Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1389–1406
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Moment-based tail index estimation

Tucker McElroy[a], Dimitris N. Politis[b,*]

[a] *US Bureau of the Census, University of California, San Diego, USA*
[b] *University of California, San Diego, USA*

Received 20 June 2003; accepted 28 April 2006
Available online 2 June 2006

Abstract

A general method of tail index estimation for heavy-tailed time series, based on examining the growth rate of the logged sample second moment of the data was proposed and studied in Meerschaert and Scheffler (1998. A simple robust estimator for the thickness of heavy tails. J. Statist. Plann. Inference 71, 19–34) as well as Politis (2002. A new approach on estimation of the tail index. C. R. Acad. Sci. Paris, Ser. I 335, 279–282). To improve upon the basic estimator, we introduce a scale-invariant estimator that is computed over subsets of the whole data set. We show that the new estimator, under some stronger conditions on the data, has a polynomial rate of consistency for the tail index. Empirical studies explore how the new method compares with the Hill, Pickands, and DEdH estimators.

Keywords: Heavy tails; Domains of attraction; Subsampling

1. Introduction

Let $X_1, \ldots, X_n$ be an observed stretch of a linear-dependent time series satisfying

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j} \tag{1}$$

for all $t \in \mathbb{Z}$, where $\{Z_t\}$ is *iid* (independent and identically distributed) from some continuous distribution F; the case where $\psi_j = 0$ for $j \neq 0$ is the special case of $\{X_t\}$ being *iid*, and is considered in detail throughout Section 2. We assume that F belongs to $D(\alpha)$, the domain of attraction of an α-stable law; however, the heavy-tail index α is unknown and must be estimated from the data. In this context, there exist sequences a_n and b_n such that $a_n^{-1}(\sum_{t=1}^{n} Z_t - b_n) \stackrel{\mathscr{L}}{\Longrightarrow} S_\alpha$, where S_α denotes a generic α-stable law with unspecified scale, location and skewness, and $\alpha \in (0, 2)$; it is always true that we can write $a_n = n^{1/\alpha} L(n)$ for some slowly varying function L. If L is either constant or asymptotically tends to a nonzero constant, we say that F is in the normal domain of attraction, denoted by $\mathrm{ND}(\alpha)$. We restrict to the case that $\alpha < 2$ to ensure that the variance of the data is always infinite.

Estimators of the tail index are often constructed from extreme order statistics—see Csörgő et al. (1985) for a general class of such estimators. The well-known Hill estimator H_q falls into this class. A challenging problem lies in

* Corresponding author.
E-mail address: politis@math.ucsd.edu (D.N. Politis).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.04.002

choosing the number of order statistics q to be used in practice; see Embrechts et al. (1997) and the references therein for a discussion of this topic. The work of Politis (2002) presents an alternative estimation approach that is based on empirically examining the growth of appropriately chosen diverging statistics. A prime example of such a statistic is given by the sample variance that diverges to infinity in the absence of a finite second moment. As a matter of fact, a consistent—albeit with logarithmic rate—tail index estimator can be constructed by simply taking the ratio of the logarithm of the sample variance to the logarithm of the sample size; see Meerschaert and Scheffler (1998). For a survey on nonparametric methods for heavy-tailed data, see Meerschaert and Scheffler (2003); these authors also extend their methodology to heavy-tailed random vectors.

In this paper, a new class of tail index estimators are presented, which are in the same philosophy as the estimators of Meerschaert and Scheffler (1998) and Politis (2002); in particular, they do not rely on extreme order statistics. The main estimator is an effort to improve on the convergence rate of the tail index estimator of Meerschaert and Scheffler (1998), hereafter referred to as the MS statistic. Our approach is simple and intuitive, and can also be generalized to rate estimation settings other than the heavy-tail problem. We also propose some ways to improve upon the basic form of the new estimators, and give some finite-sample simulation results. Section 2 covers the theoretical results that establish the asymptotics of the tail index estimator, while Section 3 deals with the more practical issues of how to conduct inference for α, and presents the result of several simulation studies. All proofs are placed in the Appendix.

2. Theory

2.1. Motivation for the new approach

The theory developed in this section motivates the construction of the tail index estimators that we explore empirically in Section 3. In order to facilitate simple proofs, the results in this section are presented for *iid* data, i.e., the case $\psi_j = 0$ for $j \neq 0$ in (1). Some notations that are consistently used in this paper: $\mathbb{E}$ denotes the expectation operator, whereas $\mathbb{D}$ is the variance operator (D for dispersion). By the notation $\mathbb{D}[A, B]$, we denote the covariance between variables A and B. Also, when we write a random variable without a subscript, we indicate a common version.

Let us define the sum of squares process $S_n(X^2) = \sum_{t=1}^n X_t^2$. It is well-known (see, for example Theorem 4.2 of Davis and Resnick (1985) for the $MA(\infty)$ case) that $S_n(X^2)$ diverges (when $\alpha < 2$) at rate a_n^2, i.e., the normalized partial sums of squares $U_n = a_n^{-2} S_n(X^2)$ converge in distribution to a nondegenerate random variable. Since $a_n = n^{1/\alpha} L(n)$, the rate of divergence of $S_n(X^2)$ may give crucial information about α. So define $\zeta = 1/\alpha$ and $M = L^2$, and consider the identity

$$\log S_n(X^2) = 2\zeta \log n + \log M(n) + \varepsilon_n, \tag{2}$$

where the random variables

$$\varepsilon_n = \log(a_n^{-2} S_n(X^2))$$

can be thought of as "residuals" in a regression of $\log S_n(X^2)$ on $\log n$. This is the basic motivation behind the regression estimator of Politis (2002), as well as the MS statistic:

$$\hat{\zeta}_n^{\mathrm{MS}} = \log^+ S_n(X^2)/(2 \log n).$$

Here $\log^+ x = \max\{\log x, 0\}$. This differs from the statistic defined by Meerschaert and Scheffler (1998) in that the sample second moments rather than the sample variance is computed; the centering makes no difference to asymptotics when $\alpha < 2$. In this paper we shall use log rather than $\log^+$; the latter function has the advantage of disregarding the behavior of X^2 near zero, whose negative values of $\log X^2$ we essentially explore. However, the logarithm allows for the intuitive decomposition given by (2). As suggested in Meerschaert and Scheffler (1998), these estimators can be easily adapted to handle the case that $\alpha \geqslant 2$ by computing fourth or six sample moments; thus, we define

$$\hat{\zeta}_n^{\mathrm{BAS}r} = \log S_n(X^{2r})/(2r \log n),$$

where the integer r is taken sufficiently large such that the $2r$th moment of the data's distribution does not exist. The notation denotes a "basic" estimator, dependent on a user-defined integer r. The following result is similar to

Lemma 2.1 of Politis (2002), but goes deeper in producing convergence of certain moments, which will be important for our later results.

Proposition 1. *Assume that the probability density function f_X of X exists and is bounded, and also that $f_X(x)/x$ is bounded in a neighborhood of zero. Also assume that the distribution $F \in D(\alpha)$ for some $\alpha \in (0, 2)$, and that $\{X_t\}$ are independent. Defining the normalized sample second moments by $U_n = a_n^{-2} S_n(X^2)$, then $U_n \overset{\mathcal{L}}{\Longrightarrow} U$ and $\log U_n \overset{\mathcal{L}}{\Longrightarrow} \varepsilon$ as $n \to \infty$, where U has the distribution of an $\alpha/2$ positively skewed stable random variable, and ε has the distribution of $\log U$. The random variable U has scale parameter C depending on α and the scale of the original data. In addition, for any $p < \alpha/2$, we have*

$$\begin{aligned} &\mathbb{E}[U_n^p] \to \mathbb{E}[U^p], \\ &\mathbb{E}[\varepsilon_n^j] \to \mathbb{E}[\varepsilon^j], \end{aligned} \tag{3}$$

for $j = 1, 2$. Moreover, the probability density functions (pdfs) of the random variables U_n exist and converge uniformly to the pdf of U.

Remark 1. The weak convergences in this proposition are also true for the MA(∞) case (as demonstrated in Davis and Resnick, 1985), but the uniform integrability results (3) have not been established by the authors for the case when serial correlation is present. The last statement of the theorem is not new, but is stated for enunciation.

There are many distributions in $D(\alpha)$—these distributions are characterized by their regularly varying tail probabilities, as discussed in Theorem 2.2.8 of Embrechts et al. (1997). Thus, Proposition 1 covers a wide range of distributions. As pointed out in Meerschaert and Scheffler (1998), the trade-off of this nonparametric approach is a slow convergence rate for the estimator $\hat{\zeta}^{\mathrm{MS}}$. This paper investigates some variants of $\hat{\zeta}^{\mathrm{MS}}$ and $\hat{\zeta}^{\mathrm{BAS1}}$ in the hopes of obtaining a faster convergence rate.

2.2. *Some growth estimators*

We first investigate the properties of $\hat{\zeta}^{\mathrm{BAS1}}$; these are quite similar in spirit to results obtained for $\hat{\zeta}^{\mathrm{MS}}$ in Meerschaert and Scheffler (1998), but we use log instead of $\log^+$.

Theorem 1. *Under the same assumptions as Proposition 1, we have the following convergences as $n \to \infty$:*

$$\begin{aligned} &\hat{\zeta}_n^{\mathrm{BAS1}} \xrightarrow{P} \zeta, \\ &\mathbb{E}\hat{\zeta}_n^{\mathrm{BAS1}} \to \zeta, \\ &\mathrm{MSE}(\hat{\zeta}_n^{\mathrm{BAS1}}) \to 0, \\ &2\log n(\hat{\zeta}_n^{\mathrm{BAS1}} - \zeta - c_n) \overset{\mathcal{L}}{\Longrightarrow} \varepsilon, \end{aligned}$$

for some sequence $c_n \to 0$ (if $F \in \mathrm{ND}(\alpha)$, we can take c_n equal to a constant divided by $\log n$). As usual, $\mathrm{MSE}(\hat{\zeta}) = \mathbb{E}[\hat{\zeta} - \zeta]^2$. The variable ε is defined in Proposition 1.

The proof of Theorem 1 reveals that the bias of $\hat{\zeta}^{\mathrm{BAS1}}$ decreases quite slowly; by altering the estimator, we hope to produce a statistic with smaller bias. Toward that end, consider the following simple "centered" estimator of ζ:

$$\hat{\zeta}_n^{\mathrm{CEN}} = \frac{\log S_n(X^2) - \log S_{\sqrt{n}}(X^2)}{2\log\sqrt{n}}. \tag{4}$$

Note that

$$\hat{\zeta}_n^{\mathrm{CEN}} = \zeta + \frac{\log(M(n)/M(\sqrt{n}))}{\log n} + \frac{\varepsilon_n - \varepsilon_{\sqrt{n}}}{\log n}$$

by Eq. (2). The differencing may reduce the bias of the BAS1 estimator.[1] To see this, observe that

$$\mathrm{Bias}(\hat{\zeta}_n^{\mathrm{BAS1}}) = \frac{\log M(n)}{2\log n} + \frac{\mathbb{E}\varepsilon_n}{2\log n},$$

$$\mathrm{Bias}(\hat{\zeta}_n^{\mathrm{CEN}}) = \frac{\log(M(n)/M(\sqrt{n}))}{\log n} + \frac{\mathbb{E}\varepsilon_n - \mathbb{E}\varepsilon_{\sqrt{n}}}{\log n}.$$

To see why this may be helpful, suppose for concreteness that $F \in \mathrm{ND}(\alpha)$ and that $|\mathbb{E}[\varepsilon] - \mathbb{E}[\varepsilon_n]| = \mathrm{O}(n^{-p})$ for some $p > 0$; then it is easy to show that $\mathrm{Bias}(\hat{\zeta}_n^{\mathrm{CEN}}) = \mathrm{O}(n^{-p/2}/\log n)$ whereas $\mathrm{Bias}(\hat{\zeta}_n^{\mathrm{BAS1}}) = \mathrm{O}(1/\log n)$. Thus, we may expect $\hat{\zeta}^{\mathrm{CEN}}$ to have superior bias properties in small samples. As for the overall mean squared error, in the $\mathrm{ND}(\alpha)$ case we have

$$\mathrm{MSE}(\hat{\zeta}_n^{\mathrm{BAS1}}) = \frac{1}{4(\log n)^2}\mathbb{E}[\varepsilon_n^2],$$

$$\mathrm{MSE}(\hat{\zeta}_n^{\mathrm{CEN}}) = \frac{1}{4(\log\sqrt{n})^2}\mathbb{E}[(\varepsilon_n - \varepsilon_{\sqrt{n}})^2].$$

Now $\mathbb{E}[(\varepsilon_n - \varepsilon_{\sqrt{n}})^2]$ is difficult to analyze, but may well be smaller than $\mathbb{E}[\varepsilon_n^2]$ for small and large samples. In the case where $F \notin \mathrm{ND}(\alpha)$, the bias $\log(M(n)/M(\sqrt{n}))/\log n$ of $\hat{\zeta}^{\mathrm{CEN}}$ may be smaller than the corresponding term $\log M(n)/2\log n$ for $\hat{\zeta}^{\mathrm{BAS1}}$. For example, if $M(x) = \log x$, the respective biases are

$$\frac{\log 2}{\log n} \quad \text{and} \quad \frac{\log\,\log n}{2\log n}$$

showing that the bias of $\hat{\zeta}^{\mathrm{CEN}}$ is of smaller order. Of course, there are many choices of $M(n)$ for which this does not hold.

Since the logarithm of any slowly-varying function is $\mathrm{o}(\log n)$, it is easy to see that $\hat{\zeta}^{\mathrm{CEN}}$ will be consistent for ζ as $n \to \infty$. Now to further optimize this estimator's properties, we chop the data $\{X_1, \ldots, X_n\}$ into nonoverlapping blocks of size b^2, where b is to be determined. Then $\hat{\zeta}_{b^2}^{\mathrm{CEN}}$ is computed over each separate block, and finally is averaged across all $K = \lfloor n/b^2 \rfloor$ blocks. Let the $\hat{\zeta}_{b^2}^{\mathrm{CEN}}$ statistic evaluated on data points with indices $\{(k-1)b^2+1, \ldots, kb^2\}$ be denoted by $\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}$ for $k = 1, \ldots, K$; then our block-smoothed estimator is

$$\hat{\zeta}_{b,n}^{\mathrm{SCEN}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}.$$

Under our assumption of independence on the $\{X_t\}$ series, the variance will be greatly reduced from the averaging. The following theorem describes the asymptotic properties (as both b and $n \to \infty$) of the block-smoothed estimator $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$.

Theorem 2. *Under the same assumptions as Proposition* 1, *the following conclusions hold as* $b \to \infty$:

(i) *The squared bias of each block estimator* $\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}$ *tends to zero and the variance of* $\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}$ *is* $\mathrm{O}((\log b)^{-2})$.

(ii) *The squared bias of the block-smoothed estimator* $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$ *is* $\mathrm{o}(1)$ *as* b *and* n *tend to infinity, and its variance is* $\mathrm{O}(b^2 n^{-1}(\log b)^{-2})$ *as* b *and* n *tend to infinity.*

(iii) *The MSE of* $\hat{\zeta}_{b,n}^{\mathrm{SCEN}} \to 0$ *as* $b, n \to \infty$, *so long as* b *is chosen such that* $b/(\log b\sqrt{n}) \to 0$.

Remark 2. This bound on the error is not sharp, and in some special cases it can be substantially improved—for example, if we assume that $F \in \mathrm{ND}(\alpha)$; see Theorem 3 of Section 2.4 below. In practice, b must be small relative to n, e.g., $b = n^{1/3}$.

[1] This question, under the further assumption that $F \in \mathrm{ND}(\alpha)$ is explored in Section 2.4.

2.3. Variance reduction via additional smoothing

In the construction of $\hat{\zeta}_b$, data points $X_1, \dots, X_{b^2}$ were compared with the data points $X_1, \dots, X_b$; however, this was a somewhat arbitrary choice, since we could have used $X_{b+1}, \dots, X_{2b}$ or $X_{2b+1}, \dots, X_{3b}$ and so forth for the secondary subsample. A block of size b^2 contains precisely b disjoint blocks of size b. If we smooth over these additional smaller blocks, a reduction of the MSE of $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$ can be achieved; in addition, this innovation should be robust against data anomalies arising from the subjectivity inherent in selecting only the first block.

For any integer n and series $\{A_t\}$, we define

$$S_n^{(j)}(A) = \sum_{t=(j-1)n+1}^{jn} A_t \tag{5}$$

for $j = 1, 2, \dots, n$, and define a smoothed block estimator by

$$\hat{\zeta}_{b^2}^{\mathrm{RCEN}} = \frac{1}{b} \sum_{j=1}^{b} \frac{\log S_{b^2}(X^2) - \log S_b^{(j)}(X^2)}{\log b}. \tag{6}$$

Let the statistic $\hat{\zeta}_{b^2}^{\mathrm{RCEN}}$ evaluated on data points $\{(k-1)b^2 + 1, \dots, kb^2\}$ be denoted by $\hat{\zeta}_{b^2}^{R\mathrm{CEN}(k)}$; then our resulting estimator is

$$\hat{\zeta}_{b,n}^{\mathrm{SRCEN}} = \frac{1}{K} \sum_{k=1}^{K} \hat{\zeta}_{b^2}^{R\mathrm{CEN}(k)}.$$

The following proposition is useful for determining the second-moment asymptotics of our estimators of ζ.

Proposition 2. *Under the same assumptions as Proposition* 1, *the following results hold*:

$$\mathbb{E}[(\varepsilon_{b^2} - \varepsilon_b)^2] \to 2\mathbb{D}[\varepsilon],$$

$$\mathbb{D}[\varepsilon_{b^2}, \varepsilon_b] \to 0,$$

$$\varepsilon_{b^2} - \varepsilon_b \overset{\mathscr{L}}{\Longrightarrow} Y_1 - Y_2$$

as $b \to \infty$. *The random variables* Y_1 *and* Y_2 *have the same distribution as* ε *given in Proposition* 1, *and are independent*.

Remark 3. Note that ε_{b^2} and ε_b are asymptotically uncorrelated. An intuitive explanation for this is that the proportion of random variables common to both quantities is $1/b$, which tends to zero as $b \to \infty$.

As a corollary to this proposition, we have the finite sample and asymptotic variances of the basic (BAS1), centered (CEN), and robust centered (RCEN) estimators:

Corollary 1. *Under the same assumptions as Proposition* 1, *we have the following formulas for the variance of our estimators*:

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{BAS1}}] = \frac{\mathbb{D}[\varepsilon_{b^2}]}{4(\log b^2)^2},$$

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{CEN}}] = \frac{\mathbb{D}[\varepsilon_{b^2} - \varepsilon_b]}{4(\log b)^2},$$

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{RCEN}}] = \frac{\mathbb{D}[\varepsilon_{b^2} - \varepsilon_b] - (1 - 1/b)\mathbb{D}[\varepsilon_b]}{4(\log b)^2}.$$

Asymptotically, as $b \to \infty$, these become

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{BAS1}}] \sim \frac{\mathbb{D}[\varepsilon]}{16(\log b)^2},$$

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{CEN}}] \sim \frac{\mathbb{D}[\varepsilon]}{2(\log b)^2},$$

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{RCEN}}] \sim \frac{\mathbb{D}[\varepsilon]}{4(\log b)^2}.$$

This result indicates that $\hat{\zeta}_{b,n}^{\mathrm{SRCEN}}$ is preferable to $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$ in terms of variance. Their bias is of course identical. Both have more variance than $\hat{\zeta}_{n}^{\mathrm{BAS1}}$, though the order of the variance is the same. Hence, the focus of our empirical studies in Section 3 will be an MSE comparison of SRCEN and BAS1.

2.4. Stronger assumptions and bias reduction

Under some stronger assumptions on F, it is possible to prove that the bias of the CEN and RCEN estimators can be dramatically superior to the bias of BAS1, as alluded to in the discussion after Theorem 1. In particular, we will obtain a polynomial rate of decay for these new estimators, which is superior to the logarithmic rate of BAS1.

Proposition 3. *Suppose that the assumptions of Proposition* 1 *hold, and also assume that $F \in \mathrm{ND}(\alpha)$. Further suppose that the cdf F_{X^2} and F_U satisfy*

$$\int_0^\infty |F_{X^2}(x) - F_U(x)|\,\mathrm{d}x < \infty. \tag{7}$$

Then there exists $p > 0$ such that, as $n \to \infty$,

$$\mathbb{E}[\varepsilon_n] - \mathbb{E}[\varepsilon] = \mathrm{O}(n^{-p}).$$

The proposition does not assert that this is the optimal rate; determination of the optimal rate is a difficult problem, related to the theory of local limit theorems. Putting everything together, the following theorem gives the asymptotic orders of bias and variance for the BAS1, SCEN, and SRCEN estimators. For wise choices of the block size b, as a function of n, a polynomial rate of convergence is obtained for the MSE of the latter two estimators.

Theorem 3. *Under the same assumptions as Proposition* 3, *the asymptotic biases of* BAS1, CEN, *and* RCEN *are given by*

$$\mathrm{Bias}(\hat{\zeta}_{b^2}^{\mathrm{BAS1}}) = \frac{\mathbb{E}[\varepsilon_{b^2}]}{4\log b} = \mathrm{O}(1/\log b),$$

$$\mathrm{Bias}(\hat{\zeta}_{b^2}^{\mathrm{CEN}}) = \frac{\mathbb{E}[\varepsilon_{b^2} - \varepsilon_b]}{2\log b} = \mathrm{O}(1/b^p \log b),$$

$$\mathrm{Bias}(\hat{\zeta}_{b^2}^{\mathrm{RCEN}}) = \frac{\mathbb{E}[\varepsilon_{b^2} - \varepsilon_b]}{2\log b} = \mathrm{O}(1/b^p \log b).$$

The asymptotic biases of $\hat{\zeta}_{n}^{\mathrm{BAS1}}$, $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$, and $\hat{\zeta}_{b,n}^{\mathrm{SRCEN}}$ are given by

$$\mathrm{Bias}(\hat{\zeta}_{n}^{\mathrm{BAS1}}) = \frac{\mathbb{E}[\varepsilon_n]}{2\log n} = \mathrm{O}(1/\log n),$$

$$\mathrm{Bias}(\hat{\zeta}_{b,n}^{\mathrm{SCEN}}) = \frac{\mathbb{E}[\varepsilon_{b^2} - \varepsilon_b]}{2\log b} = \mathrm{O}(1/b^p \log b),$$

$$\mathrm{Bias}(\hat{\zeta}_{b,n}^{\mathrm{SRCEN}}) = \frac{\mathbb{E}[\varepsilon_{b^2} - \varepsilon_b]}{2\log b} = \mathrm{O}(1/b^p \log b).$$

Given the asymptotic variances for these estimators stated in Corollary 1, *the choice of block size b equal to a constant times* $n^{1/2(1+p)}$ *gives an* MSE of *order*

$$\frac{1}{n^{p/(p+1)}(\log n)^2}$$

for SCEN *and* SRCEN, *whereas* $\mathrm{MSE}(\hat{\zeta}_n^{\mathrm{BAS1}}) = \mathrm{O}(1/(\log n)^2)$. *Here*, $p > 0$ *is any number satisfying the conclusions of Proposition* 3.

Remark 4. This theorem tells us that, for an appropriate choice of b, SCEN and SRCEN have MSE converging to zero at polynomial rate, which is superior to BAS1. This rate $b = n^{1/2(1+p)}$ (up to a constant) is obtained by equating the asymptotic order of variance and squared bias. Note that this is not the optimal rate, since the order of convergence in Proposition 3 is not sharp. More practically, it is difficult to see how p may be determined from the data. Since an unknown constant also figures into this result, it is imperative to empirically investigate the choice of b, relative to n, that produces the lowest MSE. These issues are investigated in Section 3. By examination of the proof of Proposition 3, one might conjecture that smaller values of α allow one to take p larger, whereas lighter tails force a smaller value of p. This is borne out, to some extent, in the studies of Section 3.

2.5. Moving averages

Here we return to our original linear model (1), in order to indicate which results can be easily extended from the *iid* case. It is quite possible that all the previous results are valid for infinite order moving averages, but we only discuss those results that are easily proved.

Theorem 4. *Assume that* $F \in D(\alpha)$ *for some* $\alpha \in (0, 2)$, *and that the linear model* (1) *holds for a sequence of coefficients* $\{\psi_j\}$ *satisfying*

$$\sum_{j\in\mathbb{Z}} |\psi_j|^\delta < \infty$$

for some $\delta < \alpha$, $\delta \leqslant 1$. *Then the following results hold true*:

- *From Proposition* 1, *as* $n \to \infty$, *for any* $p < \alpha/2$

$$U_n \overset{\mathscr{L}}{\Longrightarrow} U,$$

$$\varepsilon_n \overset{\mathscr{L}}{\Longrightarrow} \varepsilon.$$

- *From Theorem* 1, *as* $n \to \infty$

$$\hat{\zeta}_n^{\mathrm{BAS1}} \xrightarrow{P} \zeta,$$

$$2\log n(\hat{\zeta}_n^{\mathrm{BAS1}} - \zeta) \overset{\mathscr{L}}{\Longrightarrow} \varepsilon.$$

- *From Proposition* 2, *as* $b \to \infty$

$$\varepsilon_{b^2} - \varepsilon_b \overset{\mathscr{L}}{\Longrightarrow} Y_1 - Y_2.$$

- *The formulas of Corollary* 1 *and Theorem* 3 *remain true*, *but the asymptotic statements have not been validated.*

These results in the moving average case follow essentially from the weak convergence results—Theorems 4.1 and 4.2 of Davis and Resnick (1985). What is needed to extend the remaining results is a local limit theorem for moving averages—such a result would be a worthwhile contribution, and would at once facilitate extending the proofs of the *iid* case.

3. Empirical studies

In this section, we wish to explore the small-sample properties of BASr and SRCEN. Theorem 3 indicates that, at least for $F \in \mathrm{ND}(\alpha)$ with $\alpha < 2$, SCEN and SRCEN should have a faster rate of convergence than BAS1. And SRCEN should have lower variance (but comparable bias) than SCEN, according to Corollary 1. In order to assess our new estimators in a broader context, it is desirable to compare them to other widely used tail index estimators, such as the Pickands (1975), DEdH (Dekkers et al., 1989), and Hill (1975) estimators.

We now return to our original $2m+1$-dependent model (1). As a referee pointed out, it is known that long memory can handicap tail index estimators, so we focus on time series models with short-range dependence. For example, the R/S statistic, which is used to measure long memory (see Beran, 1994), contains a sum of squares similar to the MS and other tail index estimators of this paper. Our empirical findings indicate that our SRCEN estimator experiences some sensitivity to serial correlation, similar to the Pickands, DEdH, and Hill estimators; however, it is generally more robust to serial dependence than these latter estimators. We have tried to frame the SRCEN statistic as an extension of Meerschaert and Scheffler (1998), in the effort to obtain an estimator with faster asymptotic rate of convergence (while preserving the robustness of the original MS statistic). The simulation results bear this out to a large extent.

The theoretical results of Section 2 consider the asymptotics of an estimator for ζ; to obtain an estimator for α, we should apply the functional $f(x) = 1/x$. In this case, the asymptotic MSE will be scaled by the factor $(f'(\zeta))^2 = \zeta^{-4}$.

3.1. An example

In order to illustrate how our tail index estimator SRCEN might be used in practice, we considered the log returns of the Dow Jones Industrial Average (DJIA), measured in 40 s increments from approximately 6:30 AM to 1:00 PM of December 15, 2003. There were 585 observations, which reduced to 584 after the log return transformation

$$X_t \mapsto Y_t = \log X_t - \log X_{t-1}.$$

The Y_t series is covariance stationary with evidence of heavy-tails. We ran our SRCEN procedure, together with BAS1, the MS statistic, and the Pickands, DEdH, and Hill estimators. Like the latter three estimators, one plots SRCEN versus various choices of the block size b, which produces a range of estimates—see Embrechts et al. (1997). BAS1 produced the value -1.084659, demonstrating its sensitivity to the scale of the data. Thus, the MS produces zero; hence it is little better. Note that SRCEN, Pickands, DEdH, and Hill are all scale invariant, which is an excellent property for finite samples. Figs. 1–4 show the performance of Pickands, Hill, and DEdH versus a choice of order statistics. Finally, SRCEN is plotted against block choice b.

The Pickands plot is horrible, and the Hill estimator is not much better. Both DEdH and SRCEN do reasonably well on this data—which is modelled as an $AR(1)$—giving a tail index value between 1.65 and 1.75. For SRCEN, we looked at block sizes 3–17 (since at $b = 18$, K drops to 1).

3.2. Simulations

Simulation studies were conducted on a variety of models, block sizes, and estimators. The following tables summarize the empirical root MSE calculations for 1000 Monte Carlo simulations of particular heavy-tailed models with sample size $n = 1000$. The symmetric α-stable and Student T distributions both belong to $\mathrm{ND}(\alpha)$; in the latter case, the degrees of freedom parameter is equal to α. We also consider a logGamma and symmetric Burr distribution; the former is in $D(\alpha)\backslash\mathrm{ND}(\alpha)$. The Burr has cumulative distribution function $1 - (k/k + x^{\tau})^{\alpha}$ for $x > 0$, and has tail index

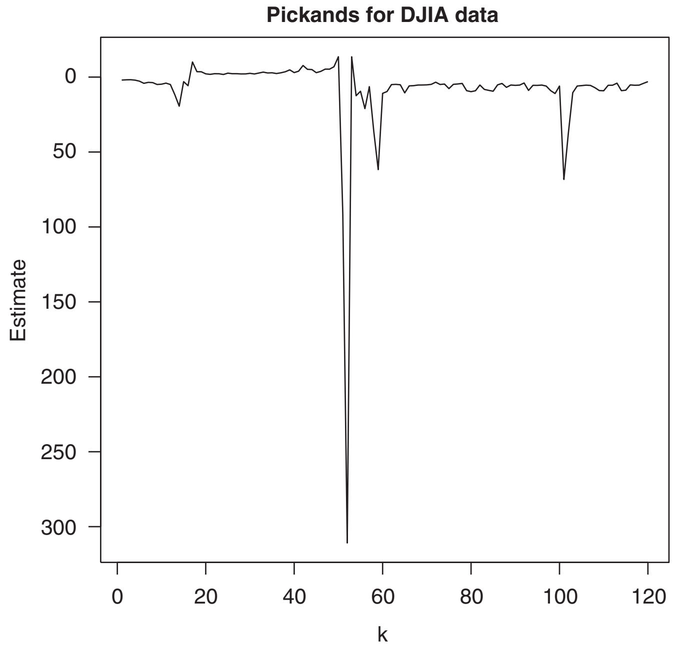

Fig. 1. Pickands estimator.

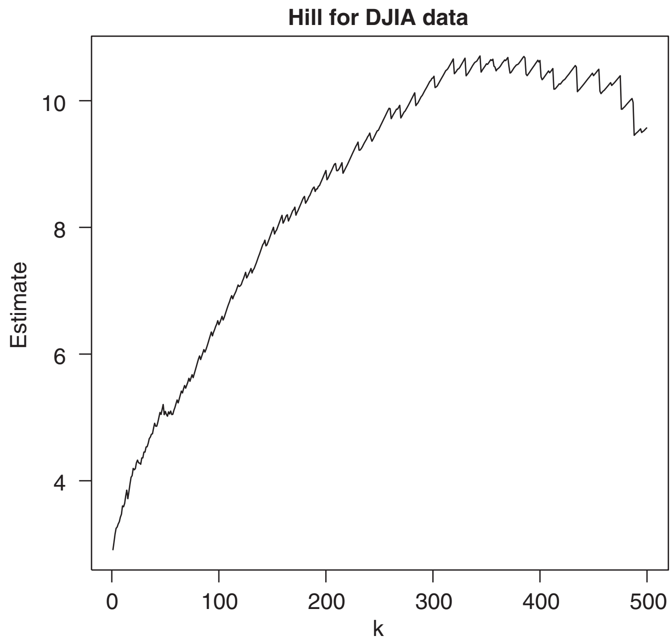

Fig. 2. Hill estimator.

$\alpha\tau$; we let $k=1$ and $\tau=1$ in our simulations below. Whereas for the above distributions the calculation of a_n is simple (using Proposition 2.2.13 of Embrechts et al., 1997), the case of the logGamma is more complicated; we summarize its properties in Proposition 4 below.

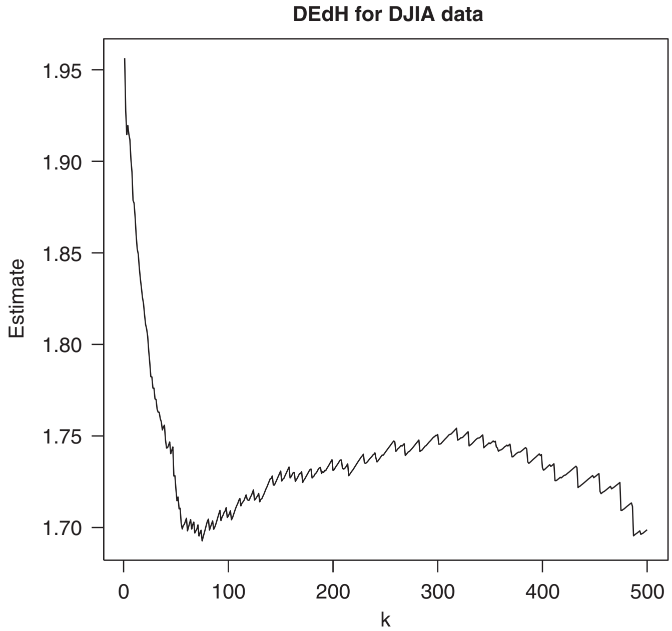

Fig. 3. DEdH estimator.

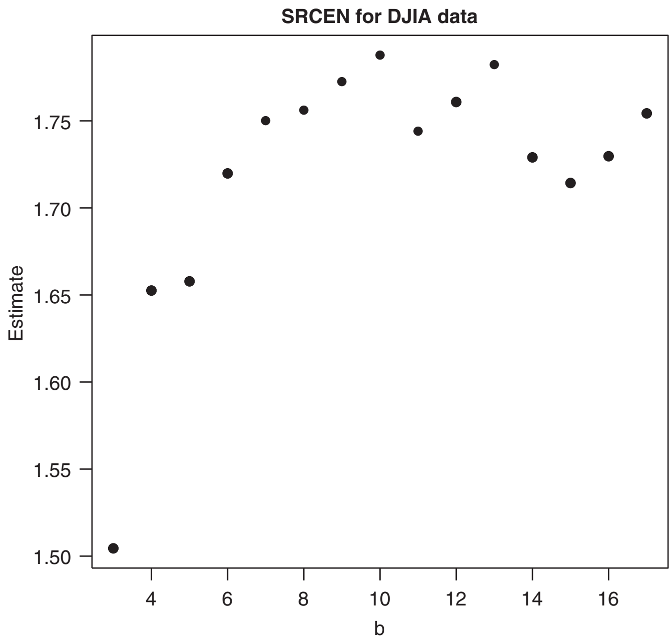

Fig. 4. SRCEN estimator.

Proposition 4. *Let Y be Gamma distributed with shape β and rate α. Then $X = \exp Y$ is heavy tailed of index α, and the rate of convergence of partial sums a_n is given by*

$$a_n = n^{1/\alpha}(\log n)^{(\beta-1)/\alpha}\Gamma(\beta)^{-1/\alpha}, \tag{8}$$

where $\Gamma(x)$ denotes the Gamma function. Hence the logGamma is in $D(\alpha)\backslash\mathrm{ND}(\alpha)$ whenever $\beta \neq 1$.

Table 1
Comparison of root MSEs, *iid* α-stable

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0317	0.0743	0.1089	0.1321	0.1468	0.1506	0.1682
SRCEN, $b=7$	0.0230	0.0542	0.0796	0.0883	0.0976	**0.1056**	0.1127
SRCEN, $b=10$	0.0293	0.0722	0.1012	0.1159	0.1248	0.1220	**0.1016**
SRCEN, $b=14$	0.0385	0.0939	0.1285	0.1514	0.1622	0.1502	0.1160
Hill, $k=50$	0.0292	0.0792	0.1196	0.1599	0.1976	0.2870	0.1987
Hill, $k=100$	0.0202	0.0523	0.0814	0.1032	0.1357	0.2962	0.1999
Hill, $k=200$	**0.0162**	**0.0406**	**0.0608**	**0.0716**	**0.0940**	0.2799	0.1999

Table 2
Comparison of root MSEs, *iid* Student T

d.o.f.	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0316	0.0787	0.1118	0.1306	0.1574	0.2274	0.3572
SRCEN, $b=7$	0.0259	0.0602	0.0865	0.0902	0.1092	0.2072	0.3632
SRCEN, $b=10$	0.0323	0.0732	0.1099	0.1122	0.1245	0.1903	0.3166
SRCEN, $b=14$	0.0411	0.0957	0.1384	0.1483	0.1583	0.1894	0.2864
Hill, $k=50$	0.0323	0.0775	0.1291	0.1568	0.1939	0.2116	0.1902
Hill, $k=100$	0.0207	0.0518	0.0902	0.1057	0.1231	0.1436	**0.1717**
Hill, $k=200$	**0.0147**	**0.0361**	**0.0576**	**0.0689**	**0.0893**	**0.1422**	0.2288

Table 3
Comparison of root MSEs, *iid* Student T

d.o.f.	2.0	2.2	2.5	2.8	3.0	3.2	3.5
BAS2	0.2946	0.3541	0.4603	0.6090	0.6996	0.8127	—
SRCEN, $b=7$	0.2500	0.3244	0.4570	0.6373	0.7546	0.8878	—
SRCEN, $b=10$	0.2422	0.2880	0.3857	0.5306	0.6324	0.7451	0.9452
SRCEN, $b=14$	0.2851	0.3077	0.3735	0.4742	0.5486	0.6540	0.8292
Hill, $k=50$	0.2937	0.3116	**0.3356**	**0.4131**	**0.4462**	**0.5008**	**0.5945**
Hill, $k=100$	**0.2200**	**0.2649**	0.3444	0.4695	0.5449	0.6584	0.8122
Hill, $k=200$	0.3021	0.4032	0.5516	0.7326	0.8594	0.9973	—

Table 4
Comparison of root MSEs, $MA(7)$ α-stable

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	**0.0311**	**0.0848**	0.1558	0.2104	0.2610	0.3299	0.4124
SRCEN, $b=7$	0.0742	0.1341	0.1483	**0.1325**	**0.1073**	**0.1110**	0.2017
SRCEN, $b=10$	0.0594	0.1112	**0.1374**	0.1385	0.1315	0.1270	0.1666
SRCEN, $b=14$	0.0579	0.1166	0.1536	0.1619	0.1660	0.1559	**0.1490**
Hill, $k=50$	0.1089	0.2735	0.4316	0.4641	0.4642	0.4044	0.2190
Hill, $k=100$	0.0639	0.1511	0.2495	0.2982	0.3598	0.3704	0.1980
Hill, $k=200$	0.0381	0.0904	0.1502	0.1796	0.2451	0.3358	0.1974

Our study investigates several competing methods, tested on 10 different time series models, which are displayed in Tables 1–10. Note that by Cline (1983), an $MA(\infty)$ process will be in $D(\alpha)$ if its *iid* inputs are in $D(\alpha)$.

Table 5
Comparison of root MSEs, $AR1$ α-stable

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	**0.0310**	**0.0767**	**0.1201**	0.1501	0.1905	0.2491	0.3281
SRCEN, $b=7$	0.2343	0.2218	0.1889	0.1457	**0.1074**	**0.0958**	0.1760
SRCEN, $b=10$	0.1556	0.1630	0.1571	**0.1420**	0.1302	0.1171	0.1460
SRCEN, $b=14$	0.1179	0.1473	0.1685	0.1653	0.1645	0.1469	**0.1321**
Hill, $k=50$	0.2695	0.3434	0.4071	0.4426	0.4616	0.4012	0.2262
Hill, $k=100$	0.1531	0.2088	0.2580	0.3103	0.3582	0.3778	0.2022
Hill, $k=200$	0.0896	0.1292	0.1680	0.2055	0.2483	0.3453	0.1978

Table 6
Comparison of root MSEs, *iid* LogGamma, shape $=\frac{1}{2}$

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0608	0.1358	**0.1843**	**0.1952**	**0.1784**	**0.1342**	0.1849
SRCEN, $b=7$	0.0583	0.1453	0.2172	0.2396	0.2243	0.1411	**0.0950**
SRCEN, $b=10$	0.0568	0.1396	0.2132	0.2411	0.2345	0.1660	0.1049
SRCEN, $b=14$	0.0633	0.1481	0.2248	0.2545	0.2549	0.1971	0.1227
Hill, $k=50$	0.0527	0.1337	0.2112	0.2602	0.3100	0.3266	0.1804
Hill, $k=100$	**0.0491**	**0.1225**	0.1966	0.2428	0.2981	0.3314	0.1862
Hill, $k=200$	0.0546	0.1353	0.2139	0.2782	0.3318	0.3829	0.1982

Table 7
Comparison of root MSEs, *iid* LogGamma, shape $=2$

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0603	0.1520	0.2433	0.3228	0.3884	0.5141	0.6620
SRCEN, $b=7$	**0.0323**	**0.0812**	**0.1307**	0.1733	0.2130	0.3056	0.4219
SRCEN, $b=10$	0.0334	0.0857	0.1325	0.1749	0.2070	0.2895	0.3980
SRCEN, $b=14$	0.0369	0.0923	0.1385	0.1831	0.2121	0.2785	0.3808
Hill, $k=50$	0.0331	0.0835	0.1355	0.1736	**0.2017**	**0.2597**	**0.3036**
Hill, $k=100$	0.0349	0.0861	0.1385	**0.1718**	0.2095	0.2605	0.3144
Hill, $k=200$	0.0403	0.1002	0.1593	0.1999	0.2403	0.2983	0.3575

Table 8
Comparison of root MSEs, $MA(7)$ LogGamma, shape $=2$

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0578	0.1353	0.2043	0.2528	0.3062	0.3890	0.5264
SRCEN, $b=7$	0.0442	**0.0950**	0.1369	0.1473	**0.1373**	**0.09489**	**0.1541**
SRCEN, $b=10$	**0.0373**	0.0868	**0.1277**	**0.1431**	0.1476	0.1214	0.1662
SRCEN, $b=14$	0.0403	0.0985	0.1487	0.1711	0.1776	0.1549	0.1776
Hill, $k=50$	0.0956	0.2252	0.3954	0.4374	0.4625	0.4024	0.2617
Hill, $k=100$	0.0536	0.1358	0.2349	0.2965	0.3450	0.3483	0.2101
Hill, $k=200$	0.0407	0.0995	0.1525	0.1937	0.2434	0.3132	0.1889

Since the $AR(1)$ model has a representation as an $MA(\infty)$ process, all the models below have marginal distributions in $D(\alpha)$.

- Table 1: *iid* α-stable with low tail index.
- Table 2: *iid* Student T with low degrees of freedom.

Table 9
Comparison of root MSEs, *iid* Burr, $k = 1$

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	0.0316	0.0712	0.1203	0.1470	0.1800	0.2533	0.3531
SRCEN, $b = 7$	0.0258	0.0592	0.0805	0.1066	0.1596	0.3042	0.4942
SRCEN, $b = 10$	0.0324	0.0736	0.1018	0.1156	0.1469	0.2621	0.4313
SRCEN, $b = 14$	0.0406	0.0934	0.1326	0.1466	0.1575	0.2335	0.3855
Hill, $k = 50$	0.0318	0.0774	0.1224	0.1564	0.1730	**0.2164**	**0.2624**
Hill, $k = 100$	0.0210	0.0509	0.0831	**0.1030**	**0.1334**	0.2175	0.3310
Hill, $k = 200$	**0.0144**	**0.0348**	**0.0689**	0.1149	0.1805	0.3226	0.4917

Table 10
Comparison of root MSEs, $AR(1)$ Burr, $k = 1$

α	0.2	0.5	0.8	1.0	1.2	1.5	1.8
BAS1	**0.0315**	**0.0788**	**0.1346**	0.1855	0.2361	0.3465	0.4878
SRCEN, $b = 7$	0.2365	0.2303	0.1751	**0.1157**	**0.0874**	0.2190	0.4308
SRCEN, $b = 10$	0.1571	0.1731	0.1534	0.1215	0.1146	**0.2146**	0.4016
SRCEN, $b = 14$	0.1197	0.1527	0.1571	0.1528	0.1474	0.2162	0.3733
Hill, $k = 50$	0.2686	0.3574	0.4078	0.4391	0.4560	0.3918	0.2623
Hill, $k = 100$	0.1557	0.2187	0.2581	0.2990	0.3315	0.3221	**0.2410**
Hill, $k = 200$	0.0908	0.1346	0.1665	0.1929	0.2109	0.2347	0.2432

- Table 3: *iid* Student T with moderate degrees of freedom.
- Table 4: $MA(7)$ α-stable with low tail index.
- Table 5: $AR(1)$ α-stable with low tail index.
- Table 6: *iid* LogGamma with shape 0.5 and low tail index.
- Table 7: *iid* LogGamma with shape 2 and low tail index.
- Table 8: $MA(7)$ LogGamma with shape 2 and low tail index.
- Table 9: *iid* Symmetric Burr with $k = 1$, $\tau = 1$ and low tail index.
- Table 10: $AR(1)$ Symmetric Burr with $k = 1$, $\tau = 1$ and low tail index.

The methods we consider are the BAS1, SRCEN, and Hill estimators; results for the Pickands, and DEdH estimators were generally inferior to the Hill estimator, and are not presented here. In the third model, the tail index α (which is equal to the degrees of freedom) is between 2 and 4, so we use the BAS2 estimator and the analogue for SRCEN. Models 4, 5, 8, and 10 introduce serial correlation; since the 7-dependence in the fourth and eighth models do not exceed the block choices $b = 7, 10, 14$, we expect SRCEN to do reasonably well. This is because, in order to obtain independence of the various blocks, the block size must be greater than 7 for a 7-dependent process. The fifth and 10th models have correlation at all lags, and so none of the methods are expected to perform very well. For the $MA(7)$, we used all coefficients equal to $1/7$, whereas for the $AR(1)$ we used $\phi = 0.8$. The block choices $b = 7, 10, 14$ for SRCEN correspond to $k = 20, 10, 5$, respectively, and represent low, moderate, and high values of b relative to the sample size. For the Hill estimator, we chose three order statistics representing low, moderate, and high values—50, 100, 200. Recall that the absolute value of the data should be taken before using these latter three estimators. For the Hill estimator, we utilized a truncation technique to make it more competitive—in the case of the low-tail index, any values larger than 2 were set equal to 2; the values are set to 4 in the case of Table 3. This procedure makes comparisons with SRCEN more fair, since the latter has an "automatic" maximum of 2. Note that the simulation studies are somewhat biased in favor of BASr and SRCEN, since we assume that a suitable value of r (either 1 or 2) is known.

The best result in each column is presented in bold. Given the range of values of the parameter, a root MSE of 0.1 seems fairly reasonable. Whenever the root MSE was greater than 1, its value was omitted from the table. The following statements summarize the findings of the simulation study:

- Hill is superior to BASr and SRCEN for many of the cases of serial independence.
- BASr and SRCEN are competitive with Hill when serial correlation is present.
- SRCEN provides an alternative to BASr that is often more accurate and precise.
- SRCEN is upward biased for low α and downward biased for high α. Bias goes down as b increases (bias is not shown in tables), and variance goes up.
- SRCEN and BASr perform fairly well in the $D(\alpha)\backslash \mathrm{ND}(\alpha)$ case of the logGamma.
- All estimators are sensitive to serial dependence.

These simulations indicate that a case can be made for using SRCEN on real data. It compares favorably with BASr in Monte Carlo studies. But for any particular data set, SRCEN has the advantage over BASr of being scale invariant; this was demonstrated saliently through the DJIA example. The Hill is a good competitor, as neither Hill nor SRCEN are superior across the board. For independent data, the Hill estimator is generally better, especially when $\alpha \leqslant 1$. Overall SRCEN seems to handle serial correlation better; hence it may be useful when serial dependence is observed, but no particular time series model is used. One drawback is that, like the use of BASr, the user must select a value of r. In practice, SRCEN can be used much as Hill and order-statistics-based methods—namely plotting the estimates versus a reasonable selection of block sizes b, as demonstrated in the example.

Acknowledgments

Many thanks are due to Professor Holger Drees for his comments to one of the authors at the Prague EMS Meeting, August 2002. The authors are also indebted to the comments of the referees and editor, who identified some errors in an earlier version of this paper.

Appendix

Proof of Proposition 1. Since $\alpha < 2$ and $F \in D(\alpha)$, the tail probabilities of X_t are regularly varying of index $-\alpha$—see Theorem 2.2.8 of Embrechts et al. (1997). It follows from Theorem 4.2 of Davis and Resnick (1985) that $U_n = a_n^{-2}\sum_{t=1}^n X_t^2$ converges weakly to a positively skewed $\alpha/2$ stable variable, which we denote by U. The weak convergence of ε follows by the continuous mapping theorem.

Next, we prove the last assertion. Since $f_X(x)/x$ is bounded in a neighborhood of zero, f_{X^2} is a bounded function ($f_{X^2}(x) = (f_X(\sqrt{x}) + f_X(-\sqrt{x}))/\sqrt{x}$). Hence the pdf of U_1 is bounded, and it now follows from Gnedenko (1953/1954)—also see Macht and Wolf (1989) for an English translation of these Russian results—that the pdf of U_n converges uniformly to the pdf of U.

Now let $p < \alpha/2$ be a positive exponent. Then $\mathbb{E}|U_n^p| < \infty$ for all n, since the tail probabilities of U_n are regularly varying too, by Cline (1983). Clearly $\mathbb{E}|U^p| < \infty$ as well. Since these random variables are positive, the absolute values are superfluous. Then we claim that $\mathbb{E}U_n^p \to \mathbb{E}U^p$ as $n \to \infty$, which will be proved by establishing uniform integrability of the sequence U_n^p. Let $\varepsilon > 0$ be so small such that $p(1+\varepsilon) < \alpha/2$, so

$$\mathbb{E}[U_n^{p(1+\varepsilon)}] = \int_0^\infty \mathbb{P}[U_n > t^{1/p(1+\varepsilon)}]\,\mathrm{d}t \leqslant 1 + \int_1^\infty \mathbb{P}[U_n > t^{1/p(1+\varepsilon)}]\,\mathrm{d}t.$$

By Lemma 2 of Meerschaert and Scheffler (1998), we can choose $\delta > 0$ such that $\alpha/2 > \delta + p(1+\varepsilon)$ so that

$$\mathbb{P}[U_n > x] \leqslant C x^{-\alpha/2+\delta}$$

for all x and $n \geqslant 1$, for some positive constant C. Applying this yields

$$\sup_n \mathbb{E}[U_n^{p(1+\varepsilon)}] \leqslant 1 + C\int_1^\infty t^{-(\alpha/2-\delta)/p(1+\varepsilon)}\,\mathrm{d}t < \infty,$$

so that U_n^p is uniformly integrable.

In order to show the convergence of the log moments, we first invoke the following simple formula for the expectation of an absolutely continuous random variable V with support on all of $\mathbb{R}$:

$$\mathbb{E}[V] = \int_0^\infty \mathbb{P}[V > t]\,\mathrm{d}t - \int_{-\infty}^0 \mathbb{P}[V < t]\,\mathrm{d}t. \tag{9}$$

Letting $V = \varepsilon_n$, we see that

$$\int_0^\infty \mathbb{P}[\varepsilon_n > t]\,\mathrm{d}t = \int_0^\infty \mathbb{P}[U_n > e^t]\,\mathrm{d}t \to \int_0^\infty \mathbb{P}[U > e^t]\,\mathrm{d}t$$

as $n \to \infty$ by the same arguments used above for the convergence of moments. The other tail is trickier. We calculate

$$\int_{-\infty}^0 \mathbb{P}[\varepsilon_n < t] - \mathbb{P}[\varepsilon < t]\,\mathrm{d}t = \int_0^\infty \mathbb{P}[U_n < e^{-t}] - \mathbb{P}[U < e^{-t}]\,\mathrm{d}t = \int_0^\infty \int_0^{e^{-t}} f_{U_n}(x) - f_U(x)\,\mathrm{d}x\,\mathrm{d}t, \tag{10}$$

which tends to zero as $n \to \infty$ by the uniform convergence of the pdfs and the integrability of e^{-t}. For the second moment, we write

$$\mathbb{E}[\varepsilon_n^2] = \int_0^\infty \mathbb{P}[\varepsilon_n^2 > t]\,\mathrm{d}t = \int_0^\infty \mathbb{P}[\varepsilon_n > \sqrt{t}]\,\mathrm{d}t + \int_0^\infty \mathbb{P}[\varepsilon_n < -\sqrt{t}]\,\mathrm{d}t$$

and apply the previous arguments to obtain $\mathbb{E}[\varepsilon_n^2] \to \mathbb{E}[\varepsilon^2]$, which concludes the proof. □

Proof of Theorem 1. Using the decomposition (2), we have

$$\hat{\zeta}^{\mathrm{BAS1}} = \zeta + \frac{\log M(n)}{2\log n} + \frac{\varepsilon_n}{2\log n}. \tag{11}$$

Since $\varepsilon_n \overset{\mathscr{L}}{\Longrightarrow} \varepsilon$ by Proposition 1, we obtain the weak convergence result of the theorem by letting $c_n = \log M(n)/2\log n$. We observe that if $F \in \mathrm{ND}(\alpha)$, then we make take L asymptotic to a nonzero constant, call it c. Then $c_n = \log|c|/\log n$. Consistency follows at once from this weak convergence, and the fact that $c_n \to 0$ as $n \to \infty$. This is because the logarithm of any slowly varying function must diverge to infinity (if it does at all) slower than the logarithm itself. Using the representation of a slowly varying function M—see Theorem A3.12 of Embrechts et al. (1997)—we can write

$$M(x) = c(x)\exp\left\{\int_z^x \eta(u)/u\,\mathrm{d}u\right\}$$

for a function c tending to a constant c_0 and some positive number z, and a function η tending to zero. Then

$$\frac{\log M(n)}{\log n} = \frac{\log c(n)}{\log n} + \frac{\int_z^n \eta(u)/u\,\mathrm{d}u}{\log n} \to 0 \tag{12}$$

$n \to \infty$, using L'Hopital's rule on the second term if necessary.

Taking expectations in (11) yields

$$\mathbb{E}[\hat{\zeta}^{\mathrm{BAS1}}] = \zeta + \frac{\log M(n)}{2\log n} + \frac{\mathbb{E}\varepsilon_n}{2\log n}$$

which tends to ζ, since $\mathbb{E}[\varepsilon_n] \to \mathbb{E}[\varepsilon]$ by Proposition 1. As for the MSE,

$$\mathbb{E}[(\hat{\zeta}^{\mathrm{BAS1}} - \zeta)^2] = \frac{1}{4\log n^2}(\log M(n)^2 + 2\mathbb{E}[\varepsilon_n]\log M(n) + \mathbb{E}[\varepsilon_n^2])$$

which tends to zero, utilizing $\mathbb{E}[\varepsilon_n^2] \to \mathbb{E}[\varepsilon^2]$. □

Proof of Theorem 2. We refer to (4), which gives the important decomposition of the estimator, and replace n everywhere by b^2. Each block estimator $\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}$ has bias

$$\mathbb{E}[\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}] - \zeta = \frac{\log(M(b^2)/M(b))}{2\log b} + \mathbb{E}\left[\frac{\varepsilon_{b^2} - \varepsilon_b}{2\log b}\right] \tag{13}$$

so that the "deterministic bias" due to the slowly varying function is

$$\frac{\log(M(b^2)/M(b))}{2\log b}.$$

For any slowly varying function L, this ratio is o(1) as $b \to \infty$, which is seen by applying the arguments of (12). The second term in (13) is $O(1/\log b)$ by Proposition 1, but may be $o(1/\log b)$ for some distributions F. The variance of the block estimator is

$$\mathbb{D}[\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}] = (2\log b)^{-2}\mathbb{D}[\varepsilon_{b^2} - \varepsilon_b]. \tag{14}$$

This is bounded by $\sup_n \mathbb{D}[\varepsilon_n](\log b)^{-2}/2$ which is $O((\log b)^{-2})$, since the sequence ε_n has convergent first and second moments. Clearly, the bias for $\hat{\zeta}_{b,n}^{\mathrm{SCEN}}$ has the same asymptotics as the bias of $\hat{\zeta}_{b^2}^{\mathrm{CEN}(k)}$, but the variance will be

$$\mathbb{D}[\hat{\zeta}_{b,n}^{\mathrm{SCEN}}] = \frac{1}{K}(2\log b)^{-2}\mathbb{D}[\varepsilon_{b^2} - \varepsilon_b]$$

from (14). Now $K = \lfloor n/b^2 \rfloor$, so the variance has order $b^2/(n(\log b)^2)$. This completes the proof. $\square$

Proof of Proposition 2. We begin with the last assertion. Let $W_{b^2}(X^2) = \sum_{t=b+1}^{b^2} X_t^2$, which is independent of $S_b(X^2)$. Write

$$\varepsilon_{b^2} - \varepsilon_b = \log\left(\frac{a_{b^2}^{-2} S_{b^2}}{a_b^{-1} S_b}\right) = \log\left(\frac{a_{b^2}^{-2} S_{b^2}}{a_{b^2}^{-2} W_{b^2}}\right) + \log\left(\frac{a_{b^2}^{-2} W_{b^2}}{a_{b^2}^{-1} S_b}\right) = C + D;$$

we show that $C = C_b$ is uniformly square integrable and converges weakly to zero. Since $a_b^{-1} S_b$ and $a_{b^2}^{-2} W_{b^2}$ are independent and bounded in probability,

$$\frac{a_{b^2}^{-2} S_{b^2}}{a_{b^2}^{-2} W_{b^2}} = \frac{a_{b^2}^{-2} S_b}{a_{b^2}^{-2} W_{b^2}} + 1 = 1 + O_P(a_b^{-1})$$

so that the logarithm of the left-hand side converges weakly to zero as $b \to \infty$. Let $A = \log(a_{b^2}^{-2} S_{b^2})$ and $B = \log(a_{b^2}^{-2} W_{b^2})$, and consider $\mathbb{E}|A - B|^{2+\delta}$ for any $\delta > 0$; let $\delta < 2$ so that $(2+\delta)/4 < 1$. Then

$$\begin{aligned}\mathbb{E}|A - B|^{2+\delta} &= \mathbb{E}(|A - B|^4)^{(2+\delta)/4} \leqslant \mathbb{E}|A|^{2+\delta} + 4^{(2+\delta)/4}(\mathbb{E}|A|^{2+\delta})^{3/4}(\mathbb{E}|B|^{2+\delta})^{1/4} \\ &\quad + 6^{(2+\delta)/4}(\mathbb{E}|A|^{2+\delta})^{1/2}(\mathbb{E}|B|^{2+\delta})^{1/2} + 4^{(2+\delta)/4}(\mathbb{E}|A|^{2+\delta})^{1/4}(\mathbb{E}|B|^{2+\delta})^{3/4} + \mathbb{E}|B|^{2+\delta}\end{aligned}$$

follows from the triangle and Hölder inequalities. Now both $A = A_b$ and $B = B_b$ have bounded $2 + \delta$ moment, by a direct adaption of the argument of Proposition 1. Therefore $\sup_b \mathbb{E}|A_b - B_b|^{2+\delta} < \infty$, yielding uniform square integrability of $\log(a_{b^2}^{-2} S_{b^2}/a_{b^2}^{-2} W_{b^2})$. So C_b tends to zero in probability, mean, and mean square. Clearly, $\mathbb{E}[(\varepsilon_{b^2} - \varepsilon_b)^2] = \mathbb{E}C^2 + 2\mathbb{E}CD + \mathbb{E}D^2$. Now we just showed that $\mathbb{E}[C_b^2] \to 0$ as $b \to \infty$, and $\mathbb{E}[CD] \leqslant \sqrt{\mathbb{E}[C^2]\mathbb{E}[D^2]} \to 0$ as well, since $\mathbb{E}[D^2]$ is bounded. In fact, $D = \log(a_{b^2}^{-2} W_{b^2}) - \log(a_{b^2}^{-1} S_b) = G - H$, where G and H are independent random variables. By Proposition 1, both G_b and H_b tend (jointly) weakly, in mean, and in mean square to independent logged positive $\alpha/2$ stable random variables, which will be denoted by Y_1 and Y_2. Hence $\varepsilon_{b^2} - \varepsilon_b \overset{\mathscr{L}}{\Longrightarrow} Y_1 - Y_2$, and

$$\mathbb{E}D^2 = \mathbb{E}G^2 - 2\mathbb{E}G\mathbb{E}H + \mathbb{E}H^2 \to \mathbb{E}Y_1^2 - 2\mathbb{E}Y_1\mathbb{E}Y_2 + \mathbb{E}Y_2^2 = 2\mathbb{D}\varepsilon.$$

This shows the first assertion. Finally, the second assertion of the proposition follows from the identity

$$\mathbb{D}[\varepsilon_{b^2}, \varepsilon_b] = \tfrac{1}{2}(\mathbb{E}[\varepsilon_{b^2}^2] + \mathbb{E}[\varepsilon_b^2] - \mathbb{E}[(\varepsilon_{b^2} - \varepsilon_b)^2]) - \mathbb{E}[\varepsilon_{b^2}]\mathbb{E}[\varepsilon_b]$$

and matching up the various limits calculated here and in Proposition 1. □

Proof of Corollary 1. The formulas for BAS1 and CEN are immediate from previous calculations. Using the decomposition (4), we see that the variance is

$$\mathbb{D}\left[\varepsilon_{b^2} - \frac{1}{b}\sum_{j=1}^{b}\varepsilon_b^{(j)}\right] = \frac{1}{b^2}\sum_{i,j}\mathbb{D}[\varepsilon_{b^2} - \varepsilon_b^{(i)}, \varepsilon_{b^2} - \varepsilon_b^{(j)}]$$
$$= \mathbb{D}[\varepsilon_{b^2}] - 2\,\mathbb{D}[\varepsilon_{b^2}, \varepsilon_b] + \frac{1}{b}\mathbb{D}[\varepsilon_b].$$

The asymptotic formulas follow from Proposition 2. □

Proof of Proposition 3. By Satybaldina (1972), it follows that under (7),

$$\sup_x |F_{U_n}(x) - F_U(x)| = \mathrm{O}(n^{1-2/\alpha}).$$

Next we write

$$\mathbb{E}\varepsilon_n - \mathbb{E}\varepsilon = \int_0^\infty \mathbb{P}([\varepsilon_n > t] - \mathbb{P}[\varepsilon > t])\,\mathrm{d}t - \int_{-\infty}^0 (\mathbb{P}[\varepsilon_n < t] - \mathbb{P}[\varepsilon < t])\,\mathrm{d}t. \tag{15}$$

Focus on the first term. Let b_n be a positive sequence tending to infinity as $n \to \infty$. Then the first term can be broken up as

$$\int_0^{b_n} \mathbb{P}[\varepsilon_n > t] - \mathbb{P}[\varepsilon > t]\,\mathrm{d}t + \int_{b_n}^\infty \mathbb{P}[\varepsilon_n > t]\,\mathrm{d}t - \int_{b_n}^\infty \mathbb{P}[\varepsilon > t]\,\mathrm{d}t,$$

the first term of which is $\int_0^{b_n} \mathbb{P}[U_n < e^t] - \mathbb{P}[U < e^t]\,\mathrm{d}t = \mathrm{O}(b_n\, n^{1-2/\alpha})$. Using Lemma 2 of Meerschaert and Scheffler (1998), we have

$$\int_{b_n}^\infty \mathbb{P}[\varepsilon_n > t]\,\mathrm{d}t = \int_{e^{b_n}}^\infty \mathbb{P}[U_n > x]/x\,\mathrm{d}x \leqslant C\int_{e^{b_n}}^\infty x^{-(1+\alpha/2)+\delta}\,\mathrm{d}x$$

for any $\delta > 0$ and some constant $C > 0$, for all $n \geqslant 1$. Naturally, we choose $\delta < \alpha/2$ to ensure integrability of the bounding function. Thus the bound is $C\exp\{(\delta - \alpha/2)b_n\}/(\alpha/2 - \delta)$. The same argument provides a similar bound for $\int_{b_n}^\infty \mathbb{P}[\varepsilon > t]\,\mathrm{d}t$. If we let $b_n = \log(n)$, for example, we obtain a $\mathrm{O}(\log(n)\, n^{1-2/\alpha})$ term and a $\mathrm{O}(n^{\delta-\alpha/2})$ term. For the second term of (15), we decompose as follows:

$$\int_{-b_n}^0 \mathbb{P}[\varepsilon_n < t] - \mathbb{P}[\varepsilon < t]\,\mathrm{d}t + \int_{-\infty}^{-b_n} \mathbb{P}[\varepsilon_n > t]\,\mathrm{d}t - \int_{-\infty}^{-b_n} \mathbb{P}[\varepsilon > t]\,\mathrm{d}t.$$

The first term here is handled analogously to the above argument. Using the boundedness of the pdf for all n, we have

$$\int_{-\infty}^{-b_n} \mathbb{P}[\varepsilon_n < t]\,\mathrm{d}t = \int_0^{e^{-b_n}} \mathbb{P}[U_n < x]/x\,\mathrm{d}x = \int_0^{e^{-b_n}} \frac{1}{x}\int_0^x f_{U_n}(u)\,\mathrm{d}u\,\mathrm{d}x \leqslant \sup_u f_{U_n}(u)\, e^{-b_n}$$

with a similar bound for $\int_{-\infty}^{-b_n} \mathbb{P}[\varepsilon < t]\,\mathrm{d}t$. Letting $b_n = \log(n)$ gives polynomial order bounds for all the terms. We see that the power p in Proposition 3 can be taken (approximately) to be the minimum of $2/\alpha - 1$ and $\alpha/2$, which are equal at $\alpha = \sqrt{5} - 1$. □

Proof of Theorem 3. The bias calculations for BAS1, CEN, and RCEN follow from earlier calculations and Proposition 3. Of course the bias is the same for SCEN and SRCEN, compared to CEN and RCEN. In order to determine a sensible choice for b as a function of n for SCEN and SRCEN, we simply equate the orders of variance and squared bias:

$$\frac{b^2}{n(\log b)^2} = \frac{1}{b^{2p}(\log b)^2}.$$

Solving for b yields $b = n^{1/2(1+p)}$, and the rate of the MSE will then be $n^{-p/(p+1)}(\log n)^{-2}$. $\square$

Proof of Proposition 4. Basic methods give the pdf of X as

$$f_X(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\log x)^{\beta-1}x^{-(\alpha+1)} \quad \text{for } x \geqslant 1.$$

By Karamata's Theorem (see Theorem A.3.6 of Embrechts et al., 1997),

$$G(x) = \mathbb{P}[X \geqslant x] \sim \frac{\alpha^{\beta-1}}{\Gamma(\beta)}x^{-\alpha}(\log x)^{\beta-1}$$

as $x \to \infty$. This has the form of a slowly varying function multiplied by $x^{-\alpha}$, indicating that the logGamma is in $D(\alpha)$. Substituting (8) into $G(x)$ yields

$$G(a_n) \sim \frac{1}{n}\left(1 + (\beta - 1)\frac{\log\, \log n}{\log n} - \frac{\log \Gamma(\beta)}{\log n}\right)^{\beta-1} \sim \frac{1}{n}$$

which by Proposition 2.2.13 of Embrechts et al., 1997, indicates that a_n is the rate of normalization for the partial sums of *iid* logGamma random variables. $\square$

References

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman & Hall, New York.

Cline, D., 1983. Infinite series of random variables with regularly varying tails. Technical Report 83-24, Institute of Applied Mathematics and Statistics, University of British Columbia.

Csörgő, S., Deheuvels, P., Mason, D.M., 1985. Kernel estimates of the tail index of a distribution. Ann. Statist. 13, 1050–1077.

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.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, Berlin.

Gnedenko, B., 1953/1954. On the local limit theorem for independent and identically distributed summands (in Russian). Wiss. Zt. Humboldt-Univ. Berlin, Math.-Nat. Reihe 3, 287–293.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.

Macht, W., Wolf, W., 1989. The local limit problem and Hölder-continuity. Probab. Theory Relat. Fields 82, 295–305.

Meerschaert, M., Scheffler, H., 1998. A simple robust estimator for the thickness of heavy tails. J. Statist. Plann. Inference 71, 19–34.

Meerschaert, M., Scheffler, H., 2003. Nonparametric methods for heavy-tailed vector data: a survey with applications from finance and hydrology. In: Akritas, M., Politis, D.N. (Eds.), Recent Advances and Trends in Nonparametric Statistics. Elsevier Science, Amsterdam, pp. 265–279.

Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.

Politis, D., 2002. A new approach on estimation of the tail index. C. R. Acad. Sci. Paris, Ser. I 335, 279–282.

Satybaldina, K., 1972. Absolute estimates of the rate of convergence to stable laws. Theory Probab. Appl. 17, 726–728.