(\frac{1}{t}\right)\right)-G_n^-\left(\frac{1}{t}\right)\right]\right| \leqslant \frac{\log n}{\sqrt{n}} \qquad \text{a.s.}$$

Therefore, for all $4 \leqslant t \leqslant n_1(\log n_1)^2$,

$$\left|tG_n^-\left(\frac{1}{t}\right)-1\right| \leqslant \frac{2\sqrt{t}\log n}{\sqrt{n}} \qquad \text{a.s.} \tag{3.6}$$

Let F_n denote the empirical distribution function of $\mathscr{X}_n$, $U_n = (1/(1-F_n))^-$. Now we use (3.1), (3.5), (3.6),

$$|\log y| \leqslant 2\,|y-1| \qquad \text{for all} \quad 1/2 \leqslant y \leqslant 2$$

$$|y^{-\rho}-1| \leqslant (-\rho)(2^{-\rho-1} \vee 2^{1+\rho})\,|y-1| \qquad \text{for all} \quad 1/2 \leqslant y \leqslant 2$$

and

$$\begin{aligned}
\log U_n(t) &= \log F_n^-\left(1-\frac{1}{t}\right) \\
&\overset{d}{=} \log F^-\left(G_n^-\left(1-\frac{1}{t}\right)\right) \\
&= \log U\left(\frac{1}{1-G_n^-\left(1-\frac{1}{t}\right)}\right) \\
&\overset{d}{=} \log U\left(\frac{t}{tG_n^-\left(\frac{1}{t}\right)}\right).
\end{aligned}$$

From this we conclude that for any $0<\varepsilon<1$ there exists $t_0>4$ such that for $t_0<t<n_1(\log n_1)^2$ and $t_0<tx<n_1(\log n_1)^2$,

$$\frac{\log U_n(tx)-\log U_n(t)-\gamma\log x}{\frac{A(tx)}{\rho}}$$

$$\stackrel{d}{=}\frac{\log U\left(\frac{tx}{txG_n^-(\frac{1}{tx})}\right)-\log U(tx)-\gamma\log\left(\frac{1}{txG_n^-(\frac{1}{tx})}\right)}{\frac{A(tx)}{\rho}}\frac{A(tx)}{A(t)}$$

$$-\frac{\log U\left(\frac{t}{tG_n^-(\frac{1}{t})}\right)-\log U(t)-\gamma\log\left(\frac{1}{tG_n^-(\frac{1}{t})}\right)}{A(t)/\rho}$$

$$+\frac{\log U(tx)-\log U(t)-\gamma\log x}{A(t)/\rho}$$

$$+\frac{\gamma\log\left(\frac{1}{txG_n^-(\frac{1}{tx})}\right)}{A(t)/\rho}-\frac{\gamma\log\left(\frac{1}{tG_n^-(\frac{1}{t})}\right)}{A(t)/\rho}$$

$$\leqslant\left[(1+\varepsilon)\left(txG_n^-\left(\frac{1}{tx}\right)\right)^{-\rho}e^{\varepsilon\,|\log(txG_n^-(1/tx))|}-1\right](1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}$$

$$-(1-\varepsilon)\left(tG_n^-\left(\frac{1}{t}\right)\right)^{-\rho}e^{-\varepsilon\,|\log(tG_n^-(1/t))|}+1+(1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}$$

$$-1+\left|\frac{\gamma\rho}{A(t)}\right|2\left(\left|txG_n^-\left(\frac{1}{tx}\right)-1\right|+\left|tG_n^-\left(\frac{1}{t}\right)-1\right|\right)\qquad\text{a.s.}$$

$$\leqslant(1+\varepsilon)\left[\left(txG_n^-\left(\frac{1}{tx}\right)\right)^{-\rho}-1\right]e^{\varepsilon\,|\log(txG_n^-(1/tx))|}(1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}$$

$$+(1+\varepsilon)\,|e^{\varepsilon\,|\log(txG_n^-(1/tx))|}-1|\,(1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}$$

$$+\varepsilon(1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}-(1-\varepsilon)\left[\left(tG_n^-\left(\frac{1}{t}\right)\right)^{-\rho}-1\right]$$

$$\times e^{-\varepsilon\,|\log(tG_n^-(1/t))|}-(1-\varepsilon)[e^{-\varepsilon\,|\log(tG_n^-(1/t))|}-1]-\varepsilon$$

$$+(1+\varepsilon)\,x^\rho e^{\varepsilon\,|\log x|}-1+\left|\frac{\gamma\rho}{A(t)}\right|\frac{4\sqrt{t}\log n}{\sqrt{n}}(\sqrt{x}+1)\qquad\text{a.s.}$$

$$\leqslant (1+\varepsilon)(-\rho)(2^{-\rho-1} \vee 2^{1+\rho}) \left| txG_n^-\left(\frac{1}{tx}\right) - 1 \right| e^{\varepsilon \log 2}(1+\varepsilon)\, x^\rho$$

$$\times e^{\varepsilon\,|\log x|} + 4\varepsilon e^{\varepsilon \log 2}(1+\varepsilon)\, x^\rho e^{\varepsilon\,|\log x|} + (1+\varepsilon)^2\, x^\rho e^{\varepsilon\,|\log x|} - 1$$

$$+ (1-\varepsilon)(-\rho)(2^{-\rho-1} \vee 2^{1+\rho}) \left| tG_n^-\left(\frac{1}{t}\right) - 1 \right| e^{\varepsilon \log 2}$$

$$+ 4\varepsilon(1-\varepsilon)\, e^{\varepsilon \log 2} - \varepsilon + \left| \frac{\gamma\rho}{A(t)} \right| \frac{4\sqrt{t}\log n}{\sqrt{n}} (\sqrt{x}+1) \qquad \text{a.s.}$$

$$\leqslant \left[(-\rho)(2^{-\rho+1} \vee 2^{3+\rho}) + 2 \left| \frac{\gamma\rho}{A(t)} \right| \right] \frac{2\sqrt{t}\log n}{\sqrt{n}} (\sqrt{x}+1)$$

$$+ (1+9\varepsilon)(1+\varepsilon)\, x^\rho e^{\varepsilon\,|\log x|} - 1 + 7\varepsilon \qquad \text{a.s.} \tag{3.7}$$

Similarly,

$$\frac{\log U_n(tx) - \log U_n(t) - \gamma \log x}{A(t)/\rho}$$

$$\geqslant -\left[(-\rho)(2^{-\rho+1} \vee 2^{3+\rho}) + 2 \left| \frac{\gamma\rho}{A(t)} \right| \right] \frac{2\sqrt{t}\log n}{\sqrt{n}} (\sqrt{x}+1)$$

$$+ (1-9\varepsilon)(1-\varepsilon)\, x^\rho e^{-\varepsilon\,|\log x|} - 1 - 7\varepsilon \qquad \text{a.s.} \tag{3.8}$$

Inequalities (3.7) and (3.8) are valid in probability with t replaced by Y_{n_1, n_1-k_1} and tx replaced by Y_{n_1, n_i+1} $(i=1, ..., k_1)$ since

$$4 \leqslant Y_{n_1, n_1-i+1} \leqslant Y_{n_1, n_1} \qquad (i=1, ..., k_1) \qquad \text{in probability,}$$

and

$$\frac{Y_{n_1, n_1}}{(n_1(\log n_1)^2)} \to 0 \qquad \text{in probability}$$

for $n_1 \to \infty$ and $k_1/n_1 \to 0$.

We now minimize

$$E((M_{n_1}^*(k_1) - 2(\gamma_{n_1}^*(k_1))^2)^2 \mid \mathscr{X}_n).$$

Note that, conditionally, given $\mathscr{X}_n$, P_{n_1} is once again a normalized m of i.i.d. random variables from an exponential distribution. Hence, when n_1 increases, the distribution of P_{n_1} approaches a normal one. The case is similar for Q_{n_1}.

We proceed as in the proof of Theorem 2 and use

$$\gamma^*_{n_1}(k_1) \stackrel{d}{=} \gamma + \frac{\gamma P_{n_1}}{\sqrt{k_1}} + d_1 A(Y_{n_1, n_1-k_1}) + o_p(A(n_1/k_1)) + O\left(\frac{\log n \sqrt{n_1/k_1}}{\sqrt{n}}\right),$$

$$(\gamma^*_{n_1}(k_1))^2 \stackrel{d}{=} \gamma^2 + \frac{2\gamma^2 P_{n_1}}{\sqrt{k_1}} + 2\gamma d_1 A(Y_{n_1, n_1-k_1}) + o_p\left(A\left(\frac{n_1}{k_1}\right)\right) + O\left(\frac{\log n \sqrt{n_1/k_1}}{\sqrt{n}}\right),$$

and

$$M^*_{n_1}(k_1) \stackrel{d}{=} 2\gamma^2 + \frac{\gamma^2 Q_{n_1}}{\sqrt{k_1}} + d_2 A(Y_{n_1, n_1-k_1}) + o_p\left(A\left(\frac{n_1}{k_1}\right)\right) + O\left(\frac{\log n \sqrt{n_1/k_1}}{\sqrt{n}}\right).$$

Note that the term $\log n \sqrt{n_1/k_1}/\sqrt{n} = o(1/\sqrt{k_1})$, so that it can be neglected in the minimization process. The statement of Theorem 4 follows. ▮

Proof of Corollary 5. The proof follows easily from Theorem 2 and Theorem 4 and the fact that

$$t^{1/(2\rho-1)} s^-(1/t) \to (-2\rho c^2)^{1/(1-2\rho)}. \quad ▮$$

Proof of Theorem 6. Since $k^*_{1,0} \in \mathrm{RV}_{-2\rho/(1-2\rho)}$ in probability, we have

$$\frac{\log k^*_{1,0}}{\log n_1} \xrightarrow{p} \frac{-2\rho}{1-2\rho}$$

(see Proposition 1.7.1 of Geluk and de Haan, 1987); i.e.,

$$\frac{\log k^*_{1,0}}{-2\log n_1 + 2\log k^*_{1,0}} \xrightarrow{p} \rho. \tag{3.9}$$

Write the result of Corollary 5 for $k^*_{1,0}$ and $k^*_{2,0}$,

$$\frac{k^*_{1,0}}{\bar{k}_0} \Big/ \left(\frac{n_1}{n}\right)^{2\rho/(2\rho-1)} \xrightarrow{p} 1,$$

$$\frac{k^*_{2,0}}{\bar{k}_0} \Big/ \left(\frac{n_2}{n}\right)^{2\rho/(2\rho-1)} \xrightarrow{p} 1.$$

Hence

$$\bar{k}_0 k^*_{2,0}/(k^*_{1,0})^2 \xrightarrow{p} 1, \tag{3.10}$$

and by Corollary 3

$$\frac{(k^*_{1,0}(n_1))^2}{k^*_{2,0}(n_2)\, k_0(n)} \xrightarrow{p} \left(1-\frac{1}{\rho}\right)^{2/(1-2\rho)}.$$

An application of the estimate of ρ from (3.9) gives the result. ∎

Proof of Corollary 7. We now have a random sequence $\hat{k}_0(n)$ with the property

$$\lim_{n\to\infty} \frac{\hat{k}_0(n)}{k_0(n)} = 1 \qquad \text{in probability.}$$

Theorem 4.1 of Hall and Welsh (1985) now guarantees that $\gamma_n(\hat{k}_0(n))$ achieves the optimal rate. ∎

4. SIMULATION AND ESTIMATION

We investigate the performance of our fully automatic estimation procedure by means of Monte Carlo experiments and by an application to some financial data sets, i.e., the stock-price index S&P 500 and foreign exchange quote data. The sample sizes are typical for current financial data sets, ranging from 2,000 to 20,000 observations. The sample size in the Monte Carlo experiments were chosen to be equally large.

4.1. *Simulations*

We evaluate the performance of our estimators for γ, ρ, and $k_0(n)$ on the basis of pseudo i.i.d. random numbers from the Student-t and type-II extreme value distributions in addition to two cases of dependent data. The tail index $1/\gamma$ equals the degrees of freedom in the case of the Student-t distribution. Recall that the type-II extreme value distribution reads $\exp[-x^{-1/\gamma}]$. We focus on $1/\gamma = 1$, 4, and 11. For the Student-t distribution $\rho/\gamma = -2$, while for the extreme value distribution $\rho = -1$.

In addition to the i.i.d. data, we also investigate the performance of our estimator for dependent data. From Hsing (1991), Resnick and Starica (1998), and Embrechts *et al.* (1997) we know that the Hill estimator is consistent for dependent data like the ARMA processes and ARCH-type processes. We focus on two stochastic processes.

First, the MA(1) process $Y_t = X_t + X_{t-1}$, where the X_t are i.i.d. Student-t with $1/\gamma = 3$ degrees of freedom is considered. The first- and second-order parameters of the tail expansion of Y_t can be computed by standard calculus methods. The interest in this process derives from the fact that while $\gamma_n(k)$ is biased upward for Student-t distributions, the bias switches sign for the marginal distribution of Y; i.e., the c-parameter in the $A(t)$ function switches sign.

The other stochastic process exhibits conditional heteroscedasticity. Financial time series return data typically have the fair game property with dependence only in the second moment; see e.g. Bollerslev *et al.* (1992) and Embrechts *et al.* (1997). The following stochastic volatility process is typical for the processes that are used to model financial return data:

$$Y_t = U_t X_t H_t,$$

$$U_t \sim \text{i.i.d. discrete uniform on } -1, 1,$$

$$X_t = \sqrt{57/Z_t}, \qquad Z_t \sim \chi_{(3)} \qquad \text{i.i.d.},$$

$$H_t = 0.1 Q_t + 0.9 H_{t-1}, \qquad Q_t \sim N(0, 1), \qquad \text{i.i.d.}$$

The X_t and Z_t are chosen such that the marginal distribution Y_t has a Student-t with a three degrees of freedom distribution. This allows us to evaluate the performance of our procedure.

The results of the Monte Carlo experiments are reported in Tables I and II for sample sizes of 2,000 and 20,000, respectively. Each table is based on 250 simulations per distribution. For the choice of the tuning parameter n_1 we use the procedure described at the end of Section 2. Hence, for $n = 2{,}000$ we searched over the interval from $n_1 = 600$ to $n_1 = 1700$ in increments of 100. The number of bootstrap resamples was 1,000. In the larger sample with size $n = 20{,}000$ we searched from $n_1 = 2{,}000$ to $n_1 = 15{,}000$, in increments of 1,000, using 500 bootstrap resamples for each n_1. The grid size could be made much finer, and the number of resamples larger, for a specific data set in order to increase the precision. For each distribution we report the true value of the parameter, the mean, the standard error (s.e.), and the root mean squared error (RMSE). We report estimates for γ and $-\rho$, while $\hat{k}_0(n)$ is reported relative to $k_0(n)$.

From the Tables I and II we see that the estimator for the inverse tail index γ performs well in terms of bias and standard error for both the larger and the smaller sample sizes. Evidently, in most cases the bias and standard error are lower for the larger sample size $n = 20{,}000$. The only exception to decent performance in terms of bias is the Student-t with 11 degrees of freedom, since it is heavily upward biased in the smaller sample.

TABLE I

Monte Carlo Experiment with $n=2{,}000$

Distribution	Parameters	True	Mean	S.E.	RMSE
Student(1)	γ	1.000	1.004	0.106	0.106
	$-\rho$	2.000	1.332	0.362	0.768
	$\hat{k}(n)/k_0(n)$	1.000	0.874	0.426	0.444
Student(4)	γ	0.250	0.296	0.074	0.087
	$-\rho$	0.500	0.562	0.235	0.242
	$\hat{k}_0(n)/k_0(n)$	1.000	1.133	0.988	0.995
Student(11)	γ	0.091	0.170	0.050	0.094
	$-\rho$	0.182	0.374	0.173	0.258
	$\hat{k}_0(n)/k_0(n)$	1.000	1.386	1.114	1.177
Extreme(1)	γ	1.000	1.035	0.095	0.101
	$-\rho$	1.000	2.140	0.818	1.402
	$\hat{k}_0(n)/k_0(n)$	1.000	1.342	0.732	0.806
Extreme(4)	γ	0.250	0.259	0.024	0.025
	$-\rho$	1.000	2.138	0.817	1.400
	$\hat{k}_0(n)/k_0(n)$	1.000	1.339	0.732	0.805
Extreme(11)	γ	0.091	0.094	0.009	0.010
	$-\rho$	1.000	2.137	0.824	1.403
	$\hat{k}_0(n)/k_0(n)$	1.000	1.338	0.735	0.808
MA(1)	γ	0.333	0.322	0.089	0.090
	$-\rho$	0.667	0.621	0.279	0.282
	$\hat{k}_0(n)/k_0(n)$	1.000	2.544	2.260	2.733
Stochastic volatility	γ	0.333	0.368	0.083	0.090
	$-\rho$	0.667	0.663	0.252	0.252
	$\hat{k}_0(n)/k_0(n)$	1.000	1.041	0.827	0.826

This occurs even though the RMSE does not vary that much with γ for the Student-t class. Thus for some applications the RMSE criterion may give too low a weight to the bias. The method also works well for the two stochastic processes.

The estimates for the second-order parameter ρ are less precise than those for the first-order parameter (after rescaling the standard error by the true parameter value). The tail observations are naturally more informative about the leading terms of the expansion at infinity. Because $\hat{k}_0(n)$ depends on $\hat{\rho}$, it is not surprising to see that the same observation applies to $\hat{k}_0(n)/k_0(n)$. As was predicted on the basis of the theoretical parameters, the MA(1) γ-estimate is downward biased, while it is upward biased for the Student-t model.

Another way to evaluate our procedures is to see how the performance changes as the sample size is increased by the factor 10 if we move from 2,000 to 20,000 observations. From theory we know that the asymptotic

TABLE II

Monte Carlo Experiment with $n = 20{,}000$

Distribution	Parameters	True	Mean	S.E.	RMSE
Student(1)	γ	1.000	1.009	0.037	0.038
	$-\rho$	2.000	1.519	0.253	0.543
	$\hat{k}(n)/k_0(n)$	1.000	1.023	0.372	0.372
Student(4)	γ	0.250	0.283	0.029	0.044
	$-\rho$	0.500	0.646	0.126	0.193
	$\hat{k}_0(n)/k_0(n)$	1.000	1.562	1.038	1.179
Student(11)	γ	0.091	0.146	0.033	0.064
	$-\rho$	0.182	0.423	0.118	0.269
	$\hat{k}_0(n)/k_0(n)$	1.000	2.379	2.235	2.631
Extreme(1)	γ	1.000	1.026	0.033	0.042
	$-\rho$	1.000	1.940	0.417	1.028
	$\hat{k}_0(n)/k_0(n)$	1.000	1.635	0.722	0.960
Extreme(4)	γ	0.250	0.257	0.008	0.011
	$-\rho$	1.000	1.939	0.415	1.026
	$\hat{k}_0(n)/k_0(n)$	1.000	1.629	0.715	0.951
Extreme(11)	γ	0.091	0.093	0.063	0.004
	$-\rho$	1.000	1.942	0.414	1.028
	$\hat{k}_0(n)/k_0(n)$	1.000	1.632	0.719	0.956
MA(1)	γ	0.333	0.321	0.044	0.046
	$-\rho$	0.667	0.766	0.201	0.224
	$\hat{k}_0(n)/k_0(n)$	1.000	3.977	2.732	4.037
Stochastic volatility	γ	0.333	0.357	0.030	0.038
	$-\rho$	0.667	0.744	0.134	0.154
	$\hat{k}_0(n)/k_0(n)$	1.000	1.281	0.768	0.816

bias and RMSE should drop by a factor $10^{-\rho/(1-2\rho)}$, while the squared root of the ratio of the asymptotically optimal number of highest order statistics k_0 should increase by the same factor. In Table III we report the ratios that are implied by comparing the numbers from Tables I and II. The RMSE and upper order statistics ratios are close to the true factor. The bias ratio is less favorable. There are two cases where the bias deteriorated in the larger sample.

4.2. *Asset Return Data*

The financial data sets we examine have been widely studied in the area of finance. The use of high frequency data in financial research and applications has become standard. For example, some data sets studied in the special issue of the *Journal of Empirical Finance* edited by Baillie and Dacorogna (1997) were larger than 1.5 million, and the sample sizes of the data sets studied in Embrechts *et al.* (1997, Chap. 6) are of the order of magnitude of 10,000. Nevertheless, even though several aspects of these

TABLE III

Asymptotic Ratios

Distribution	True factor	Bias ratio	RMSE ratio	Root of the $\hat{k}_0(n)$ ratio
Student(1)	2.51	0.44	2.78	2.71
Student(4)	1.78	1.39	1.97	2.09
Student(11)	1.36	1.43	1.46	1.78
Extreme(1)	2.15	1.35	2.41	2.37
Extreme(2)	2.15	1.28	2.27	2.38
Extreme(11)	2.15	1.50	2.50	2.38
MA(1)	1.93	0.92	1.96	2.41
Stochastic volatility	1.93	1.46	2.37	2.14

high frequency data are by now well understood, the distribution of tail events has received comparatively little attention in the finance literature. On the other hand, this is of clear importance for such applications as risk management. Here we describe the shape of the tails for two such data sets.

We selected daily returns from the S&P 500 stock index with 18,024 observations from 1928 to 1997, and data extracted from all quotes on the DM–Dollar contract from September 1992 to October 1993. The quote data was supplied by Olsen and Associates who continuously collect these data from the markets. The number of quotes is over 1.5 million, and these quotes are irregularly spaced throughout the year. The quotes were aggregated into 52,588 10-minute return observations. The data and the aggregation procedures are described by Danielsson and de Vries (1997). In order to examine the change in the tail properties of the data over the time interval we decided to create subsamples of the first 2,000 and last 2,000 observations for both data sets in addition to using the first and last 20,000 observations on the foreign exchange rate data, and the entire stock index data set. In the estimation procedure we employed the same grid for n_1 as was used in the simulations; the number of bootstrap resamples, however, was increased to 5,000.

Let P_t be the price at time t of a financial asset like equity or foreign exchange. The compound return on holding such an asset for one period is $\log(P_{t+1}/P_t)$. Hence, returns are denomination free. Therefore returns on different assets can be directly compared. One dimension along which the asset returns can be compared in order to assess their relative risk characteristics is by means of the tail index. Financial corporations are required to use large data sets on past returns to evaluate the risk on their trading portfolio. The minimum required capital stock of these financial institutions is determined on the basis of this risk. The capital requirement ensures that banks can meet the incidental heavy losses that are so characteristic of the financial markets. The frequency of these large losses can be

TABLE IV

Descriptive Statistics

Series	Annualized mean return	Annualized standard error	Skewness	Kurtosis
DM/US				
First 2,000	0.842	0.209	0.70	7.98
Last 2,000	0.431	0.131	0.78	12.82
First 20,000	0.377	0.144	0.31	10.85
Last 20,000	0.051	0.116	−0.01	17.35
S&P500				
First 2,000	−0.080	0.343	0.22	5.33
Last 2,000	0.115	0.117	−0.45	4.53
All 18,024	0.053	0.179	−0.49	22.71

analyzed by means of extreme value theory; see e.g. Jansen and de Vries (1991) for an early example of this approach and Embrechts *et al.* (1997) for a more recent treatment. In this analysis, the measurement of γ is very important because it indicates the shape and heaviness of the distribution of returns. It is the essential input for predictions of out-of-sample losses; see de Haan *et al.* (1994).

In Table IV we report some descriptive statistics. The mean return and standard error of the returns have been annualized because the magnitude in the high frequency returns is typically very small (for the daily return data we assumed 250 trading days per year). As the tables shows, all data exhibit a high kurtosis which points to peakedness in the center of the return distribution and heavy tails. The main results are reported in Table V. We see that the tails are indeed heavy. The $1/\gamma$ estimates show that the number of bounded moments hovers around 3 to 4. The shorter samples necessarily give less precise estimates of γ, but the results for the subsamples appear to be consistent with the large sample results. As was the case in the simulation experiments there is more variation in the $\hat{\rho}$ and $\hat{k}_0(n)$. The table yields an interesting impression concerning the first- and second-order tail indices; It appears that both γ and ρ are about equal for either asset. An economic explanation for this observation might be that arbitrage induces similar tail shapes and hence similar risk properties. The equality of γ across different assets has been suggested before. But due to the fact that this observation depends on the more or less arbitrary choices of $k(n)$, no firm conclusion regarding this observation could be reached. The current method overcomes this problem.

TABLE V

Lower Tail Parameters

Series	$\hat{\gamma}$	$-\hat{\rho}$	$\hat{k}_0(n)$
DM/US			
First 2,000	0.10	9.93	10
Last 2,000	0.35	1.93	29
First 20,000	0.27	1.70	187
Last 20,000	0.30	2.01	64
S&P500			
First 2,000	0.33	1.45	57
Last 2,000	0.24	2.06	13
All 18,024	0.32	1.85	96

REFERENCES

1. R. T. Baillie and M. M. Dacorogna, Special issue on high frequency data in finance, Part 1, *J. Empirical Finance* **4** (1997), 69–293.
2. T. Bollerslev, R. Y. Chou, and K. F. Kroner, ARCH modeling in finance, *J. Econometrics* **52** (1992), 5–59.
3. S. Csörgő, P. Deheuvels, and D. M. Mason, Kernel estimator of the tail index of a distribution, *Ann. Statist.* **13** (1985), 1050–1077.
4. J. Danielsson and C. G. de Vries, Tail index and quantile estimation with very high frequency data, *J. Empirical Finance* **4** (1997), 241–257.
5. R. Davis and S. I. Resnick, Tail estimates motivated by extreme value theory, *Ann. Statist.* **12** (1984), 1467–1487.
6. L. de Haan, D. W. Jansen, K. Koedijk, and C. G. de Vries, Safety first portfolio selection, extreme value theory and long run asset risks, *in* "Extreme Value Theory and Applications" (J. Galambos, J. Lechner, E. Simiu, and N. Macri, Eds.), pp. 471–487, 1994.
7. L. de Haan and S. I. Resnick, A simple asymptotic estimate for the index of a stable distribution, *J. Roy. Statist. Soc. Ser. B* **42** (1980), 83–87.
8. A. L. M. Dekkers and L. de Haan, Optimal choice of sample fraction in extreme-value estimation, *J. Multivariate Anal.* **47**, No. 2 (1993), 173–195.
9. H. Drees and E. Kaufmann, Selection of the optimal sample fraction in univariate extreme value estimation, *Stochast. Process. Appl.* **75** (1998), 149–172.
10. P. Embrechts, C. Kuppelberg, and T. Mikosch, "Modelling Extreme Events," Springer-Verlag, New York/Berlin, 1997.
11. J. Geluk and L. de Haan, Regular variation, extensions and Tauberian theorems, CWI Tract 40, Amsterdam, 1987.
12. P. Hall, On some simple estimates of an exponent of regular variation, *J. Roy. Statist. Soc. Ser. B* **42** (1982), 37–42.
13. P. Hall, Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *J. Multivariate Anal.* **32** (1990), 177–203.

14. P. Hall and A. H. Welsh, Adaptive estimate of parameters of regular variation, *Ann. Statist.* **13** (1985), 331–341.
15. B. M. Hill, A simple general approach to inference about the tail of a distribution, *Ann. Statist.* **3** (1975), 1163–1174.
16. T. Hsing, On tail index estimation using dependent data, *Ann. Statist.* **19** (1991), 1547–1569.
17. D. W. Jansen and C. G. de Vries, On the frequency of large stock returns: putting booms and busts into perspective, *Rev. Econ. Statist.* **73** (1991), 18–24.
18. D. M. Mason, Law of large numbers for sum of extreme values, *Ann. Probab.* **10** (1982), 754–764.
19. V. V. Petrov, "Sums of Independent Random Variables," Springer-Verlag, New York, 1975.
20. J. Pickands, III, Statistical inference using extreme order statistics, *Ann. Statist.* **3** (1975), 119–131.
21. S. Resnick and C. Starica, Tail index estimation for dependent data, *Ann. Appl. Probab.* **8** (1998), 1156–1183.
22. G. Shorack and J. Wellner, "Empirical Processes with Applications to Statistics," Wiley, New York, 1986.