A simple robust estimation method for the thickness of heavy tails

Mark M. Meerschaert [a,*], Hans-Peter Scheffler [b,1]

[a] *Department of Mathematics, University of Nevada, Reno, NV 89557, USA*
[b] *Fachbereich Mathematik, Universität Dortmund, 44221 Dortmund, Germany*

Abstract

We present a simple general method for estimating the thickness of heavy tails based on the asymptotics of the sum. The method works for dependent data, and only requires that the centered and normalized partial sums are stochastically compact. For data in the domain of attraction of a stable law our estimator is asymptotically log stable, consistent and asymptotically unbiased, and converges in the mean-square sense to the index of regular variation. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

Heavy tail probability distributions are important in applications to electrical engineering, finance, geology, hydrology, and physics. See, for example, Brockwell and Davis (1991), Feller (1971), Hosking and Wallis (1987), Janicki and Weron (1994), Jansen and de Vries (1991), Leadbetter et al. (1980), Loretan and Phillips (1994), Mandelbrot (1982), McCulloch (1997), Mittnik and Rachev (1995), Nikias and Shao (1995) and Resnick and Stărică (1995). Several different heavy tail distributions are commonly used, including the stable laws, the type II extreme value distributions, and the generalized Pareto laws. In many applications it is impossible to determine a priori which family of distributions is appropriate, leading to a need for robust estimators of the distributional parameters. Hill (1975) proposed the most popular robust estimator, based on the asymptotics of extreme values. McCulloch (1997) shows that Hill's

---

* Corresponding author. E-mail: mcubed@unr.edu.

estimator can yield misleading results when applied to stable data. In this paper we propose an alternative robust estimator based on the asymptotics of the sum. Our estimator performs about as well as Hill's estimator in most practical situations, and it performs best in exactly those situations in which Hill's estimator is most likely to fail.

We say that a random variable X has heavy tails if there is some $\alpha > 0$ such that $E|X|^\delta < \infty$ for $0 < \delta < \alpha$ and $E|X|^\delta = \infty$ for $\delta > \alpha$. The parameter α measures tail thickness. In this paper we present a new method for estimating the parameter α for a heavy tail distribution. Our method is based on the asymptotics of the sum. For heavy tail data, these asymptotics depend only on the tail index α, and not on the exact form of the distribution. This makes our estimator very robust. It even works for dependent data, or when the centered and normalized partial sums are only stochastically compact. Since the asymptotics of the sum are well known, theoretical properties of our estimator are relatively easy to check. When the data belong to the domain of attraction of a stable law with index α, our estimator is asymptotically log stable. This allows us to construct confidence intervals and hypothesis tests. Our estimator is also simple to compute, and perform as well or better than existing estimators in most practical applications.

We illustrate our estimation method in a simple special case. Suppose that X_n are i.i.d. strictly positive random variables belonging to the domain of normal attraction of some α stable random variable $Y > 0$ (this requires $\alpha < 1$) and write $S_n = X_1 + \cdots + X_n$. Then we have

$$n^{-1/\alpha} S_n \Rightarrow Y,$$

$$\ln S_n - \frac{1}{\alpha} \ln n \Rightarrow \ln Y, \tag{1.1}$$

$$\ln n \left(\frac{\ln S_n}{\ln n} - \frac{1}{\alpha} \right) \Rightarrow \ln Y$$

and so $\ln S_n / \ln n$ is a consistent, asymptotically log-stable estimator of $1/\alpha$. Although the rate of convergence is slow, the limiting distribution is rather narrow, and we have found that the estimator is useful in most situations where asymptotic methods would be appropriate.

2. Asymptotic results

In this section we present asymptotic results for our estimation method. First, we discuss the estimation of the tail thickness parameter α for data in the domain of attraction of a stable law. Suppose that $X, X_1, X_2, X_3, \ldots$ are independent, identically distributed random variables on $\mathbb{R}^1$. We say that X belongs to the domain of attraction of a stable random variable Y with index α if there exist real constants $a_n > 0$ and b_n such that $a_n^{-1}(X_1 + \cdots + X_n - nb_n) \Rightarrow Y$, and we write $X \in \mathrm{DOA}(\alpha)$. If $0 < \alpha < 2$ then $P[|X| > t]$ varies regularly with index $-\alpha$ and we can take $a_n = \sup\{t : nP[|X| > t] \geqslant 1\}$, $b_n = EXI(|X| \leqslant a_n)$, see, for example, Feller (1971) XVII. In the special case $a_n = n^{1/\alpha}$

we say that X belongs to the domain of normal attraction of Y and we write $X \in$ DONA(α). Define

$$\hat{\gamma}_n = \frac{\ln_+ \sum_{i=1}^n (X_i - \bar{X}_n)^2}{2\ln n}, \tag{2.1}$$

where $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$ is the sample mean and $\ln_+(x) = \max\{\ln x, 0\}$. Our first result shows that $\hat{\gamma}_n$ is a consistent estimator of $1/\alpha$ which converges in L_2 norm and whose asymptotic distribution can be expressed as a simple function of an $\alpha/2$ stable law.

Theorem 1. *If $X \in$ DOA(α) for some $0 < \alpha < 2$ then*:
  (i) $\hat{\gamma}_n \xrightarrow{P} 1/\alpha$;
  (ii) $E(\hat{\gamma}_n) \to 1/\alpha$;
  (iii) $E(\hat{\gamma}_n - 1/\alpha)^2 \to 0$; *and furthermore,*

$$2\ln n(\hat{\gamma}_n - 1/\alpha - c_n) \Rightarrow \ln Y^{(2)} \tag{2.2}$$

for some $\alpha/2$ stable law $Y^{(2)}$ and some $c_n \to 0$ (if $X \in$ DONA(α) we can take $c_n = 0$).

Proof. Following Feller (1971) VIII.8 we define $V_\eta(t) = E|X|^\eta I(|X| > t)$ and $U_\zeta(t) = E|X|^\zeta I(|X| \leqslant t)$. If $X \in$ DOA(α) for $0 < \alpha < 2$ then $P[X^2 > t] = P[|X| > \sqrt{t}] = V_0(\sqrt{t})$ varies regularly with index $\alpha/2$, hence $X^2 \in$ DOA($\alpha/2$) and, in fact, $a_n^{-2} \sum X_i^2 \Rightarrow Y^{(2)}$ centered stable with index $\alpha/2$ and skewness 1. Note that no centering is required since $\alpha/2 < 1$.

Lemma 1. *Define $\bar{S}_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ is the sample mean. If $X \in$ DOA(α) for $0 < \alpha < 2$ then*

$$a_n^{-2} \bar{S}_n \Rightarrow Y^{(2)}, \tag{2.3}$$

where $Y^{(2)}$ is centered stable with index $\alpha/2$ and skewness 1.

Note that $\bar{S}_n - \sum X_i^2 = -n\bar{X}_n^2$. Here and below the sum is taken over $i = 1, \ldots, n$. Then it will suffice to show that $na_n^{-2}\bar{X}_n^2 \to 0$ in probability. Since $X \in$ DOA(α) we have $na_n^{-1}(\bar{X}_n - b_n) \Rightarrow Y$ stable with index α, hence $n^{1/2}a_n^{-1}(\bar{X}_n - b_n) \to 0$ in probability. If $1 < \alpha < 2$ then $b_n \to EX$ and since a_n varies regularly with index $1/\alpha$ we get $n^{1/2}a_n^{-1}b_n \to 0$. If $0 < \alpha \leqslant 1$ then $U_1(a_n)$ varies regularly with index $(1 - \alpha)(1/\alpha)$ and so again $|n^{1/2}a_n^{-1}b_n| \leqslant n^{1/2}a_n^{-1}U_1(a_n) \to 0$. Then $n^{1/2}a_n^{-1}\bar{X}_n \to 0$ in probability, and Eq. (2.3) follows easily.

Lemma 2. *For all $\delta > 0$, for some $K_0 > 0$ we have*

$$P\left[\sum_{i=1}^n X_i^2 \geqslant a_n^2 t\right] \leqslant K_0 t^{-\alpha/2+\delta} \tag{2.4}$$

for all $n \geqslant 1$ and all $t > 0$.

Define $S_{nn} = \sum X_i^2 I(X_i^2 \leqslant a_n^2 t)$, $E_n = \{X_i^2 > a_n^2 t \; \exists \; 1 \leqslant i \leqslant n\}$, $G_n = \{S_{nn} \geqslant a_n^2 t\}$ so that $\{\sum X_i^2 \geqslant a_n^2 t\} \subseteq E_n \cup G_n$. Then $P[\sum X_i^2 \geqslant a_n^2 t] \leqslant P(E_n) + P(G_n)$. Here $P(E_n) \leqslant nP[X^2 > a_n^2 t] = nP[X > a_n\sqrt{t}] = nV_0(a_n\sqrt{t}) = nV_0(a_n) \cdot V_0(a_n\sqrt{t})/V_0(a_n)$ where $nV_0(a_n) \to C$ so that $nV_0(a_n) \leqslant C_1$ for all $n \geqslant 1$ for some $C_1 \geqslant C$. Without loss of generality, $a_n \geqslant 1$ for all n and then Potter's theorem (see, for example, Bingham et al., 1987, p. 25) implies that there exist $t_0 > 0$ and $A > 0$ such that $V_0(a_n\sqrt{t})/V_0(a_n) \leqslant At^{-\alpha/2+\delta}$ for all $n \geqslant 1$ and all $t \geqslant t_0$. Hence, $P(E_n) \leqslant C_1 A t^{-\alpha/2+\delta}$ for all $n \geqslant 1$ and all $t \geqslant t_0$. Next, we have $P(G_n) \leqslant (a_n^2 t)^{-1} E S_{nn} = n(a_n^2 t)^{-1} U_2(a_n\sqrt{t}) = U_2(a_n\sqrt{t})/(a_n\sqrt{t})^2 V_0(a_n\sqrt{t}) \cdot V_0(a_n\sqrt{t})/V_0(a_n) \cdot nV_0(a_n) \leqslant C_2 \cdot At^{-\alpha/2+\delta} \cdot C_1$ for all $n \geqslant 1$ and all $t \geqslant t_0$, since $I_{n,t} = U_2(a_n\sqrt{t})/(a_n\sqrt{t})^2 V_0(a_n\sqrt{t}) \to \alpha/(2-\alpha)$ by Karamata's Theorem (see, for example, Feller, 1971, p. 283) and so $I_{n,t} \leqslant C_2$ for all $n \geqslant 1$ and all $t \geqslant t_0$ for some $C_2 \geqslant \alpha/(2-\alpha)$. Then Eq. (2.4) holds for all $n \geqslant 1$ and all $t \geqslant t_0$ with $K_0 = C_1 A(1 + C_2)$. Enlarge K_0 if necessary so that $K_0 t^{-\alpha/2+\delta} > 1$ for all $0 < t < t_0$ to finish the proof.

Lemma 3. *Define* $V_n(t) = P[\bar{S}_n \geqslant a_n^2 t]$ *where* $\bar{S}_n$ *is as in Lemma* 1. *For all* $\delta > 0$, *for some* $C > 0$ *we have*

$$V_n(t) \leqslant Ct^{-\alpha/2+\delta} \tag{2.5}$$

for all $n \geqslant 1$ *and all* $t > 0$.

Write $V_n(t) = P[\bar{S}_n \geqslant a_n^2 t] = P[\sum X_i^2 - n\bar{X}_n^2 \geqslant a_n^2 t] \leqslant P[\sum X_i^2 \geqslant a_n^2 t/2] + P[n\bar{X}_n^2 \geqslant a_n^2 t/2]$. Apply Lemma 2. The large deviations theorem of Heyde (1967) yields that $P[n\bar{X}_n^2 \geqslant a_n^2 t/2] = P[(\sum X_i)^2 \geqslant na_n^2 t/2] = P[|\sum X_i| \geqslant a_n\sqrt{nt/2}] \leqslant C_0 nP[|X| \geqslant a_n\sqrt{nt/2}] = C_0 nV_0(a_n\sqrt{nt/2}) \leqslant C_0 nV_0(a_n\sqrt{nt_0/2})$ for all $n \geqslant 1$ and $t \geqslant t_0$. But $nV_0(a_n\sqrt{nt_0/2}) = nV_0(a_n) \cdot V_0(a_n\sqrt{nt_0/2})/V_0(a_n)$ where $nV_0(a_n) \leqslant C_1$ and $V_0(a_n\sqrt{nt_0/2})/V_0(a_n) \leqslant A(nt_0/2)^{-\alpha/2+\delta}$ by Potter's theorem. Then Eq. (2.5) holds for $n \geqslant 1$ and $t \geqslant t_0$ with $C = (K_0 + C_1 A)2^{\alpha/2-\delta}$. Enlarge C if necessary so that $Ct^{-\alpha/2+\delta} > 1$ for all $0 < t < t_0$ to finish the proof.

Lemma 4. *Define* $V_n(t)$ *as in Lemma* 3. *Then*

$$E\hat{\gamma}_n = \frac{1}{2\ln n} \int_{a_n^{-2}}^{\infty} V_n(s) \frac{ds}{s}. \tag{2.6}$$

Recall that $\bar{S}_n = \sum (X_i - \bar{X}_n)^2$ and that $Y^{(2)}$ is stable so it has a density. From Lemma 1 we obtain $V_n(t) = P[\bar{S}_n \geqslant a_n^2 t] \to P[Y^{(2)} \geqslant t]$ for all $t > 0$ as $n \to \infty$. Since $t^{-\alpha/2} P[Y^{(2)} \geqslant t] \to K'$ as $t \to \infty$ we also have $P[Y^{(2)} \geqslant t] \leqslant Kt^{\alpha/2}$ for all $t > 0$. Define $\tilde{V}_n(t) = P[\bar{S}_n \geqslant t]$ and integrate by parts to obtain

$$E(\hat{\gamma}_n) = \frac{1}{2\ln n} \int \ln_+ \bar{S}_n \, dP$$

$$= \frac{1}{2\ln n} \int_{\bar{S}_n \geqslant 1} \ln \bar{S}_n \, dP$$

$$= \frac{-1}{2\ln n} \int_1^\infty \ln t \, d\tilde{V}_n(t)$$

$$= \frac{-1}{2\ln n} \ln t \, \tilde{V}_n(t) \Big|_1^\infty + \frac{1}{2\ln n} \int_1^\infty \tilde{V}_n(t) \frac{dt}{t},$$

where the first term vanishes since for n fixed $\tilde{V}_n(t) = V_n(a_n^{-2}t) \leqslant C a_n^{\alpha-2\delta} t^{-\alpha/2+\delta}$ by virtue of Lemma 3. Now, substitute $t = a_n^2 s$ to obtain Eq. (2.6).

Now, we establish (ii). Write $a_n = n^{1/\alpha} \ell_n$ and note that ℓ_n is slowly varying. In view of Lemma 1 and the fact that $Y^{(2)}$ is stable (so it has a density) we get for every $\varepsilon > 0$ a constant $c > 0$ such that $1 - \varepsilon \leqslant V_n(c) = P[a_n^{-2}\bar{S}_n \geqslant c] \leqslant 1$ for all large n. Apply Lemma 4 to write $E\hat{\gamma}_n = I_n + J_n$ where

$$I_n = \frac{1}{2\ln n} \int_{a_n^{-2}}^c V_n(s) \frac{ds}{s} \quad \text{and} \quad J_n = \frac{1}{2\ln n} \int_c^\infty V_n(s) \frac{ds}{s}$$

and compute that

$$I_n \leqslant \frac{1}{2\ln n} \int_{a_n^{-2}}^c \frac{ds}{s} = \frac{\ln c + 2\ln a_n}{2\ln n} = \frac{\ln c}{2\ln n} + \frac{1}{\alpha} + \frac{\ln \ell_n}{2\ln n},$$

while

$$I_n \geqslant \frac{1}{2\ln n} V_n(c) \int_{a_n^{-2}}^c \frac{ds}{s} \geqslant (1 - \varepsilon) \left(\frac{1}{\alpha} + \frac{\ln c}{2\ln n} + \frac{\ln \ell_n}{2\ln n} \right).$$

Then as $n \to \infty$ we have $(1-\varepsilon)/\alpha \leqslant \liminf I_n \leqslant \limsup I_n \leqslant 1/\alpha$, and since $\varepsilon > 0$ is arbitrary it follows that $I_n \to 1/\alpha$. Recall from the proof of Lemma 4 that $V_n(s) \to P[Y^{(2)} \geqslant s] \leqslant K s^{\alpha/2}$. Then by Lemma 3 we can apply dominated convergence to obtain

$$\int_c^\infty V_n(s) \frac{ds}{s} \to \int_c^\infty P[Y^{(2)} \geqslant s] \frac{ds}{s} < \infty$$

and so $J_n \to 0$ which completes the proof of (ii).

The proof of (iii) is quite similar to (ii). Note that $E(\hat{\gamma}_n - 1/\alpha)^2 = E\hat{\gamma}_n^2 - 2\alpha^{-1}E\hat{\gamma}_n + \alpha^{-2}$ so in view of (ii) it is enough to show that $E\hat{\gamma}_n^2 \to \alpha^{-2}$. Integrate by parts to obtain

$$E(\hat{\gamma}_n^2) = \frac{1}{(2\ln n)^2} \int_{\bar{S}_n \geqslant 1} (\ln \bar{S}_n)^2 \, dP$$

$$= \frac{-1}{(2\ln n)^2} \int_1^\infty (\ln t)^2 \, d\tilde{V}_n(t)$$

$$= \frac{2}{(2\ln n)^2} \int_1^\infty (\ln t) \, \tilde{V}_n(t) \frac{dt}{t}$$

and substitute $t = a_n^2 s$ to obtain

$$E\hat{\gamma}_n^2 = \frac{2}{(2\ln n)^2} \int_{a_n^{-2}}^{\infty} \ln(a_n^2 s) V_n(s) \frac{\mathrm{d}s}{s}$$

$$= \frac{2\ln a_n}{\ln n} \frac{1}{2\ln n} \int_{a_n^{-2}}^{\infty} V_n(s) \frac{\mathrm{d}s}{s} + \frac{2}{(2\ln n)^2} \int_{a_n^{-2}}^{\infty} \ln(s) V_n(s) \frac{\mathrm{d}s}{s}$$

$$= A_n + B_n.$$

Since $\ln a_n / \ln n \to \alpha^{-1}$ we get $A_n = (2\ln a_n / \ln n) I_n \to 2\alpha^{-2}$. Substituting $s = \mathrm{e}^t$ we also have for any $b > 0$ that

$$B_n = \frac{2}{(2\ln n)^2} \int_{-2\ln a_n}^{\infty} t V_n(\mathrm{e}^t) \, \mathrm{d}t$$

$$= \frac{2}{(2\ln n)^2} \int_{-2\ln a_n}^{-b} t V_n(\mathrm{e}^t) \, \mathrm{d}t + \frac{2}{(2\ln n)^2} \int_{-b}^{\infty} t V_n(\mathrm{e}^t) \, \mathrm{d}t$$

$$= B_n^{(1)} + B_n^{(2)},$$

where by Lemma 3

$$|B_n^{(2)}| \leqslant \frac{2}{(2\ln n)^2} \int_{-b}^{\infty} |t| C \mathrm{e}^{-(\alpha/2+\delta)t} \, \mathrm{d}t,$$

which tends to zero as $n \to \infty$. Given $\varepsilon > 0$ choose $b > 0$ so that $1 - \varepsilon \leqslant V_n(\mathrm{e}^{-b}) \leqslant 1$ for all large n. Then

$$B_n^{(1)} \leqslant \frac{2}{(2\ln n)^2} \int_{-2\ln a_n}^{-b} t \, \mathrm{d}t = \frac{b^2 - (2\ln a_n)^2}{(2\ln n)^2} \to \frac{-1}{\alpha^2},$$

while

$$B_n^{(1)} \geqslant \frac{2}{(2\ln n)^2} V_n(\mathrm{e}^{-b}) \int_{-2\ln a_n}^{-b} t \, \mathrm{d}t \geqslant (1 - \varepsilon) \frac{b^2 - (2\ln a_n)^2}{(2\ln n)^2} \to (1 - \varepsilon) \frac{-1}{\alpha^2}$$

and since $\varepsilon > 0$ is arbitrary it follows that $B_n^{(1)} \to -\alpha^{-2}$. Then $E\hat{\gamma}_n^2 = A_n + B_n \to 2\alpha^{-2} - \alpha^{-2}$ which completes the proof of (iii).

To obtain Eq. (2.2) start with Eq. (2.3) and apply the continuous mapping theorem to get $\ln \bar{S}_n - 2\ln a_n \Rightarrow \ln Y^{(2)}$. Note that $P[|\ln \bar{S}_n - \ln_+ \bar{S}_n| > \varepsilon] = P[\ln \bar{S}_n > \varepsilon, \bar{S}_n < 1] = P[\bar{S}_n < \mathrm{e}^{-\varepsilon}] = P[a_n^{-2} \bar{S}_n < a_n^{-2} \mathrm{e}^{-\varepsilon}] \to 0$ as $n \to \infty$ in view of Lemma 1, so we also have $\ln_+ \bar{S}_n - 2\ln a_n \Rightarrow \ln Y^{(2)}$ or equivalently

$$2\ln n \left(\frac{\ln_+ \bar{S}_n}{2\ln n} - \frac{2\ln a_n}{2\ln n} \right) \Rightarrow \ln Y^{(2)},$$

where $a_n = n^{1/\alpha} \ell_n$, and this is equivalent to Eq. (2.2) with $c_n = \ln \ell_n / \ln n$. If $X \in$ DONA(α) then $\ell_n = 1$ so $c_n = 0$. Finally, note that (i) follows immediately from (ii), (iii), or Eq. (2.2).

For sake of completeness we now consider the remaining case where $X \in \mathrm{DOA}(\alpha)$ for $\alpha = 2$. Then the limit law Y is normal, and we have that either $EX^2 < \infty$ and the central limit theorem applies, or else $EX^2 I(|X| \leqslant t) \to \infty$ is slowly varying and we can take $a_n = \sup\{t: nt^{-2} EX^2 I(|X| \leqslant t) \geqslant 1\}$, $b_n = EX$. In practical applications the following result is used only to rule out the case of a normal limit, and so we have not tried to be as comprehensive as in the case $0 < \alpha < 2$ which was considered above. To ensure that the limit in Eq. (2.8) below is well defined we can set $\ln 0 = 0$.

Theorem 2. *If $X \in \mathrm{DOA}(2)$ then $\hat{\gamma}_n \to \frac{1}{2}$ in probability, and furthermore*:
  (i) *if the variance σ^2 of X is finite then*

$$2\ln n(\hat{\gamma}_n - 1/2) \xrightarrow{\mathrm{P}} \ln \sigma^2; \qquad (2.7)$$

  (ii) *if the variance of X is infinite and $P[|X| > t]$ varies regularly then*

$$2\ln n(\hat{\gamma}_n - 1/2 - C_n) \Rightarrow \ln |Y_0| \qquad (2.8)$$

for some stable law Y_0 with index 1 and some $C_n \xrightarrow{\mathrm{P}} 0$.

Proof. In the finite variance case it is well known that $\bar{S}_n/n \to \sigma^2$ in probability. Then Eq. (2.7) follows easily, which in turn implies that $\hat{\gamma}_n \to \frac{1}{2}$ in probability. In the remaining case we have as in the proof of Theorem 1 that $a_n^{-2} \sum (X_i^2 - b_n^{(2)}) \Rightarrow Y_0$ where the limit is 1 stable, a_n^2 varies regularly with index 1, and $b_n^{(2)} = EX^2 I(X^2 \leqslant a_n^2)$. Without loss of generality, we can assume that $EX = 0$, and then $na_n^{-2} \bar{X}_n^2 \to 0$ in probability which implies that $a_n^{-2}(\bar{S}_n - nb_n^{(2)}) \Rightarrow Y_0$. Since $na_n^{-2} b_n^{(2)} \to \infty$ is slowly varying it is not hard to check that $\ln |Z_n| / \ln n \to 0$ in probability where $Z_n = (\bar{S}_n - nb_n^{(2)})/\bar{S}_n$. It follows that Eq. (2.8) holds with $C_n = (\ln |Z_n| - \ln \ell_n)/\ln n$, and again this implies consistency.

Next, we generalize the results of the Theorem 1 to the case where the normalized partial sums are only stochastically compact. Suppose that $X, X_1, X_2, X_3, \ldots$ are independent, identically distributed random variables on $\mathbb{R}^1$ which belong to the domain of semistable attraction of some (b, c) semistable random variable Y. This means that there exists $k_n \to \infty$ with $k_{n+1}/k_n \to c > 1$, $\tilde{a}_n > 0$ and $\tilde{b}_n$ such that

$$\tilde{a}_n^{-1}(X_1 + \cdots + X_{k_n}) - \tilde{b}_n \Rightarrow Y \qquad (2.9)$$

and we will write $X \in \mathrm{DOSA}(\alpha)$ where $\alpha = \log c/\log b$. This is the most general framework in which an i.i.d. sum of random variables can be usefully approximated by a limit distribution. In the special case $\tilde{a}_n = n^{1/\alpha}$ we say that X belongs to the domain of normal semistable attraction of Y and we write $X \in \mathrm{DONSA}(\alpha)$. Assume that $0 < \alpha < 2$ so that Y is nonnormal. Then $P[|X| > t] = O(R(t))$ where R varies regularly with index $-\alpha$, see Shimizu (1970). Also, $E|X|^\beta = \infty$ for all $\beta > \alpha$, see Scheffler (1994). Thus, the parameter α measures tail thickness. Although the normalized sequence of partial sums converges in distribution only along a subsequence, the following embedding result shows that the entire sequence is stochastically compact.

Lemma 5. *Suppose that $X \in \mathrm{DOSA}(\alpha)$ with $0 < \alpha < 2$ and Eq. (2.9) holds, where the distribution v of the limit Y is (b, c) semistable. Then there exist a_n regularly varying with index $1/\alpha$ and b_n such that $a_{k_n} = \tilde{a}_n$ and $a_n^{-1}(X_1 + \cdots + X_n - nb_n)$ is stochastically compact, with every limit point of the form $\lambda^{-1/\alpha} v^\lambda$ for some $\lambda \in [1, c]$.*

Proof. Since v is the limit distribution of a triangular array, it is infinitely divisible. Then for any $\lambda > 0$ we can define the convolution power v^λ to be the probability distribution with characteristic function $\psi(t)^\lambda$, where $\psi(t)$ is the characteristic function of v. Write $n = \lambda_n k_{p_n}$ where $k_{p_n} \leqslant n < k_{p_{n+1}}$. Since $k_{n+1}/k_n \to c$ it follows that λ_n is relatively compact and every limit point lies in $[1, c]$. Define

$$a_n = \lambda_n^{1/\alpha} \tilde{a}_{p_n} \quad \text{and} \quad b_n = \lambda_n^{1-1/\alpha} \tilde{b}_{p_n}.$$

Since Y is (b, c) semistable we have $\tilde{a}_{n+1}/\tilde{a}_n \to b$ and then it is easy to check using $b = c^{1/\alpha}$ that a_n varies regularly with index $1/\alpha$. Given any subsequence choose a further subsequence along which $\lambda_n \to \lambda \in [1, c]$. Let μ denote the distribution of X, and use characteristic functions to check that $a_n^{-1} \mu^n * \delta(b_n) \to \lambda^{-1/\alpha} v^\lambda$ along this subsequence.

Theorem 3. *If $X \in \mathrm{DOSA}(\alpha)$ for some $0 < \alpha < 2$ then:*

(i) $\hat{\gamma}_n \xrightarrow{\mathrm{P}} 1/\alpha$;

(ii) $E(\hat{\gamma}_n) \to 1/\alpha$;

(iii) $E(\hat{\gamma}_n - 1/\alpha)^2 \to 0$;

and furthermore the sequence $2\ln n(\hat{\gamma}_n - 1/\alpha - c_n)$ is stochastically compact with limit set $\{\ln Y^{(\lambda)} : 1 \leqslant \lambda \leqslant c\}$ where $\mathcal{L}(Y^{(\lambda)}) = \lambda^{-2/\alpha} v_2^\lambda$ for some (b^2, c) semistable distribution v_2 and some $c_n \to 0$ (if $X \in \mathrm{DONSA}(\alpha)$ we can take $c_n = 0$).

Proof. The proof is quite similar to Theorem 1 above, and so we only sketch the argument. Apply Lemma 5 and let ϕ denote the Lévy measure of v. Since $X \in \mathrm{DOSA}(\alpha)$ we also have $X^2 \in \mathrm{DOSA}(\alpha/2)$. Moreover, the standard convergence criteria for triangular arrays can be used to check that if $a_n \sum X_i - b_n$ converges in distribution along a subsequence to $\lambda^{-1/\alpha} v^\lambda$ then $a_n^{-2} \sum X_i^2 \Rightarrow Y^{(\lambda)}$ along the same subsequence. The limit distribution is $\lambda^{-2/\alpha} v_2^\lambda$ where v_2 is (b^2, c) semistable with Lévy measure $\phi_2(t, \infty) = \phi\{x : |x| > \sqrt{t}\}$. Recall that $\bar{S}_n - \sum X_i^2 = -n\bar{X}_n^2$ and argue as before that $na_n^{-2}\bar{X}_n^2 \to 0$ in probability. Then $a_n^{-2}\bar{S}_n$ is stochastically compact with limit points $Y^{(\lambda)}$. Note that $\ln Y^{(\lambda)}/\ln n \to 0$ in probability uniformly in $\lambda \in [1, c]$. The argument for (ii) and (iii) is similar to Theorem 1 above, using the fact that $P[|X| > t] = \mathrm{O}(R(t))$ where R varies regularly with index $-\alpha$. If $a_n^{-2}\bar{S}_n \Rightarrow Y^{(\lambda)}$ along a subsequence, the continuous mapping theorem applies to yield $-2\ln a_n + \ln_+ \bar{S}_n \Rightarrow \ln Y^{(\lambda)}$, which is equivalent to the asserted stochastic compactness result with $c_n = \ln \ell_n / \ln n$ where $a_n = n^{-1/\alpha} \ell_n$. If $X \in \mathrm{DONSA}(\alpha)$ then $\ell_n = 1$ so $c_n = 0$. Finally, note that (i) follows easily from (ii) or (iii).

Next, we show that our estimator is consistent and asymptotically log stable even for dependent data. Suppose that $Z, Z_1, Z_2, Z_3, \ldots$ are independent, identically distributed random variables on $\mathbb{R}^1$. We consider the infinite moving average $X_t = \sum_j c_j Z_{t-j}$ where

$Z \in \text{DOA}(\alpha)$ for some $0 < \alpha < 2$. As before we let $\bar{S}_n = \sum (X_i - \bar{X}_n)^2$ where $\bar{X}_n = n^{-1} \sum X_i$ is the sample mean and the sums are taken over $i = 1, \dots, n$. Davis and Resnick (1985) show that in this case $a_n^{-2} \bar{S}_n \Rightarrow Y^{(2)}$ where a_n varies regularly with index $1/\alpha$ and $Y^{(2)}$ is centered stable with index $\alpha/2$ and skewness 1.

Theorem 4. *If* $X_t = \sum c_j Z_{t-j}$ *where* $Z \in \text{DOA}(\alpha)$ *for some* $0 < \alpha < 2$ *then* $\hat{\gamma}_n \xrightarrow{\text{P}} 1/\alpha$ *and furthermore*

$$2\ln n(\hat{\gamma}_n - 1/\alpha - c_n) \Rightarrow \ln Y^{(2)} \tag{2.10}$$

for some $\alpha/2$ *stable law* $Y^{(2)}$ *and some* $c_n \to 0$ *(if* $Z \in \text{DONA}(\alpha)$ *we can take* $c_n = 0$*).*

Proof. Since $Y^{(2)}$ is stable it has a density, and since it has skewness 1 it is almost surely positive. Then continuous mapping applied to the result of Davis and Resnick (1985) yields $-2\ln a_n + \ln \bar{S}_n \Rightarrow \ln Y^{(2)}$. The remaining argument is exactly the same as in Theorem 1.

Our estimator is also consistent and asymptotically log stable even for nonstationary time-series models. Consider the periodic moving average $X_t = \sum c_j(t) Z_{t-j}$ where $Z \in \text{DOA}(\alpha)$ for some $0 < \alpha < 2$ and $c_j(t)$ are all periodic with the same period v. Anderson and Meerschaert (1997) show that in this case $a_n^{-2} \bar{S}_n \Rightarrow Y^{(2)}$ where a_n varies regularly with index $1/\alpha$ and $Y^{(2)}$ is centered stable with index $\alpha/2$ and skewness 1. $\square$

Theorem 5. *If* $X_t = \sum c_j(t) Z_{t-j}$ *where* $Z \in \text{DOA}(\alpha)$ *for some* $0 < \alpha < 2$ *and* $c_j(t)$ *are all periodic with the same period* $v \geq 1$ *then* $\hat{\gamma}_n \xrightarrow{\text{P}} 1/\alpha$ *and furthermore,*

$$2\ln n(\hat{\gamma}_n - 1/\alpha - c_n) \Rightarrow \ln Y^{(2)} \tag{2.11}$$

for some $\alpha/2$ *stable law* $Y^{(2)}$ *and some* $c_n \to 0$ *(if* $Z \in \text{DONA}(\alpha)$ *we can take* $c_n = 0$*).*

Proof. The proof is identical to that of Theorem 4, using Anderson and Meerschaert (1997) in place of Davis and Resnick (1985).

The limit distribution of our estimator $\hat{\gamma}_n$ in Eq. (2.2) has a nonzero mean, and in practical applications it is advantageous to correct for this. By recentering our estimator so that the limit has mean zero we arrive at

$$\hat{\beta}_n = \frac{\gamma + \ln_+ \sum_{i=1}^n (X_i - \bar{X}_n)^2}{2(\gamma + \ln n)}, \tag{2.12}$$

where $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$ is the sample mean, $\gamma \doteq 0.5772$ is Euler's constant, and $\ln_+(x) = \max\{\ln x, 0\}$. Our next result shows that $\hat{\beta}_n$ and $\hat{\gamma}_n$ have the same asymptotics. In the special case where $n^{-1/\alpha} \sum_i (X_i - b_n) \Rightarrow Y$ and the stable limit Y has scale $\sigma_2 = 1$, we will write $X \in \text{DONA}(\alpha, 1)$. If $X \in \text{DONA}(\alpha)$ this can always be arranged by a simple rescaling of the data. $\square$

Theorem 6. *If $X \in \mathrm{DOA}(\alpha)$ for some $0 < \alpha < 2$ then*:

(i) $\hat{\beta}_n \xrightarrow{\mathrm{P}} 1/\alpha$;

(ii) $E(\hat{\beta}_n) \to 1/\alpha$;

(iii) $E(\hat{\beta}_n - 1/\alpha)^2 \to 0$; *and furthermore, for some $\tilde{c}_n \to 0$ we have*

$$2\ln n(\hat{\beta}_n - 1/\alpha - \tilde{c}_n) \Rightarrow \ln Y_0, \tag{2.13}$$

where Y_0 is $\alpha/2$ stable and $E \ln Y_0 = 0$ (if $X \in \mathrm{DONA}(\alpha, 1)$ we can take $\tilde{c}_n = 0$).

Proof. Kanter (1975) showed that a centered stable law $Y^{(2)}$ with index $\beta = \alpha/2$, skewness 1, and scale factor $\sigma_2 = 1$ is identically distributed with $(a(\theta)/W)^{(1-\beta)/\beta}$ where W is exponential with mean 1, $a(\theta) = \sin((1 - \beta)\theta)(\sin(\beta\theta))^{\beta/(1-\beta)}/(\sin\theta)^{1/(1-\beta)}$, and θ is uniform over $[0, \pi]$. Then it is easy to compute that $E \ln Y^{(2)} = \gamma(1 - \beta)/\beta$. (Several different parameterizations of the family of stable laws are commonly used. The dispersion C and scale factor σ of a β-stable law are related to the alternative scale factor σ_2 by $\sigma_2^\beta = \Gamma(1 - \beta)C = \sigma^\beta/\cos(\pi\beta/2)$, see for example, Weron, 1996.) Now, compute that $\hat{\beta}_n = A_n + B_n\hat{\gamma}_n$ where

$$A_n = \frac{\gamma}{2(\gamma + \ln n)} \to 0 \quad \text{and} \quad B_n = \frac{2\ln n}{(\gamma + \ln n)} \to 1$$

as $n \to \infty$. It follows easily that (ii) and (iii) hold. Apply Eq. (2.2) and define $Y_0 = Y^{(2)}/C$ where $\ln C = E \ln Y^{(2)}$. Then Y_0 is also centered stable with the same index, scale, and skewness and furthermore, $E \ln Y_0 = 0$. Note that the norming constants a_n can always be chosen so that the limit $Y^{(2)}$ in Eq. (2.2) has $\sigma_2 = 1$. (Altering a_n also changes c_n.) Now, $2(\gamma + \ln n)(\hat{\beta}_n - 1/\alpha - \tilde{c}_n) \Rightarrow \ln Y_0$ is algebraically equivalent to Eq. (2.2) where $\tilde{c}_n = c_n \ln n/(\gamma + \ln n) = \ln \ell_n/(\gamma + \ln n)$. If $X \in \mathrm{DONA}(\alpha, 1)$ then $\ell_n = 1$ so $\tilde{c}_n = 0$. Since $(\gamma + \ln n)/\ln n \to 1$ this is equivalent to Eq. (2.13). Consistency (i) follows immediately from (ii), (iii), or Eq. (2.13).

Finally, we obtain asymptotic confidence intervals for $1/\alpha$. We restrict our attention to the case $X \in \mathrm{DONA}(\alpha)$. Note that the confidence intervals are asymmetric, since the limit in Eq. (2.2) or Eq. (2.13) is skewed. Precise quantiles for stable distributions with skewness 1 can be obtained from the tables of McCulloch and Panton (1997).

Theorem 7. *Under the conditions of Theorem 1, for any $0 < q < p < 1$ we have*

$$P\left[\hat{\gamma}_n - \frac{\ln y_p}{2\ln n} \leqslant \frac{1}{\alpha} \leqslant \hat{\gamma}_n - \frac{\ln y_q}{2\ln n}\right] \to p - q \tag{2.14}$$

as $n \to \infty$ where $P[Y^{(2)} \leqslant y_q] = q$.

Proof. Rewrite the event in Eq. (2.14) as $\ln y_q \leqslant 2\ln n(\hat{\gamma}_n - \alpha^{-1}) \leqslant \ln y_p$ and use the weak convergence (2.2) to see that as $n \to \infty$ the probability of this event converges to $P[\ln y_q \leqslant \ln Y^{(2)} \leqslant \ln y_p] = P[y_q \leqslant Y^{(2)} \leqslant y_p] = p - q$, since the stable limit $Y^{(2)}$ has a density.

3. Comparison with Hill's estimator

Our purpose in this section is to evaluate the performance of our estimator against that of Hill's estimator in the context of real-data analysis applications. Hill (1975) considered the Pareto case $P[X > t] = Ct^{-\alpha}$ for $t \geqslant D$ where D is known and calculated that

$$\hat{H}_r = r^{-1} \sum_{i=1}^{r} \ln X_{(i)} - \ln X_{(r+1)} \tag{3.1}$$

is the conditional maximum likelihood estimator of $1/\alpha$ conditional on $X_{(r+1)} \geqslant D$ where $X_{(1)} \geqslant X_{(2)} \geqslant \cdots$ are the order statistics of a random sample $X_1 \cdots X_n$. For data which are approximately Pareto in the tail, one should choose r small enough so that only the Pareto-like tail is represented. A typical application of Hill's estimator to real-data involves 1000–3000 observations and estimates α using the upper 10% of the data or less, see, for example, Hill (1975), Jansen and de Vries (1991), Loretan and Phillips (1994), or Resnick and Stărică (1995). We consider three commonly used heavy tail distributions, and compare both estimators in each case. Our overall conclusion is that our estimator performs about as well as Hill's estimator in most cases, and substantially better in some cases involving stable data. We also consider a distribution which belongs to the domain of attraction of a stable law but not to the domain of normal attraction, and we explain why neither estimator works well in this case. Finally, we compare asymptotic confidence intervals for both estimators.

First, we performed 1000 simulation trials in which 1000 standard Pareto random variables with $\alpha = 1.5$ were generated, and computed both estimators. Using our estimator $1/\hat{\beta}_n$ we found a sample mean of 1.52, a sample standard deviation of 0.18, and an interquartile range of $[1.42, 1.65]$. Using $r = 50$ for Hill's estimator $1/\hat{H}_r$ we found a sample mean of 1.54, a sample standard deviation of 0.23, and an interquartile range of $[1.38, 1.67]$. Using $r = 100$ for Hill's estimator we found a sample mean of 1.48, a sample standard deviation of 0.14, and an interquartile range of $[1.38, 1.66]$. Repeated simulation for different values of α produced similar results. Our estimator is about as good as Hill's estimator in this case. (Note that if the distribution of the data is known to be Pareto, then $r = 999$ is optimal, which results in much better performance for Hill's estimator. We are assuming that the distribution is unknown, so that the experimenter will choose a value of r between 50 and 100 in order to capture the tail behavior. Also note that it is not necessary to simulate to obtain the distribution of Hill's estimator in the Pareto case, since it was computed by Hill, 1975.)

Hall (1982) calculated that the optimal value of r for Hill's estimator when applied to data which satisfy

$$P[X > t] = Ct^{-\alpha}(1 + O(t^{-\delta})) \tag{3.2}$$

for $\delta > 0$ is $r = n^{2\delta/(2\delta+\alpha)}$. One important case in which this formula is valid is the type II extreme value distribution $P[X > t] = 1 - \exp(-Ct^{-\alpha})$. A Taylor expansion shows that Eq. (3.2) holds with $\delta = \alpha$ and so the optimal value is $r = n^{2/3}$. We repeated
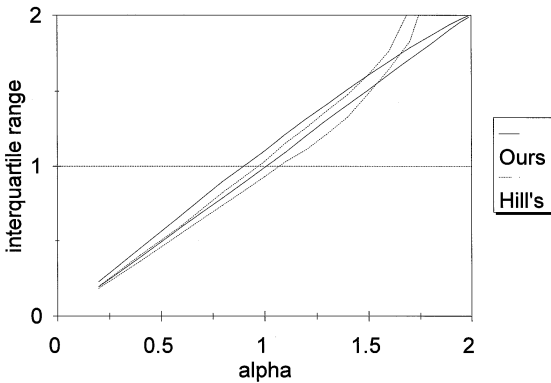
Fig. 1. Comparison of interquartile range.

the simulation of the above paragraph with 1000 trials of $n = 1000$ observations using $r = 100$ for Hill's estimator. In the case $\alpha = 0.5$ we found a sample mean of 0.49 and a sample standard deviation of 0.05 for Hill's estimator while the same simulation produced a sample mean of 0.51 and a sample standard deviation of 0.08 for our estimator. The interquartile ranges were [0.46, 0.57] for our estimator and [0.46, 0.53] for Hill's estimator. Repeating the simulation with $\alpha = 1.5$ yielded a sample mean of 1.48 and a sample standard deviation of 0.151 for Hill's estimator while the same simulation produced a sample mean of 1.55 and a sample standard deviation of 0.177 for our estimator. The interquartile ranges were [1.44, 1.68] for our estimator and [1.37, 1.57] for Hill's estimator. Hill's estimator performs a bit better in this case, but both estimators yield good results.

McCulloch (1997) applies Hill's estimator to simulated stable data, using the simulation method of Chambers et al. (1976). He performs 100 trials with $n = 3000$ and $r = 50$ for various values of α and tabulates the results. These values are chosen to typify several econometric applications including Loretan and Phillips (1994). We repeated that simulation with 100 trials of $n = 3000$ for our estimator and compared to the results of McCulloch for Hill's estimator. Fig. 1 compares the simulated interquartile range for both estimators. It is apparent that Hill's estimator has a significant bias for $1.5 < \alpha < 2$ while our estimator becomes increasingly accurate as α approaches 2. For example, when $\alpha = 1.8$ the median value of Hill's estimator as calculated by McCulloch is 2.45 and the interquartile range is [2.21, 2.85]. For our estimator the median is 1.84 and the interquartile range is [1.80, 1.87]. For $\alpha = 1.9$ the corresponding values are 3.59 and [3.16, 4.00] for Hill's estimator, and 1.93 and [1.91, 1.94] for our estimator. Loretan and Phillips argue that certain US stock returns and currency exchange rate returns cannot be modeled by a stable distribution because Hill's estimator yields values of α significantly greater than 2 (for a stable law only $0 < \alpha \leqslant 2$ are allowed). McCulloch counters that Hill's estimator is significantly biased in this case, yielding estimates of α greater than 2 when the true α is less than 2. Hall (1982) notes that the value of r in Hill's estimator should be chosen to balance between bias and

accuracy. A larger value of r yields more accuracy (i.e. a smaller variance) but larger bias. For stable data with α near 2.0, the choice of $r = 50$ results in an unacceptably large bias. Our estimator produces much better results than Hill's estimator in this case, and it does not require choosing an appropriate number of order statistics r.

We also compared the two estimators in the case where $X > 0$ belongs to a stable domain of attraction but Eq. (3.2) does not apply. Writing $G(t) = P[X > t]$ we have from LePage et al. (1981) that the ith largest order statistic $X_{(i)}$ from a sample of size n is identically distributed with $G^{-1}(\Gamma_i / \Gamma_{n+1})$ where $\Gamma_i = Z_1 + \cdots + Z_i$ and Z_i are iid standard exponential. We set $G^{-1}(t) = t^{-1/\alpha}(-\ln t)$ and repeated the simulation described above using $\alpha = 1.5$ and $n = 1000$. The surprising result was that both estimators produced estimates of α that were consistently low, generally between 1.0 and 1.1. The reason for this anomaly can be understood in terms of the LePage representation. The dominant order statistics correspond to the smallest indices i, for which Γ_i / Γ_{n+1} is on the order of $\frac{1}{1000}$. In the range $0.0005 < t < 0.01$ the graph of G^{-1} lies between that of t^{-1} and $t^{-1/1.1}$ and all three graphs are fairly close together. The data are consistent with the model $P[X > t] = Ct^{-\alpha}$ for $\alpha \approx 1.1$ and no α estimator is likely to perform much better than ours in this case. Data analysis will typically give poor estimates of α in cases where Eq. (3.2) fails to hold. This is also true for stochastically compact data, using the series representation of Meerschaert and Scheffler (1996).

Finally, we compare asymptotic confidence bands for both estimators. Assume that Eq. (3.2) holds with $\delta = \alpha$ as in the case of the type II extreme value distribution or the stable distribution. Hall (1982) shows that in this case $\alpha \sqrt{r}(\hat{H}_r - 1/\alpha) \Rightarrow Z$ standard normal as $r \to \infty$ with $r/n^{2/3} \to 0$. Then with probability approximately $1 - 2q$ we will have that $z_q/(\alpha \sqrt{r}) \leqslant \hat{H}_r - 1/\alpha \leqslant z_{1-q}/(\alpha \sqrt{r})$ where $P[Z \leqslant z_q] = q$. Theorem 6 in the previous section shows that in this case $2\ln n(\hat{\beta}_n - 1/\alpha) \Rightarrow \ln Y_0$ where Y_0 is stable with index $\alpha/2$, skewness 1, $\sigma_2 = 1$, and $E \ln Y_0 = 0$. Then with probability approximately $1 - 2q$ we will have that $y_q/(2\ln n) \leqslant \hat{\beta}_n - 1/\alpha \leqslant y_{1-q}/(2\ln n)$ where $P[Y_0 \leqslant y_q] = q$. The quantiles of Y can be obtained from the tables of McCulloch and Panton (1997). Setting $\alpha = 1.5$ and $r = n^{2/3}$ (the optimal rate for Hill's estimator in this case) we compare 90% confidence bands for both estimators as a function of the sample size n. The results are illustrated in Fig. 2, where we plot the confidence bands in terms of α for readability. Once again we conclude that our estimator is about as good as Hill's estimator for typical sample sizes. Indeed, the skewness of our estimator is an advantage when n is small, allowing a one-sided test to reject $\alpha = 2$ at the 95% level even for reasonably small samples.

4. Remarks

The main open problem in estimating the tail thickness parameter α is robustness. If we assume that the data are distributed according to some known heavy tail distribution like the generalized Pareto, stable, or type II extreme value distribution then the maximum likelihood estimator (MLE) can be used to estimate the distributional
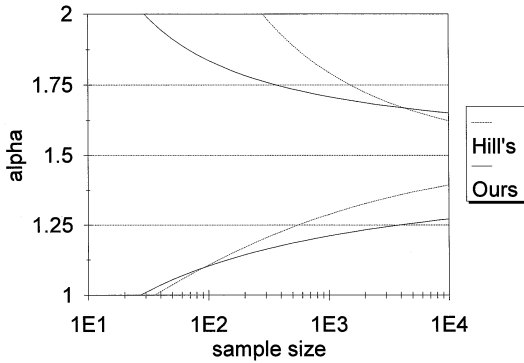
Fig. 2. Comparison of 90% confidence bands.

parameters. The MLE uses all of the data, unlike Hill's estimator which only uses the largest order statistics. For type II extreme value distributions the MLE computation is standard. For generalized Pareto distributions the log-likelihood function has a singularity, but the MLE can still be computed as a local maximum, see for example, Hosking and Wallis (1987). For stable laws the problem is complicated by the fact that the densities cannot be written in closed form. Nolan (1997) has solved this problem using the integral representation of Zolotarev (1986) for the stable density.

The MLE is closely tailored to a particular distributional form, and there is no reason to expect robustness against a much wider collection of distributions. Hence, there is a tradeoff between robustness and efficiency. The $\ln n$ rate of convergence for our estimator is slow, and this seems unavoidable. Hall (1982) remarks that if we only assume $X \in \mathrm{DOA}(\alpha)$ for $0 < \alpha < 2$ then it is not possible to obtain a rate of convergence faster than $\ln n$ for Hill's estimator. Hall (1982) also shows that if Eq. (3.2) holds for some $\delta > 0$, then $\sqrt{r}(\hat{H}_r - 1/\alpha)$ is asymptotically normal as $n \to \infty$, where $r = n^{2\delta/(2\delta+\alpha)}$ yields the optimal rate of convergence. The second-order condition (3.2) is stronger than assuming $X \in \mathrm{DONA}(\alpha)$, and if $\delta > 0$ is small then even for $X \in \mathrm{DONA}(\alpha)$ the rate of convergence is very slow. The comparison in Fig. 2 assumes $\delta = \alpha$. If $\delta < \alpha$ then our estimator compares even more favorably to Hill's estimator.

Quite a few different methods for estimating the tail index α appear in the literature. Many of these assume a particular distributional form, for example, the estimators of Nikias and Shao (1995) and McCulloch (1986) for stable laws or Hosking and Wallis (1987) for generalized Pareto distributions. Like the MLE, these estimators use all of the data. Several other estimators are based on the largest order statistics. Dekkers et al. (1989) present a refinement of Hill's estimator based on the first and second sample moments of $\ln X_i$ using the largest order statistics. See also Drees (1995), Falk (1995) and Wei (1995) and others referenced there. We have not attempted a systematic comparison of our method with any of these alternative tail estimators. These estimators obtain their robustness from the asymptotics of extreme values, and they only use a vanishingly small fraction of the data. Our estimator obtains robustness from the central limit theorem for heavy tail sums, using all of the data. One advantage

of our estimator is that it does not require calculating how many order statistics should be used.

Various extensions and refinements of the asymptotic results presented in this paper are also possible. For example, it is straightforward to obtain convergence results for $\hat{\alpha}_n = 1/\hat{\gamma}_n$ using Theorem 1 above. The convergence $\hat{\alpha}_n \to \alpha$ in probability is immediate, and the delta method yields $2\ln n(\hat{\alpha}_n - \alpha - d_n) \Rightarrow -\alpha^2 \ln Y^{(2)}$ where $\alpha^{-1} + c_n = (\alpha + d_n)^{-1}$ so that $d_n \to 0$. It is also possible to apply our estimator in cases where we suspect that $2 < \alpha < 4$, as in the applications of Jansen and de Vries (1991) or Loretan and Phillips (1994). If $P[|X| > x]$ varies regularly with index α then $X^2 \in \text{DOA}(\alpha/2)$, see for example, Mandrekar and Meerschaert (1994). So if X has heavy tails with index $2 < \alpha < 4$, we can estimate the tail index by applying our estimation procedure to the squared data.

References

Anderson, P., Meerschaert, M., 1997. Periodic moving averages or random variables with regularly varying tails. Ann. Statist. 25, 771–785.

Bingham, N., Goldie, C., Teugels, J., 1987. Regular variation, Encyclopedia of Mathematics and its Applications, vol. 27. Cambridge University Press, Cambridge.

Brockwell, P., Davis, R., 1991. Time Series: Theory and Methods, 2nd ed. Springer, New York.

Chambers, J., Mallows, C., Stuck, B., 1976. A method for simulating stable random variables. J. Amer. Statist. Assoc. 71, 340–344.

Davis, R., Resnick, S., 1985. Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. 13, 179–195.

Dekkers, A., Einmahl, J., de Haan, L., 1989. A moment estimator for the index of an extreme value distribution. Ann. Statist. 17, 1833–1855.

Drees, H., 1995. Refined Pickands estimators of the extreme value index. Ann. Statist. 23, 2013–2035.

Falk, M., 1995. On testing the extreme value index via the POT-method. Ann. Statist. 23, 2013–2035.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications, vol. II, 2nd ed. Wiley, New York.

Hall, P., 1982. On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. B 44, 37–42.

Heyde, C., 1967. On large deviation problems for sums of random variables which are not attracted to the normal law. Ann. Math. Statist. 38, 1575–1578.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 1163–1173.

Hosking, J., Wallis, J., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339–349.

Janicki, A., Weron, A., 1994. Can one see α stable variables and processes? Statist. Sci. 9, 109–126.

Jansen, D., de Vries, C., 1991. On the frequency of large stock market returns: putting booms and busts into perspective. Rev. Econ. Statist. 23, 18–24.

Kanter, M., 1975. Stable densities under change of scale and total variational inequalities. Ann. Probab. 3, 697–707.

Leadbetter, M., Lindgren, G., Rootzén, H., 1980. Extremes and Related Properties of Random Sequences and Processes. Springer, New York.

LePage, R., Woodroofe, M., Zinn, J., 1981. Convergence to a stable distribution via order statistics. Ann. Probab. 4, 624–632.

Loretan, M., Phillips, P., 1994. Testing the covariance stationarity of heavy tailed time series. J. Empirical Finance 1, 211–248.

Mandelbrot, B., 1982. The Fractal Geometry of Nature. W.H. Freeman, San Francisco.

Mandrekar, V., Meerschaert, M., 1994. Sample moments and symmetric statistics. Stochastic Analysis on Infinite Dimensional Spaces, Pitman Research Notes in Mathematics, vol. 310. Pitman, London, pp. 197–210.

McCulloch, J., 1986. Simple consistent estimators of stable distribution parameters. Comm. Statist. Simul. Comput. 15, 1109–1136.

McCulloch, J., 1997. Measuring tail thickness to estimate the stable index α: a critique. J. Business Econ. Statist. 15, 74–81.

McCulloch, J., Panton, D., 1997. Precise tabulation of the maximally-skewed stable distributions and densities. Comput. Statist. Data Anal. 23, 307–320.

Meerschaert, M., Scheffler, H.P., 1996. Series representation for semistable laws and their domains of semistable attraction. J. Theoretical Probab. 9, 931–959.

Mittnik, S., Rachev, S., 1995. Modelling Financial Assets with Alternative Stable Models. Wiley, New York.

Nikias, C., Shao, M., 1995. Signal Processing with Alpha Stable Distributions and Applications. Wiley, New York.

Nolan, J., 1997. Maximum likelihood estimation of stable parameters, preprint.

Resnick, S., Stărică, C., 1995. Consistency of Hill's estimator for dependent data. J. Appl. Probab. 32, 139–167.

Scheffler, H.P., 1994. Domains of semi-stable attraction of nonnormal semi-stable laws. J. Multivariate Anal. 51, 432–444.

Shimizu, R., 1970. On the domain of partial attraction of semi-stable distributions. Ann. Inst. Statist. Math. 22, 245–255.

Wei, X., 1995. Asymptotically efficient estimation of the index of regular variation. Ann. Statist. 23, 2036–2058.

Weron, R., 1996. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. Stat. Probab. Lett. 28, 165–171.

Zolotarev, V., 1986. One-Dimensional Stable Distributions, Translations of Mathematical Monographs, vol. 65. American Mathematical Society, Providence, Rhode Island.