The Annals of Statistics
2000, Vol. 28, No. 1, 254–274

HOW TO MAKE A HILL PLOT

BY HOLGER DREES,[1] LAURENS DE HAAN[2] AND SIDNEY RESNICK[3]

University of Cologne, Erasmus University and Cornell University

An abundance of high quality data sets requiring heavy tailed models necessitates reliable methods of estimating the shape parameter governing the degree of tail heaviness. The Hill estimator is a popular method for doing this but its practical use is encumbered by several difficulties. We show that an alternative method of plotting Hill estimator values is more revealing than the standard method unless the underlying data comes from a Pareto distribution.

1. Introduction. It is now common in diverse fields such as insurance [McNeil (1997), Resnick (1997a)], finance and economics [Jansen and de Vries (1991)], computer science and telecommunications [Leland, Taqqu, Willinger and Wilson (1994), Resnick (1997b)] to encounter large, high quality data sets for which appropriate models require heavy tailed distributions. By a heavy tailed distribution we mean a distribution F, which satisfies

$$1 - F(x) \sim x^{-\alpha} L(x), \qquad x \to \infty, \ \alpha > 0, \tag{1.1}$$

where L is a slowly varying function satisfying $\lim_{t\to\infty} L(tx)/L(t) = 1$, for all $x > 0$.

A basic statistical calibration problem is to estimate the shape parameter α based on a sample from the process $(X_n)_{n\in\mathbb{N}}$, assumed to be a stationary sequence whose marginal, one-dimensional distribution is F which satisfies (1.1). A popular estimator of the so-called extreme value index $\gamma := \alpha^{-1}$ based on $X_1, \dots, X_n$ is the Hill estimator obtained as follows. Order the observations as $X_{(1)} \geq \cdots \geq X_{(n)}$ and then the Hill estimator based on $k+1$ upper order statistics is

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

for $k = 1, \dots, n-1$. This estimator is consistent for γ in the sense that if $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence, which means

$$k_n \longrightarrow \infty, \qquad k_n/n \longrightarrow 0,$$

Received July 1998; revised July 1999.

[1]Supported in part by Netherlands Organization for Scientific Research through the Netherlands Mathematical Research Foundation and European Union TMR Grant ERB-FMRX-CT 960095.

[2]Supported in part by European Union TMR Grant ERB-FMBI-CT 961465.

[3]Supported in part by NSF Grant DMS-97-04982 and by Office of Naval Research Contract N00014-93-1-0043.

AMS 1991 *subject classifications*. Primary 60G75; secondary 62G05, 62G20.

Key words and phrases. Estimation, extreme value index, heavy tails, Hill estimator, order statistics, Pareto tails, regular variation.

then

$$H_{k_n, n} \xrightarrow{P} \gamma,$$

provided $\{X_n\}$ is a stationary sequence satisfying one of a broad set of assumptions such as $\{X_n\}$ is iid [Mason (1982)], $\{X_n\}$ can be written as a finite or infinite order moving average [Resnick and Stărică (1995)], $\{X_n\}$ satisfies mixing conditions [Rootzen, Leadbetter, de Haan (1990)] or if $\{X_n\}$ is an ARCH(1) process [Resnick and Stărică (1998)], a bilinear process [Davis and Resnick (1996), Resnick and van den Berg (2000)] or consists of random variables defined on a Markov chain [Resnick and Stărică (1998)].

Although consistency holds for all intermediate sequences $(k_n)_{n\in\mathbb{N}}$, the performance of $H_{k,n}$ strongly depends on the particular choice of the number k of order statistics. Under suitable second-order conditions, a sequence $(k_n^{\text{opt}})_{n\in\mathbb{N}}$ can be determined such that the asymptotic mean squared error of the Hill estimator is minimized. The practical usefulness of knowing k_n^{opt} is limited by the fact that k_n^{opt} is determined only up to asymptotic equivalence, only providing a solution minimizing asymptotic mse and there is little guidance available about finite sample behavior. Furthermore, k_n^{opt} depends on unknown parameters of F (see Theorem 1 below) and hence has to be replaced by an asymptotically equivalent data-driven choice $\hat{k}_n^{\text{opt}}$ using, for example, a subsample bootstrap method [see Danielsson, de Haan, Peng and de Vries (1998)] or a sequential approach where the estimator for the optimal number is defined in terms of certain stopping times [see Drees and Kaufmann (1998)]. Both procedures require the choice of certain parameters and the choices are arbitrary. For the sequential procedure one must choose the threshold r_n and the parameter ξ, while the bootstrap method requires the choice of the subsample size n_1 and of the range of k-values in which one searches for the minimum of the bootstrap statistic. [The latter remark also applies analogously to the heuristically motivated procedure introduced by Beirlant, Vynckier and Teugels (1996).]

For these methods, the choices do not matter asymptotically, but influence the performance of the resulting adaptive Hill estimator for finite samples. It is advisable to have computationally less demanding methods for a variety of applied purposes as well as for the purpose of checking whether these automatic procedures yield a reasonable number k.

Thus, in practice, it is advisable to construct a plot of the points $\{(k, H_{k,n}), 1 \le k \le n-1\}$ called a Hill plot and then the value of γ is inferred from a stable region in the graph. This is sometimes difficult since the plot may be volatile and/or may not spend a large portion of the display space in the neighborhood of γ. In fact, it is becoming increasingly clear that the traditional Hill plot is most effective only when the underlying distribution is Pareto or very close to Pareto. For the Pareto distribution,

$$1 - F(x) = \left(\frac{x}{\sigma}\right)^{-\alpha}, \qquad x > \sigma > 0,$$

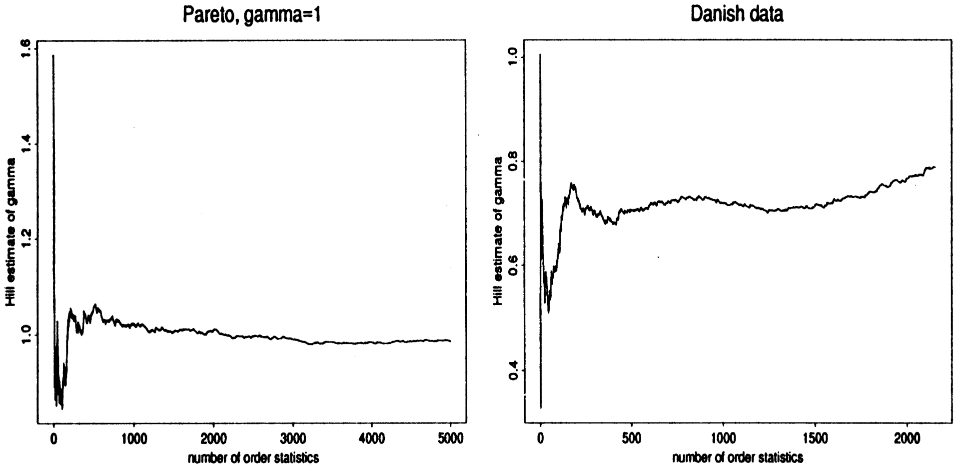

FIG. 1. *Hill plots of* 5000 *Pareto observations,* $\gamma = 1$ (*left*) *and the Danish loss data* (*right*).

one expects the Hill plot to be close to γ for the right side of the plot, since the Hill estimator $H_{n-1,n}$ is the maximum likelihood estimator in the Pareto model. This is borne out in practice. When only (1.1) holds, however, the Hill estimator is only an approximate maximum likelihood estimator based on observations which are exceedances over $X_{(k+1)}$ divided by the threshold $X_{(k+1)}$ and it is less clear what portion of the plot is most accurate.

Two examples where Hill plotting works well are shown in Figure 1. The left plot is a Hill plot for 5000 iid observations from the Pareto distribution with $\alpha = \gamma = 1$. Notice the right side of the graph clearly indicates the correct value of 1. The right plot is the Hill plot of the Danish large fire insurance claim data [see Resnick (1997a)] showing that sometimes the Hill plot can be quite clear and informative for real data.

What do we do when the Hill plot is not so informative? C. Stărică [Resnick and Stărică (1997)] has suggested a simple device called *alt* (alternative) plotting. Instead of plotting $\{(k, H_{k,n}), 1 \le k \le n-1\}$, we construct the *altHill* plot by plotting $\{(\theta, H_{\lceil n^\theta \rceil, n}), 0 \le \theta < 1\}$; that is, one uses a logarithmic scale for the k-axis. (Here $\lceil n^\theta \rceil$ denotes the smallest integer greater than or equal to n^θ.) This has the effect of stretching the left half of the Hill plot and giving more display space to smaller values of k. This will clearly not be beneficial when the underlying distribution is Pareto, but as the following plots show, is beneficial in a wide variety of circumstances.

Figure 2 displays on the left the traditional Hill plot corresponding to a sample of size 5000 from the symmetric stable ($\alpha = 0.2, \gamma = 5$) distribution alongside the alt-plot which is more revealing. The information in the alt-plot would be further enhanced by applying a smoothing procedure given in Resnick and Stărică (1997). One would have to be paranormal to discern with confidence the true value from the Hill plot. Figure 3 shows on the left

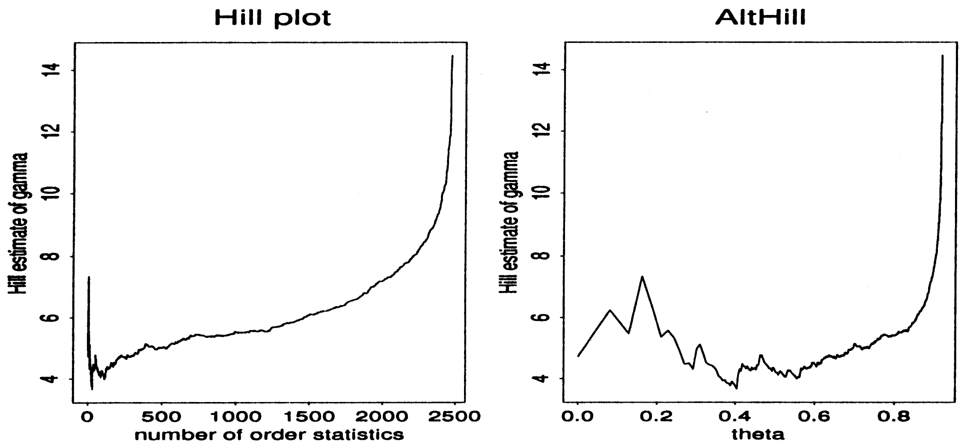

FIG. 2. *Hill and altHill plot of stable observations,* $\gamma = 5$.

a Hill plot from a sample of size 5000 from the distribution of the random variable $g(U)$ where $g(x) = x^{-1}/\log x^{-1}$ and U is uniform on (0,1). For this logarithmically perturbed Pareto distribution, F satisfies

$$1 - F(x) = g^{-1}(x) \sim x^{-1}/\log x, \qquad x \to \infty,$$

since $g(ax^{-1}/\log x) \sim x/a$ for all $a > 0$. The right-hand alt-plot shows more clearly the true value of $\gamma = 1$. Finally Figure 4 compares the traditional Hill plot with the alt-plot for a real teletraffic data set consisting of interarrival times of packets in an ISDN network. The alt-plot makes plausible an estimate of $\alpha = 1.1$; the traditional Hill plot is rather uninformative.

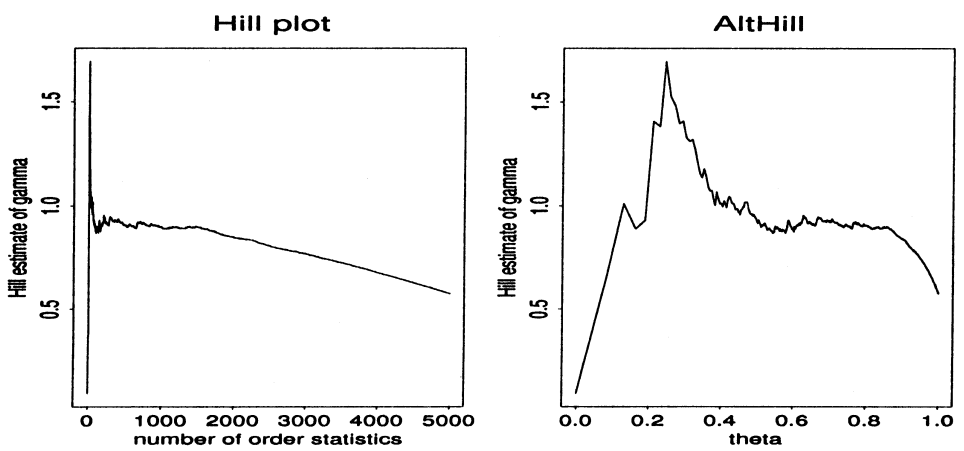

FIG. 3. *Hill and altHill plot for the logarithmically perturbed Pareto,* $\gamma = 1$.

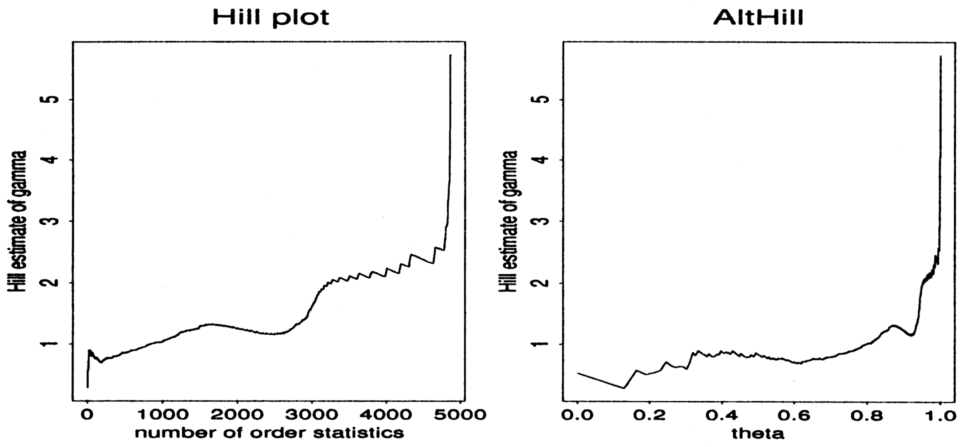

FIG. 4. *Hill and altHill plot of ISDN2.*

The engineering conclusion we emphasize in this paper is that for iid observations whose common distribution has a tail satisfying a second-order condition, alt-plotting is superior. See Theorem 2 and the accompanying discussion. For the Pareto distribution, the traditional Hill plot is preferred. However, one never knows in practice what second-order tail conditions apply so our firm recommendation is to produce both the traditional and altHill plots and compare them. Hill and altHill plotting are useful complimentary methods which can be added to the heavy tailed analyst's tool box to accompany time-honored methods using QQ, residual life and various extreme value plots.

We quantify superiority of one plotting method over another in terms of the occupation time of the plots in a neighborhood of the true value of γ. The percentage PERHILL of time the Hill plot up to $H_{l,n}$ spends in an ε-neighborhood of the true value is defined as

$$\text{PERHILL}(\varepsilon, n, l) := \frac{1}{l}\sum_{i=1}^{l} 1_{\{|H_{i,n}-\gamma|\le\varepsilon\}}$$

and the percentage PERALT of time that the alt-plot up to $H_{\lceil n^u\rceil,n}$ spends in the ε-neighborhood is

$$\text{PERALT}(\varepsilon, n, u) = \frac{1}{u}\int_0^u 1_{\{|H_{\lceil n^\theta\rceil,n}-\gamma|\le\varepsilon\}}\,d\theta.$$

Note that for $u = \log(l+1)/\log n$ both statistics are based on the same set $\{H_{i,n}, 1 \le i \le l\}$. Asymptotic results for these two quantities are given in Section 2 which show the superiority of the alt method, unless the distribution is Pareto-like, provided $l = l_n$ constitutes a suitable intermediate sequence. In order to capture as much of the whole Hill plot or alt-plot as possible, we will choose l_n such that n/l_n tends to infinity slower than every power of n, for example, $l_n = n/\log n$.

We would prefer results not limited by l or u and have achieved this in the Pareto case. See Theorem 3. However, the regular variation condition (1.1) and its second-order refinement (2.2) controls behavior only in the right tail and hence only affects the Hill plot away from the origin. To control that part of the Hill plot corresponding to order statistics not determined by the right tail, one needs an assumption on the central part and the left tail. We are loath to assume anything about these parts of the distribution for what is essentially a right tail estimation problem and hence in Theorem 2 are left with the alternative of giving results for the plots restricted by l and u.

2. Results. In the sequel, we assume that iid random variables $\{X_n, n \in \mathbb{N}\}$, with common distribution function F are observed. In order to derive the asymptotics of the PERHILL and PERALT statistics, we need second-order conditions on the underlying distribution. Recall that (1.1) holds if and only if the quantile function $U(t) := F^{\leftarrow}(1 - 1/t)$ satisfies

$$\log U(tx) - \log U(t) \to \gamma \log x \tag{2.1}$$

as $t \to \infty$. A more precise second-order assumption which strengthens (2.1) is the following condition:

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^\rho - 1}{\rho}, \qquad x > 0, \tag{2.2}$$

for some $\rho \le 0$ and some function A: $(0, \infty) \to \mathbb{R}$ which ultimately is of constant sign. Then, necessarily, $|A|$ is regularly varying with index ρ. Although condition (2.2) is actually stronger than (2.1), it holds for almost all usual textbook distributions satisfying (2.1), including the extreme value distributions ($\rho = -1$), Student's t_ν-distributions ($\alpha = 1/\nu$, $\rho = -2/\nu$), the loggamma distribution ($\rho = 0$) and stable distributions ($\rho = -1$, except for the Cauchy distribution where $\rho = -2$). The only well-known exceptions are the Pareto distributions, for which the numerator of the left-hand side of (2.2) vanishes. For further discussion of this condition and its relation to other second-order conditions, we refer to Dekkers and de Haan (1993), de Haan and Stadtmüller (1996), de Haan, Peng and Pereira (1997) and Drees (1998b).

Most important for our investigations of the asymptotic behavior of the PERHILL and PERALT statistics will be the following approximation of the Hill process, which is of interest on its own.

THEOREM 1. *Under condition* (2.2), *there exist versions of* $H_{i,n}, 1 \le i \le n-1, n \in \mathbb{N}$ *and a standard Brownian motion* W *such that for all intermediate sequences* $(j_n)_{n\in\mathbb{N}}$ *and* $(l_n)_{n\in\mathbb{N}}$,

$$H_{i,n} - \left(\gamma + \gamma \frac{W(i)}{i} + \frac{A(n/i)}{1-\rho}\right) = O\left(\frac{\log i}{i}\right) + o(A(n/i)) \quad a.s. \tag{2.3}$$

uniformly for $j_n \le i \le l_n$. *Moreover, there exist iid standard exponential random variables* ξ_n^*, $n \in \mathbb{N}$, *such that for* $S_i^* := \sum_{n=1}^{i} \xi_n^*$ *one has*

$$H_{i,n} - \gamma \frac{S_i^*}{i} = O(A(n/i)) \quad a.s. \tag{2.4}$$

uniformly for $1 \le i \le l_n$ *as* $n \to \infty$.

Kaufmann and Reiss (1998) established closely related approximations of the Hill process under the assumption that U is normalized regularly varying, but these results are not directly applicable for our purposes, since for small i their bounds, which do not depend on i, may be of larger order than the statistic $H_{i,n} - \gamma$ which is to be approximated. See also Mason and Turova (1994).

Often it is more convenient to parametrize the Hill process continuously.

COROLLARY 1. *Let* $(k_n)_{n\in\mathbb{N}}$ *denote an arbitrary intermediate sequence. Under condition* (2.2), *there exists a sequence of Brownian motions* W_n, *such that for the versions of* $H_{i,n}$ *used in Theorem* 1 *one has*

$$\begin{aligned} &\sup_{t_n \le t \le T_n} (t^{1/2} \wedge t^{\rho-\iota}) \left| H_{\lceil k_n t \rceil, n} - \left(\gamma + k_n^{-1/2} \gamma \frac{W_n(t)}{t} + A(n/k_n) \frac{t^{-\rho}}{1-\rho} \right) \right| \\ &\qquad = o_P\big(k_n^{-1/2} + A(n/k_n)\big) \end{aligned} \tag{2.5}$$

for all $\iota > 0$ *and all* $t_n \to 0$, $T_n \to \infty$ *satisfying* $k_n t_n \to \infty$ *and* $k_n T_n / n \to 0$. *Moreover*,

$$\begin{aligned} &\sup_{0 < t \le T_n} (h(t) \wedge t^{\rho-\iota}) \left| H_{\lceil k_n t \rceil, n} - \left(\gamma + k_n^{-1/2} \gamma \frac{W_n(t)}{t} + A(n/k_n) \frac{t^{-\rho}}{1-\rho} \right) \right| \\ &\qquad = o_P\big(k_n^{-1/2} + A(n/k_n)\big) \end{aligned} \tag{2.6}$$

if $t \mapsto t/h(t)$ *is an upper class function of a standard Brownian motion, for example, if*

$$\lim_{t\to 0} h(t)(\log\ \log(1/t)/t)^{1/2} = 0.$$

Note that (2.5) is less accurate than (2.3) for large t and that (2.6) is less accurate for both small and large t. According to Corollary 1, under condition (2.2) the middle part of the Hill plot can be approximated by the shifted graph of a power function that is perturbed by small random fluctuations if the sample size is sufficiently large. This fact can be employed to check graphically whether condition (2.2) is fulfilled. For example, the Hill plot based on stable random variables ($\rho = -1$) should have a linear trend, which indeed shows up in the plots of Figure 2.

From Corollary 1 it is easily seen that the optimal rate of convergence (in terms of the asymptotic mean squared error) is obtained if $k_n^{1/2}|A(n/k_n)|$ tends to a positive constant, for example,

$$k_n^{1/2}|A(n/k_n)| \longrightarrow 1. \tag{2.7}$$

Observe that, according to Theorem 1.5.12 of Bingham, Goldie and Teugels (1987), relation (2.7) is satisfied by an intermediate sequence, which is unique up to asymptotic equivalence [see discussion item (2) after Theorem 2]. The resulting Hill estimator is asymptotically biased; that is, the limiting distribution of the standardized estimator is not centered. This effect, however, is common in nonregular non- or semiparametric estimation problems like density estimation or regression.

Since the rate of convergence of the optimal Hill estimator is $k_n^{-1/2}$, it is natural to examine the asymptotic behavior of PERHILL and PERALT for a neighborhood shrinking with this rate towards the true value γ.

THEOREM 2. *Suppose that* $(k_n)_{n\in\mathbb{N}}$ *and* $(l_n)_{n\in\mathbb{N}}$ *are intermediate sequences satisfying* (2.7) *and* $l_n/k_n \to \infty$, *and let* $u_n := \log(l_n+1)/\log n$. *Then for* $\rho < 0$ *we have*

$$\begin{aligned}\frac{l_n}{k_n} PERHILL(k_n^{-1/2}\varepsilon, n, l_n) &= \frac{1}{k_n}\sum_{i=1}^{l_n} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} \\ &\stackrel{d}{\to} \int_0^\infty 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)|\le\varepsilon\}}\,dt\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}\log(l_n+1) PERALT(k_n^{-1/2}\varepsilon, n, u_n) &= \log n \int_0^{u_n} 1_{\{k_n^{1/2}|H_{\lceil n^\theta\rceil,n}-\gamma|\le\varepsilon\}}\,d\theta \\ &\stackrel{d}{\to} \int_0^\infty 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)|\le\varepsilon\}}\,\frac{dt}{t},\end{aligned} \tag{2.9}$$

where the limit random variables are finite almost surely. If, in addition, $|A|$ *is eventually decreasing, then we have for* $\rho = 0$,

$$\frac{l_n}{k_n} PERHILL(k_n^{-1/2}\varepsilon, n, l_n) \begin{cases} \stackrel{d}{\to} \int_0^\infty 1_{\{|\gamma W(t)/t+1|\le\varepsilon\}}\,dt, & \text{if } \varepsilon < 1, \\ \stackrel{P}{\to} \infty, & \text{if } \varepsilon > 1, \end{cases} \tag{2.10}$$

and

$$\log(l_n+1) PERALT(k_n^{-1/2}\varepsilon, n, u_n) \begin{cases} \stackrel{d}{\to} \int_0^\infty 1_{\{|\gamma W(t)/t+1|\le\varepsilon\}}\,\frac{dt}{t}, & \text{if } \varepsilon < 1, \\ \stackrel{P}{\to} \infty, & \text{if } \varepsilon > 1, \end{cases} \tag{2.11}$$

where the limits are finite a.s. if $\varepsilon < 1$.

DISCUSSION. (1) The limiting random variables can be expressed in terms of the local time of a standard Brownian motion defined by

$$\begin{aligned}L_t^a &:= \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_0^t 1_{(a-\varepsilon,\,a+\varepsilon)}(W(s))\,ds \\ &= 2\Big((W(t)-a)^+ - a1_{(-\infty,0)}(a) + \int_0^t 1_{\{W(s)>a\}}\,dW(s)\Big).\end{aligned}$$

According to Revuz and Yor [(1991), Example (VI.1.15)], one has a.s.,

$$\int_0^\infty 1_{\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\varepsilon\}} t^{-\sigma}\,dt = \int_{-\infty}^{\infty}\int_0^\infty 1_{\{|\gamma a/t+t^{-\rho}/(1-\rho)|\le\varepsilon\}} t^{-\sigma}\,dL_t^a\,da.$$

(2) From Theorem 2 we have

$$\text{PERHILL}(k_n^{-1/2}\varepsilon, n, l_n) = O_p\left(\frac{k_n}{l_n}\right)$$

and

$$\text{PERALT}(k_n^{-1/2}\varepsilon, n, u_n) = O_p\left(\frac{1}{\log(l_n+1)}\right)$$

if $\rho < 0$, or $\rho = 0$ and $\varepsilon < 1$. Hence, if n/l_n is of smaller order than every positive power of n (for example, $l_n = n/\log n$), then the rate of convergence to 0 is faster for PERHILL, confirming the claimed superiority of the altHill plot. To see this, recall that $k_n^{1/2}|A(n/k_n)| \to 1$, which is equivalent to $\tilde{A}(n/k_n) \sim n^{1/2}$ where $\tilde{A}(t) := t^{1/2}/|A(t)|$ is a $(1/2-\rho)$-varying function. According to Theorem 1.5.12 of Bingham, Goldie and Teugels (1987), there exists an asymptotically unique inverse $\tilde{A}^{\leftarrow}$, such that $k_n/n \sim 1/\tilde{A}^{\leftarrow}(n^{1/2})$ is a $-1/(1-2\rho)$-varying function of n. Hence k_n/l_n converges to 0 at a faster rate than the slowly varying function $1/\log(l_n+1) \sim 1/\log n$.

This provides a comparison between the two plotting methods when the second-order condition (2.2) holds. However, as mentioned above, this excludes a result for the important case of the Pareto distribution, for which we expect traditional Hill plotting to be superior. For the Pareto distributions, we have the following result.

THEOREM 3. *Suppose F is Pareto, and $n > l_n \ge k_n \to \infty$. Then*

$$\begin{aligned}(2.12)\qquad &\frac{\log(l_n+1)}{\log k_n}\left(1-PERALT(k_n^{-1/2}\varepsilon, n, u_n)\right)\\ &= \frac{\log n}{\log k_n}\int_0^{u_n} 1_{\{k_n^{1/2}|H_{\lceil n^\theta\rceil,n}-\gamma|>\varepsilon\}}\,d\theta \xrightarrow{P} 1,\end{aligned}$$

where again $u_n := \log(l_n+1)/\log n$. If, in addition, $l_n/k_n \to c \in [1,\infty]$, then

$$\begin{aligned}(2.13)\qquad \frac{l_n}{k_n}\left(1-PERHILL(k_n^{-1/2}\varepsilon, n, l_n)\right) &= \frac{l_n}{k_n}\left(\frac{1}{l_n}\sum_{i=1}^{l_n} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|>\varepsilon\}}\right)\\ &\xrightarrow{d} \int_0^c 1_{\{|\gamma W(t)/t|>\varepsilon\}}\,dt,\end{aligned}$$

where the limit is finite a.s.

DISCUSSION. (1) When $c = \infty$, the percentage of time the Hill plot spends outside the neighborhood of γ is

$$1 - \text{PERHILL}(k_n^{-1/2}\varepsilon, n, l_n) = O_P(k_n/l_n) = o_P(1).$$

In contrast, the corresponding percentage for the altHill plot,

$$1 - \text{PERALT}(k_n^{-1/2}\varepsilon, n, u_n) = O_P\left(\frac{\log k_n}{\log l_n}\right),$$

converges to 0 in probability only if k_n is of smaller order than any positive power of n, that is, if one considers a very large neighborhood of γ. Moreover, even in this case $\log k_n / \log l_n$ converges to 0 more slowly than k_n/l_n.

(2) When $1 \le c < \infty$,

$$\frac{1 - \text{PERHILL}(k_n^{-1/2}\varepsilon, n, l_n)}{1 - \text{PERALT}(k_n^{-1/2}\varepsilon, n, u_n)} \xrightarrow{d} \frac{1}{c}\int_0^c 1_{\{|\gamma W(t)/t| > \varepsilon\}}\, dt =^d \int_0^1 1_{\{|\gamma W(t)/t| > c^{1/2}\varepsilon\}}\, dt,$$

where the limiting random variable is almost surely less than 1.

(3) In particular, one may choose $l_n = n - 1$, that is, one considers the whole Hill and altHill plot. If, in addition, one chooses $k_n = n - 1$, so that in analogy to Theorem 2, $k_n^{-1/2}$ equals the optimal rate of convergence, then

$$\text{PERHILL}(k_n^{-1/2}\varepsilon, n, n-1) \xrightarrow{d} \int_0^1 1_{\{|\gamma W(t)/t| \le \varepsilon\}}\, dt,$$

whereas

$$\text{PERALT}(k_n^{-1/2}\varepsilon, n, 1) \xrightarrow{P} 0.$$

Hence, the traditional Hill plot spends a nonvanishing percentage of time in the neighborhood of the true value, while almost the whole altHill plot lies outside the neighborhood, thus confirming the superiority of the Hill plot for the Pareto distribution.

3. A maximal occupation time estimator of the extreme value index. In this section we will formalize the heuristic idea to infer γ from a stable region in the Hill or altHill plot. If one assumes that condition (2.2) holds for some $\rho < 0$, then, according to Theorem 1, the sequence of Hill estimators $H_{i,n}$ can be approximated by the sum of the unknown extreme value index γ, a weighted Brownian motion and a monotone bias function, which converges to 0 as i/n converges to 0. Consequently, for $i = O(k_n)$ (that is, if the bias does not dominate the random error) the standardized error $i^{1/2}|H_{i,n} - \gamma|/\gamma$ is stochastically bounded, so that for a sufficiently large constant $a > 0$ and a consistent initial estimator $\tilde{\gamma}_n$ of γ a large percentage of Hill estimators $H_{i,n}$ satisfies $i^{1/2}|H_{i,n} - \gamma| \le a\tilde{\gamma}_n$. On the other hand, for all $\gamma > 0$ the condition $i^{1/2}|H_{i,n} - \gamma| \le a\tilde{\gamma}_n$ is violated for most $i \gg k_n$ due to the increasing bias term. Hence, it is natural to estimate the extreme value index by the value γ such that the time one of the Hill plots spends in such neighborhoods of γ is maximized. Since, in view of Theorem 2, it is advisable to use the altHill plot, we define

$$\hat{\gamma}_n := \arg\max_{\gamma\in\mathbb{R}} \int_0^{u_n} 1_{\{\lceil n^\theta\rceil^{1/2}|H_{\lceil n^\theta\rceil,\,n}-\gamma|\le a\tilde{\gamma}_n\}}\,d\theta$$

for some $u_n \in (0,1)$. Notice that the normalization of the neighborhood $[\gamma - a\tilde{\gamma}_n\lceil n^\theta\rceil^{-1/2}, \gamma + a\tilde{\gamma}_n\lceil n^\theta\rceil^{-1/2}]$ with the estimated asymptotic standard deviation $\tilde{\gamma}_n\lceil n^\theta\rceil^{-1/2}$ of the Hill estimator $H_{\lceil n^\theta\rceil,\,n}$ automatically ensures that for the optimal order $n^\theta \sim k_n$ the width of the interval is of the order $k_n^{-1/2}$ considered in Section 2.

Employing Theorem 1 and the ideas of the proof of Theorem 2, one can show by lengthy computations that $\hat{\gamma}_n$ is consistent for γ with the optimal rate up to a factor that is of smaller order than any positive power of n:

$$\hat{\gamma}_n - \gamma = o_P(k_n^{-1/2+\delta}) \quad \text{for all } \delta > 0.$$

However, for practical applications the finite sample behavior of $\hat{\gamma}_n$ is more important. To investigate this performance, we have drawn $n = 100, 200, 500$ and 1000 iid random variables according to the Cauchy distribution ($\gamma = 1, \rho = -2$), a Fréchet distribution $F_1(x) = \exp(-x^{-1})(\gamma = 1, \rho = -1)$, t_ν-distributions with $\nu = 4$ and 10 degrees of freedom ($\gamma = 1/\nu$, $\rho = -2/\nu$) and a loggamma distribution with density proportional to $(\log x)^2 x^{-4} 1_{[1,\,\infty)}(x)\,(\gamma = 1/3, \rho = 0)$. In addition, we have simulated random variables according to the Pareto distribution with parameter $\gamma = 1$, which does not meet the second-order condition (2.2). In the definition of $\hat{\gamma}_n$ the parameters are chosen as $u_n = [n^+/2]$, $\tilde{\gamma}_n = H_{[2\sqrt{n^+}],\,n}$ and $a = 1.5$ with n^+ denoting the number of positive observations.

The simulation results are displayed in Table 1. For each sample size and each distribution, the first figure gives the empirical root mean squared error (RMSE) of the estimator $\hat{\gamma}_n$ based on 1000 simulations. In the second lines, this empirical RMSE is divided by the minimum of the empirical RMSE of all Hill estimators based on a deterministic number of order statistics; the pertaining optimal number will be denoted by $k_n^{\text{opt, sim}}$. Hence, these figures measure the loss of efficiency of the new estimator compared with the best possible Hill estimator. Note, however, that the latter cannot be used in applications, since $k_n^{\text{opt, sim}}$ depends on the unknown underlying distribution. In contrast, in the third lines the RMSE of $\hat{\gamma}_n$ is compared with the empirical RMSE of a real estimator, namely of the adaptive Hill estimator based on the data driven choice $\hat{k}_n^{\text{opt}}$ of the number of order statistics that is defined in Drees and Kaufmann [(1998), Section 3]. (For the Pareto distribution these figures are omitted, because the estimator $\hat{k}_n^{\text{opt}}$ does not make sense in that case.)

Obviously, the RMSE of $\hat{\gamma}_n$ decreases as the sample size increases and, as expected, the speed of convergence is lowest for the loggamma distribution, where the optimal rate of convergence $k_n^{-1/2}$ is a slowly varying function of n.

TABLE 1
RMSE($\hat{\gamma}_n$), RMSE($\hat{\gamma}_n$) /RMSE($H_{k_n^{\mathrm{opt\,sim}},n}$) and RMSE($\hat{\gamma}_n$)/RMSE($H_{\hat{k}_n^{\mathrm{opt}},n}$)

d.f.	γ	ρ	n **=100**	n **=200**	**500**	**1000**
Cauchy	1	−2	0.26	0.21	0.17	0.14
			1.03	1.12	1.32	1.43
			0.86	0.88	1.03	1.08
F_1	1	−1	0.22	0.18	0.15	0.13
			1.06	1.09	1.33	1.44
			0.83	0.81	1.02	1.03
t_4	0.25	−0.5	0.22	0.18	0.13	0.10
			1.28	1.26	1.26	1.20
			0.94	0.98	0.99	0.94
t_{10}	0.1	−0.2	0.27	0.22	0.16	0.13
			1.50	1.52	1.44	1.37
			1.01	1.04	1.02	1.00
loggamma	1/3	0	0.12	0.10	0.09	0.08
			0.96	0.98	1.04	1.08
			0.81	0.88	0.93	0.88
Pareto	1	—	0.23	0.17	0.13	0.10
			2.37	2.41	2.88	3.14

For sample size $n = 100$ and 200, the loss of efficiency of $\hat{\gamma}_n$ compared with the best possible Hill estimator is small to moderate if $|\rho|$ is not too small, that is, for the Cauchy, Fréchet and t_4-distribution, and, somewhat surprisingly, also in the case of the loggamma distribution, where $\rho = 0$. For the t_{10}-distribution, where ρ is negative but close to 0 [a case that is known to be the most difficult one, see, for example, Beirlant, Vynckier and Teugels (1999) or Danielsson, de Haan, Peng and de Vries (1998)], the RMSE of $\hat{\gamma}_n$ is about 50% higher than the RMSE of the optimal Hill estimator $H_{k_n^{\mathrm{opt\,sim}},n}$. As the sample size increases, the relative efficiency is improved for the t-distributions, while it deteriorates for the Cauchy and Fréchet distribution.

In most cases the new estimator performs better than the adaptive Hill estimator with estimated optimal number of order statistics for small sample sizes, and the RMSE of both estimators are about the same for sample size $n = 500$ and 1000. In the case of the loggamma distribution, however, the new estimator is clearly superior for all sample sizes.

As one may expect from Theorem 3, the maximal occupation time estimator $\hat{\gamma}_n$ based on the altHill plot shows a poor performance for the Pareto distribution, with its RMSE being about twice to three times as big as the RMSE of the optimal Hill estimator.

Finally, it is worth mentioning that one can improve the performance of the maximal occupation time estimator significantly if one chooses the constant a larger for large $|\rho|$ and smaller if ρ is close to 0. For instance, for $a = 3$ the fraction RMSE($\hat{\gamma}_n$)/RMSE($H_{k_n^{\mathrm{opt\,sim}},n}$) ranges from 0.75 to 1.11 for the Fréchet distribution and from 0.85 to 1.20 for the Cauchy distribution. The corresponding figures for the t_ν-distribution and $a = 1$ are 1.06–1.11 ($\nu = 4$) and 1.15–1.25 ($\nu = 10$). However, such an approach would require a data driven choice of a and thus a more detailed investigation of the asymptotic behavior of $\hat{\gamma}_n$, which would go beyond the scope of the present paper and will be considered elsewhere.

4. Proofs.

PROOF OF THEOREM 1. We take up the approach used by Kaufmann and Reiss (1998). Denote by ξ_n, $n \in \mathbb{N}$, iid standard exponential random variables and define $S_i := \sum_{n=1}^{i} \xi_n$. Recall that $U(S_{n+1}/S_i)$, $1 \le i \le n$, are versions of the order statistics $X_{(i)}$, $1 \le i \le n$ [Reiss (1989), Corollary 1.6.9].

Next note that (2.2) implies

$$\sup_{x \ge 1} x^{-\iota} \left| \log\bigl(U(tx)/U(t)\bigr) - \left(\gamma \log x + A(t)\frac{x^{\rho}-1}{\rho}\right)\right| = o(A(t)) \tag{4.1}$$

for all $\iota > 0$. This is a direct consequence of Lemma 2.1 of Drees (1998a), where in case $\rho = 0$, we use the fact that (2.2) is equivalent to the Π-variation of $\log(t^{-\gamma}U(t))$. Hence applying (4.1) with $t = S_{n+1}/S_{i+1}$ and $x = S_{i+1}/S_j$ yields

$$\begin{aligned}\left(\frac{S_j}{S_{i+1}}\right)^{\iota}&\left|\log\frac{U(S_{n+1}/S_j)}{U(S_{n+1}/S_{i+1})} - \left(\gamma\log\frac{S_{i+1}}{S_j} + A\left(\frac{S_{n+1}}{S_{i+1}}\right)\frac{(S_{i+1}/S_j)^{\rho}-1}{\rho}\right)\right| \\ &= o\left(A\left(\frac{S_{n+1}}{S_{i+1}}\right)\right) \quad \text{a.s.}\end{aligned} \tag{4.2}$$

uniformly for $1 \le j \le i \le l_n$. The strong law of large numbers and the uniform convergence theorem for regularly varying functions yield

$$\frac{A(S_{n+1}/S_{i+1})}{A(n/i)} \longrightarrow 1 \quad \text{a.s.} \tag{4.3}$$

uniformly for $j_n \le i \le l_n$. The law of iterated logarithm gives

$$\max(|S_i/i - 1|, |S_{i+1}/i - 1|) = O((\log\log(3i)/i)^{1/2}),$$

and thus

$$\left|\frac{(S_{i+1}/S_j)^{\rho}-1}{\rho} - \frac{(i/j)^{\rho}-1}{\rho}\right| = O\bigl((i/j)^{\rho}(\log\log(3j)/j)^{1/2}\bigr) \quad \text{a.s.} \tag{4.4}$$

uniformly for $1 \le j \le i < \infty$.

Combining (4.2)–(4.4) and the strong law of large numbers, we arrive at

$$\left(\frac{j}{i}\right)^{\iota}\left|\log\frac{U(S_{n+1}/S_j)}{U(S_{n+1}/S_{i+1})} - \left(\gamma\log\frac{S_{i+1}}{S_j} + A\left(\frac{n}{i}\right)\frac{(i/j)^{\rho}-1}{\rho}\right)\right| = o\left(A\left(\frac{n}{i}\right)\right) \quad \text{a.s.}$$

uniformly for $1 \le j \le i \le l_n$. Consequently,

$$\frac{1}{i}\sum_{j=1}^{i}\log\frac{U(S_{n+1}/S_j)}{U(S_{n+1}/S_{i+1})} = \gamma\frac{1}{i}\sum_{j=1}^{i}\log\frac{S_{i+1}}{S_j} + A\Big(\frac{n}{i}\Big)\frac{1}{i}\sum_{j=1}^{i}\frac{(i/j)^\rho - 1}{\rho}$$
$$+o(A(n/i)) \quad \text{a.s.}$$

uniformly for $j_n \le i \le l_n$. Since $\xi_j^* := j\log(S_{j+1}/S_j)$ defines a sequence of iid exponential random variables [Reiss (1989), Corollary 1.6.11], the famous Komlós–Major–Tusnády approximation of the partial sum process by a Brownian motion combined with the facts that $S_i^* := \sum_{j=1}^{i}\log(S_{i+1}/S_j) = \sum_{j=1}^{i}\xi_j^*$ and $\sum_{j=1}^{i}((i/j)^\rho - 1)/(i\rho) \to 1/(1-\rho)$ yields (2.3) [cf. Kaufmann and Reiss (1998), proof of Theorem 1].

Using $A(S_{n+1}/S_{i+1})/A(n/i) = O(1)$ a.s. uniformly for $1 \le i \le l_n$ instead of (4.3), one obtains the second assertion. □

PROOF OF COROLLARY 1. First note that (2.3) implies

$$\sup_{t_n\le t\le T_n}\bigg(t^{1/2}\wedge\frac{A(n/k_n)}{A(n/\lceil k_n t\rceil)}\bigg)\bigg|H_{\lceil k_n t\rceil,n} - \bigg(\gamma + \gamma\frac{W(\lceil k_n t\rceil)}{\lceil k_n t\rceil} + \frac{A(n/\lceil k_n t\rceil)}{1-\rho}\bigg)\bigg|$$
$$= o(k_n^{-1/2} + A(n/k_n)) \quad \text{a.s.}$$

For all $\iota > 0$, the Potter bounds [Bingham, Goldie and Teugels (1987), Theorem 1.5.6] yield

$$\frac{1}{2}\big(t^{\rho-\iota}\wedge t^{\rho+\iota}\big) \le \frac{A(n/k_n)}{A(n/\lceil k_n t\rceil)} \le 2\big(t^{\rho-\iota}\vee t^{\rho+\iota}\big) \tag{4.5}$$

for sufficiently large n and all $t_n \le t \le T_n$, so that

$$t^{1/2}\wedge\frac{A(n/k_n)}{A(n/\lceil k_n t\rceil)} \ge \frac{1}{2}\big(t^{1/2}\wedge t^{\rho-\iota}\big).$$

Moreover, the uniform convergence theorem gives

$$\sup_{t\ge s} t^{\rho-\iota}|A(n/\lceil k_n t\rceil) - t^{-\rho}A(n/k_n)| = o(A(n/k_n))$$

for all $s > 0$, and hence, by a standard diagonal argument, there exists a sequence $s_n \to 0$ such that

$$\sup_{t\ge s_n} t^{\rho-\iota}|A(n/\lceil k_n t\rceil) - t^{-\rho}A(n/k_n)| = o(A(n/k_n)).$$

On the other hand, in view of (4.5), we have

$$\sup_{t\le s_n} t^{1/2}\big(|A(n/\lceil k_n t\rceil)| + |t^{-\rho}A(n/k_n)|\big) = o(A(n/k_n)).$$

To sum up, we have shown that

$$\sup_{t_n \le t \le T_n} (t^{1/2} \wedge t^{\rho-\iota}) \left| H_{\lceil k_n t \rceil, n} - \left(\gamma + k_n^{-1/2} \gamma \frac{W_n(\lceil k_n t \rceil / k_n)}{\lceil k_n t \rceil / k_n} + A\left(\frac{n}{k_n}\right) \frac{t^{-\rho}}{1-\rho} \right) \right|$$
$$= o(k_n^{-1/2} + A(n/k_n)) \quad \text{a.s.,}$$

where

$$W_n(t) := k_n^{-1/2} W(k_n t) \tag{4.6}$$

is a Brownian motion.

Since $(W_n(t)/t)_{t \ge 1}$ is uniformly continuous and convergence in quadratic mean implies convergence in probability, for the proof of (2.5) it suffices to verify that

$$E\left(\sup_{t_n \le t \le 1} t^{1/2} \left| \frac{W_n(\lceil k_n t \rceil / k_n)}{\lceil k_n t \rceil / k_n} - \frac{W_n(t)}{t} \right| \right)^2 = o(1). \tag{4.7}$$

For this, we use the following series respresentation of a Brownian motion [see, e.g., Breiman (1968), Proposition 12.24]:

$$W_n =^d \left(\frac{t}{\pi^{1/2}} Y_0 + \left(\frac{2}{\pi}\right)^{1/2} \sum_{j=1}^{\infty} \frac{\sin(jt)}{j} Y_j \right)_{t \in [0,1]},$$

where Y_j, $j \ge 0$, are independent standard normal random variables. Hence the left-hand side of (4.7) equals.

$$\sup_{t_n \le t \le 1} t \frac{2}{\pi} \sum_{j=1}^{\infty} \left(\frac{\sin(j \lceil k_n t \rceil / k_n)}{j \lceil k_n t \rceil / k_n} - \frac{\sin(jt)}{jt} \right)^2. \tag{4.8}$$

Because $x d/dx(\sin x / x) = \cos x - \sin x / x$ is bounded and $|j \lceil k_n t \rceil / k_n - jt| \le j/k_n$, by the mean value theorem (4.8) is of the order

$$O\left(\sup_{t_n \le t \le 1} t \left(\sum_{j=1}^{k_n} \left(\frac{j}{k_n} \frac{1}{jt} \right)^2 + \sum_{j=k_n+1}^{\infty} \left(\frac{1}{jt} \right)^2 \right) \right) = O\left(\frac{1}{k_n t_n} \right) = o(1).$$

To prove the second assertion, choose t_n such that $k_n t_n \to \infty$ but $\sup_{0 \le t \le t_n} h(t) k_n^{1/2} \to 0$, which is possible because of $\lim_{t \to 0} t^{-1/2} h(t) = 0$. Then, the definition of h and (2.4) ensure that

$$\sup_{0 \le t \le t_n} h(t) \left| H_{\lceil k_n t \rceil, n} - \left(\gamma + \gamma k_n^{-1/2} \frac{W_n(t)}{t} + A(n/k_n) \frac{t^{-\rho}}{1-\rho} \right) \right|$$
$$= o(k_n^{-1/2} + A(n/k_n)) \quad \text{a.s.} \qquad \square$$

PROOF OF THEOREM 2. First we prove that the limit random variables are finite a.s. For (2.8) and (2.10) this is an immediate consequence of $\lim_{t\to\infty} W(t)/t = 0$ a.s., whereas for the limit random variables in (2.9) and (2.11), in addition, one has to take into account that, for all $a > 0$,

$$
\begin{aligned}
&E\int_0^a 1_{\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\varepsilon\}}\frac{dt}{t}\\
&\quad=\int_0^a \Phi\biggl(\frac{t^{1/2-\rho}}{\gamma(1-\rho)}+\frac{\varepsilon t^{1/2}}{\gamma}\biggr)-\Phi\biggl(\frac{t^{1/2-\rho}}{\gamma(1-\rho)}-\frac{\varepsilon t^{1/2}}{\gamma}\biggr)\frac{dt}{t}\\
&\quad\le\int_0^a (2\pi\gamma^2)^{-1/2}2\varepsilon t^{-1/2}\,dt<\infty.
\end{aligned}
\tag{4.9}
$$

Next we will show that, due to the large bias, for i being large compared with k_n one has $k_n^{1/2}|H_{i,n}-\gamma| > \varepsilon$ with large probability if $\rho < 0$, or $\rho = 0$ and $\varepsilon < 1$. Pick some (small) $\delta > 0$. For $\rho < 0$, choose M sufficiently large such that $P\{\sup_{t\ge M}|W(t)|/t \ge 1\} \le \delta/2$ and $M^{-\rho/2}/(2(1-\rho)) \ge \varepsilon+\gamma+1$. Then, for $Mk_n \le i \le l_n$, (2.3), (4.6), (2.7) and the Potter bounds (4.5) imply

$$
\begin{aligned}
k_n^{1/2}|H_{i,n}-\gamma| &= \left|\gamma\frac{W_n(i/k_n)}{i/k_n}+\frac{A(n/i)(1+o(1))}{|A(n/k_n)|(1-\rho)}+O\Bigl(k_n^{1/2}\frac{\log i}{i}\Bigr)\right|\\
&\ge \frac{M^{-\rho/2}}{2(1-p)}-\gamma-\frac{1}{2}>\varepsilon
\end{aligned}
\tag{4.10}
$$

with probability greater than $1-\delta$ for sufficiently large n. Likewise, in case of $\rho = 0$ the monotonicity of $|A|$ yields

$$
k_n^{1/2}|H_{i,n}-\gamma| \ge 1-\gamma\frac{|W_n(i/k_n)|}{i/k_n}+o(1)>\varepsilon
\tag{4.11}
$$

with probability greater than $1-\delta$ if M is chosen such that $P\{\sup_{t\ge M}|W(t)|/t \ge (1-\varepsilon)/(2\gamma)\} \le \delta/2$.

Moreover, the finiteness of the limit random variables in (2.8) and (2.10) shows that

$$
\lim_{M\to\infty}\int_M^\infty 1_{\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\varepsilon\}}\,dt = 0 \qquad \text{a.s.}
$$

Hence, for the convergence of the normalized PERHILL statistic for $\rho < 0$, or $\rho = 0$ and $\varepsilon < 1$, it suffices to prove that for all $M < \infty$,

$$
\begin{aligned}
\frac{1}{k_n}\sum_{i=1}^{\lceil Mk_n\rceil} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} &= \int_0^M 1_{\{k_n^{1/2}|H_{\lceil k_n t\rceil,n}-\gamma|\le\varepsilon\}}\,dt+O(k_n^{-1})\\
&\xrightarrow{d}\int_0^M 1_{\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\varepsilon\}}\,dt.
\end{aligned}
$$

This, however, follows easily from (2.6), which implies that

$$
\sup_{m\le t\le M}|k_n^{1/2}(H_{\lceil k_n t\rceil,n}-\gamma)-(\gamma W_n(t)/t\pm t^{-\rho}/(1-\rho))|\xrightarrow{P}0
$$

for all $0 < m < M < \infty$, by a continuous mapping argument. For one has, for all $\delta > 0$, with probability tending to 1,

$$\int_{\delta/2}^{M} 1_{\{|\gamma W_n(t)/t + t^{-\rho}/(1-\rho)| \le \varepsilon - \delta\}}\, dt - \frac{\delta}{2} \le \int_0^M 1_{\{k_n^{1/2}|H_{\lceil k_n t\rceil, n} - \gamma| \le \varepsilon\}}\, dt$$
$$\le \int_{\delta/2}^{M} 1_{\{|\gamma W_n(t)/t + t^{-\rho}/(1-\rho)| \le \varepsilon + \delta\}}\, dt + \frac{\delta}{2},$$

where the left- and the right-hand side converge to

$$I(\varepsilon) := \int_0^M 1_{\{|\gamma W_n(t)/t + t^{-\rho}/(1-\rho)| \le \varepsilon\}}\, dt$$

as $\delta \to 0$, since the map $\varepsilon \mapsto I(\varepsilon)$ is continuous.

Next, we turn to the limit behavior of the PERHILL statistic if $\rho = 0$ and $\varepsilon > 1$. Choose M such that $P\{\sup_{t \ge M} |W(t)|/t > (\varepsilon - 1)/(2\gamma)\} < \delta/2$, and note that for all $K > 1$ the uniform convergence theorem gives $\sup_{Mk_n \le i \le MKk_n} |A(n/i)/A(n/k_n) - 1| \to 0$. Hence one has with probability greater than $1 - \delta$

$$k_n^{1/2}|H_{i,n} - \gamma| \le 1 + o(1) + (\varepsilon - 1)/2 < \varepsilon \tag{4.12}$$

for $Mk_n \le i \le MKk_n$ and sufficiently large n, so that

$$\frac{l_n}{k_n}\mathrm{PERHILL}(k_n^{-1/2}\varepsilon, n, l_n) \ge \frac{\lceil MKk_n\rceil - \lfloor Mk_n \rfloor}{k_n} \longrightarrow M(K-1).$$

Since $K > 1$ and $\delta > 0$ are arbitrary, it follows that the left-hand side converges to ∞ in probability.

Now we examine the asymptotics of

$$\log(l_n + 1)\mathrm{PERALT}(k_n^{-1/2}\varepsilon, n, u_n) = \sum_{i=1}^{l_n} \log\frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n} - \gamma| \le \varepsilon\}}. \tag{4.13}$$

In case of $\rho = 0$ and $\varepsilon > 1$ we obtain from (4.12) that with probability greater than $1 - \delta$,

$$\log(l_n + 1)\mathrm{PERALT}(k_n^{-1/2}\varepsilon, n, u_n) \ge \log\frac{\lceil MKk_n\rceil}{\lfloor Mk_n\rfloor} \longrightarrow \log K$$

for all $K > 1$ and hence (2.11).

If $\rho < 0$, or $\rho = 0$ and $\varepsilon < 1$, then again (4.10) and (4.11), respectively, in combination with $\lim_{M\to\infty}\int_M^\infty 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)| \le \varepsilon\}}\, t^{-1}\, dt = 0$ a.s. show that it suffices to prove that

$$\sum_{i=1}^{\lceil Mk_n\rceil} \log\frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n} - \gamma| \le \varepsilon\}} \xrightarrow{d} \int_0^M 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)| \le \varepsilon\}} \frac{dt}{t} \tag{4.14}$$

for all $M > 0$.

Because in the expression for the PERALT statistic the weights $1/k_n$ occurring in the definition of PERHILL are replaced with the weights $\log(1+1/i)$, which are of larger order for $i = o(k_n)$, an additional argument is needed for the integral over $(0, m)$, $m \downarrow 0$. In view of (2.4), for fixed i, $H_{i,n}$ is asymptotically gamma distributed with shape and scale parameter i. Since this distribution is continuous, it follows that $P\{k_n^{1/2}|H_{i,n} - \gamma| \le \varepsilon\} = P\{H_{i,n} \in [\gamma - \varepsilon k_n^{-1/2}, \gamma + \varepsilon k_n^{-1/2}]\} \to 0$, so that $\sum_{i=1}^{i_0} \log((i+1)/i) 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} \xrightarrow{P} 0$ for all fixed i_0. Thus, a standard diagonal argument proves that there exists an intermediate sequence $(j_n)_{n\in\mathbb{N}}$ such that

$$\sum_{i=1}^{j_n} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} \xrightarrow{P} 0. \tag{4.15}$$

Next note that by (2.3), (2.7) and (4.5) for all $\delta > 0$ there exists $C > 0$ such that with probability greater than $1 - \delta$,

$$\begin{aligned}\gamma \frac{|W(i)|}{i} &\le |H_{i,n} - \gamma| + k_n^{-1/2} \left| \frac{A(n/i)(1+o(1))}{A(n/k_n)(1-\rho)} \right| + C \frac{\log i}{i} \\ &\le |H_{i,n} - \gamma| + 2k_n^{-1/2}(i/k_n)^{-\rho-\iota} + C \frac{\log i}{i}\end{aligned}$$

for all $j_n \le i \le mk_n + 1$ and sufficiently large n. Use the inequalities $\log(1 + x) \le x$ for $x > 0$ and $\Phi'(x) \le (2\pi)^{-1/2}$ for all $x \in \mathbb{R}$ to obtain

$$\begin{aligned}&\sum_{i=j_n+1}^{\lfloor mk_n \rfloor} \log \frac{i+1}{i} P\left\{ \gamma \frac{|W(i)|}{i} \le \varepsilon k_n^{-1/2} + 2k_n^{-1/2}(i/k_n)^{-\rho-\iota} + C \frac{\log i}{i} \right\} \\ &\quad\le 2(2\pi\gamma^2)^{-1/2} \sum_{i=j_n+1}^{\lfloor mk_n \rfloor} i^{-1}\big(\varepsilon(i/k_n)^{1/2} + 2(i/k_n)^{1/2-\rho-\iota} + Ci^{-1/2}\log i\big) \\ &\quad\le \text{const.}\,(m^{1/2} + m^{1/2-\rho-\iota}) + o(1) \longrightarrow 0\end{aligned}$$

as $m \downarrow 0$. Hence, it follows that for all $\delta > 0$ one can find $m > 0$ such that with probability greater than $1 - \delta$ one has

$$\sum_{i=j_n+1}^{\lfloor mk_n \rfloor} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} \le \delta \tag{4.16}$$

for sufficiently large n.

In view of (4.14)–(4.16) and (4.9), it remains to prove that for all $0 < m < M < \infty$ one has

$$\begin{aligned}\sum_{i=\lfloor mk_n \rfloor}^{\lceil Mk_n \rceil} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n}-\gamma|\le\varepsilon\}} &= \int_m^M 1_{\{k_n^{1/2}|H_{\lceil k_n t\rceil,n}-\gamma|\le\varepsilon\}} \frac{dt}{t} + O(k_n^{-1}) \\ &\xrightarrow{d} \int_m^M 1_{\{|\gamma W_n(t)/t + t^{-\rho}/(1-\rho)|\le\varepsilon\}} \frac{dt}{t},\end{aligned}$$

but this follows by the continuous mapping argument mentioned above. □

PROOF OF THEOREM 3. Following the lines of the proof of Theorem 1, one can show that for suitable versions of $H_{i,n}$,

$$H_{i,n} = \gamma + \gamma \frac{W(i)}{i} + O\left(\frac{\log(i+1)}{i}\right) \quad \text{a.s.} \tag{4.17}$$

uniformly for $1 \le i \le n-1$.

Since $\sup_{t \ge M} |W(t)/t| \to 0$ a.s. as $M \to \infty$, for each $\delta > 0$ one can pick a large M such that one has eventually with probability greater than $1-\delta$,

$$k_n^{1/2} |H_{i,n} - \gamma| \le \gamma \frac{|W_n(i/k_n)|}{i/k_n} + \varepsilon/2 \le \varepsilon \tag{4.18}$$

for all $Mk_n \le i \le l_n$ with W_n defined in (4.6). Hence, by similar arguments as in the proof of Theorem 2,

$$\begin{aligned}
&\frac{\log(l_n+1)}{\log k_n}\bigl(1 - \mathrm{PERALT}(k_n^{-1/2}\varepsilon, n, u_n)\bigr) \\
&\quad = \frac{1}{\log k_n}\Biggl(\sum_{i=1}^{\lceil Mk_n\rceil \wedge l_n} \log\frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n}-\gamma| > \varepsilon\}} \\
&\qquad\qquad + \sum_{i=(\lceil Mk_n\rceil \wedge l_n)+1}^{l_n} \log\frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n}-\gamma| > \varepsilon\}}\Biggr) \\
&\quad = \frac{\log(\lceil Mk_n\rceil + 1)}{\log k_n} - \frac{1}{\log k_n}\left(\int_0^{M\wedge(l_n/k_n)} 1_{\{\gamma|W(t)|/t \le \varepsilon\}} \frac{dt}{t} + o(1)\right),
\end{aligned}$$

from which assertion (2.12) is obvious.

Because of (4.18), for the examination of PERHILL, one may restrict oneself to $1 \le i \le Mk_n \wedge l_n$. Similarly, as in the proof of Corollary 1, one may deduce from (4.17) that

$$\sup_{0 < t \le M\wedge(l_n/k_n)} h(t) \left| k_n^{1/2}\bigl(H_{\lceil k_n t\rceil, n} - \gamma\bigr) - \gamma \frac{W_n(t)}{t} \right| = o(1).$$

Thus we obtain assertion (2.13) using the continuous mapping argument of the proof of Theorem 2 and the a.s. finiteness of $\int_0^c 1_{\{|\gamma W(t)/t| > \varepsilon\}}\, dt$, which is immediate from $\lim_{t\to\infty} W(t)/t = 0$ a.s. □

Acknowledgments. Sidney Resnick gratefully acknowledges the hospitality of the Department of Statistics, University of North Carolina and particularly Professor Ross Leadbetter. Sidney Resnick and Laurens de Haan also gratefully acknowledge the hospitality of the University of Lisbon. Parts of this paper were written while Holger Drees visited the University of Rotterdam. The simulation study was started while Holger Drees and Laurens de Haan visited Chalmers University, Gothenburg. The hospitality of the Department of Mathematical Statistics and Holger Rootzén are gratefully acknowledged. The final draft was assembled in July 1998 while Sidney Resnick was a guest of the Tinbergen Institute, Erasmus University Rotterdam.

REFERENCES

BEIRLANT, J., DIERCKX, G., GOEGEBEUR, Y. and MATTHYS, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2** 177–200.

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.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C. G. (1998). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal*. To appear.

DAVIS, R. and RESNICK, S. (1996). Limit theory for bilinear processes with heavy tailed noise. *Ann. Appl. Probab.* **6** 1191–1210.

DE HAAN, L., PENG, L. and PEREIRA, T. T. (1997). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*. To appear.

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** 173–195.

DREES, H. (1998a). On smooth statistical tail functionals. *Scand. J. Statist.* **25** 187–210.

DREES, H. (1998b). Estimating the extreme value index. Habilitation thesis, Univ. Cologne.

DREES, H. and KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.* **75** 149–172.

JANSEN, D. and DE VRIES, C. (1991). On the frequency of large stock returns: putting booms and busts into perspective. *Rev. Econom. Statist.* **73** 18–24.

KAUFMANN, E. and REISS, R.-D. (1998). Strong approximation of the Hill estimator process. *Statist. Probab. Lett.* **39** 347–354.

LELAND, W., TAQQU, M., WILLINGER, W. and WILSON, D. (1994). On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Networking* **2** 1–15.

MASON, D. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10** 754–764.

MASON, D. and TUROVA, T. S. (1994). Weak convergence of the Hill estimator process. In *Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simin, eds.) 419–431. Kluwer, Dordrecht.

MCNEIL, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull.* **27** 125–139.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer, New York.

RESNICK, S. (1997a). Discussion of the Danish data on large fire insurance losses. *Astin Bull.* **27** 139–151.

RESNICK, S. (1997b). Heavy tail modeling and teletraffic data. *Ann. Statist.* **25** 1805–1869.

RESNICK, S. and STĂRICĂ, C. (1995). Consistency of Hill's estimator for dependent data. *J. Appl. Probab.* **32** 139–167.

RESNICK, S. and STĂRICĂ, C. (1997). Smoothing the Hill estimator. *Adv. in Appl. Probab.* **29** 271–293.

RESNICK, S. and STĂRICĂ, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Probab.* **8** 1156–1183.

RESNICK, S. and VAN DEN BERG, E. (2000). Sample correlation behavior for the heavy tailed general bilinear process. *Stochastic Model* **16**. To appear.

REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion.* Springer, Berlin.

ROOTZEN, H., LEADBETTER, M. and DE HAAN, L. (1990). Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Dep. Statistics, Univ. North Carolina.

H. DREES
MATHEMATICAL INSTITUTE
UNIVERSITY OF COLOGNE
WEYRTAL 86 90
50931 COLOGNE
GERMANY
E-MAIL: hdrees@mi.uni-koeln.de

L. DE HAAN
ERASMUS UNIVERSITY
ECONOMETRIC INSTITUTE
P.O. BOX 1738
3000 DR ROTTERDAM
THE NETHERLANDS
E-MAIL: ldhaan@few.eur.nl

S. RESNICK
SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: sid@orie.cornell.edu