Econometric Theory, **18**, 2002, 1019–1039. Printed in the United States of America.
DOI: 10.1017.S026646660218501X

ROBUST INFERENCE FOR THE MEAN IN THE PRESENCE OF SERIAL CORRELATION AND HEAVY-TAILED DISTRIBUTIONS

Tucker McElroy and Dimitris N. Politis
University of California, San Diego

The problem of statistical inference for the mean of a time series with possibly heavy tails is considered. We first show that the self-normalized sample mean has a well-defined asymptotic distribution. Subsampling theory is then used to develop asymptotically correct confidence intervals for the mean *without* knowledge (or explicit estimation) either of the dependence characteristics, or of the tail index. Using a symmetrization technique, we also construct a distribution estimator that combines robustness and accuracy: it is higher-order accurate in the regular case, while remaining consistent in the heavy tailed case. Some finite-sample simulations confirm the practicality of the proposed methods.

1. INTRODUCTION

Consider a strictly stationary time series $\{X_t, t \in \mathbb{N}\}$ that exhibits long range dependence. We are interested in estimating the mean of this time series under the assumption that the marginal distributions are (or may be) heavy tailed. We focus on the sample mean $\bar{X} := n^{-1}\sum_{i=1}^{n} X_i$ as an estimator for $\theta := \mathbb{E}X_t$, which is assumed to be finite. Our point of view is one of generality: the sample mean is a ubiquitous estimate of location; in particular, it is generally consistent for θ even if the regularity condition of finite variance breaks down. Nevertheless, statistical inference (confidence intervals and tests) for θ is based on the distribution of $\bar{X}$, which is crucially affected by dependence and/or heavy tails. The purpose of this paper is to provide a way of consistently estimating the distribution of the (normalized) sample mean *without* knowledge (or explicit estimation) of either the dependence or the heavy-tailed index (this tail index measures the heaviness of the tails); we will achieve this using the subsampling methodology (see Politis and Romano, 1994, Politis, Romano, and Wolf, 1999).

This research was partially supported by NSF grants DMS 97-03964 and 01-04059. Many thanks are due to the editor and the referees for helpful comments that led to a significant improvement of this paper. Address correspondence to: Dimitris N. Politis, Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Mail Code 0112, La Jolla, CA 92093-0112, USA; e-mail: politis@euclid.ucsd.edu.

 0266-4666/02 $9.50

It turns out that the normalized sum of independent and identically distributed (i.i.d.) heavy-tailed random variables converges weakly to a non-normal limit (a stable law); thus it satisfies a non–central limit theorem. To develop confidence intervals for θ, we need the quantiles of this stable law, which, unfortunately, are generally unknown, because both the scale and the index of stability (the heavy-tailed index) will generally be unknown. The recourse is to use subsampling methodology to estimate the limit quantiles. A second practical problem is that the rate of convergence of the sum is generally unknown (it is not the common $\sqrt{n}$ that occurs in the central limit theorem), which prevents us from forming the correct statistic. This is solved by self-normalization, i.e., by dividing by some appropriate measure of scale, such as the square root of the sample variance. If this is done, the limit is no longer a stable random variable but has a well-defined continuous cumulative distribution function (c.d.f.), so that subsampling theory can still be applied.

The basic idea of subsampling is to compute the desired statistic (in our case the sample mean) over smaller subsets—*subsamples*—of the data. This gives us several estimates of the statistic instead of just one. If we then form the full root $(\hat{\theta} - \theta)/\hat{\sigma}$ on the subsamples, this will approximate the limit distribution, just as the root evaluated on the whole data set would. By then averaging appropriately over the various subsamples (which must be contiguous sets for stationary data), we get an empirical distribution that approximates the true limiting distribution and also the finite sampling distribution. For more details, See section 2 of this paper and Politis et al. (1999).

If the sequence $\{X_t\}$ is i.i.d. (and maybe heavy tailed), then the limit distribution of the normalized sample mean is well studied by Logan, Mallows, Rice, and Shepp (1973). Romano and Wolf (1999) use the Logan et al. (1973) limit theory to show that subsampling works in this case (for more details, see Politis et al., 1999, Ch. 10). Nothing is yet known if the series $\{X_t\}$ is not independent; this is the purpose of the paper at hand. To elaborate, we will prove that $\sqrt{n}[(\bar{X} - \theta)/\hat{\sigma}]$ has a limit distribution with a well-defined continuous c.d.f.; here $\hat{\sigma}$ is a statistic that measures the scale or size of the random variables. In Logan et al. (1973), both $(n/a_n)(\bar{X} - \theta)$ and $(\sqrt{n}/a_n)\hat{\sigma}$ are shown to have nondegenerate limit distributions for an i.i.d. time series; they both depend upon an unknown rate a_n (that corresponds to the unknown tail index α) and thus cannot be used statistically. In that paper, the authors form the ratio of these two quantities and prove that the quotient converges to a continuous limit distribution. Because the factor a_n cancels in the division, the unknown rate is eliminated, and the resulting object $\sqrt{n}[(\bar{X} - \theta)/\hat{\sigma}]$ is statistically useful. We prove the corresponding limit theorems in a time series context and thus generalize the work of Logan et al. (1973). Other literature on this topic includes LePage, Woodroofe, and Zinn (1981), Davis and Resnick (1985, 1986), Resnick (1986, 1987), Phillips and Solo (1992), and Davis and Hsing (1995). The first paper explores the convergence of self-normalized sums of heavy-tailed random variables. The next two papers are primarily concerned with the limit behavior of sample autocorrelations for this linear model, whereas Resnick (1986, 1987) considers point process techniques used to prove many of these results. Phillips and Solo (1992) develop asymptotic results for moving average time series via a decomposition of the linear process. Davis and Hsing (1995) examine models with long range dependence that are not linear.

Because the limit distribution depends on the unknown tail index parameter α, one cannot calculate its quantiles, which are necessary for the formation of confidence intervals. For the unnormalized statistic one could explicitly estimate the parameter α, e.g., by Hill's estimator (namely, Resnick, 1997) and the scale of the limiting stable distribution (which would involve estimation of the filter coefficients), for the formation of the desired confidence intervals. Alternatively, subsampling with estimated rate may be used, as in Bertail, Politis, and Romano (1999). However, these options are only viable when the data are in the "normal" domain of attraction of a stable law; if the data are in a more general domain of attraction (which is the general case we consider in this paper), the rate of convergence depends crucially upon an unknown slowly varying function (see Davis and Resnick, 1985). Thus estimation of α via the Hill estimator is insufficient to determine the rate; neither has subsampling for estimated rate been shown to work in this more general case (see the discussion in Politis et al., 1999). Therefore, there is no alternative methodology to what we present here.

Our approach in this paper involves self-normalization; this bypasses the issue of the unknown rate of convergence but brings in a new difficulty: the limiting distribution is no longer stable. Subsampling, nonetheless, can still be applied because a well-defined limiting distribution exists; hence, this difficulty is alleviated also. Note that estimation of α via the Hill estimator is of no practical use here, because the quantiles of the limiting distribution are unknown analytically, and moreover the limit distribution depends upon the unknown filter coefficients coming from the model. Thus, there seems to be no other feasible method of solving this problem.

Finally, we consider the notion of accuracy; using a symmetrization technique, we construct a distribution estimator that combines robustness and accuracy. It is higher order accurate in the regular case (i.e., when the data have finite variance and possibly higher order moments), while remaining consistent in the heavy-tailed case (provided the marginal distribution of X_1 is symmetric). This justifies the term *robust* in the title of the paper. Note that there are many situations wherein the bootstrap fails but subsampling is valid (for a discussion of the contrasts, see Politis et al., 1999). The failure of the bootstrap for the sample mean of infinite variance data is an example of the bootstrap's limitations (for background, see Athreya, 1987; Knight, 1989; Hall, 1990). Thus subsampling, not resampling (i.e., the bootstrap), is the appropriate method for this scenario.

This paper is organized as follows. Section 2 contains background material and our theoretical results; Section 3 contains the proofs, in some detail; and Section 4 contains some finite-sample simulation studies.

2. THEORY

2.1. Background

We consider a *linear* time series:

$$X_t := \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j} \quad \forall t \in \mathbb{N},$$

where $\{Z_t\}$ is an i.i.d. sequence, and for some filter coefficients $\{\psi_j\}$ satisfying a summability condition.

Throughout the paper, we require that the filter $\{\psi_j\} \in l_1$ (see Brockwell and Davis, 1991, Ch. 13) to ensure that the sum converges almost surely; also we impose that the i.i.d. random variables $\{Z_t\}$ (for any $t \in \mathbb{Z}$) are *heavy-tailed*; i.e., they satisfy the following two properties for some $\alpha \in (1,2)$:

$$\mathbb{P}[|Z_t| > x] = x^{-\alpha}L(x), \tag{1}$$

$$\frac{\mathbb{P}[Z_t > x]}{\mathbb{P}[|Z_t| > x]} \to p \qquad \frac{\mathbb{P}[Z_t \leq -x]}{\mathbb{P}[|Z_t| > x]} \to q \tag{2}$$

as $x \to \infty$. Here p and q are between 0 and 1 and add up to 1. We will say throughout this paper that such a random variable is *heavy tailed* with tail thickness parameter α. The term $L(x)$ is a "slowly varying" function, i.e., $L(ux)/L(x) \to 1$ as $x \to \infty$ for any fixed u. An example of a slowly varying function is the logarithmic function. Note that it easily follows that the right and left tails of Z_t behave like

$$\mathbb{P}[Z_t > x] \sim px^{-\alpha}L(x), \qquad \mathbb{P}[Z_t \leq -x] \sim qx^{-\alpha}L(x),$$

where $\sim$ denotes that the ratio tends to unity as $x \to \infty$. For a reference on heavy-tailed linear time series, see Cline (1983).

A random variable that satisfies the preceding conditions is said to have "heavy tails." Such random variables are in the domain of attraction of an α-stable law; i.e., if we take an i.i.d. sequence of such Z_t's, then there exist real constants $a_n > 0$ and b_n such that

$$a_n^{-1}\left(\sum_{t=1}^{n} Z_t - b_n\right) \stackrel{\mathcal{L}}{\Longrightarrow} S,$$

where S is an α-stable law. We then write $\{Z_t\} \in D(\alpha)$, which means that Z_t is in the α-stable domain of attraction. It is common to see condition (2) written as $Z_t \sim F$ (which means, in this context, that Z_t has c.d.f. $F(\cdot)$) and as $x \to \infty$,

$$x^\alpha(1 - F(x)) \to p, \qquad x^\alpha F(-x) \to q.$$

An example of the preceding situation is given by Z_t's in the "normal" domain of attraction of an α-stable law, which means that we can take $a_n = n^{1/\alpha}$. If the Z_t's are themselves α-stable, then the hypotheses are certainly satisfied. If in addition they are symmetric (written Z_t is sαs), then X_t has the law of a sαs also, but scaled by $(\sum_j |\psi_j|^\alpha)^{1/\alpha}$. (This quantity is finite, because $\{\psi_j\} \in l_p$ for $p \in [0,1]$.) Note that no generality is lost if one places a constant $C > 0$ on the right-hand side of equations (1) and (2); this constant is called the "dispersion" of Z_t (written $\mathrm{disp}(Z_t)$). If Z_t has the $S_\alpha(\sigma, \beta, \mu)$ law, i.e., Z_t is α-stable with scale σ, skewness β, and location μ (see Samorodnitsky and Taqqu, 1994), then we can take $p = (\beta + 1)/2$ and $q = (\beta - 1)/2$, and the dispersion is

$$C = \begin{cases} \sigma/\Gamma(1-\alpha)\cos\dfrac{\pi\alpha}{2} & \alpha \neq 1 \\ 2\sigma/\pi & \alpha = 1 \end{cases}.$$

However, for simplicity we will assume $C = 1$ throughout this paper.

There are a few facts about the choice of a_n: first, the sequence should satisfy

$$n\mathbb{P}[|Z_1| > a_n x] \to x^{-\alpha} \tag{3}$$

as $n \to \infty$ for every positive x. It is easy to check that $a_n := \inf\{x : \mathbb{P}[|Z_1| > x] \leq n^{-1}\}$ satisfies this condition. Condition (3) will be very important in what follows. A change of variable argument, using (1), implies that $a_n = n^{1/\alpha}L(n)$ (not necessarily the same slowly varying function in (1)); thus the "normal" domain of attraction has $L \equiv 1$. Given this, a suitable choice for b_n is $\mathbb{E}[Z_1; |Z_1| \leq a_n]$. This definition is interesting—it suggests a "natural" truncation for Z_1.

Notice that because $\{\psi_j\} \in l_p$, they are also in l_α because $p \leq 1 < \alpha$, so $(\sum_j |\psi|^\alpha)^{1/\alpha} < \infty$. The following notation will be used: Ψ will denote the whole sequence of $\{\psi_j\}$, and Ψ_p will denote its l_p norm. It is true that $\{X_t\}$ forms a strictly stationary sequence, because applying a shift operator to the law for the Z-series does not affect the distribution. Now because we are taking $\alpha > 1$, the mean does exist, and we shall call it $\eta := \mathbb{E}(Z_t)$. Thus $\mathbb{E}X_t = \psi_\infty \cdot \eta =: \theta$, where $\psi_\infty := \sum_{j\in\mathbb{Z}} \psi_j$.

We are interested in the following "self-normalized" statistic:

$$S_n := \frac{\sum_{t=1}^{n} (X_t - \theta)}{\sqrt{\sum_{t=1}^{n} X_t^2}}.$$

The term *self-normalized* refers to the fact that the denominator essentially estimates the convergence rate a_n of the numerator. If S_n converges to a nondegenerate limit with continuous c.d.f., then we can develop confidence intervals for θ via subsampling techniques (see Corollary 2, which follows).

When α is not less than 2, the situation is quite different. From the central limit theorem, a partial sum of any i.i.d. collection of $\mathbb{L}_2$ random variables has a stable limit, which is the normal distribution (this is the $\alpha = 2$ stable law). Thus $D(2)$ contains all $\mathbb{L}_2$ random variables, but it also contains some infinite variance random variables. By contrast, the other domains of attraction were characterized by tail behavior. Our general assumption on F is that either it has finite variance (and obeys the central limit theorem) or it has infinite variance and actually lies in $D(\alpha)$ for $\alpha \in (1,2]$.

2.2. Asymptotic Behavior of the Self-Normalized Statistic from Dependent Data

In this section we investigate the limiting behavior of S_n, which was defined in the previous section to be

$$S_n := \frac{\sum_{t=1}^{n} (X_t - \theta)}{\sqrt{\sum_{t=1}^{n} X_t^2}}.$$

The following theorem gives the main result of the paper.

THEOREM 1. *Consider the case that F is heavy tailed, such that it is in $D(\alpha)$ with $\alpha \in (1,2)$, and consider a linear time series with summable filter coefficients. In addition assume conditions (1)–(3) hold also. Then:*

$$S_n \stackrel{\mathcal{L}}{\Rightarrow} \frac{\psi_\infty}{\Psi_2} \frac{S}{\sqrt{S_2}} =: S(2),$$

where S is an α-stable random variable and S_2 has a totally right skewed (thus positive) $\alpha/2$-stable law; in particular, the limit random variable has a continuous nondegenerate distribution. In the case that F is heavy tailed but $\alpha = 2$, *then the same result holds if we center the denominator of S_n, i.e., replace X_t by $X_t - \overline{X}$.*

Remark. Resnick (1986) proves a similar result for i.i.d. data (which is the special case where $\Psi = \{\ldots 0, \psi_0 = 1, 0, \ldots\}$), under the hypotheses that the numerator is uncentered and $\alpha \in (0,1)$. If Z is symmetric, his result holds for $\alpha = 1$ also. So the preceding theorem is the natural generalization of Resnick's result to a linear time series with $\alpha \in (1,2)$.

Define the following "centered version" of S_n:

$$T_n := \frac{\sum_{t=1}^{n}(X_t - \theta)}{\sqrt{\sum_{t=1}^{n}(X_t - \overline{X})^2}} = \sqrt{n}\,\frac{\overline{X} - \theta}{\hat{\sigma}_n},$$

where $\hat{\sigma}_n := \sqrt{(1/n)\sum_{t=1}^{n}(X_t - \overline{X})^2}$ is an estimate of scale. Then we obtain the following corollary that includes the regular finite variance case as a possibility.

COROLLARY 1. *Let F be either a heavy-tailed distribution with $\alpha \in (1,2]$ or an $\mathbb{L}_2$ random variable and consider a linear time series with summable filter coefficients. So in the case that F is heavy tailed, assume conditions (1)–(3) hold. Then $T_n \stackrel{\mathcal{L}}{\Rightarrow} T$, where the random variable T is either ψ_∞/Ψ_2 times a standard normal (in the finite variance case) or is $S(2)$ (for the heavy-tailed case).*

2.3. Subsampling for the Self-Normalized Statistics

We now describe how the previous results can be implemented statistically via subsampling. Define the "subsampling distribution estimator" of T_n to be the following empirical distribution function (e.d.f.):

$$K_b(x) := \frac{1}{n-b+1}\sum_{i=1}^{n-b+1} 1_{\{T_{b,i} \le x\}},$$

where $T_{b,i}$ is essentially the statistic T_b evaluated on the subseries $\{X_i, \ldots, X_{b+i-1}\}$ (but with the unknown θ replaced by the estimate $\overline{X}_n$) ; in other words,

$$T_{b,i} := b^{1/2}\,\frac{\overline{X}_{b,i} - \overline{X}_n}{\hat{\sigma}_{b,i}}.$$

The precise definitions of $\overline{X}_{b,i}$ and $\hat{\sigma}_{b,i}$ are as follows:

$$\overline{X}_{b,i} := \frac{1}{b}\sum_{t=i}^{b+i-1} X_t,$$

$$\hat{\sigma}_{b,i} := \sqrt{\frac{1}{b}\sum_{t=i}^{b+i-1} (X_t - \overline{X}_{b,i})^2}.$$

The object of subsampling is to use the subsampling distribution estimator as an approximation of the limit distribution in Corollary 1; it also approximates the sampling distribution. For more details and background on these methods, see Politis et al. (1999).

Strong mixing is a sufficient condition on the dependence structure that insures the validity of subsampling. The strong mixing assumption requires that $\alpha_X(k) \to 0$ as $k \to \infty$; here $\alpha_X(k) := \sup_{A,B} |\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$, where A and B are events in the σ-fields generated by $\{X_t, t \leq l\}$ and $\{X_t, t \geq l + k\}$, respectively, for any $l \geq 0$. For references on strong mixing, see Rosenblatt (1956, 1984, 1985). General conditions for a linear series to be strong mixing are given by Withers (1981); they require that the ψ_j tend to zero fast enough (with j) and that the Z_t's have an absolutely continuous distribution. The strong mixing condition is easily seen to be satisfied if the series is an MA(m) model for some $m \in \mathbb{N}$, i.e., when only a finite number of the filter coefficients ψ_j are nonzero. In addition, if the series has an AR(1) representation, i.e., if

$$X_t = \psi X_{t-1} + Z_t$$

for some ψ bounded by one in absolute value, then the time series is strong mixing (see Pham and Tran, 1985). The following corollary is a standard application of subsampling to the sample mean problem.

COROLLARY 2. *Let F be either a heavy-tailed distribution with $\alpha \in (1,2)$ or an $\mathbb{L}_2$ random variable and consider a strong mixing linear time series with summable filter coefficients. So in the case that F is heavy tailed, assume conditions (1)–(3) hold. Then the subsampling distribution estimator K_b is consistent as an estimator of the true sampling distribution of T_n, denoted by $J_n(x) = \mathbb{P}\{T_n \leq x\}$. In other words, if $b \to \infty$ as $n \to \infty$ but with $b/n \to 0$, we have*

$$\sup_x |K_b(x) - J_n(x)| \xrightarrow{P} 0$$

and in addition

$$K_b^{-1}(t) \xrightarrow{P} J^{-1}(t)$$

for any $t \in (0,1)$; here $G^{-1}(t)$ is the t-quantile of distribution G, i.e., $G^{-1}(t) := \inf\{x : G(x) \geq t\}$.

From this result, we may use $K_b(x)$ as a c.d.f. from which to draw quantiles and develop confidence intervals with asymptotic veracity. This is done as follows:

$$1 - t = \mathbb{P}[J_n^{-1}(t/2) \le T_n \le J_n^{-1}(1-t/2)] \approx \mathbb{P}[K_b^{-1}(t/2) \le T_n \le K_b^{-1}(1-t/2)]$$
$$= \mathbb{P}\left[\bar{X} - K_b^{-1}(1-t/2)\,\frac{\hat{\sigma}}{\sqrt{n}} \le \theta \le \bar{X} - K_b^{-1}(t/2)\,\frac{\hat{\sigma}}{\sqrt{n}}\right],$$

where $1 - t$ is the confidence level. Thus the approximate equal-tailed confidence interval for θ is

$$\left[\bar{X} - K_b^{-1}(1-t/2)\cdot\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} - K_b^{-1}(t/2)\cdot\frac{\hat{\sigma}}{\sqrt{n}}\right].$$

The good thing about this procedure is that the rate a_n need neither be known nor estimated to form the preceding confidence interval. If the $\{Z_t\}$ sequence were in the normal domain of attraction, then $a_n = n^{1/\alpha}$. The problem with using this rate is that the exact value of α is unknown. One can either estimate α empirically or use self-normalization. The Hill estimator may be used to first estimate α, but its implementation requires the choice of a bandwidth parameter and is not robust under dependence (cf. Resnick, 1997). Also see Quintos, Fan, and Phillips (2000) and Resnick and Starica (1998) for some theory on the Hill estimator for dependent time series. However, in the case that the domain of attraction is not normal, there will also be a slowly varying function to estimate, which is a problem of a higher order of difficulty. Providentially, self-normalization avoids the need to estimate a_n; in essence, self-normalization implicitly performs the estimation.

2.4. The Symmetrized Situation

We now consider a special subcase wherein the statistic has a symmetric limit distribution. This will permit us to make use of a symmetrized version of the subsampling distribution estimator. In the finite variance case this will give a faster rate of convergence, although still being robust under the case that the data are heavy tailed. Thus if a practitioner believes that the data have balanced tails (i.e., that $p = q = \frac{1}{2}$ in equation (2)) and thinks that the variance exists, then the sample mean will be an optimal estimate (in terms of maximum likelihood criterion) of location, and the symmetrized subsampling distribution estimator will converge faster. But if the data really are heavy tailed, this estimator will still be consistent and will function properly.

Throughout this section, we assume that the inputs $\{Z_t\}$ are heavy-tailed random variables with tail thickness α that have *balanced* tails or in other words that $p = q = .5$ in equation (2). Equivalently, one could specify that the inputs are in the α domain of attraction of a *symmetric* α-stable distribution. For purposes of modeling, this is not an unreasonable assumption. Note that we do *not* assume that the Z_t's (or equivalently the X_t's) have a symmetric distribution, in which case the sample median might be a good measure of location; our assumption is on the tails of the distribution only. The following result specializes Theorem 1 to the symmetric case.

THEOREM 2. *Let F be either a heavy-tailed distribution with $\alpha \in (1,2]$ and balanced tails, or an $\mathbb{L}_2$ random variable, and consider a linear time series with summable filter coefficients. So in the case that F is heavy tailed, assume conditions (1)–(3) hold with $p = q = .5$. Then $T_n \stackrel{\mathcal{L}}{=} T$ as in Corollary 1, and the limit T is symmetric about zero.*

We now discuss the regular case, in which the variance is finite and perhaps even moments of order higher than two exist. Consider the following moment and mixing conditions:

$$\alpha_X(k) \le d^{-1}e^{-dk} \quad \text{for some } d > 0 \tag{4}$$

and

$$E|X_0|^s < \infty \quad \text{for some } s \ge 5. \tag{5}$$

If (4) and (5) hold true, and if the Cramér-type regularity conditions (2.3), (2.5) and (2.6) of Götze and Hipp (1983) also hold true, then the following Edgeworth expansion holds uniformly in x:

$$P(T_n \le x) = \Phi(x/|\psi_\infty|) + n^{-1/2}p(x) + O(n^{-1}), \tag{6}$$

where Φ is the standard normal c.d.f. and the $p(x)$ is a well-defined smooth function. If in addition to the preceding conditions, the conditions $A2[2]$, $\alpha 3[2]$, and $\alpha 4[2]$ of Bertail (1997) hold true, then we also have

$$K_b(x) = \Phi(x/|\psi_\infty|) + b^{-1/2}p(x) + O_P(b^{-1}). \tag{7}$$

To make use of the higher order terms in the regular case discussed previously, we will employ a notion of "robust interpolation" that was first proposed in Politis et al. (1999) in the context of i.i.d. data. For this purpose, define

$$K_b^{symm}(x) := \frac{K_b(x) + 1 - K_b(-x)}{2}$$

and

$$K_b^{rob}(x) := \sqrt{\frac{b}{n}}K_b(x) + \left(1 - \sqrt{\frac{b}{n}}\right)K_b^{symm}(x).$$

The next result shows the efficacy of K_b^{rob}.

COROLLARY 3. *Assume that $b \to \infty$ and $(b/n) \to 0$ as $n \to \infty$. Then we have*

$$\sup_x |K_b^{rob}(x) - J_n(x)| = o_P(1)$$

as $n \to \infty$. If it so happens that the inputs Z_t are i.i.d. $\mathbb{L}_2$ random variables, and the two Edgeworth expansions (6) and (7) hold true, then in addition we have

$$\sup_x |K_b^{rob}(x) - J_n(x)| = O_P(1/b). \tag{8}$$

Recall that equation (6) implies that

$$\sup_x |\Phi(x/\psi_\infty) - J_n(x)| = O\left(\frac{1}{\sqrt{n}}\right),$$

so that the error in the normal approximation to $J_n(x)$ is of order $O(1/\sqrt{n})$. Now observe that, if we let $\sqrt{n} = o(b)$ (and $b = o(n)$ as always), equation (8) implies

$$\sup_x |K_b^{rob}(x) - J_n(x)| = o_P\left(\frac{1}{\sqrt{n}}\right), \tag{9}$$

showing that $K_b^{rob}(x)$ exemplifies a higher order accuracy as compared to the normal approximation (when they are both applicable) under the corollary's assumptions.

However, $K_b^{rob}(x)$ *remains* consistent even when the normal approximation breaks down, e.g., in the presence of heavy tails; thus the title "robust" is justified for $K_b^{rob}(x)$. Another attractive feature of $K_b^{rob}(x)$ versus the normal approximation (when the latter is applicable) is that the quantity ψ_∞ must be explicitly estimated for the normal approximation to be used in practice; subsampling does not suffer from this difficulty as it implicitly (and automatically) provides an estimate of ψ_∞.

3. PROOFS

3.1. Proof of Theorem 1

The following discussion uses standard techniques (cf. Davis and Resnick, 1985, 1986). We have divided the proof into two lemmas, which handle the numerator and denominator, respectively.

LEMMA 1. *If $Z_t \sim$ i.i.d. F, then*

$$a_n^{-1}\sum_{t=1}^{n}(X_t-\theta)=o_P(1)+\psi_\infty a_n^{-1}\sum_{t=1}^{n}(Z_t-\eta).$$

Remark. From this it follows that the left-hand side has an α-stable weak limit.

LEMMA 2. *If $Z_t \sim$ i.i.d. $F \in D(\alpha)$ and $\alpha \in (1,2)$, then*

$$a_n^{-2}\sum_{t=1}^{n}X_t^2=o_P(1)+\frac{\Psi_2^2}{a_n^2}\sum_{t=1}^{n}(Z_t-\eta)^2.$$

If $\alpha = 2$, we must replace X_t by $X_t-\overline{X}$ for the result to be valid.

Remark. From this it follows that the left-hand side has a positive $\alpha/2$-stable weak limit.

Proof of Lemma 1. First restrict to the MA$(2m+1)$ case, where $\psi_j = 0$ unless $-m \le j \le m$. Then

$$a_n^{-1}\sum_{t=1}^{n}(Z_{t-j}-\eta)=o_P(1)+a_n^{-1}\sum_{t=1}^{n}(Z_t-\eta)\quad \forall -m\le j\le m$$

(namely, Davis and Resnick, 1985, p. 190). If we take the dot product of this vector with $(\psi_{-m},\dots,\psi_0,\dots,\psi_m)$ (asymptotic equivalence is preserved because dot product is a continuous mapping), we obtain

$$a_n^{-1}\sum_{t=1}^{n}\sum_{j=-m}^{m}\psi_j(Z_{t-j}-\eta)=o_P(1)+a_n^{-1}\sum_{j=-m}^{m}\psi_j\sum_{t=1}^{n}(Z_t-\eta).$$

We let $m\to\infty$ on both sides. On the right-hand side we get

$$o_P(1)+\psi_\infty a_n^{-1}\sum_{t=1}^{n}(Z_t-\eta).$$

On the left-hand side there is some justification to be made (again see Davis and Resnick, 1985, pp. 190–191), but it is valid to take the limit as $m\to\infty$, and we arrive at the statement of Lemma 1. ■

Proof of Lemma 2. It is easy to show that

$$a_n^{-2}\sum_{t=1}^{n}X_t^2=o_P(1)+a_n^{-2}\sum_{t=1}^{n}(X_t-\theta)^2$$

using the fact that $\alpha < 2$. If $\alpha = 2$, we obtain the same equivalence if X_t is replaced by $X_t-\overline{X}$. For the calculation, see the proof of Corollary 1, which follows. Expanding the right-hand side in terms of Z_t, one obtains two terms:

$$a_n^{-2} \sum_{t=1}^{n} \sum_{j \in \mathbb{Z}} \psi_j^2 (Z_{t-j} - \eta)^2$$

and

$$a_n^{-2} \sum_{t=1}^{n} \sum_{i \neq j} \psi_i \psi_j (Z_{t-j} - \eta)(Z_{t-i} - \eta).$$

The second term is $\mathbb{L}_1$ bounded by a constant times $(n/a_n^2)\Psi_1^2$, which tends to zero. The first term, with arguments similar to Lemma 1, is

$$= o_P(1) + \frac{\Psi_2^2}{a_n^2} \sum_{t=1}^{n} (Z_t - \eta)^2,$$

which proves the lemma. ■

Implementing these results, we see that

$$\left(a_n^{-1} \sum_{t=1}^{n} (X_t - \theta), a_n^{-2} \sum_{t=1}^{n} X_t^2 \right) = o_P(1) + \left(a_n^{-1} \psi_\infty \sum_{t=1}^{n} (Z_t - \eta), a_n^{-2} \Psi_2^2 \sum_{t=1}^{n} (Z_t - \eta)^2 \right).$$

By now applying the continuous function $f(x, y) := x/\sqrt{y}$ to this convergence result, we obtain

$$S_n = o_P(1) + \frac{\psi_\infty a_n^{-1} \sum_{t=1}^{n} (Z_t - \eta)}{\Psi_2 \sqrt{a_n^{-2} \sum_{t=1}^{n} (Z_t - \eta)^2}}.$$

Now the right-hand side converges by the Logan et al. (1973) result, because it only involves the i.i.d. inputs $\{Z_t\}$. The weak limit of the right-hand side is precisely

$$\frac{\psi_\infty}{\Psi_2} \frac{S}{\sqrt{S_2}} =: S(2),$$

and this proves the theorem. ■

3.2. Proofs of Corollaries 1 and 2

Proof of Corollary 1. The notations $A \overset{\mathbb{P}}{\sim} B$ will mean $A = o_P(1) + B$ (which is a reflexive and symmetric relation). We have the following equivalences in probability:

$$\frac{\sum_{t=1}^{n}(X_t-\theta)}{\sqrt{\sum_{t=1}^{n}(X_t-\overline{X})^2}} \overset{\mathbb{P}}{\sim} \frac{\sum_{t=1}^{n}(X_t-\theta)}{\sqrt{\sum_{t=1}^{n}(X_t-\theta)^2}} \overset{\mathbb{P}}{\sim} \frac{\sum_{t=1}^{n}(X_t-\theta)}{\sqrt{\sum_{t=1}^{n}X_t^2}},$$

where the second equivalence is only true in the heavy-tailed case. This statement is really two claims: the first claim is the left equivalence, and the second claim is the right equivalence.

First Claim. Let $U_n := a_n^{-1}\sum_{t=1}^{n} Y_t = (n/a_n)\overline{Y}_n$ where $Y_t := X_t - \theta$. It follows from Theorem 1 that $U_n = O_P(1)$. Hence

$$\begin{aligned} a_n^{-2}\sum_{t=1}^{n}(X_t-\overline{X}_n)^2 &= a_n^{-2}\sum_{t=1}^{n}((X_t-\theta)+(\theta-\overline{X}_n))^2 \\ &= a_n^{-2}\sum_{t=1}^{n}(X_t-\theta)^2 + \frac{-1}{n}U_n^2. \end{aligned}$$

Thus

$$a_n^{-2}\sum_{t=1}^{n}(X_t-\overline{X}_n)^2 = O_P\left(\frac{1}{n}\right) + a_n^{-2}\sum_{t=1}^{n}(X_t-\theta)^2,$$

which also holds true for the joint law:

$$\begin{aligned} \left(a_n^{-1}\sum_{t=1}^{n}(X_t-\theta), a_n^{-2}\sum_{t=1}^{n}(X_t-\overline{X}_n)^2\right) &= O_P\left(\frac{1}{n}\right) \\ &\quad + \left(a_n^{-1}\sum_{t=1}^{n}(X_t-\theta), a_n^{-2}\sum_{t=1}^{n}(X_t-\theta)^2\right). \end{aligned}$$

Applying the continuous function $f(x,y) := x/\sqrt{y}$ finishes the claim.

Second Claim. Here we suppose that F is heavy tailed with $1<\alpha<2$. Then

$$a_n^{-2}\sum_{t=1}^{n}(X_t-\theta)^2 = a_n^{-2}\sum_{t=1}^{n}X_t^2 - \frac{2\theta}{a_n^2}\sum_{t=1}^{n}X_t + \frac{n\theta^2}{a_n^2},$$

the second term is $-(2\theta/a_n)U_n = O_P(a_n^{-1}) = o_P(1)$, and the third term is $n^{1-(2/\alpha)}L(n)^{-1} = o(1)$, where $L(\cdot)$ is slowly varying. Note that we need $\alpha < 2$ for the last statement to hold true. Therefore the claim holds, by applying f again to the joint vector.

Now the case of Corollary 1 where $\alpha \in (1,2)$ is clear, because $T_n \overset{\mathbb{P}}{\sim} S_n$ from the claims; so we can apply Theorem 1 to get the desired result. Now suppose that the variance is finite, i.e., $Z_t \in \mathbb{L}_2$. Suppose that Z_t has variance σ^2; then by Theorem 7.1.2 of Brockwell and Davis (1991), we know that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \theta) \stackrel{\mathcal{L}}{=} \psi_\infty \sigma N,$$

where N is a standard normal random variable. As for the denominator, we have

$$\frac{1}{n} \sum_{t=1}^{n} (X_t - \theta)^2 \xrightarrow{P} \Psi_2^2 \sigma^2$$

by Proposition 7.3.5 of Brockwell and Davis (1991). Hence, using Slutsky's theorem (note that the preceding limit is nonzero!),

$$T_n \stackrel{\mathbb{P}}{\sim} \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (X_t - \theta)}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (X_t - \theta)^2}}$$

$$\stackrel{\mathcal{L}}{\Longrightarrow} \frac{\psi_\infty}{\Psi_2 \sigma} \sigma N = \frac{\psi_\infty}{\Psi_2} N.$$

This proves the corollary. ∎

Proof of Corollary 2. This result follows immediately from our Theorem 1 and Theorem 12.2.2 of Politis et al. (1999) for self-normalized statistics under strong mixing—one of our standing hypotheses on the time series. Once we write T_n in the form $\tau_n[(\hat{\theta} - \theta)/\hat{\sigma}]$, we see that $T_n = \sqrt{n}[(\overline{X} - \theta)/\hat{\sigma}_n]$ so that $\tau_n = \sqrt{n}$. Hence $\tau_b/\tau_n = \sqrt{b/n} \to 0$, which is an important condition of Theorem 12.2.2 in Politis et al. (1999). Finally, we note that the limit law in Theorem 1 has no point masses, because the random variables S and S_2 have continuous c.d.f.s. Thus the limit c.d.f. $J(x)$ is continuous, and the stated sup-norm convergence holds. ∎

3.3. Proof of Theorem 2

Henceforth "symmetric" will mean symmetric about zero, unless another center is specified. Now in the finite variance case, the limit is normal and hence is symmetric, and the theorem holds (and the tails need not be balanced). So now we examine the case where F is heavy tailed with $1 < \alpha < 2$ with balanced tails.

The first assertion—convergence—is true from Corollary 1, as our hypotheses are only stronger. From the proof of Theorem 1 we may write T_n in the following manner:

$$T_n = o_P(1) + \frac{\psi_\infty a_n^{-1} \sum_{t=1}^{n} (Z_t - \eta)}{\Psi_2 \sqrt{a_n^{-2} \sum_{t=1}^{n} (Z_t - \eta)^2}}.$$

The limit of the numerator has skewness equal to $(p - q)\text{sign}(\sum_{j\in\mathbb{Z}} \psi_j)$ (see Davis and Hsing, 1995, Example 5.5); here p and q are the tail-balance parameters defined in equation (2). The limit random variable will also have location zero, as desired (this is independent of the tail-balance issue). The denominator does not affect skewness, because it and its limit are positive random variables. So we obtain a symmetric limit by imposing that F is in a *symmetric* α-stable domain of attraction or equivalently by demanding that $p = q$. This proves the theorem. ∎

Proof of Corollary 3. This follows from our Theorem 2 and Theorem 10.3.1 of Politis et al. (1999).

4. SIMULATION STUDIES

Herein we present the results of simulating the linear time series model. We specialize to having F be the c.d.f. for an sαs random variable; for simplicity we take η (and thus θ too) to be zero. Two AR(1) processes and an MA(11) were considered separately, with the parameter α ranging between 1.2, 1.5, and 1.8. Then 10,000 time series were generated of length 100 and 1,000, each for a different subsampling length. The length b of the subsampling block was varied between 1 and 25. Confidence intervals for the mean (see Sect. 2.3) were developed from the computed subsampling distribution estimator at the 90%, 95%, and 99% confidence levels, and we recorded the fraction of times that the constructed interval contained the mean of zero. We have done this for each subsampling length, and Tables 1–3 give the accuracy for certain values of b.

The AR(1) processes are simply

$$X_t = .5\,X_{t-1} + Z_t,$$

$$X_t = .9\,X_{t-1} + Z_t,$$

respectively, for $t = 1,2,\ldots,n$, and $X_0 := Z_0$; the sequence $\{Z_t\}$ are i.i.d. sαs random variables with zero mean. It is elementary that the preceding time series can also be written as

$$X_t = \sum_{j=0}^{\infty} (0.5)^j Z_{t-j}\, X_t = \sum_{j=0}^{\infty} (0.9)^j Z_{t-j},$$

which fits our model assumptions in this paper.

TABLE 1. AR(1) model

$n = 100$; $\alpha = 1.2$				$n = 1{,}000$; $\alpha = 1.2$			
b	90%	95%	99%	b	90%	95%	99%
1	.9670	.9820	1.0000	1	.9453	.9456	.9519
2	.8679	.9294	.9905	4	.7425	.8214	.9138
4	.7269	.8355	.9133	8	.6624	.7429	.9135
8	.6896	.7317	.8005	12	.6423	.7105	.8777
12	.6490	.6920	.7424	16	.6273	.7004	.8635
16	.6114	.6480	.6939	24	.6193	.6891	.7962
$n = 100$; $\alpha = 1.5$				$n = 1{,}000$; $\alpha = 1.5$			
b	90%	95%	99%	b	90%	95%	99%
1	.9912	.9935	.9999	1	.9508	.9633	.9995
2	.9686	.9820	.9994	4	.8982	.9371	.9569
4	.8735	.9359	.9772	8	.8144	.8766	.9820
8	.7829	.8325	.8905	12	.7941	.8568	.9710
12	.7378	.7841	.8395	16	.7639	.8334	.9493
16	.7106	.7526	.8066	24	.7514	.8121	.9086
$n = 100$; $\alpha = 1.8$				$n = 1{,}000$; $\alpha = 1.8$			
b	90%	95%	99%	b	90%	95%	99%
1	.9925	.9927	1.0000	1	.9999	1.0000	1.0000
2	.9930	.9937	1.0000	4	.9510	.9788	.9980
4	.9544	.9841	.9953	8	.8969	.9436	.9940
8	.8709	.9130	.9541	12	.8656	.9243	.9839
12	.8251	.8668	.9179	16	.8540	.9071	.9750
16	.7949	.8393	.8916	24	.8395	.8949	.9568

Note: Entries are empirical averages of the equal-tailed subsampling confidence intervals for the mean; the nominal (target) coverage level is given at the top of each column. These data were generated by an AR(1) model of parameter .5.

The MA(11) process is

$$X_t = \sum_{j=0}^{10} \psi_j Z_{t+j}$$

for $t = 1, 2, \ldots, n$, and for coefficients $\{\psi_j\}$ equal to

0.03, 0.05, 0.07, 0.10, 0.15, 0.20, 0.15, 0.10, 0.07, 0.05, 0.03.

Both models are strong mixing (see the discussion in Sect. 2.3).

TABLE 2. AR(1) model

$n = 100$; $\alpha = 1.2$				$n = 1{,}000$ $\alpha = 1.2$			
b	90%	95%	99%	b	90%	95%	99%
1	.9681	.9827	.9999	1	.9393	.9399	.9461
2	.8656	.9274	.9941	4	.7376	.8201	.9147
4	.7312	.8429	.9133	8	.6647	.7418	.9176
8	.6886	.7352	.7959	12	.6407	.7154	.8904
12	.6438	.6825	.7360	16	.6291	.7062	.8682
16	.6108	.6458	.6972	24	.6183	.6914	.8057
$n = 100$; $\alpha = 1.5$				$n = 1{,}000$; $\alpha = 1.5$			
b	90%	95%	99%	b	90%	95%	99%
1	.9912	.9935	.9999	1	.9463	.9572	.9991
2	.9724	.9851	.9991	4	.8951	.9311	.9509
4	.8717	.9339	.9747	8	.8230	.8827	.9877
8	.7882	.8330	.8955	12	.7899	.8558	.9709
12	.7427	.7836	.8402	16	.7723	.8402	.9520
16	.7066	.7460	.8000	24	.7554	.8164	.9186
$n = 100$; $\alpha = 1.8$				$n = 1{,}000$; $\alpha = 1.8$			
b	90%	95%	99%	b	90%	95%	99%
1	.9951	.9955	1.0000	1	.9999	1.0000	1.0000
2	.9941	.9950	.9999	4	.9494	.9771	.9990
4	.9510	.9811	.9961	8	.8861	.9400	.9951
8	.8734	.9150	.9595	12	.8632	.9158	.9840
12	.8257	.8677	.9185	16	.8465	.9016	.9735
16	.7882	.8319	.8856	24	.8313	.8885	.9595

Note: Entries are empirical averages of the equal-tailed subsampling confidence intervals for the mean; the nominal (target) coverage level is given at the top of each column. These data were generated by an AR(1) model of parameter .9.

These simulation studies show that the optimal block size b depends both on α and on the degree of dependence. The MA(11) model has less dependence than the AR(1) model, but the optimal block lengths were larger for the former model. In general, lower values of α resulted in undercoverage, especially in the AR(1) model; the optimal block size tended to be quite small (sometimes 1 or 2). For $\alpha = 1.8$, the coverage was optimized for larger values of b. Thus there appears to be some relationship between α and the optimal b. The MA(11) model had coverages closer to the true percentage and thus performed better

Table 3. MA(11) model

$n = 100$; $\alpha = 1.2$				$n = 1{,}000$; $\alpha = 1.2$			
b	90%	95%	99%	b	90%	95%	99%
1	.9992	.9998	.9999	1	.9446	.9464	1.0000
2	.9973	.9995	.9998	4	.8903	.9381	.9693
4	.9748	.9956	.9973	8	.8239	.9032	.9988
8	.8981	.9648	.9746	12	.7707	.8579	.9888
12	.8483	.8944	.9100	16	.7285	.8179	.9697
16	.7768	.8183	.8307	24	.6906	.7748	.9321
$n = 100$; $\alpha = 1.5$				$n = 1{,}000$; $\alpha = 1.5$			
b	90%	95%	99%	b	90%	95%	99%
1	.9995	.9999	1.0000	1	.9999	1.0000	1.0000
2	.9991	.9995	1.0000	4	.9423	.9439	.9999
4	.9860	.9982	.9998	8	.9251	.9401	.9998
8	.9284	.9740	.9891	12	.8984	.9280	.9991
12	.8919	.9267	.9492	16	.8650	.9127	.9949
16	.8286	.8645	.8899	24	.8255	.8833	.9827
$n = 100$; $\alpha = 1.8$				$n = 1{,}000$; $\alpha = 1.8$			
b	90%	95%	99%	b	90%	95%	99%
1	.9999	1.0000	1.0000	1	.9831	.9949	1.0000
2	.9998	1.0000	1.0000	4	.9696	.9811	1.0000
4	.9942	.9988	.9999	8	.9606	.9710	1.0000
8	.9608	.9855	.9939	12	.9424	.9597	1.0000
12	.9186	.9452	.9639	16	.9240	.9521	.9992
16	.8765	.9035	.9226	24	.9041	.9430	.9959

Note: Entries are empirical averages of the equal-tailed subsampling confidence intervals for the mean; the nominal (target) coverage level is given at the top of each column. This is the case of data generated by an MA(11) model.

than the AR(1); thus the amount of dependence also seems to affect the optimal block size.

Our simulation results are encouraging but also demonstrate the need for a practical way of picking the subsampling size b because the finite-sample performance of the confidence intervals depends on b. Some insights on the theory and practice of optimally picking b are available in the regular case of finite second moments (see Hall, Horowitz, and Jing, 1996; Politis et al., 1999). Little is known in the case treated here where dependence, heavy tails, and self-normalization complicate the picture considerably. Future work will focus on this important topic.

REFERENCES

Athreya, K. (1987) Bootstrap of the mean in the infinite variance case. *Annals of Statistics* 15, 724–731.

Beran, J. (1994) *Statistics for Long-Memory Processes*. New York: Chapman and Hall.

Beran, J., R. Sherman, M. Taqqu, & W. Willinger (1995) Long-range dependence in variable-bit rate video traffic. *IEEE Transactions and Communications* 43, 1566–1579.

Bertail, P. (1997) Second order properties of an extrapolated bootstrap without replacement: The i.i.d. and strong mixing cases. *Bernoulli* 3, 149–179.

Bertail, P., D. Politis, & J. Romano (1999) On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association* 94, 569–580.

Brockwell, P. & R. Davis (1991) *Time Series: Theory and Methods*. New York: Springer.

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.

Davis, R. & T. Hsing (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability* 23, 879–917.

Davis, R. & S. Resnick (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* 13, 179–195.

Davis, R. & S. Resnick (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558.

Götze, F. & C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 211–239.

Hall, P. (1990) Asymptotic properties of the bootstrap for heavy-tailed distributions. *Annals of Probability* 18, 1342–1360.

Hall, P., J.L. Horowitz, & B.-Y. Jing (1996) On blocking rules for the bootstrap with dependent data. *Biometrika* 50, 561–574.

Knight, K. (1989) On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics* 17, 1168–1175.

LePage, R., M. Woodroofe, & J. Zinn (1981) Convergence to a stable distribution via order statistics. *Annals of Probability* 9, 624–632.

Logan, B.F., C.L. Mallows, S.O. Rice, & L.A. Shepp (1973) Limit distributions of self-normalized sums. *Annals of Probability* 1, 788–809.

Pham, T.D. & L.T. Tran (1985) Some mixing properties of time series models. *Stochastic Processes and Their Applications* 19, 297–303.

Phillips, P. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Politis, D. & J. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22, 2031–2050.

Politis, D., J. Romano, & M. Wolf (1999) *Subsampling*. New York: Springer.

Quintos, C., Z. Fan, & P. Phillips (2000) Structural Change in Tail Behavior and the Asian Financial Crisis. Preprint, New York University.

Resnick, S. (1986) Point processes, regular variation, and weak convergence. *Advances in Applied Probability* 18, 66–138.

Resnick, S. (1987) *Extreme Values, Regular Variation, and Point Processes*. New York: Springer–Verlag.

Resnick, S. (1997) Special Invited Paper: Heavy tail modeling and teletraffic data. *Annals of Statistics* 25, 1805–1849.

Resnick, S. & C. Starica (1998) Tail index estimation for dependent data. *Annals of Applied Probability* 8, 1156–1183.

Romano, P. & M. Wolf (1999) Subsampling inference for the mean in the heavy-tailed case. *Metrika* 50, 55–69.

Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences* 42, 43–47.

Rosenblatt, M. (1984) Asymptotic normality, strong mixing, and spectral density estimates. *Annals of Probability* 12, 1167–1180.

Rosenblatt, M. (1985) *Stationary Sequences and Random Fields*. Boston: Birkhäuser.

Samorodnitsky, G. & M. Taqqu (1994) *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.

Withers, C.S. (1981) Conditions for linear processes to be strong mixing. *Probability Theory and Related Fields* 57, 477–480.