HOW TO MAKE A HILL PLOT

HOLGER DREES, LAURENS DE HAAN, AND SIDNEY RESNICK

Abstract. 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 becoming increasingly common to encounter large, high quality data sets for which appropriate models require heavy tailed distributions. Examples abound from the fields of insurance (McNeil, 1997; Resnick, 1997), finance, economics (Jansen and de Vries, 1991) computer science and telecommunications (Leland et al, 1994).

By a heavy tailed distribution we mean a distribution F, assumed for convenience to concentrate on $[0, \infty)$, which satisfies

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

where L is a slowly varying function satisfying

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1, \quad \forall x > 0.$$

Suppose $(X_n)_{n \in \mathbb{N}}$ is a stationary sequence whose marginal, one-dimensional distribution is F, so that

$$P\{X_n > x\} = 1 - F(x).$$

A basic statistical calibration problem is to estimate the shape parameter α which is the negative of the index of regular variation. A popular estimator of $\gamma := \alpha^{-1}$ is the Hill estimator obtained as follows. Suppose one observes $X_1, \ldots, X_n$ and orders these observations as

$$X_{(1)} \geq \cdots \geq X_{(n)}.$$

AMS 1991 *subject classifications.* Primary 60G70; secondary 60F15.

Key words and phrases. heavy tails, regular variation, estimation, Pareto tails, Hill estimator, order statistics .

Sidney Resnick's research was partially supported by NSF grant DMS-97-04982 at Cornell University and by Office of Naval Research Contract N00014-93-1-0043 at University of North Carolina at Chapel Hill. Grateful acknowledgement is made for hospitality to the Department of Statistics, University of North Carolina and particularly to Professor Ross Leadbetter. Laurens de Haan was partially supported by the European Union TMR grant ERBFMBICT961465. 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, supported by the Netherlands Organization for Scientific Research through the Netherlands Mathematical Research Foundation. The final draft was assembled in July 1998 while Sidney Resnick was a guest of the Tinbergen Institute, Erasmus University Rotterdam.

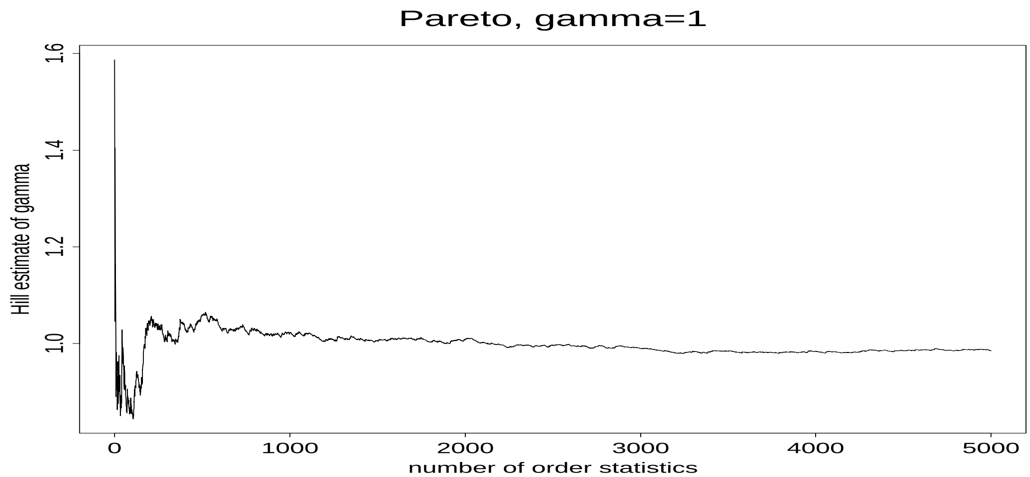

Figure 1. Hill plot of 5000 Pareto observations, γ=1.

The Hill estimator based on $k+1$ upper order statistics is

$$H_{k,n} := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} \tag{1.2}$$

for $k = 1, \dots, n-1$.

The Hill estimator is consistent for γ in the following sense. If $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence, that is,

$$k_n \longrightarrow \infty, \quad k_n/n \longrightarrow 0, \tag{1.3}$$

then

$$H_{k_n,n} \xrightarrow{P} \gamma, \tag{1.4}$$

provided either

(i) $\{X_n\}$ is iid (Mason, 1982) or
(ii) $\{X_n\}$ can be written as a finite or infinite order moving average process (Resnick and Stărică, 1995) or
(iii) $\{X_n\}$ satisfies mixing conditions (Rootzen, Leadbetter, de Haan (1990)) or
(iv) $\{X_n\}$ is an ARCH(1) process (Resnick and Stărică, 1998), a bilinear process (Davis and Resnick (1997), Resnick and Van den Berg (1998)) or consists of random variables defined on a Markov chain (Resnick and Stărică, 1998).

Because of condition (1.3) on the number of order statistics, it is not clear how to apply this consistency result (1.4). One can try to choose an optimal k which minimizes asymptotic mean square error. See Danielsson et al. (1997) and Drees and Kaufmann (1998). However, what is usually done in practice is to construct a Hill plot, defined as

$$\{(k, H_{k,n}^{-1}), 1 \le k \le n-1\}$$

and then to infer the value of γ 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

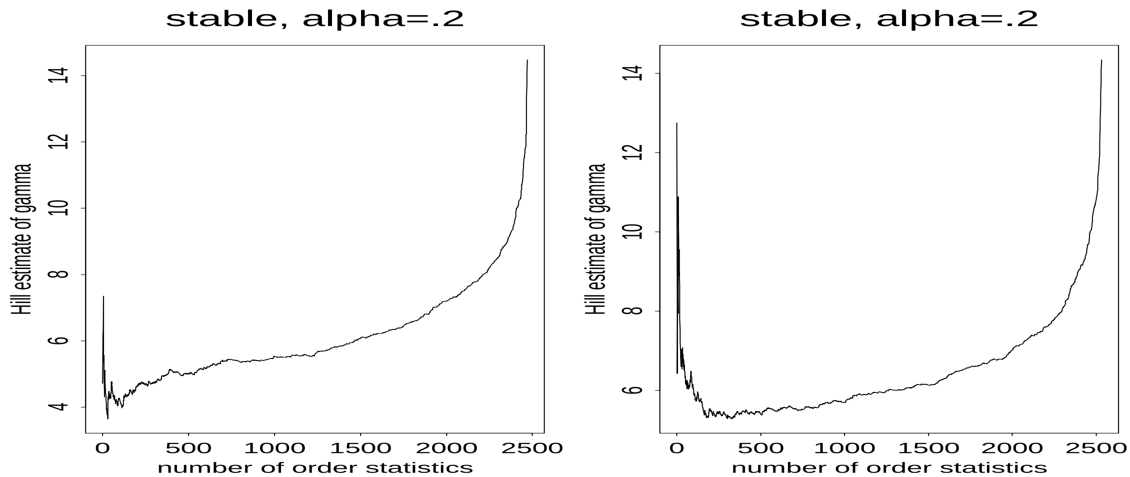

FIGURE 2. Hill plot of two independent samples from stable, $\alpha = 0.2$.

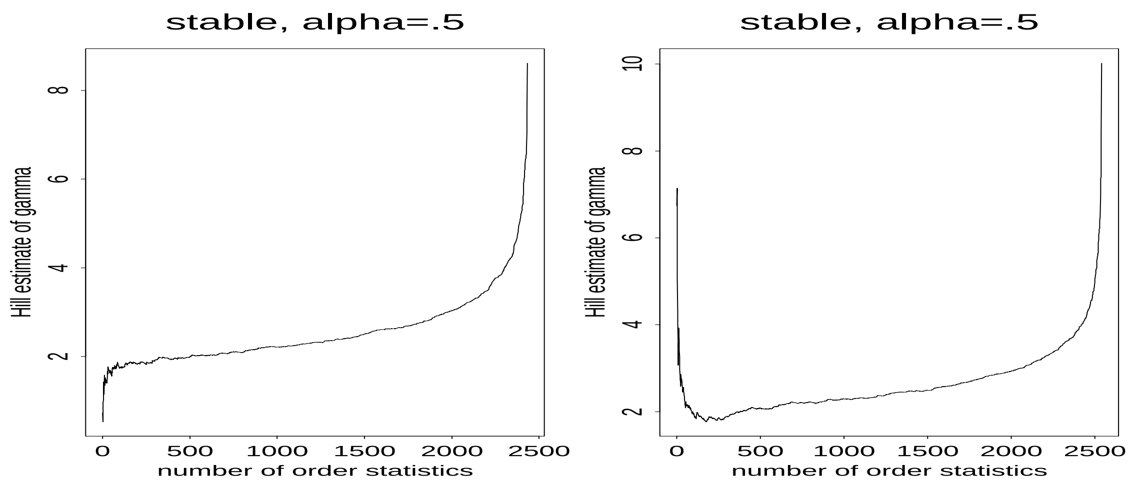

FIGURE 3. Hill plot of two independent samples from stable, $\alpha = 0.5$.

effective only when the underlying distribution is Pareto or very close to Pareto. For the Pareto distribution,

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

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 born 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.

Figure 1 is a Hill plot for 5000 iid observations from the Pareto distribution with $\alpha = 1$. Notice the right side of the graph clearly indicates the correct value of 1. However, Figures 2 and 3 each display two independent samples of size 5000 from stable distributions with parameters 0.2 and

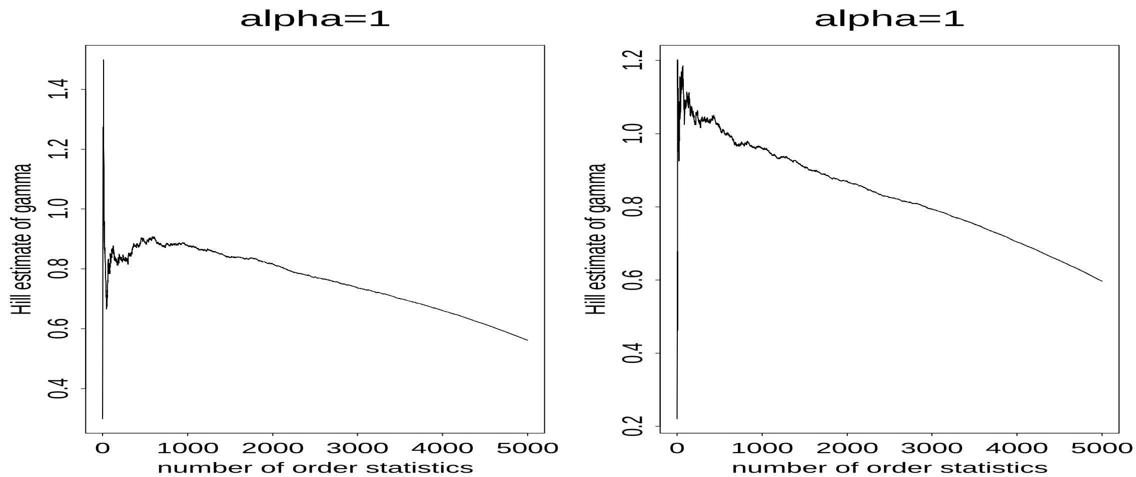

FIGURE 4. Hill plot of two independent samples from logarimithmically perturbed Pareto, $\alpha = 1$.

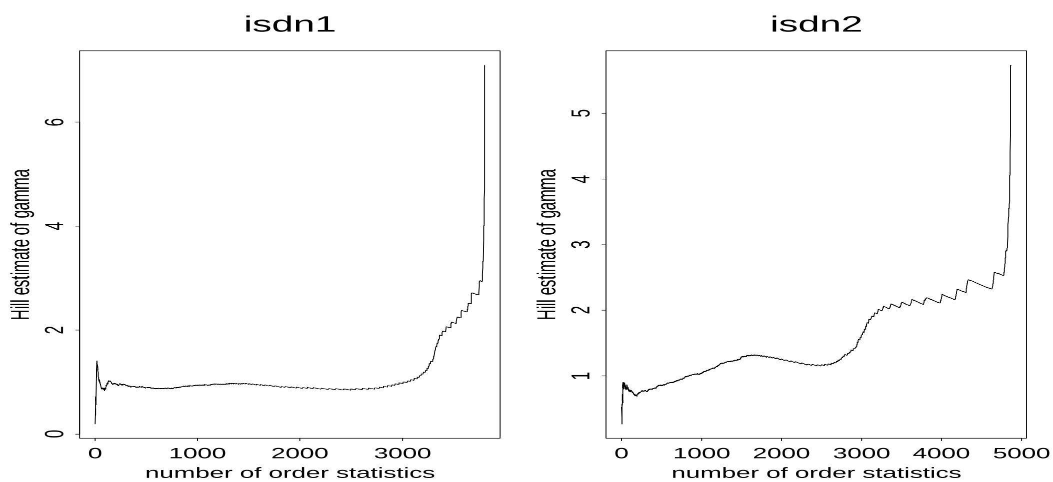

FIGURE 5. Hill plot of interarrival times of packets to a sever.

0.5 and due to the small amount of time the plots spend in the vicinity of the correct values, one would have to be paranormal to discern with confidence the true values. Figure 4 gives two Hill plots for samples of size 5000 from the distribution of the random variable $U^{-1} \log U^{-1}$ where U is uniform on $[0, 1]$. For this distribution F satisfies

$$1 - F(x) \sim x^{-1}/\log x, \quad x \to \infty.$$

Figure 5 is a Hill plot for two real teletraffic data sets of length 3802 and 4867 respectively, consisting of interarrival times of packets in a network. Finally Figure 6 is the Hill plot of the Danish large fire insurance claim data of length 2156. This example shows that sometimes the Hill plot can be quite clear and informative for real data.

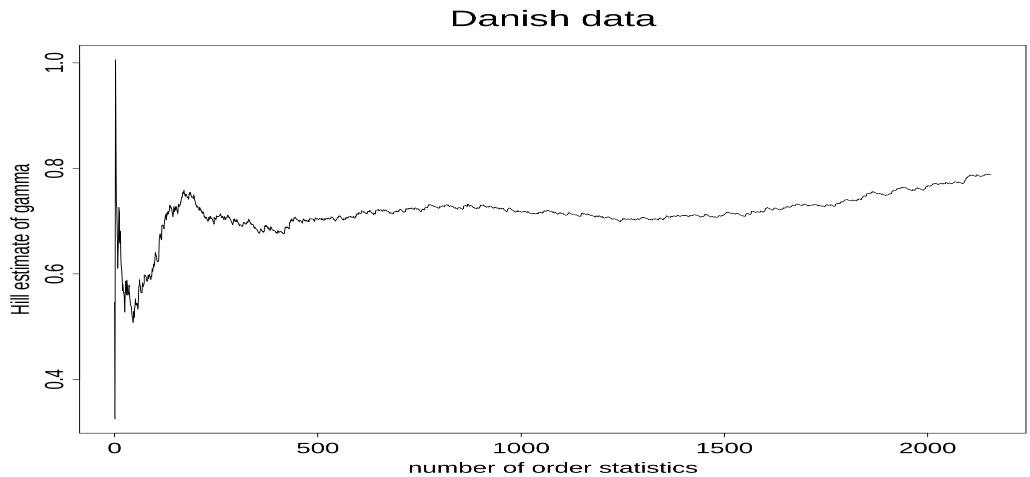

FIGURE 6. Hill plot of Danish loss data.

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 $\{(k, H_{k,n}^{-1}), 1 \leq k \leq n-1\}$, we construct the *altHill* plot by plotting

$$\{(\theta, H_{\lceil n^\theta \rceil,n}^{-1}), 0 \leq \theta < 1\},$$

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 7 redisplays the traditional Hill plot for sample of size 5000 from the stable($\alpha = 0.2$) distribution alongside the alt plot which is more revealing since the plot spends more time in the neighborhood of the true value. The information in the alt plot would be further enhanced by applying a smoothing procedure given in Resnick and Starica (1997).

Corresponding to Figure 4 we observe in Figure 8 that the alt plot more clearly shows the correct value of $\alpha = 1$. Finally Figure 9 compares the traditional Hill with the alt plots for the second ISDN data set; the alt plot makes plausible an estimate of $\alpha = 1.1$.

The engineering conclusion we emphasize in this paper is that for iid observations whose common distribution has a tail satisfying a second order condition, altplotting is superior. See Theorem 2 and the accompanying discussion. For the Pareto distribution, the traditional Hill plot is preferred. We quantify superiority 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

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

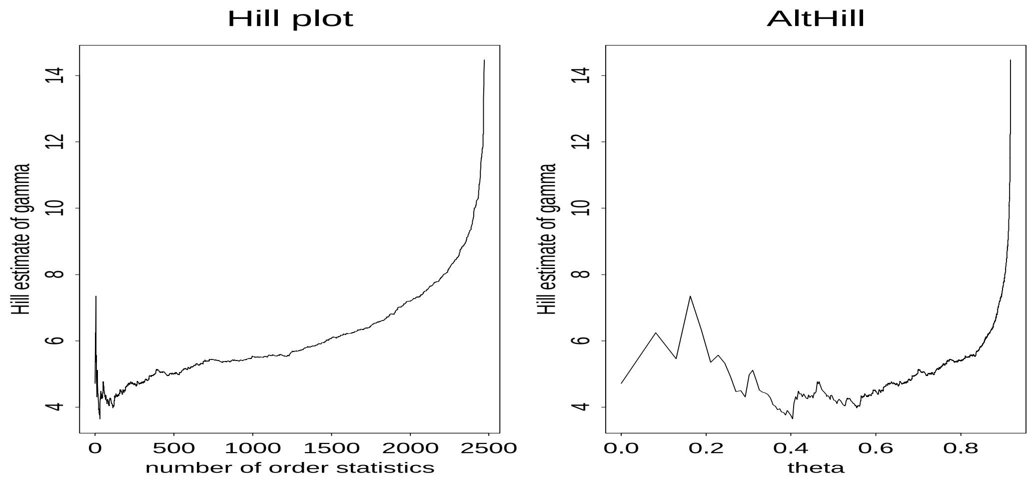

FIGURE 7. Hill and altHill plot of stable observations, $\alpha = 0.2$.

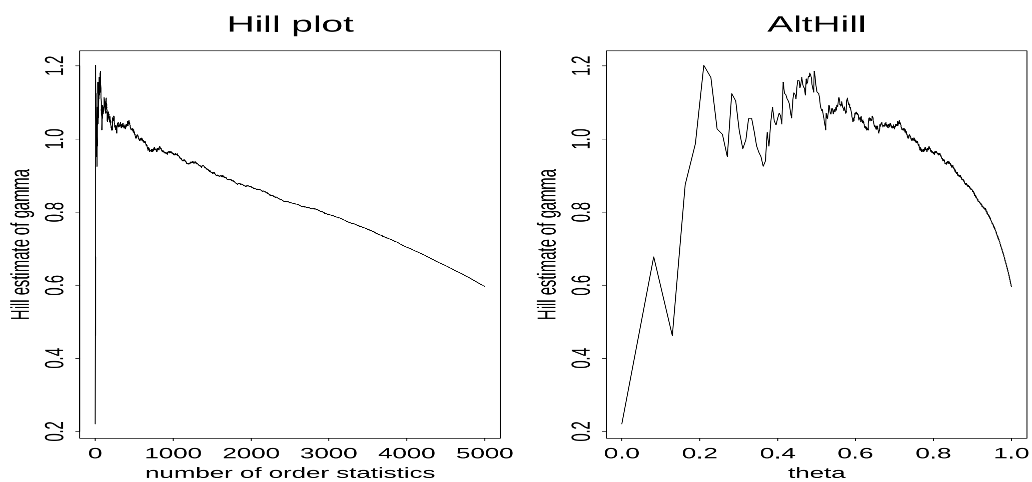

FIGURE 8. Hill and altHill plot corresponding to Figure 4, $\alpha = 1$.

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

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

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

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.

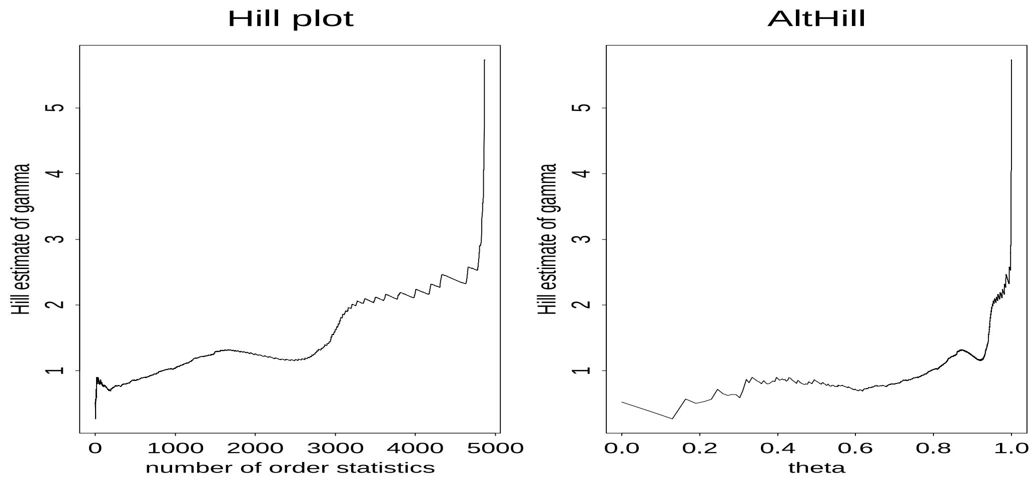

FIGURE 9. Hill and altHill plot of ISDN2.

To control that part of the Hill plot corresponding to order statistics not determined by the right tail, one needs a left tail assumption. We are loath to assume anything about the left tail 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 $\rho \leq 0$ and some function $A : (0, \infty) \to \mathbb{R}$ which ultimately is of constant sign. Then, necessarily, $|A|$ is regularly varying with index ρ. 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) and de Haan et al. (1997).

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 $H_{i,n}$, $1 \leq i \leq n-1$, $n \in \mathbb{N}$, and a standard Brownian motion W such that for all intermediate sequences $(j_n)_{n\in\mathbb{N}}$ and $(l_n)_{n\in\mathbb{N}}$*

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

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*

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

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 the conditions of Theorem 1, there exists a sequence of Brownian motions* W_n, *such that*

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

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,*

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

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

$$\lim_{t\to 0} h(t) (\log\log(3/t)/t)^{1/2} = 0.)$$

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.

From Corollary 1 it is easily seen that the optimal rate of convergence which minimizes the asymptotic mean squared error is obtained if the standard deviation and the bias are balanced, i.e., if

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

Therefore, it is natural to examine the asymptotic behavior of PERHILL and PERALT for a neighborhood shrinking with the rate $k_n^{-1/2}$ towards the true value γ. Observe that, according to Theorem 1.5.12 of Bingham et al. (1987), relation (2.7) is satisfied by an intermediate sequence, which is unique up to asymptotic equivalence (see discussion item (2) after Theorem 2).

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

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

and

$$\log(l_n+1)PERALT(k_n^{-1/2}\epsilon,n,u_n) = \log n\int_0^{u_n} 1_{\{k_n^{1/2}|H_{\lceil n^\theta\rceil,n}-\gamma|\le\epsilon\}}\,d\theta$$

$$\stackrel{d}{\to}\int_0^\infty 1_{\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\epsilon\}}\frac{dt}{t} \tag{2.9}$$

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

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

and

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

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

DISCUSSION:

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

$$L_t^a := \lim_{\epsilon\downarrow 0}\frac{1}{2\epsilon}\int_0^t 1_{(a-\epsilon,a+\epsilon)}(W(s))\,ds = 2\big((W(t)-a)^+ - a1_{(-\infty,0)}(a) + \int_0^t 1_{\{W(s)>a\}}\,dW(s)\big).$$

According to Revuz and Yor (1991), Ex. (VI.1.15), one has

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

(2) From Theorem 2, we have

$$\mathrm{PERHILL}(k_n^{-1/2}\epsilon,n,l_n) = O_p\Big(\frac{k_n}{l_n}\Big)$$

and

$$\mathrm{PERALT}(k_n^{-1/2}\epsilon,n,u_n) = O_p\Big(\frac{1}{\log(l_n+1)}\Big)$$

if $\rho<0$, or $\rho=0$ and $\epsilon<1$. Hence, if n/l_n is of smaller order than every positive power of n (e.g., $l_n=n/\log n$), 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 et al. (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, this excludes a result for the important case of the Pareto distribution, for which we expect traditional Hill plotting is superior. For the Pareto distributions, we have the following result.

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

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

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

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

where the limit is finite a.s.

Discussion:

(1) Convergence (2.12) and (2.13) discuss the percentage of time the altHill and Hill plots spend outside a neighborhood of the true value γ.

(2) When $c = \infty$,

$$\left(1 - \mathrm{PERHILL}(k_n^{-1/2}\epsilon, n, l_n)\right) = O_p(k_n/l_n), \tag{2.14}$$

where $k_n/l_n \to 0$ and

$$\left(1 - \mathrm{PERALT}(k_n^{-1/2}\epsilon, n, u_n)\right) = O_p\Big(\frac{\log k_n}{\log l_n}\Big) \tag{2.15}$$

where $(\log k_n)/(\log l_n) \to 0$ more slowly than $k_n/l_n \to 0$

(3) When $1 \leq c < \infty$,

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

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

If F is a Pareto distribution, then the assertions of Theorem 3 hold true for all intermediate sequences $(k_n)_{n\in\mathbb{N}}$ and all sequences $(l_n)_{n\in\mathbb{N}}$ satisfying $l_n \to \infty$ for (2.12) and $l_n/k_n \to \infty$ for (2.13). In particular, one may choose $l_n = n-1$. In this case, the percentage of time the alt plot is outside a neighborhood of γ is

$$1 - \mathrm{PERALT}(k_n^{-1/2}\epsilon, n, 1) = O_p(\log k_n/\log n),$$

and the corresponding percentage of time for the Hill plot is

$$1 - \mathrm{PERHILL}(k_n^{-1/2}\epsilon, n, n-1) = O_p(k_n/n),$$

confirming the superiority of the Hill plot for the Pareto distribution.

3. 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

$$\sup_{x \ge 1} x^{-\iota} \Big| \log(U(tx)/U(t)) - \big(\gamma \log x + A(t) \frac{x^\rho - 1}{\rho}\big) \Big| = o(A(t)) \tag{3.1}$$

for all $\iota > 0$ This is a direct consequence of Lemma 2.1 of Drees (1998), 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 (3.1) with $t = S_{n+1}/S_{i+1}$ and $x = S_{i+1}/S_j$ yields

(3.2)
$$\Big(\frac{S_j}{S_{i+1}}\Big)^\iota \Big| \log \frac{U(S_{n+1}/S_j)}{U(S_{n+1}/S_{i+1})} - \big(\gamma \log \frac{S_{i+1}}{S_j} + A\big(\frac{S_{n+1}}{S_{i+1}}\big) \frac{(S_{i+1}/S_j)^\rho - 1}{\rho}\big) \Big| = o\Big(A\big(\frac{S_{n+1}}{S_{i+1}}\big)\Big) \text{ a.s.}$$

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

$$\frac{A(S_{n+1}/S_{i+1})}{A(n/i)} \longrightarrow 1 \qquad \text{a.s.} \tag{3.3}$$

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

$$\Big| \frac{(S_{i+1}/S_j)^\rho - 1}{\rho} - \frac{(i/j)^\rho - 1}{\rho} \Big| = O\big((i/j)^\rho (\log\log(3j)/j)^{1/2}\big) \qquad \text{a.s.} \tag{3.4}$$

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

Combining (3.2)–(3.4) and the strong law of large numbers, we arrive at

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

Consequently,

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

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 (3.3), one obtains the second assertion. □

Proof of Corollary 1. First note that (2.3) implies

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

For all $\iota > 0$, the Potter bounds (Bingham et al, 1987, Theorem 1.5.6) yield

$$\frac{1}{2}(t^{\rho-\iota} \wedge t^{\rho+\iota}) \leq \frac{A(n/k_n)}{A(n/\lceil k_n t \rceil)} \leq 2(t^{\rho-\iota} \vee t^{\rho+\iota}) \tag{3.5}$$

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

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

Moreover, the uniform convergence theorem gives

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

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

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

On the other hand, in view of (3.5), we have

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

To sum up, we have shown that

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

where

$$W_n(t) := k_n^{-1/2} W(k_n t) \tag{3.6}$$

is a Brownian motion.

Since $(W_n(t)/t)_{t \geq 1}$ is uniformly continuous and

$$\sup_{t_n \leq t \leq 1} t^{1/2} |W_n(t)| \Big| \frac{1}{\lceil k_n t \rceil / k_n} - \frac{1}{t} \Big| \leq \sup_{t_n \leq t \leq 1} |W_n(t)| t^{-1/2} (k_n t)^{-1} = o_P(1)$$

by the law of iterated logarithm and $k_n t_n \to \infty$, to obtain (2.6), it remains to prove that

$$\sup_{t_n \leq t \leq 1} t^{-1/2} |W_n(\lceil k_n t \rceil / k_n) - W_n(t)| = o_P(1). \tag{3.7}$$

To this end, define $\log_1 k_n := \log k_n$, $\log_{j+1} k_n := \log(\log_j k_n)$ and, for fixed $\epsilon > 0$, $t_{n,0} := 1$ and $t_{n,j} := 6 \log_j k_n / (\epsilon^2 k_n)$. Because $t_{n,j}$ is decreasing in j and $t_{n,j} < 0$ for sufficiently large j, there exists j_n such that $t_{n,j_n} \leq t_n < t_{n,j_n - 1}$. According to Lemma A.1.1 of Csörgő and Horváth

(1993), for some constant C one has

$$
\begin{aligned}
& P\Big\{\sup_{t_n\le t\le 1} t^{-1/2}|W_n(\lceil k_n t\rceil/k_n)-W_n(t)|>\epsilon\Big\}\\
\le\ & \sum_{j=1}^{j_n} P\Big\{\sup_{t_{n,j}\le t\le t_{n,j-1}} |W_n(\lceil k_n t\rceil/k_n)-W_n(t)|>\epsilon t_{n,j}^{1/2}\Big\}\\
\le\ & C\sum_{j=1}^{j_n} k_n t_{n,j-1}\exp(-\epsilon^2 k_n t_{n,j}/3)\\
\le\ & Ck_n\exp(-2\log k_n)+6C\epsilon^{-2}\sum_{j=2}^{j_n}\log_{j-1}k_n\exp(-2\log_j k_n)\\
=\ & o(1)+O\Big(\sum_{j=2}^{j_n}(\log_{j-1}k_n)^{-1}\Big)=o(1),
\end{aligned}
$$

where the last equality follows from $\log_{j_n-1}k_n\ge\epsilon^2k_nt_n/6\to\infty$ and $\log_j k_n/\log_{j-1}k_n\to 0$ uniformly for $2\le j\le j_n-1$, which imply $\sum_{j=2}^{j_n}(\log_{j-1}k_n)^{-1}\le(\log_{j_n-1}k_n)^{-1}\sum_{l=0}^{j_n-2}2^{-l}=o(1)$ for sufficiently large n. Thus (3.7), and the proof of (2.5) is finished.

To prove the second assertion, choose t_n such that $k_nt_n\to\infty$ but $\sup_{0\le t\le t_n}h(t)k_n^{1/2}\to 0$, which is possible because of $\lim_{t\to0}t^{-1/2}h(t)=0$. Then, the definition of h and (2.4) ensure that

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

□

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 $a>0$,

$$
\begin{aligned}
E\int_0^a 1\{|\gamma W(t)/t+t^{-\rho}/(1-\rho)|\le\epsilon\}\,\frac{dt}{t} &= \int_0^a\mathcal{N}_{(0,\gamma^2)}\Big[\frac{t^{1/2-\rho}}{1-\rho}-\epsilon t^{1/2},\frac{t^{1/2-\rho}}{1-\rho}+\epsilon t^{1/2}\Big]\,\frac{dt}{t}\\
&\le \int_0^a(2\pi\gamma^2)^{-1/2}2\epsilon t^{-1/2}\,dt\ <\ \infty. \qquad (3.8)
\end{aligned}
$$

Next we will show that, due to the large bias, for i being large compared with k_n one has $k_n^{1/2}|H_{i,n}-\gamma|>\epsilon$ with large probability if $\rho<0$, or $\rho=0$ and $\epsilon<1$. Pick some (small) $\delta>0$. For $\rho<0$, choose M sufficiently large such that $P\{\sup_{t\ge M}|W(t)|/t\ge1\}\le\delta/2$ and $M^{-\rho/2}/(2(1-\rho))\ge\epsilon+\gamma+1$. Then, for $Mk_n\le i\le l_n$, (2.3), (3.6), (2.7) and the Potter bounds (3.5) imply

$$k_n^{1/2}|H_{i,n}-\gamma|=\Big|\gamma\frac{W_n(i/k_n)}{i/k_n}+\frac{A(n/i)(1+o(1))}{|A(n/k_n)|(1-\rho)}+O\big(k_n^{1/2}\frac{\log i}{i}\big)\Big|\ge\frac{M^{-\rho/2}}{2(1-\rho)}-\gamma-\frac12>\epsilon \qquad (3.9)$$

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

$$k_n^{1/2}|H_{i,n}-\gamma| \geq 1-\gamma\frac{|W_n(i/k_n)|}{i/k_n}+o(1)>\epsilon \tag{3.10}$$

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

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

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

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

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

This, however, follows easily from (2.6), which implies that $\sup_{m\leq t\leq M}|k_n^{1/2}(H_{\lceil k_nt\rceil,n}-\gamma)-(\gamma W_n(t)/t\pm t^{-\rho}/(1-\rho))|\xrightarrow{P}0$ for all $0<m<M<\infty$, by a continuous mapping argument. For one has, for all $\delta>0$, with probability tending to 1

$$\begin{aligned}\int_{\delta/2}^M 1_{\{|\gamma W_n(t)/t+t^{-\rho}/(1-\rho)|\leq\epsilon-\delta\}}\,dt-\frac{\delta}{2} &\leq \int_0^M 1_{\{k_n^{1/2}|H_{\lceil k_nt\rceil,n}-\gamma|\leq\epsilon\}}\,dt\\ &\leq \int_{\delta/2}^M 1_{\{|\gamma W_n(t)/t+t^{-\rho}/(1-\rho)|\leq\epsilon+\delta\}}\,dt+\frac{\delta}{2}\end{aligned}$$

where the left- and the right-hand side converge to $I(\epsilon):=\int_0^M 1_{\{|\gamma W_n(t)/t+t^{-\rho}/(1-\rho)|\leq\epsilon\}}\,dt$ as $\delta\to 0$, since the map $\epsilon\mapsto I(\epsilon)$ is continuous.

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

$$k_n^{1/2}|H_{i,n}-\gamma|\leq 1+o(1)+(\epsilon-1)/2<\epsilon \tag{3.11}$$

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

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

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

$$\log(l_n+1)\mathrm{PERALT}(k_n^{-1/2}\epsilon,n,u_n)=\sum_{i=1}^{l_n}\log\frac{i+1}{i}1_{\{k_n^{1/2}|H_{i,n}-\gamma|\leq\epsilon\}}. \tag{3.12}$$

In case of $\rho = 0$ and $\epsilon > 1$ we obtain from (3.11) that with probability greater than $1 - \delta$

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

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

If $\rho < 0$, or $\rho = 0$ and $\epsilon < 1$, then again (3.9) and (3.10), respectively, in combination with $\lim_{M\to\infty} \int_M^\infty 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)| \leq \epsilon\}} t^{-1} dt = 0$ show that it suffices to prove that

$$\sum_{i=1}^{\lceil Mk_n \rceil} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n} - \gamma| \leq \epsilon\}} \xrightarrow{d} \int_0^M 1_{\{|\gamma W(t)/t + t^{-\rho}/(1-\rho)| \leq \epsilon\}} \frac{dt}{t} \tag{3.13}$$

for all $M > 0$.

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 $P\{k_n^{1/2}|H_{i,n} - \gamma| \leq \epsilon\} = P\{H_{i,n} \in [\gamma - \epsilon k_n^{-1/2}, \gamma + \epsilon k_n^{-1/2}]\} \to 0$, so that $\sum_{i=1}^{i_0} \log((i+1)/i) 1_{\{k_n^{1/2}|H_{i,n}-\gamma| \leq \epsilon\}} \xrightarrow{P} 0$ 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

$$\sum_{i=1}^{j_n} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n} - \gamma| \leq \epsilon\}} \xrightarrow{P} 0. \tag{3.14}$$

Next note that by (2.3), (2.7) and (3.5) for all $\delta > 0$ there exists $C > 0$ such that with probability greater than $1 - \delta$

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

for all $j_n \leq i \leq mk_n + 1$ and sufficiently large n. Since

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

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

$$\sum_{i=j_n+1}^{\lfloor mk_n \rfloor} \log \frac{i+1}{i} 1_{\{k_n^{1/2}|H_{i,n} - \gamma| \leq \epsilon\}} \leq \delta \tag{3.15}$$

for sufficiently large n.

In view of (3.13)–(3.15) and (3.8), it remains to prove that for all $0 < m < M < \infty$ one has

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

yet 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 $H_{i,n}$

$$H_{i,n} = \gamma + \gamma\frac{W(i)}{i} + +O\big(\frac{\log(i+1)}{i}\big) \quad \text{a.s.} \tag{3.16}$$

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 $1-\delta$

$$k_n^{1/2}|H_{i,n}-\gamma| \le \gamma\frac{|W_n(i/k_n)|}{i/k_n} + \epsilon/2 \le \epsilon \tag{3.17}$$

for all $Mk_n \le i \le l_n$ with W_n defined in (3.6). Hence, by similar arguments as in the proof of Theorem 2

$$\begin{aligned}&\frac{\log(l_n+1)}{\log k_n}\big(1-\text{PERALT}(k_n^{-1/2}\epsilon, n, u_n)\big) \\ &= \frac{1}{\log k_n}\Bigg(\sum_{i=1}^{\lceil Mk_n\rceil\wedge l_n}\log\frac{i+1}{i}1_{\{k_n^{1/2}|H_{i,n}-\gamma|>\epsilon\}} + \sum_{i=(\lceil Mk_n\rceil\wedge l_n)+1}^{l_n}\log\frac{i+1}{i}1_{\{k_n^{1/2}|H_{i,n}-\gamma|>\epsilon\}}\Bigg) \\ &= \frac{\log(\lceil Mk_n\rceil+1)}{\log k_n} - \frac{1}{\log k_n}\Bigg(\int_0^{M\wedge(l_n/k_n)} 1_{\{\gamma|W(t)|/t\le\epsilon\}}\frac{dt}{t} + o(1)\Bigg),\end{aligned}$$

from which assertion (2.12) is obvious.

Because of (3.17), 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 (3.16) that

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

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

4. Concluding Remark

It may be possible to tune the scaling of the Hill plot's horizontal axis to suit the distribution. In practice, this would require a procedure to estimate, at least approximately, the second order behavior of the distribution tail $1-F$. It may also be possible to use the idea of occupation time in a strip to improve the estimation of γ. Investigations are underway.

References

[1] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge University Press, Cambridge.

[2] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, Chichester.

[3] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (1997). Using a bootstrap method to choose the sample fraction in tail index estimation. Preprint, Erasmus University Rotterdam.

[4] Davis, R. and Resnick, S. (1996). Limit theory for bilinear processes with heavy tailed noise. *Ann. Applied Probability,* 6, 1191–1210.

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

[6] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.* **25**, 187–210.

[7] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Processes Appl.* **75**, 149–172.

[8] de Haan, L., Peng, L. and Pereira, T.T. (1997). A bootstrap-based method to achieve optimality in estimating the extreme-value index. Preprint, Erasmus University Rotterdam and University of Lisbon.

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

[10] Jansen, D. and de Vries, C. (1991). On the frequency of large stock returns: putting booms and busts into perspective. *The Review of Economics and Statistics,* LXXIII, No 1, 18–24.

[11] Kaufmann, E. and Reiss, R.-D. (1998). Strong approximation of the Hill estimator process. To appear in *Statist. Probab. Lett.*

[12] Leland, W., Taqqu, M., Willinger, W., Wilson, D. (1994) On the self–similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2, 1–15.

[13] Mason, D. (1982). Laws of large numbers for sums of extreme values. *Ann. Probability,* 10, 754–764.

[14] Mason, D. and Turova, T. (1994) Weak convergence of the Hill estimator process. *Extreme Value Theory and Applications;* edited by Galambos, J., Lechner, J. and Simiu, E. Kluwer Academic Publishers, Dordrecht, Holland 419–432.

[15] McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory *Astin Bulletin,* 27, 125–139.

[16] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics.* Springer, New York.

[17] Resnick, S. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin*, 27, 139–151.

[18] Resnick, S. and Stărică, C. (1995). Consistency of Hill's estimator for dependent data. *J. Applied Probability,* 32, 139–167.

[19] Resnick, S. and Stărică, C. (1997). Smoothing the Hill estimator, *Advances in Applied Probability,* 29, 271–293.

[20] Resnick, S. and Stărică, C. (1998). Tail index estimation for dependent data. Available as TR1174.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html; to appear Ann. Appl. Probability.

[21] Resnick, S. and Van den Berg, E. (1998). Sample correlation behavior for the heavy tailed general bilinear process. Technical report.

[22] Revuz, D. and Yor, M.. (1991) *Continuous Martingales And Brownian Motion.* Springer, Berlin.

[23] 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, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260.

Holger Drees, Mathematical Institute, University of Cologne, Weyertal 86–90, 50931 Cologne, Germany

E-mail address: `hdrees@mi.uni-koeln.de`

Laurens de Haan, Erasmus University, Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

E-mail address: `ldhaan@few.eur.nl`

Sidney Resnick, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853

E-mail address: `sid@orie.cornell.edu`