

Confidence intervals for the tail index

SHIHONG CHENG¹ and LIANG PENG²

¹*Department of Probability and Statistics, Peking University, Beijing, 100871, P.R. China*

²*School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.*

E-mail: peng@math.gatech.edu

One of the best-known estimators for the tail index of a heavy-tailed distribution is the Hill estimator. In this paper, confidence intervals based on the asymptotic normal approximation of Hill estimator are studied. The coverage accuracy is evaluated and the theoretical optimal choice of the sample fraction for the one-sided confidence interval is given. One surprising finding is that the order of optimal coverage accuracy for the one-sided confidence interval depends on the sign of the second-order regular variation.

Keywords: confidence interval; coverage accuracy; Hill estimator; tail index

1. Introduction

Several estimators have been proposed for the tail index of a heavy-tailed distribution (for example, Hill 1975; Pickands 1975; Hall 1982; Csörgő *et al.* 1985; Dekkers *et al.* 1989; Drees 1995; Csörgő and Viharos 1997; de Haan and Peng 1998). Since we make inferences about the tail quantity, we can only use the k upper order statistics of a sample of size n ($k/n \rightarrow 0$) in the estimation procedure. When k is small the variance of the tail index estimate is large, and the use of a large value of k introduces a large bias in the estimation. Therefore the choice of k plays an important role in the applications of extremes, for example, high quantile estimation (see de Haan and Rootzén 1993; Danielsson and de Vries 1997a), extreme tail probability estimation (see Hall and Weissman 1997), value-at-risk in finance (see Danielsson and de Vries 1997b; Danielsson *et al.* (1998); Embrechts *et al.* (1998; 1999), and some areas of engineering (see Resnick 1997). Recently, several procedures have been proposed for choosing the optimal value of k in the sense of asymptotic minimal mean squared errors (see Hall 1990; Dekkers and de Haan 1993; Beirlant *et al.* 1996; Drees and Kaufmann 1998; Danielsson *et al.* 2001). Moreover, some new estimators have been proposed to reduce the bias of Hill estimator (see Beirlant *et al.* 1999; Feuerverger and Hall 1999; Guillou and Hall 2001).

In this paper we are interested in the important statistical issue of obtaining the confidence interval for the tail index. We concentrate on the well-known tail index estimator due to (Hill 1975), defined as

$$\hat{\gamma}_n = \frac{1}{k} \sum_{i=1}^k \log X_{n,n-i+1} - \log X_{n,n-k},$$

where X_1, \dots, X_n are independent and identically distributed observations with distribution function F satisfying

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma} \quad (1.1)$$

for all $x > 0$, and $X_{n,1} \leq \dots \leq X_{n,n}$ denote the order statistics of X_1, \dots, X_n . The parameter $\gamma > 0$ is termed the tail index. For the consistency of $\hat{\gamma}_n$ we refer to Mason (1982). In order to derive the limit of $\hat{\gamma}_n$ we need a stricter condition than (1.1). Suppose, as $x \rightarrow \infty$, that

$$1 - F(x) = cx^{-1/\gamma} \{1 + bx^{-\beta} + o(x^{-\beta})\}, \quad (1.2)$$

where $c > 0$, $\beta > 0$ and $b \neq 0$. Note that (1.2) is a special case of the general second-order regular variation (see de Haan and Stadtmüller 1996).

The following proposition easily follows from Peng and Qi (1997) by noting the relation between the second-order regular variation for $1 - F(x)$ and one for the inverse of $1/(1 - F(x))$ – see the proof of Proposition 2.

Proposition 1. *Suppose (1.2) holds and $k \rightarrow \infty$, $k/n \rightarrow 0$. Then*

$$\sqrt{k}(\hat{\gamma}_n - \gamma) \xrightarrow{d} N(0, \gamma^2) \quad (1.3)$$

if and only if $k = o(n^{2\beta\gamma/(1+2\beta\gamma)})$.

Then nominal α -level confidence intervals for γ , based on the normal approximation (1.3), are

$$I_1(\alpha) = \left(0, \hat{\gamma}_n + \frac{z_\alpha \hat{\gamma}_n}{\sqrt{k}} \right) \quad \text{and} \quad I_2(\alpha) = \left(\hat{\gamma}_n - \frac{x_\alpha \hat{\gamma}_n}{\sqrt{k}}, \hat{\gamma}_n + \frac{x_\alpha \hat{\gamma}_n}{\sqrt{k}} \right),$$

where x_α and z_α are defined by

$$P(|N(0, 1)| \leq x_\alpha) = \alpha \quad \text{and} \quad P(N(0, 1) \leq z_\alpha) = \alpha.$$

In Section 2 we study the coverage accuracy for confidence intervals $I_1(\alpha)$ and $I_2(\alpha)$, and give the theoretical optimal choice of sample fraction k for $I_1(\alpha)$ in the sense of minimizing the absolute coverage error. A data-driven method for choosing the optimal sample fraction is given in Section 3 and a simulation study is presented in Section 4. All proofs are deferred to Section 5.

2. Coverage accuracy

In the study of the coverage accuracy of a confidence interval, one basic tool is the Edgeworth expansion (see Hall 1992). The following Edgeworth expansion for the Hill estimator follows from Cheng and Pan (1998).

Proposition 2. *Suppose (1.2) holds and $k \rightarrow \infty$, $k/n \rightarrow 0$. Then*

$$P\left(\frac{\sqrt{k}(\hat{\gamma}_n - \gamma)}{\gamma} \leq x\right) = \Phi(x) + \phi(x) \left\{ \frac{1 - x^2}{3\sqrt{k}} + \frac{b\beta\gamma}{(1 + \beta\gamma)c^{\beta\gamma}} \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma} \right\} + o\left(\frac{1}{\sqrt{k}} + \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma}\right) \tag{2.1}$$

uniformly on R , where $\Phi(x)$ and $\phi(x)$ denote the standard normal distribution function and density function, respectively.

Remark 1. Expansion (2.1) was used by Cheng and Pan (1998) to construct a more accurate confidence interval by taking account of $\phi(x)(1 - x^2)/(3\sqrt{k})$, but ignoring $\phi(x)(b\beta\gamma/((1 + \beta\gamma)c^{\beta\gamma}))\sqrt{k}(n/k)^{-\beta\gamma}$. In this paper we use both terms to choose the optimal sample fraction in terms of minimizing the absolute coverage error.

Based on the Edgeworth expansion above, we may show our main results which give the coverage probabilities for $I_1(\alpha)$ and $I_2(\alpha)$.

Theorem 1. *Suppose (1.2) holds and $k \rightarrow \infty$, $k/n \rightarrow 0$. Then*

$$P(\gamma \in I_1(\alpha)) = \alpha - \phi(z_\alpha) \left\{ \frac{1 + 2z_\alpha^2}{3\sqrt{k}} + \frac{b\beta\gamma}{(1 + \beta\gamma)c^{\beta\gamma}} \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma} \right\} + o\left(\frac{1}{\sqrt{k}} + \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma}\right) \tag{2.2}$$

and

$$P(\gamma \in I_2(\alpha)) = \alpha + o\left(\frac{1}{\sqrt{k}} + \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma}\right). \tag{2.3}$$

Hence the optimal value of k that minimizes the absolute value of the leading coverage error term in (2.2) is

$$k^* = \begin{cases} \left\{ \frac{(1 + 2z_\alpha^2)(1 + \beta\gamma)c^{\beta\gamma}}{3b\beta\gamma(1 + 2\beta\gamma)} \right\}^{1/(1+\beta\gamma)} n^{\beta\gamma/(1+\beta\gamma)} & \text{if } b > 0, \\ \left\{ \frac{(1 + 2z_\alpha^2)(1 + \beta\gamma)c^{\beta\gamma}}{-3b\beta\gamma} \right\}^{1/(1+\beta\gamma)} n^{\beta\gamma/(1+\beta\gamma)} & \text{if } b < 0, \end{cases} \tag{2.4}$$

which automatically satisfies the condition $k = o(n^{2\beta\gamma/(1+2\beta\gamma)})$ in Proposition 1. Furthermore, the optimal coverage accuracy for $I_1(\alpha)$ is

$$P(\gamma \in I_1(\alpha)) =$$

$$\begin{cases} \alpha - 2 \left\{ \frac{(1 + \beta\gamma)(1 + 2z_a^2)}{3(1 + 2\beta\gamma)} \right\}^{(1+2\beta\gamma)/(2(1+\beta\gamma))} \left\{ \frac{b\beta\gamma}{c^{\beta\gamma}} \right\}^{1/(2(1+\beta\gamma))} \\ \times \phi(z_a) n^{-\beta\gamma/(2(1+\beta\gamma))} (1 + o(1)) & \text{if } b > 0 \\ \alpha + o(n^{-\beta\gamma/(2(1+\beta\gamma))}) & \text{if } b < 0. \end{cases} \quad (2.5)$$

We remark from (2.4) and (2.5) that the order of the optimal coverage accuracy for $I_1(\alpha)$ depends on the sign of the second-order regular variation.

In order to obtain the optimal choice of k for $I_2(\alpha)$, we may need a stricter condition than (1.2) – that is, third-order regular variation – and we conjecture that the optimal choice of k depends on the third-order parameter as well. Hence seeking a data-driven method for the choice of k becomes much more difficult.

3. Choice of k

Since the optimal sample fraction, in terms of coverage probability, depends on some unknown quantities, we propose a plug-in estimator for the optimal sample fraction by focusing on the one-sided confidence interval in this section. Put

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

$$\hat{\rho}_n = -(\log 2)^{-1} \log \left| \frac{M_n^{(2)}(n/(2\sqrt{\log n})) - 2[M_n^{(1)}(n/(2\sqrt{\log n}))]^2}{M_n^{(2)}(n/\sqrt{\log n}) - 2[M_n^{(1)}(n/\sqrt{\log n})]^2} \right|,$$

$$\hat{\delta}_n = \left\{ M_n^{(2)} \frac{n}{\sqrt{\log n}} - 2 \left[M_n^{(1)} \frac{n}{\sqrt{\log n}} \right]^2 \right\} (1 + \hat{\rho}_n) (\sqrt{\log n})^{\hat{\rho}_n} \left\{ 2 \left[M_n^{(1)} \left(\frac{n}{\sqrt{\log n}} \right) \right]^2 \hat{\rho}_n \right\}^{-1}.$$

Then it follows from the proof of Theorem 2.1 in Peng (1998) that

$$\hat{\rho}_n \xrightarrow{P} \beta\gamma,$$

$$\hat{\delta}_n \xrightarrow{P} b\beta\gamma(1 + \beta\gamma)^{-1} c^{-\beta\gamma},$$

as $n \rightarrow \infty$. Therefore the plug-in estimator for the optimal sample fraction k^* given in (2.4) is

$$\hat{k} = \begin{cases} \left\{ \frac{1 + 2z_\alpha^2}{3\hat{\delta}_n(1 + 2\hat{\rho}_n)} \right\}^{1/(1+\hat{\rho}_n)} n^{\hat{\rho}_n/(1+\hat{\rho}_n)} & \text{if } \hat{\delta}_n > 0, \\ \left\{ \frac{1 + 2z_\alpha^2}{-3\hat{\delta}_n} \right\}^{1/(1+\hat{\rho}_n)} n^{\hat{\rho}_n/(1+\hat{\rho}_n)} & \text{if } \hat{\delta}_n < 0, \end{cases}$$

where z_α satisfies $P(N(0, 1) \leq z_\alpha) = \alpha$.

4. Simulation study

We consider two different distributions $F(x) = \exp\{-x^{-1/\gamma}\}$ ($x > 0$) and $F(x) = 1 - \frac{1}{2}x^{-1/\gamma} - \frac{1}{2}x^{-1/\gamma-\beta}$ ($x \geq 1$), which have second-order regular variation of different sign, that is different sign of b in (1.2).

In the simulation we take $\gamma = 1.0$, $\beta = 1.0/\gamma$, sample size $n = 400, 1000, 5000$ and simulate 500 times for each case. The empirical coverage probabilities of confidence intervals $I_1(0.95)$ and $I_2(0.95)$ are plotted against different values of sample fraction k in Figures 1 and 2. Figure 2 clearly shows that coverage accuracy for $I_2(0.95)$ is better than that for $I_1(0.95)$, and the optimal coverage probability for $I_1(0.95)$ in Figure 1 is much better than that in Figure 2. The reason is that the second-order regular variation in Figure 2 has positive sign.

Next we examine the behaviour of our plug-in estimator, proposed in Section 3, for the optimal sample fraction in terms of minimizing absolute coverage error. In Table 1 we report the ratio of the empirical coverage probability with estimated optimal sample fraction \hat{k} to the empirical coverage probability with theoretical optimal sample fraction k^* . We find that the ratio becomes close to 1 as the sample size increases. However, a more accurate estimator for the theoretical optimal sample fraction k^* is needed.

5. Proofs

Proof of Proposition 2. Let $U(x)$ denote the inverse function of $1/(1 - F(x))$. Note that (1.2) implies that

$$\lim_{t \rightarrow \infty} \frac{(1 - F(tx))/(1 - F(t)) - x^{-1/\gamma}}{-b\beta t^{-\beta}} = x^{-1/\gamma} \frac{x^{-\beta} - 1}{-\beta}, \quad x > 0,$$

which is equivalent to

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\gamma}{-\gamma^2 b\beta (U(t))^{-\beta}} = x^\gamma \frac{x^{-\beta\gamma} - 1}{-\beta\gamma}, \quad x > 0,$$

that is

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\gamma}{-\gamma^2 b\beta c^{-\beta\gamma} t^{-\beta\gamma}} = x^\gamma \frac{x^{-\beta\gamma} - 1}{-\beta\gamma}, \quad x > 0.$$

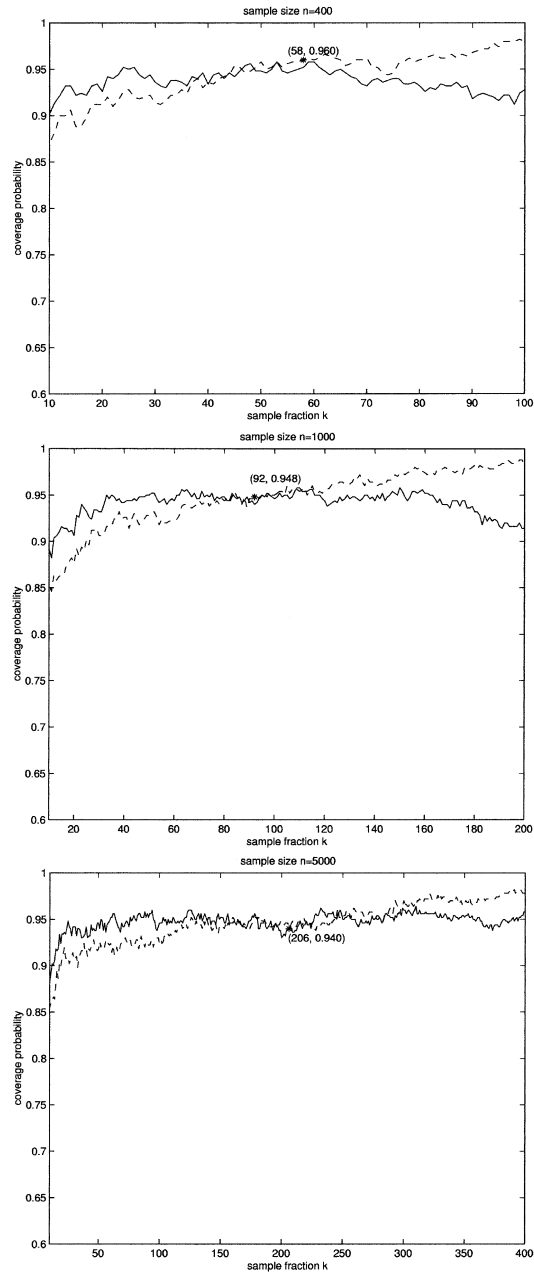


Figure 1. Coverage probabilities for $I_2(0.95)$ (solid line) and $I_1(0.95)$ (dashed line) with distribution function $F(x) = \exp\{-x^{-1}\}$, $x > 0$. The star point stands for the theoretical optimal choice of k and its corresponding empirical coverage probability.

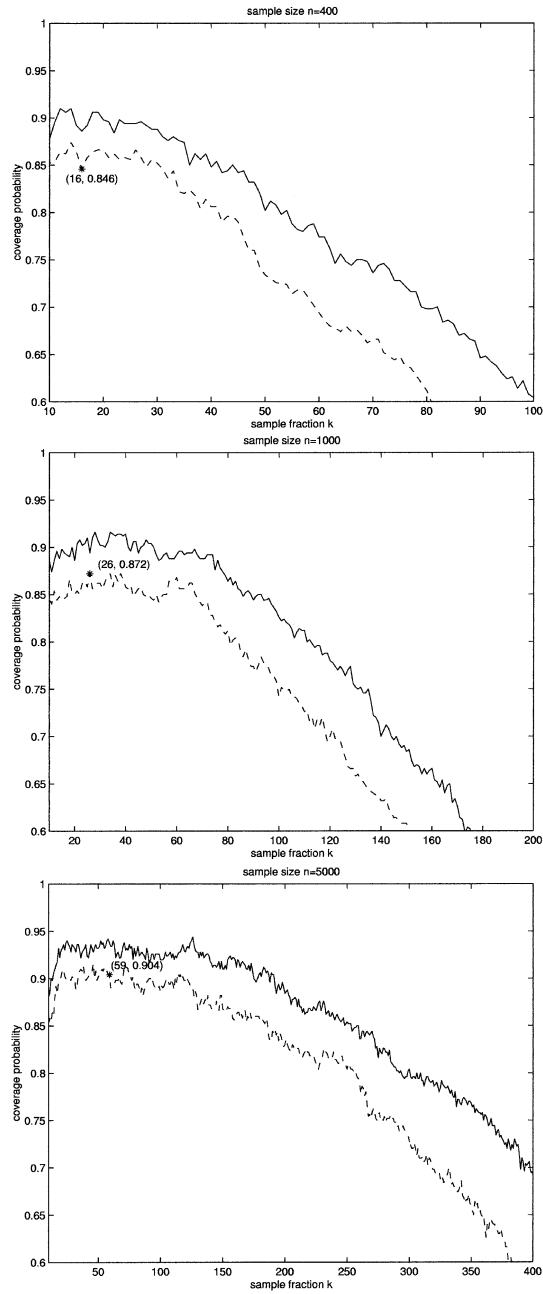


Figure 2. Coverage probabilities for $I_2(0.95)$ (solid line) and $I_1(0.95)$ (dashed line) with distribution function $F(x) = 1 - \frac{1}{2}x^{-1} - \frac{1}{2}x^{-2}$, $x > 1$. The star point stands for the theoretical optimal choice of k and its corresponding empirical coverage probability.

Table 1. Ratio of the empirical coverage probability with estimated optimal sample fraction \hat{k} to the empirical coverage probability with theoretical optimal sample fraction k^*

n	$F(x)$	
	$\exp(-x^{-1})$	$1 - \frac{1}{2}x^{-1} - \frac{1}{2}x^{-2}$
400	0.93	0.87
1000	0.97	0.90
5000	0.99	0.93

The rest of proof is similar to the proof of Theorem 1.1 in Cheng and Pan (1998). □

Proof of Theorem 1. It follows from Proposition 2 that

$$\begin{aligned}
 P(\gamma \in I_1(\alpha)) &= P(\gamma < \hat{\gamma}_n + z_\alpha \hat{\gamma}_n / \sqrt{k}) \\
 &= P\left(\hat{\gamma}_n > \frac{\gamma}{1 + z_\alpha / \sqrt{k}}\right) \\
 &= P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} > \frac{-z_\alpha}{1 + z_\alpha / \sqrt{k}}\right) \\
 &= 1 - P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} \leq \frac{-z_\alpha}{1 + z_\alpha / \sqrt{k}}\right) \\
 &= 1 - \Phi\left(\frac{-z_\alpha}{1 + z_\alpha / \sqrt{k}}\right) - \phi\left(\frac{-z_\alpha}{1 + z_\alpha / \sqrt{k}}\right) \\
 &\quad \times \left\{ \frac{1 - (-z_\alpha / (1 + z_\alpha / \sqrt{k}))^2}{3\sqrt{k}} + \frac{b\beta\gamma}{(1 + \beta\gamma)c^{\beta\gamma}} \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma} \right\} + o\left(\frac{1}{\sqrt{k}} + \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma}\right) \\
 &= \alpha - \phi(z_\alpha) \left\{ \frac{1 + 2z_\alpha^2}{3\sqrt{k}} + \frac{b\beta\gamma}{(1 + \beta\gamma)c^{\beta\gamma}} \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma} \right\} + o\left(\frac{1}{\sqrt{k}} + \sqrt{k} \left(\frac{n}{k}\right)^{-\beta\gamma}\right).
 \end{aligned}$$

Note that as $x_\alpha / \sqrt{k} < 1$,

$$\begin{aligned}
P(\gamma \in I_2(\alpha)) &= P\left(\gamma < \hat{\gamma}_n + \frac{x_\alpha \hat{\gamma}_n}{\sqrt{k}}\right) - P\left(\gamma \leq \hat{\gamma}_n - \frac{x_\alpha \hat{\gamma}_n}{\sqrt{k}}\right) \\
&= P\left(\hat{\gamma}_n > \frac{\gamma}{1 + x_\alpha/\sqrt{k}}\right) - P\left(\hat{\gamma}_n \geq \frac{\gamma}{1 - x_\alpha/\sqrt{k}}\right) \\
&= P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} > \frac{-x_\alpha}{1 + x_\alpha/\sqrt{k}}\right) - P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} \geq \frac{x_\alpha}{1 - x_\alpha/\sqrt{k}}\right) \\
&= P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} < \frac{x_\alpha}{1 - x_\alpha/\sqrt{k}}\right) - P\left(\sqrt{k} \frac{(\hat{\gamma}_n - \gamma)}{\gamma} \leq \frac{-x_\alpha}{1 + x_\alpha/\sqrt{k}}\right).
\end{aligned}$$

Thus, (2.3) follows from Proposition 2. □

Acknowledgement

We thank two reviewers for their helpful comments.

References

- Beirlant, J., Vynckier, P. and Teugels, J.L. (1996) Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, **91**, 1659–1667.
- Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999) Tail index estimation and an exponential regression model. *Extremes*, **2**(2), 177–200.
- Cheng, S. and Pan, J. (1998) Asymptotic expansions of estimators for the tail index with applications. *Scand. J. Statist.*, **25**, 717–728.
- Csörgő, S. and Viharos, L. (1997) Asymptotic normality of least-squares estimators of tail indices. *Bernoulli*, **3**(3), 351–370.
- Csörgő, S., Deheuvels, P. and Mason, D. (1985) Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050–1077.
- Danielsson, J. and de Vries, C. (1997a) Tail index and quantile estimation with very high frequency data. *J. Empir. Finance*, **4**, 241–257.
- Danielsson, J. and de Vries, C. (1997b) Value-at-risk and extreme returns. FMG Discussion Paper No. 273, Financial Markets Group, London School of Economics.
- Danielsson, J., Hartmann, P. and de Vries, C. (1998) The cost of conservatism. *RISK*, **11**(1), 101–103.
- 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 Anal.*, **76**, 226–248.
- de Haan, L. and Peng, L. (1998) Comparison of tail index estimators. *Statist. Neerlandica*, **52**(1), 60–70.
- de Haan, L. and Rootzén, H. (1993) On the estimation of high quantiles. *J. Statist. Plann. Inference*, **35**, 1–13.
- de Haan, L. and Stadtmüller, U. (1996) Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, **61**, 381–395.

- Dekkers, A.L.M. and de Haan, L. (1993) Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.*, **47**(2), 173–195.
- 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.
- Drees, H. (1995) Refined Pickands estimators of the extreme value index. *Ann. Statist.*, **23**, 2059–2080.
- Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, **75**, 149–172.
- Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1998) Living on the edge. *RISK*, **11**(1), 96–100.
- Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1999) Extreme value theory as a risk management tool. *North Amer. Actuar. J.*, **3**(2), 30–41.
- Feuerverger, A. and Hall, P. (1999) Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.*, **27**, 760–781.
- Guillou, A. and Hall, P. (2001) A diagnostic for selecting the threshold in extreme value analysis. *J. Roy. Statist. Soc. Ser. B*, **63**, 293–305.
- Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, **44**, 37–42.
- Hall, P. (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.*, **32**, 177–203.
- Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Berlin: Springer-Verlag.
- Hall, P. and Weissman, I. (1997) On the estimation of extreme tail probabilities. *Ann. Statist.*, **25**, 1311–1326.
- Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.
- Mason, D. (1982) Laws of large numbers for sums extreme values. *Ann. Probab.*, **10**, 754–764.
- Peng, L. (1998) Asymptotically unbiased estimators for the extreme-value index. *Statist. Probab. Lett.*, **38**, 107–115.
- Peng, L. and Qi, Y. (1997) Asymptotic normality of Hill estimator in a second-order submodel of regular variation. *Chinese Ann. Math. Ser. A*, **18**(5), 539–544.
- Pickands, J. III (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.
- Resnick, S. (1997) Heavy tail modeling and teletraffic data (with discussion). *Ann. Statist.*, **25**, 1805–1869.

Received January 2000 and revised March 2001