SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1066

August 1993

CONSISTENCY OF HILL'S ESTIMATOR FOR DEPENDENT DATA

by
Sidney Resnick[1]
and Cătălin Stărică[2]

[1]Research supported by NSF grant DMS-9100027 at Cornell University. Some support was also received from NSA Grant 92G-116.
[2]Supported by NSF Grant DMS-9100027 at Cornell University

Consistency of Hill's estimator for dependent data

Sidney Resnick* Cătălin Stărică†
School of ORIE Cornell University Ithaca, NY 14853

August 17, 1993

Abstract

Consider a sequence of possibly dependent random variables having the same marginal distribution F, whose tail 1-F is regularly varying at infinity with an unknown index $-\alpha < 0$ which is to be estimated. For i.i.d. data or for dependent sequences with the same marginal satisfying mixing conditions, it is well known that Hill's estimator is consistent for α^{-1} and asymptotically normally distributed. The purpose of this paper is to emphasize the central role played by the tail empirical process for the problem of consistency. This approach allows us to easily prove Hill's estimator is consistent for infinite order moving averages of independent random variables. Our method also suffices to prove that, for the case of an AR model, the unknown index can be estimated using the residuals generated by the estimation of the autoregressive parameters.

1 Introduction.

The problem of estimating the tail probability P(X > x) of a random variable for large x has obvious practical significance in diverse fields such as finance, hydrology, reliability and teletraffic engineering.

In certain cases, collected data indicates that the random variable may be heavy tailed and the index of variation has to be estimated based on a sequence $X_1, X_2, \ldots, X_n$ of observations . A well studied estimator of the index, Hill's estimator, is known to be consistent and asymptotically normal for i.i.d. samples. However, many real life applications do not provide us with independent sequences but rather with dependent, stationary data. Therefore, there is considerable interest in studying the behavior of Hill's estimator under the more general assumption of stationarity or, even more generally, assuming just a common marginal distribution. Several recent papers (Hsing (1991), Rootzen et al. (1990)) support the belief that Hill's estimator performs well even under these weaker assumptions.

Previous studies have assumed mixing conditions of one type or another for the stationary time series. These conditions can be awkward to handle and verify. Another method of studying the tail behavior for independent data was proposed by Mason (1988) and Deheuvels and Mason (1991) and investigates the behavior of the tail empirical process. In this paper we emphasize the following approach to the problem: we associate the tail empirical random measure to the sequence

*Partially supported by NSF Grant DMS-9100027 at Cornell University. Some support was also received from NSA Grant 92G-116.

†Supported by NSF Grant DMS-9100027 at Cornell University.

$X_1, X_2, \ldots, X_n$ and show that the weak convergence of the tail empirical random measure implies the consistency of Hill's estimator. We then prove the consistency of Hill's estimator for an infinite moving average sequence whose marginal distribution is regularly varying.

Some of our considerations have been motivated by the following problem. Consider a p-th order autoregressive process $\{X_n, n \geq 0\}$ with residuals with regularly varying tail probabilities of index -α. Both the stationary sequence $\{X_n\}$ and the residuals have distributions with regularly varying tails of index -α. This suggests two possible methods of estimating α: (i) apply Hill's estimator to the observed time series $X_1, X_2, \ldots, X_n$ or (ii) assuming the order of the autoregression p is known, fit coefficients of the autoregression and use this to estimate residuals. Then estimate α by applying Hill's estimator to the estimated residuals. (Methods of estimating autoregressive coefficients in the heavy tailed case have been suggested by Davis and Resnick (1985), Feigin and Resnick (1992, 1993), Mikosch, Gadrich, Klüppelberg and Adler (1993).)

An important conclusion of our paper is that both methods yield consistent procedures. We make some comments about the desirability of the procedures in Section 5, where we discuss some examples of simulated and real data. Based on these examples, it appears that applying Hill's estimator to the estimated residuals is a more satisfactory procedure. In future work, we hope to compare efficiencies of the two methods of estimation.

We now give some basic notations and assumptions. Let $\{X_n\}$ be a sequence of random variables having the same marginal distribution function F, where $\overline{F}$:= 1 - F is regularly varying at ∞, i.e., there exists an $\alpha > 0$ such that

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha} \tag{1.1}$$

for all $x > 0$. We are interested in estimating α based on observing $X_1, X_2, \ldots, X_n$. If we set

$$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}, \qquad 0 < y < 1$$

and

$$b(t) := \left(\frac{1}{1-F}\right)^{\leftarrow}(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \qquad t > 1,$$

then regular variation implies

$$\overline{F}\,(b(t)) \sim t^{-1} \tag{1.2}$$

as t $\to \infty$. For $1 \leq$ i $\leq$ n, write $X_{(i)}$ for the ith largest value of $X_1, X_2, \ldots, X_n$. For x $\in \mathbf{R}$, x_+ denotes max(x, 0). Then, with this notation, Hill's estimator is defined as:

$$H_{X,n} := \frac{1}{k} \sum_{i=1}^{k} \log\, X_{(i)} - \log\, X_{(k+1)} \tag{1.3}$$

Asymptotic properties of H_n have been studied when k depends on n in such a way that n/k $\to \infty$ as n $\to \infty$. It is useful to define an associated random measure, the tail empirical measure, which is going to play a key role.

Let E:= $(0, \infty]$ be the one point uncompactification of $[0, \infty]$ so that the compact sets of E are of the form U^c, where U $\ni$ 0 is an open set in $[0, \infty)$. Suppose $\mathcal{E}$ is the Borel σ-field on E. Define the measure μ by:

$$\mu : \mathcal{E} \to \mathbf{R}_+, \qquad \mu((x, \infty]) = x^{-\alpha}, \qquad x > 0. \tag{1.4}$$

Let $M_+(E)$ be the space of positive Radon measures on E endowed with the vague topology (Resnick (1987), Kallenberg (1983)). Let $C_K^+(E)$ be the space of continous, non-negative functions on E = $(0, \infty]$ with compact support. The vague topology on $M_+(E)$ can be generated by a countable family of seminorms

$$H = \{p_f : M_+(E) \to R_+ : p_f(\mu) = \mu(f), |f| \leq 1, f \in C_K^+(E)\}$$

(Resnick(1987), Proposition 3.17, Lemma 3.11), turning M_+(E) into a complete, separable, metric space. Convergence of $\mu_n \in M_+(E)$ to $\mu_0 \in M_+(E)$ in the vague topology is denoted $\mu_n \xrightarrow{v} \mu_0$. For x∈E and A∈ $\mathcal{E}$ define

$$\epsilon_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \in A^c \end{cases}$$

Then, for k = k(n) ≤ n, define the tail empirical measure

$$\mu_{X,n} := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{X_i/b(n/k)} \tag{1.5}$$

so that $\mu_{X,n}$ is a random element of $M_+(E)$.

In Section 2 we discuss the general methodology of showing consistency of Hill's estimator which is to show that if $\mu_{X,n}$ converges vaguely in probability to μ, then $H_{X,n} \xrightarrow{P} \alpha^{-1}$. This emphasizes the relationship between random measures and Hill's estimator which is exploited in the rest of the paper. The consistency of Hill's estimator in the i.i.d. case is quickly recovered using this method (see Mason (1982)). Section 3 applies this method to show consistency of the Hill estimator when applied to processes of the form

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \qquad -\infty < n < \infty \tag{1.6}$$

where Z_k is i.i.d., P($Z_k > x$) is regularly varying with index -α and $Z_k \geq 0$. In particular, this result applies to an autoregressive process of the form

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + Z_n \tag{1.7}$$

since such a process (under proper assumptions) has a causal representation of the form (1.6) (cf. Brockwell and Davis (1991)). An alternative method of estimating α for the AR process (1.7) given a set of consistent estimators $\hat{\phi}_1^{(n)}, \hat{\phi}_2^{(n)}, \ldots, \hat{\phi}_p^{(n)}$ of the true coefficients is to apply Hill's estimator to the estimated residuals $\hat{Z}_1^{(n)}, \hat{Z}_2^{(n)}, \ldots, \hat{Z}_n^{(n)}$ defined as

$$\hat{Z}_i^{(n)} := X_i - \sum_{j=1}^{p} \hat{\phi}_j^{(n)} X_{i-j}, \qquad i = 1, \ldots, n. \tag{1.8}$$

In Section 4 we show this procedure is also consistent by showing the tail empirical measure corresponding to $\hat{Z}_1^{(n)}, \hat{Z}_2^{(n)}, \ldots, \hat{Z}_n^{(n)}$ consistently estimates μ.

It will be assumed throughout that the following hold:
Condition I: $\{X_i\}$ is a sequence of dependent random variables having the same marginal distribution F where $\overline{F}$ is regularly varying at ∞ with index $-\alpha$. The quantile function is b(t):= $(1/1-F)^{\leftarrow}(t)$, $t > 1$
Condition II: $\{Z_i\}$ is a sequence of i.i.d positive random variables with the common distribution G where $\overline{G}$ is also regularly varying at ∞ with index $-\alpha$. The quantile function is $\flat(t) := (1/1-G)^{\leftarrow}(t)$, $t > 1$.

Define also:

$$\hat{\mu}_{X,n} := \frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/X_{(k)}} \tag{1.9}$$

where we think of $X_{(k)}$ as being the estimator of b(n/k) and

$$H^{+}_{X,n} := \frac{1}{k}\sum_{i=1}^{n} (\log\, X_i - \log\, b(n/k))_{+}. \tag{1.10}$$

2 Random measures and the consistency of Hill's estimator.

The standing assumption in this section is that Condition I holds and that, as n $\to\infty$, k/n $\to$ 0,

$$\mu_{X,n} = \frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/b(n/k)} \Rightarrow \mu \tag{2.1}$$

in $M_{+}(E)$. We are going to prove that under this assumption, Hill's estimator is consistent (Proposition 2.4). We proceed in four steps discussed as Propositions.

Proposition 2.1 *Consistency of the empirical measure given in (2.1) implies that*

$$\frac{X_{(k)}}{b(n/k)} \xrightarrow{P} 1,$$

as $n \to \infty$, $k/n \to 0$.

Proof: We have that

$$P(|\frac{X_{(k)}}{b(\frac{n}{k})} - 1| > \varepsilon) \;=\; P(X_{(k)} > (1+\varepsilon)b(\frac{n}{k})) \;+\; P(X_{(k)} < (1-\varepsilon)b(\frac{n}{k})) \;\leq$$

$$\leq\; P(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/b(\frac{n}{k})}\,(1+\varepsilon,\infty] \geq 1) \;+\; P(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/b(\frac{n}{k})}\,[1-\varepsilon,\infty] < 1)$$

But (2.1) implies that

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{X_i/b(\frac{n}{k})}\,(1+\varepsilon,\infty] \xrightarrow{P} (1+\varepsilon)^{-\alpha} < 1$$

and

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{X_i/b(\frac{n}{k})}\,[1-\varepsilon,\infty] \xrightarrow{P} (1-\varepsilon)^{-\alpha} > 1$$

and therefore the desired conclusion follows. □

Proposition 2.2 *The following results from (2.1): In* $M_+(E)$,

$$\hat{\mu}_{X,n} \stackrel{P}{\rightarrow} \mu,$$

as $n \to \infty$, $k/n \to 0$.

Proof: Define the scaling operator:

$$T : M_+(E) \times (0,\infty) \mapsto M_+(E)$$

by

$$T(\mu, x)(A) = \mu(xA).$$

From (2.1), Proposition 2.1 and Billingsley(1968), Theorem 4.4, we get joint convergence

$$(\mu_{X,n}, \frac{X_{(k)}}{b(\frac{n}{k})}) \Rightarrow (\mu, 1) \tag{2.2}$$

in $M_+(E) \times (0,\infty)$. Since

$$\hat{\mu}_{X,n}(\cdot) = \mu_{X,n}(X_{(k)}/b(\frac{n}{k}) \cdot) = T(\mu_{X,n}, X_{(k)}/b(\frac{n}{k})),$$

the conclusion will follow by the continous mapping theorem, provided that we can prove the continuity of the operator T at $(\mu, 1)$. In fact, we prove the continuity of the operator at (μ, x) where $x \neq 0$, since this is needed later. Towards this goal, let $\mu_n \stackrel{v}{\rightarrow} \mu$ and $x_n \to x$, where $\mu_n \in M_+(E)$, $x_n, x \in (0,\infty)$. It suffices to show for any y > 0 that

$$\mu_n((x_n y, \infty]) \to \mu((xy, \infty]). \tag{2.3}$$

Given $\varepsilon > 0$ and n sufficiently large we have

$$\mu_n(((x+\varepsilon)y, \infty]) \leq \mu_n((x_n y, \infty]) \leq \mu_n(((x-\varepsilon)y, \infty])$$

and letting $n \to \infty$ yields that

$$((x+\varepsilon)y)^{-\alpha} \leq \liminf_{n\to\infty} \mu_n((x_n y, \infty]) \leq \limsup_{n\to\infty} \mu_n((x_n y, \infty]) \leq ((x-\varepsilon)y)^{-\alpha}.$$

Letting $\varepsilon \to 0$ in the extremes of the inequalities gives (2.3). □

The next result shows that a simple functional of μ_n yields a consistent estimator of α^{-1}. However it is statistically unsuitable as it depends on the unknown parameter b(n/k).

Proposition 2.3 *Consistency in (2.1) implies that, as* $n \to \infty$, $k/n \to 0$,

$$H^+_{X,n} \stackrel{P}{\rightarrow} \frac{1}{\alpha}$$

Proof: Note that

$$H^+_{X,n} = \int_1^\infty \log(y)\mu_{X,n}(dy)$$

which by an integration by parts can be expressed as

$$H^+_{X,n} = \int_1^\infty \mu_{X,n}(u,\infty]\frac{du}{u}.$$

The proof consists of a converging together argument. Denote by

$$H^+_{X,n,t} = \int_1^t \mu_{X,n}(u,\infty]\frac{du}{u}$$

and then (2.1) implies

$$H^+_{X,n,t} \xrightarrow{P} \int_1^t \mu(u,\infty]\frac{du}{u} = \frac{1-t^{-\alpha}}{\alpha}$$

for any fixed t when $n \to \infty$. Also, for $t \to \infty$, $(1-t^{-\alpha})/\alpha \to 1/\alpha$. By Theorem 4.2 of Billingsley (1968), we only need check that:

$$\lim_{t\to\infty}\limsup_{n\to\infty} P(H^+_{X,n} - H^+_{X,n,t} > \varepsilon) = 0.$$

Now

$$\begin{aligned} P(H^+_{X,n} - H^+_{X,n,t} > \varepsilon) &= P\Big(\int_t^\infty \mu_{X,n}(u,\infty]\frac{du}{u} > \varepsilon\Big) \le \varepsilon^{-1}E\Big(\int_t^\infty \mu_{X,n}(u,\infty]\frac{du}{u}\Big) \\ &= \varepsilon^{-1}\int_t^\infty \frac{n}{k}\overline{F}(b(\frac{n}{k})u)\frac{du}{u}. \end{aligned}$$

By Potter's inequality (Bingham, Goldie, Teugels (1987), Theorem 1.5.6) and (1.2), for a given δ, there exists an n_0 such that, for $u \ge 1$

$$(1-\delta)u^{-\alpha-\delta} < \frac{n}{k}\overline{F}(b(\frac{n}{k})u) < (1+\delta)u^{-\alpha+\delta}$$

for any $n > n_0$. Therefore it follows that:

$$\limsup_{n\to\infty}\int_t^\infty \frac{n}{k}\overline{F}(b(\frac{n}{k})u)\frac{du}{u} < \int_t^\infty (1+\delta)u^{-\alpha+\delta-1}du.$$

Making sure that $\delta < \alpha$ and letting $t \to \infty$ completes the proof. □

The last step shows that Hill's estimator is consistent for α^{-1}.

Proposition 2.4 *If (2.1) holds then, as* $n \to \infty$, $k/n \to 0$,

$$H_{X,n} \xrightarrow{P} \frac{1}{\alpha}.$$

Proof: Since

$$H_{X,n} = \int_1^\infty \log(y)\hat{\mu}_{X,n}(dy) = \int_1^\infty \hat{\mu}_{X,n}(u,\infty]\frac{du}{u},$$

we may mimic the previous proof. By Proposition 2.2,

$$\int_1^t \hat{\mu}_{X,n}(u,\,\infty]\frac{du}{u} \stackrel{P}{\rightarrow} \frac{1-t^{-\alpha}}{\alpha}$$

for any fixed t when $n \rightarrow \infty$. Since $(1-t^{-\alpha})/\alpha \rightarrow 1/\alpha$ as $t \rightarrow \infty$, it again suffices to prove

$$\lim_{t\rightarrow\infty} \limsup_{n\rightarrow\infty} P(\int_t^\infty \hat{\mu}_{X,n}(u,\,\infty]\frac{du}{u} > \varepsilon) = 0.$$

We write

$$\int_t^\infty \hat{\mu}_{X,n}(u,\,\infty]\frac{du}{u} = \int_t^\infty \mu_{X,n}(\frac{X_{(k)}}{b(\frac{n}{k})}u,\,\infty]\frac{du}{u} = \int_{\frac{X_{(k)}}{b(n/k)}}^\infty \mu_{X,n}(s,\infty]ds.$$

Thus, for any $\delta > 0$

$$P(\int_{\frac{X_{(k)}}{b(\frac{n}{k})}t}^\infty \mu_{X,n}(u,\,\infty]\frac{du}{u} > \varepsilon) \leq P(\int_{\frac{X_{(k)}}{b(\frac{n}{k})}t}^\infty \mu_{X,n}(u,\,\infty]\frac{du}{u} > \varepsilon,\, |\frac{X_{(k)}}{b(\frac{n}{k})} - 1| \leq \delta) +$$

$$+P(|\frac{X_{(k)}}{b(\frac{n}{k})} - 1| > \delta) \leq P(\int_{(1-\delta)t}^\infty \mu_{X,n}(u,\,\infty]\frac{du}{u} > \varepsilon) + o(1).$$

Then, by the last argument in the previous proof:

$$\lim_{t\rightarrow\infty} \limsup_{n\rightarrow\infty} P(\int_{\frac{X_{(k)}}{b(\frac{n}{k})}t}^\infty \mu_{X,n}(u,\,\infty]\frac{du}{u} > \varepsilon) = 0. \square$$

3 The consistency of Hill's estimator for infinite moving averages.

In this section we prove convergence of the empirical measure associated to an infinite moving average sequence. Our approach follows the spirit of derivations of Davis and Resnick (1985); see also Section 4.5, Resnick (1987). We start with the known result for the i.i.d. samples (cf. Resnick (1986)), extend it to finite moving averages by techniques involving the continous mapping theorem and then, by a converging together argument, we prove the result for the more general setting, the infinite moving average. Let us begin with some notations and preliminaries.

We assume that Condition II holds so that Z_k are i.i.d. positive random variables with

$$P(Z_k > x) = \overline{G}(x) \sim x^{-\alpha}L(x),$$

L being a slowly varying function. Suppose, that the sequence $\{\, c_i \,\} \in \mathbf{R}^\infty$ contains at least one positive number and satisfies

$$0 < \sum_{j=0}^{\infty} |c_j|^\delta < \infty \tag{3.1}$$

for $0 < \delta < \alpha \wedge 1$. Then (cf. Cline (1983))

$$\sum_{j=0}^{\infty} c_j Z_j < \infty$$

and

$$\lim_{x\to\infty} \frac{P(\sum_{j=0}^{\infty} c_j Z_j > x)}{P(Z_1 > x)} = \sum_{\substack{j=0 \\ c_j>0}}^{\infty} c_j^{\alpha} \tag{3.2}$$

so that $\sum_{j=0}^{\infty} c_j Z_j$ also has regularly varying tail probabilities.

Define the moving average of order infinity processes, denoted MA(∞), by

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \qquad -\infty < n < \infty. \tag{3.3}$$

Causal ARMA processes can be represented in the form (3.3) (Brockwell and Davis (1991), Chapter 3). The purpose of this section is to show that applying the Hill estimator to the observed time series $X_1, X_2, \ldots, X_n$ yields a consistent estimator of α^{-1}. Note that (3.2) gives us confidence that the sample $X_1, X_2, \ldots, X_n$ contains sufficient information about α. Recall $F(x) = P(X_1 \le x)$.

Define the vector

$$\mathbf{Z}_t^{(m)} := (Z_t,\, Z_{t-1}\,,\, \ldots,\, Z_{t-m}),$$

denote basis vectors in $\mathbf{R}^{m+1}$ by

$$\mathbf{e}_0 = (1,\, 0,\, 0,\, \ldots,\, 0),\, \mathbf{e}_1 = (0,\, 1,\, 0,\, \ldots,\, 0),\, \ldots,\, \mathbf{e}_m = (0,\, 0,\, 0,\, \ldots,\, 1)$$

and define the measure $\boldsymbol{\mu}^{(m)} \in M_+([0,\infty]^{m+1} \setminus \{\mathbf{o}\})$ to be concentrated on the axes

$$\{x\mathbf{e}_i,\, x > 0\}, \qquad i = 0,\, 1,\, \ldots,\, m$$

and equal to μ on each of them so that, for A $\in \mathcal{B}([0,\infty]^{m+1} \setminus \{\mathbf{o}\})$,

$$\boldsymbol{\mu}^{(m)}(A) = \sum_{i=0}^{m} \mu(\{x \in (0,\infty] : x\mathbf{e}_i \in A\}).$$

We seek to show the tail empirical measure corresponding to $\{X_1, X_2, \ldots, X_n\}$, namely

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/b(n/k)}$$

converges to μ. We think of X_i as a functional of $(Z_i,\, Z_{i-1}, \ldots)$. An aproximation $X_i^{(m)}$ of X_i can be defined as $X_i^{(m)} = \sum_{j=0}^{m} c_j Z_{i-j}$ which we consider a functional of $(Z_i,\, Z_{i-1},\, \ldots,\, Z_{i-m})$. This suggests examining the behaviour of

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\mathbf{Z}_i^{(m)}/b(\frac{n}{k})},$$

which we do in the next proposition.

Proposition 3.1 *We have in* $M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$*, that, as* $n\to\infty$*,* $k/n\to 0$*,*

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\mathbf{Z}_i^{(m)}/b(\frac{n}{k})} \Rightarrow \boldsymbol{\mu}^{(m)}. \tag{3.4}$$

Proof: The proof goes in two steps. The first one is to prove that:

$$d(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\mathbf{Z}_i^{(m)}/b(\frac{n}{k})}, \frac{1}{k}\sum_{j=0}^{m}\sum_{i=1}^{n}\epsilon_{\mathbf{e}_j Z_i/b(\frac{n}{k})}) \xrightarrow{P} 0$$

where d is the vague metric. The proof for this step is similar to the first half of the proof of Proposition 4.26, Resnick (1987) and is omited here. The second step requires proving that

$$\frac{1}{k}\sum_{j=0}^{m}\sum_{i=1}^{n}\epsilon_{Z_i/b(\frac{n}{k})\mathbf{e}_j} \Rightarrow \boldsymbol{\mu}^{(m)}.$$

In order to prove this convergence we define the operator:

$$U : M_+(E) \mapsto M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$$

by

$$Um(A) = \sum_{i=0}^{m} m(\{x \in (0,\infty] : x\mathbf{e}_i \in A\})$$

where m $\in M_+(E)$, $A \in \mathcal{B}([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$. U is continous: let f $\in C_K^+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$ and without loss of generality supose the support of f is contained in $([0,\delta]^{m+1})^c$, where $\delta > 0$. If for $n \geq 0$ we have $m_n \in M_+(E)$ and $m_n \xrightarrow{v} m_0$ then

$$Um_n(f) = \sum_{i=0}^{m} m_n(f(\mathbf{e}_j\,\cdot\,)) \to \sum_{i=0}^{m} m_0(f(\mathbf{e}_j\,\cdot\,)) = Um_0(f),$$

since $f(\mathbf{e}_j\,\cdot\,) \in C_K^+(E)$. Thus, since we know (Resnick (1986), Proposition 5.3)

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{Z_i/b(n/k)} \Rightarrow \mu$$

in $M_+(E)$, we get by the continuous mapping theorem that

$$\begin{aligned} U(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{Z_i/b(n/k)}) &= \frac{1}{k}\sum_{j=0}^{m}\sum_{i=1}^{n}\epsilon_{Z_i/b(\frac{n}{k})\mathbf{e}_j} \\ &\Rightarrow U(\mu) = \boldsymbol{\mu}^{(m)} \end{aligned}$$

as desired. □

Proposition 3.1 is a bridge to the result describing the behavior of the tail empirical measure of $X_1^{(m)}, X_2^{(m)}, \ldots, X_n^{(m)}$.

Proposition 3.2 *If* $X_n^{(m)} = \sum_{i=0}^m c_i Z_{n-i}$ *then*

$$\frac{1}{k}\sum_{i=1}^{n} 1_{X_i^{(m)}/b(n/k)>0}\epsilon_{X_i^{(m)}/b(n/k)} \Rightarrow (\sum_{\substack{i=0\\c_i>0}}^{m} c_i^\alpha \;/\; \sum_{\substack{i=0\\c_i>0}}^{\infty} c_i^\alpha)\,\mu \tag{3.5}$$

in $M_+(E)$.

Proof: Choose m large enough that

$$\sum_{j=0}^{m} c_j^2 \;>\; 0. \tag{3.6}$$

Define the operator

$$V \;:\; M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\}) \to M_+((0,\infty])$$
$$V(\boldsymbol{\nu})(A) = \boldsymbol{\nu}\,(\{\boldsymbol{x} \in [0,\infty]^{m+1}\setminus\{\mathbf{o}\} \;:\; \sum_{i=0}^{m} c_i x_i \in A\}).$$

where $A \in \mathcal{E}$. To show that V is continuous, it suffices to show for any $f \in C_K((0,\infty])$ that

$$V^{-1}\{m \in M_+((0,\infty]) \;:\; m(f) \in O\},$$

for O open in $(0,\infty]$, is open in $M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$. Since

$$\begin{aligned} &V^{-1}\{m \in M_+((0,\infty]) \;:\; m(f) \in O\} \\ &= \{\boldsymbol{\nu} \in M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\}) \;:\; \int_{[0,\infty]^{m+1}\setminus\{\mathbf{o}\}} f(\sum_{i=0}^{m} c_i x_i)\boldsymbol{\nu}(dx_0,\,\ldots,\,dx_m) \in O\} \end{aligned}$$

it suffices, from the definition of the vague topology, to show that the function

$$f(\mathbf{c}\cdot) : (x_0,\ldots,x_m) \mapsto f(\sum_{i=0}^{m} c_i x_i)$$

belongs to $C_K([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$. Since this function is obviously continuous, it suffices to show it has compact support. We show the support is bounded away from $\mathbf{o}$. Suppose for $a>0$, that the support of f, is contained in $(a,\infty]$. If $\mathbf{o}$ were a limit point of the support of $f(\mathbf{c}\,\cdot)$, call it supp, then there would exist $\boldsymbol{x}^{(n)} \in$ supp, $\boldsymbol{x}^{(n)} \to \mathbf{o}$. In this case, $\sum_{i=0}^m c_i x_i^{(n)} \to 0$ and, for large enough n, $\sum_{i=0}^m c_i x_i^{(n)} < a$ contradicting $\boldsymbol{x}^{(n)} \in$ supp.

By the continuous mapping theorem, the previous proposition and Condition II, it follows that:

$$V(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\mathbf{Z}_i^{(m)}/b(\frac{n}{k})}) \;=\; \frac{1}{k}\sum_{i=1}^{n} 1_{[X_i^{(m)}/b(n/k)>0]}\epsilon_{X_i^{(m)}/b(\frac{n}{k})} \Rightarrow V(\boldsymbol{\mu}^{(m)})$$

where

$$V(\boldsymbol{\mu}^{(m)}) = (\sum_{\substack{i=0\\c_i>0}}^{m} c_i^\alpha)\,\mu.$$

To reach the desired conclusion of the Proposition, we must rescale the random points in the previous formula by:

$$\frac{b(\frac{n}{k})}{b(\frac{n}{k})} = \frac{(\frac{1}{1-G})^{-1}(\frac{n}{k})}{(\frac{1}{1-F})^{-1}(\frac{n}{k})}.$$

By our assumptions, $1/(1-F)$ and $1/(1-G)$ are regularly varying with index $\alpha > 0$ and from (3.2)

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{G}(x)} = \sum_{\substack{i=0\\ c_i>0}}^{\infty} c_i^{\alpha}.$$

By Proposition 0.8.(vi), Section 0.4 in Resnick (1987) it follows:

$$\lim_{n\to\infty} \frac{(\frac{1}{1-G})^{-1}(\frac{n}{k})}{(\frac{1}{1-F})^{-1}(\frac{n}{k})} = (\sum_{\substack{i=0\\ c_i>0}}^{\infty} c_i^{\alpha})^{-\alpha^{-1}}.$$

Then, by the scaling argument of Proposition 2.2, we can conclude:

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i^{(m)}/b(\frac{n}{k})} &\Rightarrow V(\boldsymbol{\mu}^{(m)}) \circ T^{-1} \\
&= (\sum_{\substack{i=0\\ c_i>0}}^{\infty} c_i^{\alpha})^{-\alpha^{-1}\alpha}(\sum_{\substack{i=0\\ c_i>0}}^{m} c_i^{\alpha})\,\mu \\
&= (\sum_{\substack{i=0\\ c_i>0}}^{m} c_i^{\alpha} \,/\, \sum_{\substack{i=0\\ c_i>0}}^{\infty} c_i^{\alpha})\,\mu
\end{aligned}$$

where T is defined as in Proposition 2.2. □

The next result extends the previous one to the infinite moving average case.

Proposition 3.3 *If* $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$ *then, as* $n \to \infty$, $n/k \to \infty$,

$$\frac{1}{k}\sum_{i=1}^{n} 1_{[X_i/b(n/k)>0]}\,\epsilon_{X_i/b(n/k)} \Rightarrow \mu.$$

in $M_+((0,\infty])$.

Proof: The proof is nothing more than a converging together argument. According to the previous proposition and to Theorem 4.2, Billingsley (1968), it is enough to check that:

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(d(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i^{(m)}/b(n/k)}, \frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i/b(n/k)}) > \varepsilon) = 0 \tag{3.7}$$

where d is the vague metric on $M_+(E)$. From the definition of d, it suffices to check that, for any f$\in C_K^+(E)$:

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(|\frac{1}{k}\sum_{i=1}^{n} f(X_i^{(m)}/b(n/k)) - \frac{1}{k}\sum_{i=1}^{n} f(X_i/b(n/k))| > \varepsilon) = 0.$$

Since f is continous with compact support, the support of f is contained in $[a, \infty]$, for some $a > 0$ and f is uniformly continous. Therefore

$$\omega_f(\theta) := \sup_{|x-y| \leq \theta,\, x,y \in E} |f(x) - f(y)| \to 0$$

when $\theta \to 0$. Let $\delta < a/2$ and define the following sets:

$$A_i = \{|X_i^{(m)}/b(n/k) - X_i/b(n/k)| \leq \delta,\, X_i^{(m)} \geq a - \delta\}$$

$$B_i = \{|X_i^{(m)}/b(n/k) - X_i/b(n/k)| \leq \delta,\, X_i^{(m)} < a - \delta\}$$

and

$$C_i = \{|X_i^{(m)}/b(n/k) - (X_i/b(n/k)| > \delta\}$$

for i = 1, ..., n. Decompose the probability as follows:

$$\begin{aligned}
&P(|\frac{1}{k}\sum_{i=1}^{n} f(X_i^{(m)}/b(n/k)) - \frac{1}{k}\sum_{i=1}^{n} f(X_i/b(n/k))| > \varepsilon) \\
&\leq P(\frac{1}{k}\sum_{i=1}^{n} |f(X_i^{(m)}/b(n/k)) - f(X_i/b(n/k))| 1_{A_i} > \varepsilon/3) \\
&\quad + P(\frac{1}{k}\sum_{i=1}^{n} |f(X_i^{(m)}/b(n/k)) - f(X_i/b(n/k))| 1_{B_i} > \varepsilon/3) \\
&\quad + P(\frac{1}{k}\sum_{i=1}^{n} |f(X_i^{(m)}/b(n/k)) - f(X_i/b(n/k))| 1_{C_i} > \varepsilon/3) \\
&\leq P(\omega_f(\delta)\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i^{(m)}/b(n/k)}[a-\delta, \infty] > \varepsilon/3) + 0 \\
&\quad + \frac{3n}{k\varepsilon} E(|f(X_1^{(m)}/b(n/k)) - f(X_1/b(n/k))| 1_{C_1})
\end{aligned}$$

where the second term is 0 because neither $X_i^{(m)}/b(n/k)$ nor $X_i/b(n/k)$ belongs to the support of f. From Proposition 3.2

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i^{(m)}/b(n/k)}[a-\delta, \infty] &\Rightarrow (\sum_{\substack{i=0 \\ c_i>0}}^{m} c_i^{\alpha} \,/\, \sum_{\substack{i=0 \\ c_i>0}}^{\infty} c_i^{\alpha})\, \mu([a-\delta, \infty]) \\
&= (\sum_{\substack{i=0 \\ c_i>0}}^{m} c_i^{\alpha} \,/\, \sum_{\substack{i=0 \\ c_i>0}}^{\infty} c_i^{\alpha})(a-\delta)^{-\alpha}
\end{aligned}$$

and thus, for δ and hence $\omega_f(\delta)$ sufficiently small

$$\limsup_{n\to\infty} P(\omega_f(\delta)\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_i^{(m)}/b(n/k)}[a-\delta, \infty] > \varepsilon/3) = 0.$$

Also, if we set M $= \sup_{x\in E} f(x) < \infty$, then

$$\frac{3n}{k\varepsilon}E(|f(X_1^{(m)}/b(n/k)) - f(X_1/b(n/k))|1_{C_1}) \leq \frac{6Mn}{k\varepsilon}P(|X_1^{(m)} - X_1| > \delta\, b(n/k))$$
$$= \frac{6Mn}{k\varepsilon}P(\sum_{j=m+1}^{\infty} c_j Z_{1-j} > \delta\, b(n/k)).$$

Applying (3.2), this bound is asymptotic, as n$\rightarrow \infty$, to

$$\frac{6M}{\varepsilon}\sum_{j=m+1}^{\infty} c_j^{\alpha}\delta^{-\alpha}$$

and, as m$\rightarrow \infty$, this last expression goes to zero. This verifies (3.7) and completes the proof. □

The next Corollary is a generalization which is needed in Section 4.

Corollary 3.1 *If Z_1 satisfies Condition II, so that Z_1 has regularly varying tail probabilities, then for any integer $p > 0$,*

$$\frac{1}{k}\sum_{i=1}^{n} 1_{[(X_i,X_{i-1},\ldots,X_{i-p})\in[0,\infty]^{p+1}\setminus\{\mathbf{o}\}]}\,\epsilon_{(X_i,X_{i-1},\ldots,X_{i-p})/b(n/k)} \Rightarrow \mu^*$$

in $M_+([0,\infty]^{p+1}\setminus\{\mathbf{o}\})$ where

$$\mu^*(A) := \sum_{i=0}^{\infty}\mu(\{x \in E : (c_i, c_{i-1}, \ldots, c_{i-p})x \in A\}) \tag{3.8}$$

for any $A \in \mathcal{B}([0,\infty]^{p+1}\setminus\{\mathbf{o}\})$ with the convention that c's of negative index are defined to be 0.

Proof: By Proposition 3.1 we know that, for any m,

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{(Z_t,Z_{t-1},\ldots,Z_{t-m})/b(\frac{n}{k})} \Rightarrow \boldsymbol{\mu}^{(m+1)},$$

in $M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$. We now follow an idea close to the one in Theorem 2.4 (ii), Davis and Resnick (1985) and define the continuous operator

$$W^{(m)}\boldsymbol{\nu}(A) = \boldsymbol{\nu}\{\boldsymbol{x} \in [0,\infty]^{m+1}\setminus\{\mathbf{o}\} : (\sum_{i=0}^{m} c_i x_i, \sum_{i=0}^{m-1} c_i x_{i+1}, \ldots, \sum_{i=0}^{m-p} c_i x_{i+p}) \in A\}$$

from $M_+([0,\infty]^{m+1}\setminus\{\mathbf{o}\})$ to $M_+([0,\infty]^{p+1}\setminus\{\mathbf{o}\})$. Then by continuity

$$W^{(m)}(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{(Z_t,Z_{t-1},\ldots,Z_{t-m})/b(\frac{n}{k})})$$
$$= \frac{1}{k}\sum_{i=1}^{n} 1_{[(X_i,X_{i-1},\ldots,X_{i-p})\in[0,\infty]^{p+1}\setminus\{\mathbf{o}\}]}\,\epsilon_{(X_i^{(m)},X_{i-1}^{(m-1)},\ldots,X_{i-p}^{(m-p)})/b(n/k)}$$
$$\Rightarrow W^{(m)}\boldsymbol{\mu}^{(m+1)}.$$

With closer inspection we note that

$$
\begin{aligned}
&W^{(m)}\boldsymbol{\mu}^{(m+1)}(A)\\
&= \boldsymbol{\mu}^{(m+1)}(\{\boldsymbol{x} \in [0,\infty]^{m+1} \setminus \{\mathbf{o}\} : (\sum_{i=0}^{m} c_i x_i,\ \sum_{i=0}^{m-1} c_i x_{i+1}, \ldots,\ \sum_{i=0}^{m-p} c_i x_{i+p}) \in A\})\\
&= \sum_{l=0}^{m} \mu(\{x \in [0,\infty]^{m+1} \setminus \{\mathbf{o}\} : (c_l,\ c_{l-1},\ \ldots,\ c_{l-p})x \in A\})
\end{aligned}
$$

Thus, if we let $m \to \infty$ as in the previous proof, the result follows. □

We now sumarize the main conclusion of this section which follows from Proposition 2.4 and Proposition 3.3.

Proposition 3.4 *If $\{X_n\}$ is an MA(∞) process given by (3.3), where $\{Z_i\}$ have regularly varying tail probabilities with index -α as described in Condition II, then the Hill estimator is consistent for α^{-1}.*

4 Estimation for AR processes using the residuals

In the following, we assume that $\{X_n\}$ is an AR(p) process defined by the p-th order autoregression:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots. \tag{4.1}$$

Recall $\{Z_n\}$ satisfy Condition II so that $\{Z_n\}$ are i.i.d., $Z_n > 0$ and

$$\overline{G}(x) = P(Z_1 > x) \sim x^{-\alpha} L(x), \qquad x \to \infty. \tag{4.2}$$

We suppose

$$\Phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i \neq 0, \qquad |z| \leq 1 \tag{4.3}$$

so that (Brockwell and Davis, (1991)) the autoregression (4.1) exists and has a stationary solution of the form

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \qquad -\infty < n < \infty \tag{4.4}$$

where

$$C(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{\Phi(z)}, \qquad |z| \leq 1. \tag{4.5}$$

We assume that we have a sequence $\hat{\boldsymbol{\phi}}^{(n)} = (\hat{\phi}_1^{(n)}, \ldots, \hat{\phi}_p^{(n)}), n \geq 1$ of consistent estimators for the coefficients of the autoregression such that:

$$\hat{\boldsymbol{\phi}}^{(n)} \xrightarrow{P} \boldsymbol{\phi} \tag{4.6}$$

where $\hat{\boldsymbol{\phi}}^{(n)}$ is based on observing $X_1, \ldots, X_n$. (Various authors have proposed estimators $\hat{\boldsymbol{\phi}}^{(n)}$ of $\boldsymbol{\phi}^{(n)}$ in the heavy tailed case. See for example, Davis and Resnick (1986), Feigin and Resnick (1992, 1993), Mikosch, Gadrich, Klüppelberg, Adler (1993)). For this sequence of estimators the residuals $\hat{Z}_t^{(n)}$ are defined as

$$\hat{Z}_t^{(n)} := X_t - \sum_{i=1}^{p} \hat{\phi}_i^{(n)} X_{t-i}. \tag{4.7}$$

Closely related to the estimated residuals defined above is the following sequence:

$$\hat{Z'}_i^{(n)} := \hat{Z}_i^{(n)} 1_{[(X_i, X_{i-1}, .., X_{i-p}) \in [0,\infty)^{p+1} \setminus \{\mathbf{o}\}]}. \tag{4.8}$$

Note that, when the sequence $\{X_t\}$ is positive, (4.8) are nothing else than the estimated residuals.

The purpose of this section is to show that the Hill estimator applied to $\hat{Z'}_1^{(n)}, \hat{Z'}_2^{(n)}, \ldots, \hat{Z'}_n^{(n)}$ yields a consistent estimator of α^{-1}. Following the outline in Section 2, we will prove that:

$$\frac{1}{k} \sum_{i=1}^{n} \epsilon_{\hat{Z'}_i^{(n)}/b(n/k)} \Rightarrow \mu$$

in $M_+(E)$. This will be shown to imply the consistency of Hill's estimator. Here is the first step.

Proposition 4.1 *If (4.1) and (4.2) hold, then, as $n \to \infty$, $n/k \to \infty$,*

$$\mu_{\hat{Z'},n} = \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\hat{Z'}_i^{(n)}/b(n/k)} \Rightarrow \mu$$

in $M_+(E)$.

Remark: Note that, if X_t are positive, Proposition 4.1 says

$$\mu_{\hat{Z},n} = \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\hat{Z}_i^{(n)}/b(n/k)} \Rightarrow \mu.$$

Proof: By Corollary 3.1 and (4.6) we have

$$\left(\frac{1}{k} \sum_{i=1}^{n} 1_{[(X_i, X_{i-1}, .., X_{i-p}) \in [0,\infty)^{p+1} \setminus \{\mathbf{o}\}]} \epsilon_{(X_i, X_{i-1}, .., X_{i-p})/b(n/k)}, (1, -\hat{\boldsymbol{\phi}})\right) \Rightarrow (\mu^*, (1, -\boldsymbol{\phi}^{(n)}))$$

in $M_+([0,\infty]^{p+1} \setminus \{\mathbf{o}\}) \times \mathbf{R}^{p+1}$. Define the operator ($A \in \mathcal{B}((0,\infty])$)

$$R(\nu, \boldsymbol{\psi})(A) = \nu(\{\boldsymbol{x} \in [0,\infty]^{p+1} \setminus \{\mathbf{o}\} : \boldsymbol{\psi} \cdot \boldsymbol{x} \in A\})$$

from $M_+([0,\infty]^{p+1} \setminus \{\mathbf{o}\}) \times R^{p+1}$ to $M_+((0,\infty])$. An argument similar to the proof of Proposition 2.2 shows that this operator is continous at $(\mu^*, (1, -\phi))$. Hence, by the continuous mapping theorem, we conclude

$$\frac{1}{k} \sum_{i=1}^{n} 1_{[(X_i, X_{i-1}, .., X_{i-p}) \in [0,\infty)^{p+1} \setminus \{\mathbf{o}\}]} \epsilon_{\hat{Z}_i^{(n)}/b(n/k)} \Rightarrow R(\mu^*, (1, -\boldsymbol{\phi})).$$

We assert that, for any A $\in \mathcal{B}((0,\infty])$

$$R(\mu^*, (1, -\boldsymbol{\phi})) = \mu(A).$$

To verify this note first of all that $c_0 = 1$, $c_i - \sum_{j=1}^{p} \phi_j c_{i-j}$, $i \geq 1$ since $C(z)\Phi(z) = 1$. So if

$$A = \{\boldsymbol{x} \in [0,\infty]^{p+1} \setminus \{\mathbf{o}\} : x_0 - \sum_{j=1}^{p} \phi_j x_j \in A\}$$

then by (3.8)

$$\begin{aligned} \mu^*(A) &= \sum_{i=0}^{\infty} \mu(\{x \in E : (c_i, c_{i-1}, \dots, c_{i-p})x \in A\} \\ &= \sum_{i=0}^{\infty} \mu(\{x \in E : (c_i - \sum_{j=1}^{p} \phi_j c_{i-j})x \in A\}) \\ &= \mu(A). \end{aligned}$$

To end the proof note that

$$\frac{1}{k} \sum_{i=1}^{n} 1_{[(X_i, X_{i-1}, \dots, X_{i-p}) \in [0,\infty)^{p+1} \setminus \{\mathbf{o}\}]} \epsilon_{\hat{Z}_i^{(n)}/\flat(n/k)}$$

$$= \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\hat{Z'}_i^{(n)}/\flat(n/k)}. \square$$

Proposition 4.2 *If (4.1) and (4.2) hold, then as $n \to \infty$ and $k/n \to 0$,*

$$(i) \quad \frac{\hat{Z'}^{(n)}_{(k)}}{\flat(n/k)} \xrightarrow{P} 1,$$

$$(ii) \quad \hat{\mu}_{\hat{Z'},n} := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\hat{Z'}_i^{(n)}/\hat{Z'}^{(n)}_{(k)}} \xrightarrow{P} \mu,$$

$$(iii) \quad H^+_{\hat{Z'},n} := \frac{1}{k} \sum_{i=1}^{n} (\log \hat{Z'}_i^{(n)} - \log \flat(n/k))_+ \xrightarrow{P} \frac{1}{\alpha},$$

$$(iv) \quad H_{\hat{Z'},n} := \frac{1}{k} \sum_{i=1}^{k} \log \hat{Z'}^{(n)}_{(i)} - \log \hat{Z'}^{(n)}_{(k+1)} \xrightarrow{P} \frac{1}{\alpha}.$$

Proof: (i) and (ii) are just Proposition 2.1 and Proposition 2.2 applied to $\hat{Z}'^{(n)}$. In the light of Proposition 2.3, to prove (iii) we must show:

$$\lim_{t\to\infty} \limsup_{n\to\infty} P(H^+_{\hat{Z'},n} - H^+_{\hat{Z'},nt} > \varepsilon) = \lim_{t\to\infty} \limsup_{n\to\infty} P(\int_t^\infty \log(y) \mu_{\hat{Z'},n}(dy) > \varepsilon) = 0.$$

Since

$$P(\int_t^\infty \log(y) \mu_{\hat{Z'},n}(dy) > \varepsilon) \leq P(\int_t^\infty \log(y) \mu_{\hat{Z},n}(dy) > \varepsilon)$$

it is enough to prove

$$\lim_{t\to\infty} \limsup_{n\to\infty} P(\int_t^\infty \log(y) \mu_{\hat{Z},n}(dy) > \varepsilon) = 0$$

which is implied, by the same Proposition 2.3, by:

$$\lim_{t\to\infty} \limsup_{n\to\infty} P(|\int_t^\infty \mu_{\hat{Z},n}(u,\,\infty]\frac{du}{u} \;-\; \int_t^\infty \mu_{Z,n}(u,\,\infty]\frac{du}{u}| \;>\; \varepsilon) = 0.$$

We have that:

$$P(|\int_t^\infty \mu_{\hat{Z},n}(u,\infty]\frac{du}{u} - \int_t^\infty \mu_{Z,n}(u,\infty]\frac{du}{u}| \;>\; \varepsilon)$$
$$\leq\; P(\int_t^\infty \frac{1}{k}\sum_{i=1}^n |\epsilon_{\hat{Z}_i^{(n)}/\flat(n/k)}(u,\infty] \;-\; \epsilon_{Z_i/\flat(n/k)}(u,\infty]|\frac{du}{u} \;>\; \frac{\varepsilon}{2}).$$

This expression can be handled as follows:

$$P(\int_t^\infty \frac{1}{k}\sum_{i=1}^n |\epsilon_{\hat{Z}_i^{(n)}/\flat(n/k)}(u,\infty] \;-\; \epsilon_{Z_i/\flat(n/k)}(u,\infty]|\frac{du}{u} \;>\; \varepsilon)$$
$$=\; P(\int_t^\infty \frac{1}{k}\sum_{i=1}^n (1_{(\hat{Z}_i^{(n)}/\flat(n/k) > u,\; Z_i/\flat(n/k) \leq u\,)}$$
$$+\, 1_{(\hat{Z}_i^{(n)}/\flat(n/k) \leq u,\; Z_i/\flat(n/k) > u\,)})\frac{du}{u} \;>\; \varepsilon)$$
$$\leq\; P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) > u,\; Z_i/\flat(n/k) \leq u)}\frac{du}{u} \;>\; \frac{\varepsilon}{2})$$
$$+\, P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) \leq u,\; Z_i/\flat(n/k) > u)}\frac{du}{u} \;>\; \frac{\varepsilon}{2}).$$

For the second term apply Chebishev's inequality:

$$P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) \leq u,\; Z_i/\flat(n/k) > u)}\frac{du}{u} \;>\; \frac{\varepsilon}{2})$$
$$\leq\; P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(Z_i/\flat(n/k) > u)}\frac{du}{u} \;>\; \frac{\varepsilon}{2})$$
$$\leq\; \frac{2n}{k\varepsilon}\int_t^\infty P(Z_i/\flat(n/k) \;>\; u)\frac{du}{u}.$$

By Potter's inequality:

$$\lim_{t\to\infty} \limsup_{n\to\infty} \int_t^\infty \frac{n}{k}P(Z_i/\flat(n/k) \;>\; u)\frac{du}{u} \;=\; 0.$$

Now we concentrate on the first term:

$$P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) > u,\; Z_i/\flat(n/k) \leq u)}\frac{du}{u} \;>\; \frac{\varepsilon}{2})$$
$$=\; P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) > u,\; Z_i/\flat(n/k) \leq u(1-\varepsilon))}\frac{du}{u} \;>\; \frac{\varepsilon}{4})$$
$$+\, P(\frac{1}{k}\sum_{i=1}^n \int_t^\infty 1_{(\hat{Z}_i^{(n)}/\flat(n/k) > u,\; Z_i/\flat(n/k) \in (u(1-\varepsilon),u])}\frac{du}{u} \;>\; \frac{\varepsilon}{4}).$$

By an argument we have already used earlier in this proof, it is easy to see that the second term has the desired behaviour:

$$P(\frac{1}{k}\sum_{i=1}^{n}\int_t^\infty 1_{(\hat{Z}_i^{(n)}/b(n/k) > u,\ Z_i/b(n/k) \in (u(1-\varepsilon),u])}\frac{du}{u} > \frac{\varepsilon}{4})$$
$$\le P(\frac{1}{k}\sum_{i=1}^{n}\int_t^\infty 1_{(Z_i/b(n/k) \ge u(1-\varepsilon))}\frac{du}{u} > \frac{\varepsilon}{4}).$$

Moreover

$$P(\frac{1}{k}\sum_{i=1}^{n}\int_t^\infty 1_{(\hat{Z}_i^{(n)}/b(n/k) > u,\ Z_i/b(n/k) \le u(1-\varepsilon))}\frac{du}{u} > \frac{\varepsilon}{4})$$
$$\le P(\int_t^\infty \frac{1}{k}\sum_{i=1}^{n}\epsilon_{(\hat{Z}_i^{(n)}-Z_i)/b(n/k)}(u\varepsilon,\infty] > \frac{\varepsilon}{4}, \bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j| < \delta)$$
$$+ P(\bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j| \ge \delta)$$
$$\le P(\int_t^\infty \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\sum_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j|X_{i-j}/b(n/k)}(u\varepsilon,\infty] > \frac{\varepsilon}{4}, \bigvee_{j=1}^{p}|\hat{\phi}_j^{(n)}-\phi_j| < \delta) + o(1)$$
$$\le P(\int_t^\infty \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\sum_{j=1}^{p}X_{i-j}/b(n/k)}(\frac{u\varepsilon}{\delta},\infty] > \frac{\varepsilon}{4}) + o(1).$$

Making use, in an already familiar way, of Chebishev's and Potter's inequality and reminding the reader that an infinite moving average of regular varying random variables is still regularly varying with the same index (cf. Cline (1983)) we conclude the proof of (iii). The proof for (iv) mimics exactly the proof of Proposition 2.4 to which we refer the reader. □

5 Simulations and data analysis

The first part of this section is devoted to illustrating our results by mean of simulated data. In the second half we analyze a real teletraffic data set, estimating its right tail behaviour.

a) Simulation: We simulated the AR(2) process

$$X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t, \quad t = 1, \ldots, 1000$$

where Z_t are Pareto distributed so that

$$P(Z_t > x) = x^{-\alpha}, x \ge 1.$$

We did this twice, once when $\alpha = 0.5$ and then for $\alpha = 0.7$. The time series plots of the two simulations are presented in Figure 1.

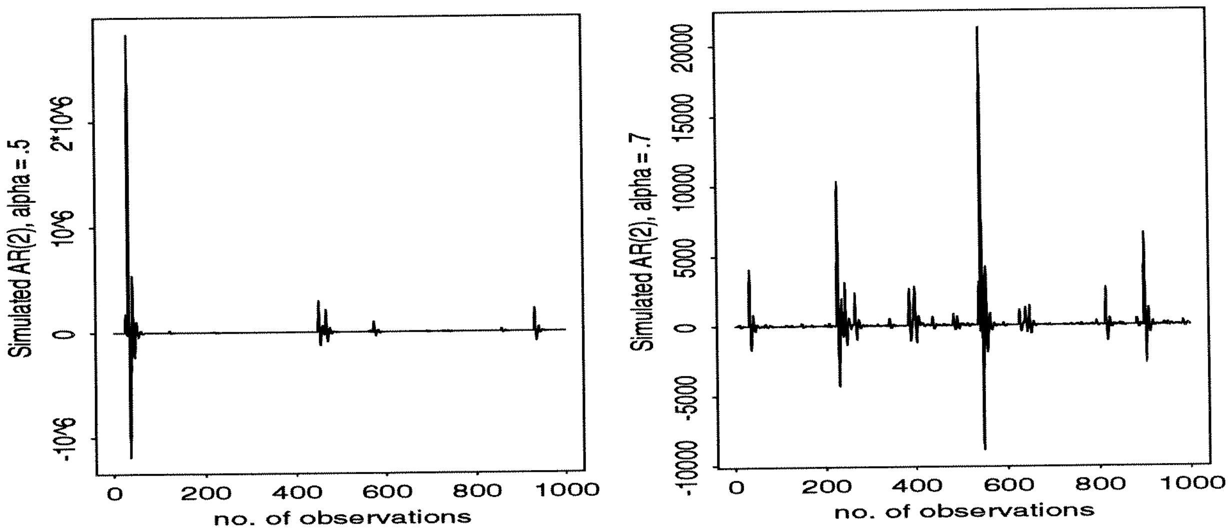

Figure 1

The AR(2) process is causal and therefore has an MA(∞) representation so that the results of Sections 3 and 4 are applicable. The coefficients ϕ_1 and ϕ_2 were estimated by the Yule-Walker method (see below for a discussion). The estimated coefficients for α = 0.5 are ϕ_1 = 1.2846 and ϕ_2 = -0.6835. In the case when α = 0.7 the estimated coefficients are ϕ_1 = 1.2792 and ϕ_2 = -0.6732.

Figure 2 gives Hill estimator plots as a function of the number of order statistics for the case $\alpha = 0.5$. In each graph, the dotted line represents the true value of α. The left graph applies the Hill estimator to the actual residuals. The middle graph applies it to the time series $\{X_t\}$ and the right graph gives the Hill plot for the estimated residuals $\{\hat{Z}_t\}$.

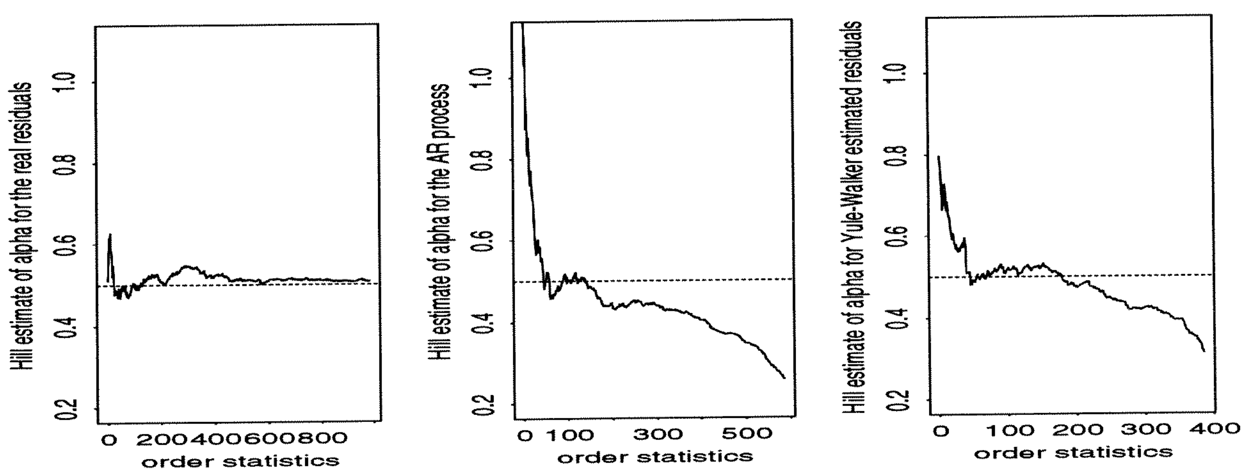

Figure 2

Figure 3 gives the same graphs for the case α = 0.7.
Note in both cases, the plot estimating α appears a bit more stable when estimating from the

estimated residuals, rather than the actual time series.

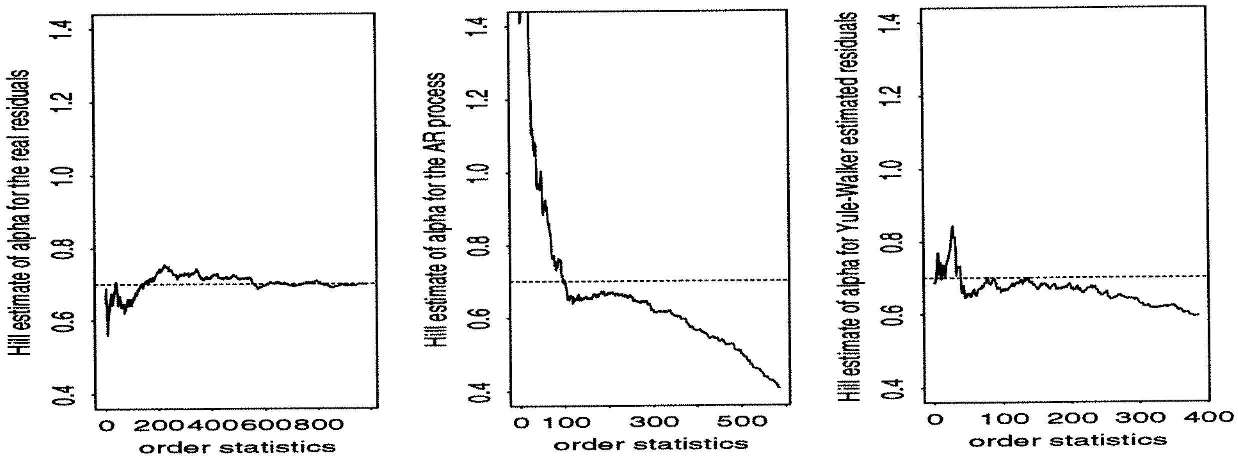

Figure 3

b) Teletraffic data:

The real data set we will analyze represents the silence periods between transmission of packets generated by a terminal during a logged-on sesion. The terminal hooks to a special type of network through a host and communicates with the network sending and receiving packets. The length of the periods between two consecutive packets generated by the terminal (silence periods) were recorded as our data. The total length of the data is 1027.

A brief examination of the data shows a tendency for taking big values with rather high probability. The Hill estimator applied to the data also suggests the appropriateness of a heavy tailed model (Figure 4).

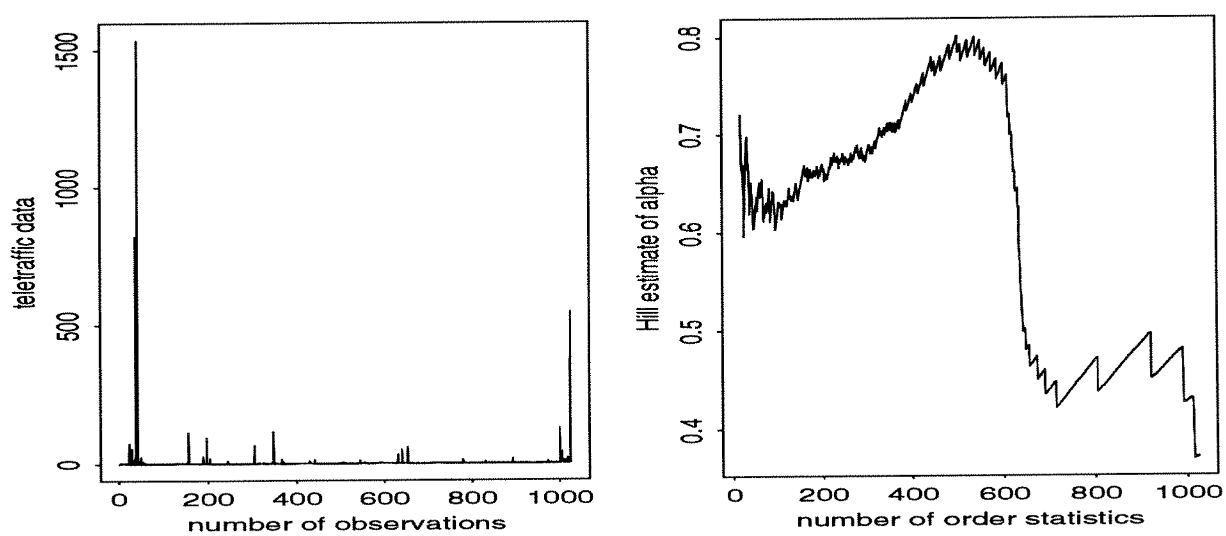

Figure 4

We now survey some model selection and estimation techniques appropriate to heavy tail data.

In the following, we assume that $\{X_t\}$ is an AR(p) process defined by the p-th order autoregression:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots. \tag{5.1}$$

Recall $\{Z_n\}$ satisfy Condition II so that $\{Z_n\}$ are i.i.d., $Z_n > 0$ and

$$\overline{G}(x) = P(Z_1 > x) \sim x^{-\alpha} L(x), \qquad x \to \infty. \tag{5.2}$$

and assume that $0 < \alpha < 2$. Also (4.3), (4.4), (4.5) hold.

Let us first review some facts about the Yule-Walker estimators for an AR process with heavy tailed innovations. If we define, for h $\geq$ 0 and m $\geq$ 1

$$\begin{aligned} \rho(h) &:= \frac{\sum_{i=0}^{\infty} c_i c_{i+h}}{\sum_{i=0}^{\infty} c_i^2}, \\ \hat{\rho}^{(n)}(h) &:= \frac{\sum_{i=0}^{n-h} X_i X_{i+h}}{\sum_{i=0}^{n} X_i^2}, \\ \mathcal{R}_m &= (R_{ij})_{i,j=1}^{m} = (\rho(i-j))_{i,j=1}^{m}, \\ \boldsymbol{\rho}_m &= (\rho(1), \ldots, \rho(m))' \end{aligned}$$

and

$$\boldsymbol{\phi}_m = (\phi(1), \ldots, \phi(m))'$$

where $\phi(j) := 0$ if $j > p$, then (Davis and Resnick (1986))

$$\hat{\rho}^{(n)}(h) \xrightarrow{P} \rho(h) \tag{5.3}$$

and

$$\mathcal{R}_m \boldsymbol{\phi}_m = \boldsymbol{\rho}_m \tag{5.4}$$

for any m $\geq$ 1.

Relation (5.3) allows us check for dependency in the data sequence and $\hat{\rho}^{(n)}$ plays here the role that the sample ACF plays for finite moment data. Thus we refer to $\hat{\rho}^{(n)}$ as the heavy tail sample ACF (HTSACF). We emphasize the difference between the sample autocorrelation function used in the heavy tail case (HTSACF), which is the uncentered sample ACF and the centered one which plays a similar role in the finite moment theory.

Note that for m = p, (5.4) is the heavy tail version of the relation defining the Yule-Walker estimates in the case of data with finite moments (Brockwell and Davis (1991), Section 8.1). Motivated by (5.3), we replace $\boldsymbol{\rho}_p$ with $\hat{\boldsymbol{\rho}}_p$ and $\mathcal{R}_p$ with $\hat{\mathcal{R}}_p$ and then solve the system to get the estimated Yule-Walker coefficients of the AR process. It is known (Davis and Resnick (1986)) that

$$\hat{\boldsymbol{\phi}}_p^{YW} := \hat{\mathcal{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \xrightarrow{P} \boldsymbol{\phi}_p. \tag{5.5}$$

For m $\geq$ 1, $\boldsymbol{\phi}_m(m)$ defines the heavy tail partial autocorrelation function (HTPACF) calculated at lag m. The name is appropriate since, for the finite moment case, when the PACF is defined using the covariance and when $\boldsymbol{\rho}_m$ stands for the usual centered autocorrelation function calculated at lag m, it is known that $\phi_m(m)$ equals the partial autocorrelation function at lag m. For an AR(p) process the HTPACF is 0 for all the lags beyond the order of the model p. Therefore, if the sample HTPACF is approximately zero beyond a threshhold, we get an indication of the possible order of the process. To estimate the HTPACF we proceed by replacing $\boldsymbol{\rho}_m$ and $\mathcal{R}_p$ by the sample estimates.

The use of the HTPACF is an exploratory tool for model selection. Another important tool in deciding on the order of the process is the AIC criterion. Let us assume that the true order is less then or equal to K. Define

$$AIC(k) := n \log(\prod_{j=1}^{k}(1 - (\hat{\phi}_j^{YW}(j))^2)) + 2k$$

and

$$\hat{p} := \arg\min_{k \leq K} AIC(k).$$

Then, for heavy tailed autoregressions, $\hat{p}$ is a consistent estimator of the true order p:

$$\hat{p} \xrightarrow{P} p$$

(cf. Knight (1989)).

We will now attempt to model the teletraffic data as an AR process. There are strong preliminary indications in Figure 4, from the Hill plot, that a heavy tailed model is appropriate. We will select the order of the best suited autoregression by means of HTPACF and AIC criterion, estimate the coefficients by the Yule-Walker method and apply Hill's estimator to the estimated residuals. To help us judge the output of this two procedures we are also going to look at the Q-Q plot of the quantiles of our data against the quantiles of a Pareto distribution, based on the assumption that data behaves almost like a Pareto beyond a certain point. The combined effort of these three methods should give us a fairly good idea about the size of the right tail of our data.

A look at HTSACF convinces us that our data is not independent (Figure 5). The left graph in Figure 5 is the HTACF and the right graph is the usual centered ACF.

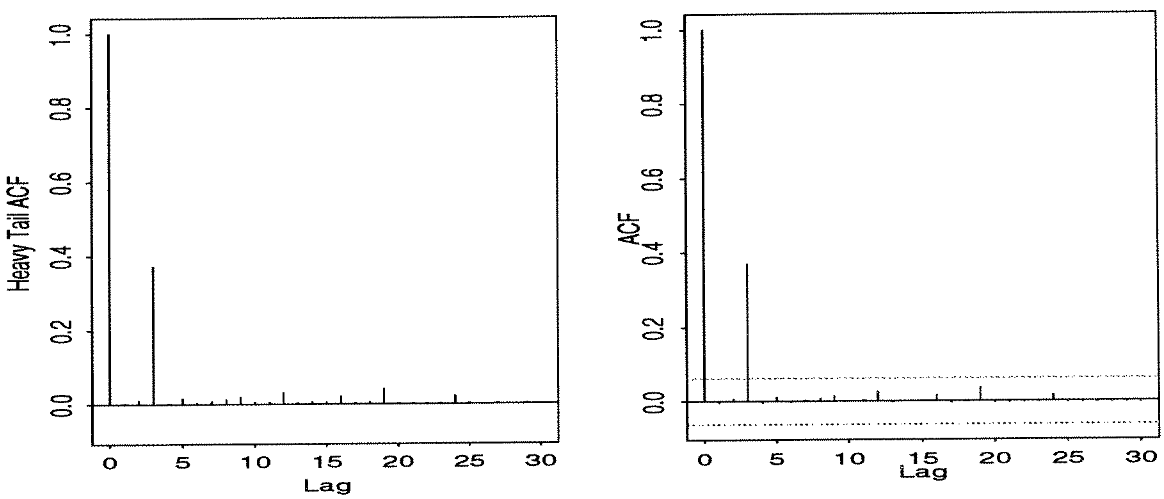

Figure 5

Figure 6 compares the HTSPACF output with the usual sample partial ACF (SPACF) for the finite moment case.

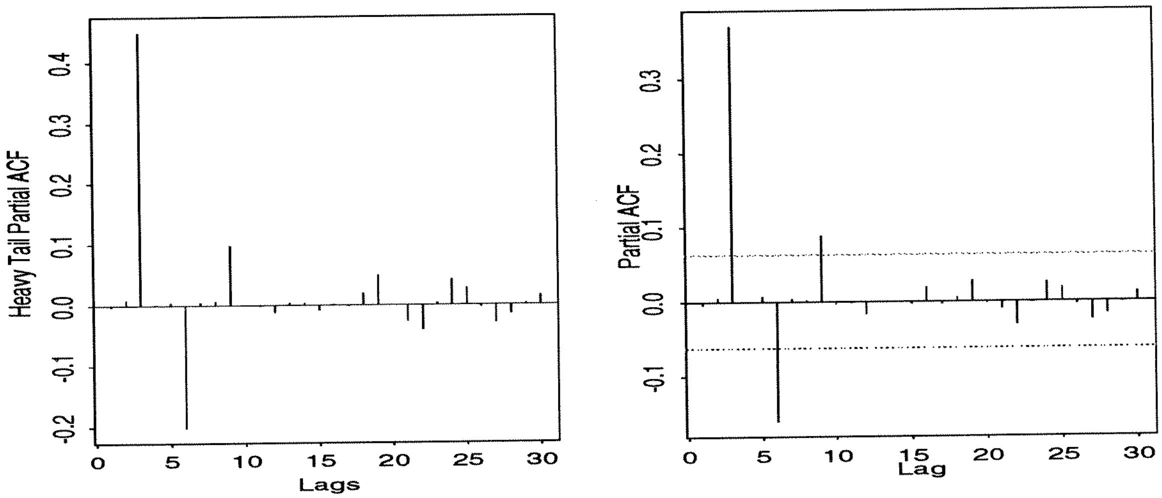

Figure 6

To help us decide on the order of the model, we plotted $AIC(k)$ vs k in Figure 7. Based on Figures 6 and 7, the order seems to be 9.

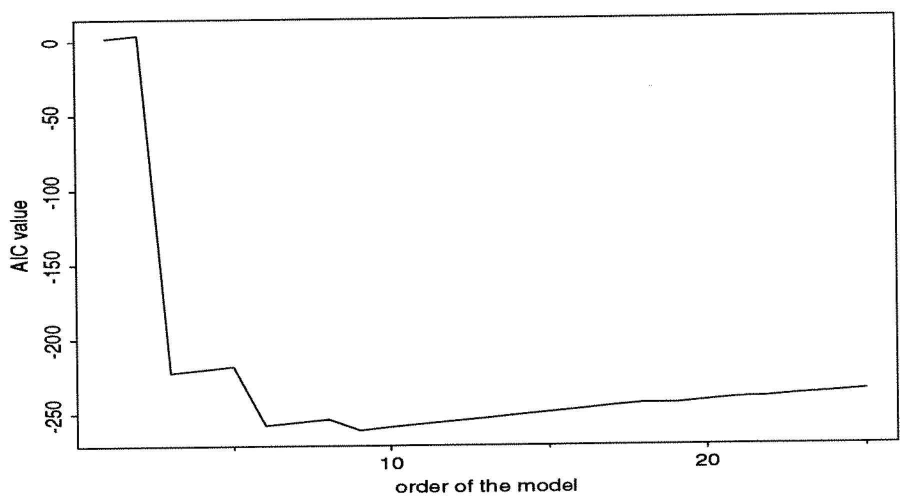

Figure 7

The next step is to estimate the coefficients by the Yule-Walker procedure and, with Proposition4.2 in mind, to apply the Hill estimator to the estimated residuals. The estimated coefficients are: ϕ_1 = -0.0004, ϕ_2 = 0.0069, ϕ_3 = 0.4442, ϕ_4 = -0.0007, ϕ_5 = 0.0064, ϕ_6 = -0.1961, ϕ_7 = 0.0058, ϕ_8 = 0.004, ϕ_9 = 0.0901.

Figure 8 shows the graph of the estimated residuals.

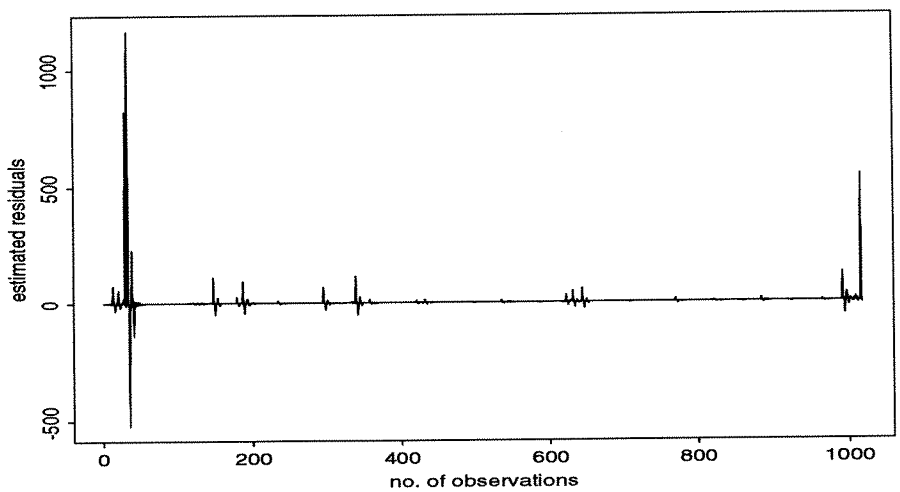

Figure 8

In Figure 9 the HTSACF of the data is plotted on the left and on the right we plot the theoretical HTACF $\rho(h)$, $0 \leq h \leq 20$, of the estimated model, obtained by solving (using ITSM by Brockwell and Davis, 1991)

$$\frac{1}{\hat{\Phi}(z)} = \frac{1}{1 - \sum_{i=1}^{9} \hat{\phi}_i z^i} = \sum_{i=0}^{\infty} c_i z^i$$

and then using

$$\rho(h) := \frac{\sum_{i=0}^{\infty} c_i c_{i+h}}{\sum_{i=0}^{\infty} c_i^2}.$$

Good agreement between the sample and theoretical HTACF provides modest evidence of goodness of fit.

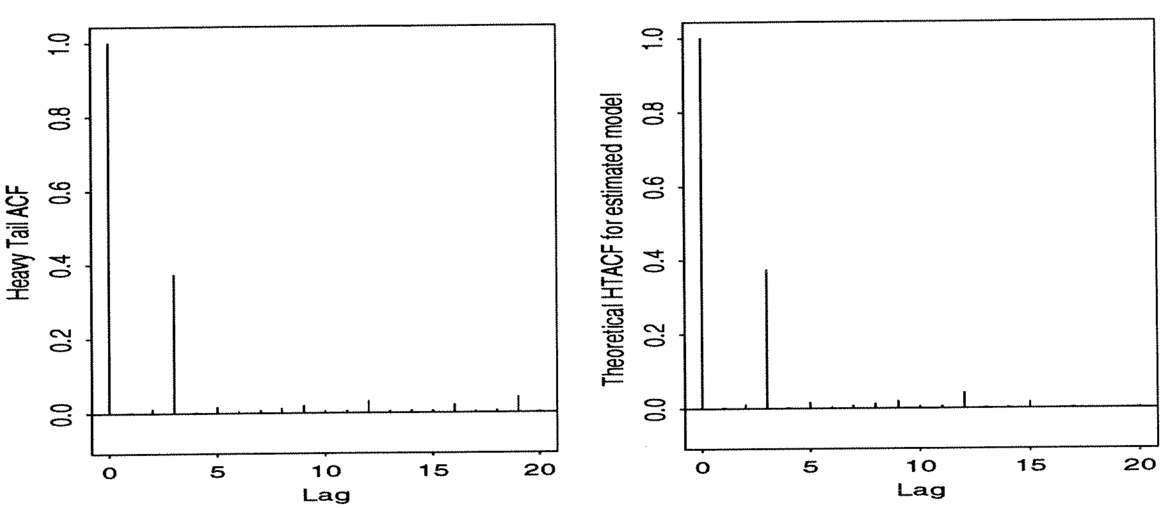

Figure 9

In Figure 10 we display the HTACF of the estimated residuals and the Hill plot based on the estimated residuals. As in part a) of this section, the Hill plot based on estimated residuals seems more stable than the one based on the actual time series. Based on Figure 9 we would guess that α was in the neighborhood of 0.6.

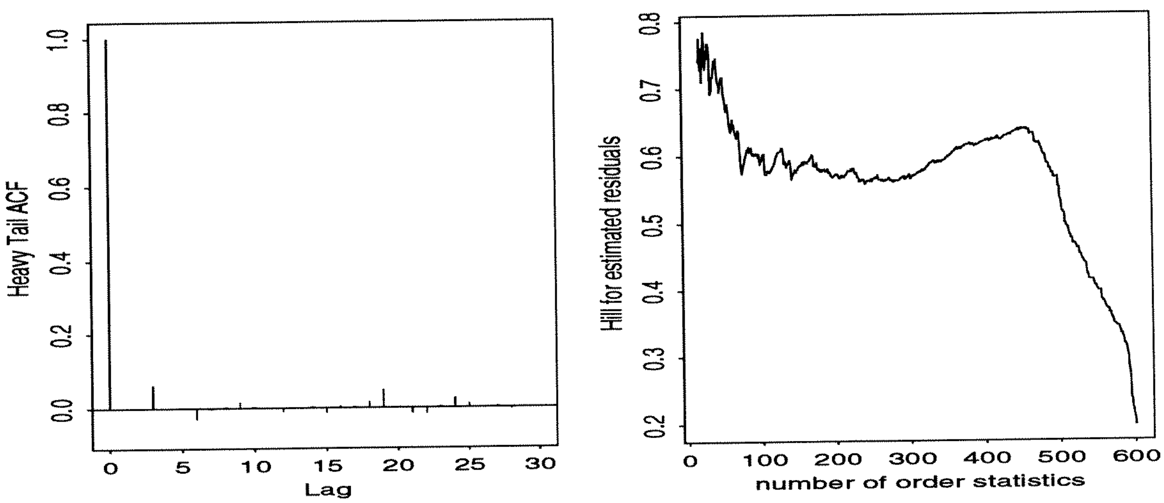

Figure 10

To confirm our estimate of α, we consider a Q-Q plot of the estimated residuals. The basis for this technique is the following. Suppose that $\hat{Z}_1$ behaves almost like a Pareto distribution beyond a certain point, i.e.

$$P(\hat{Z}_1 > x) = (\frac{x}{x_0})^{-\alpha} \tag{5.6}$$

for $x \geq x_0$. Then also

$$P(\alpha \log \frac{\hat{Z}_1}{x_0} > y) = e^{-y} = 1 - G(y), \; y > 0$$

where G is unit exponential. Let H(y) := P(log $\hat{Z}_1 \leq y$) and

$$H(y) = P(\alpha \log \frac{\hat{Z}_1}{x_0} \leq \alpha(y - \log x_0)) = G(\alpha(y - \log x_0)).$$

Therefore we conclude

$$H^{\leftarrow}(y) = \alpha^{-1} G^{\leftarrow}(y) + \log x_0.$$

So, if our assumption is correct, i.e. $\hat{Z}_1$ behaves almost like a Pareto distribution beyond a certain point, plotting

$$\{(-\log(1 - \frac{k}{n+1}), \log \hat{Z}_{(k)}), \; 1 \leq k \leq n\}$$

we should get approximately a line whose slope is α^{-1} and intercept is $\log x_0$ (see the left hand plot of Figure 11).

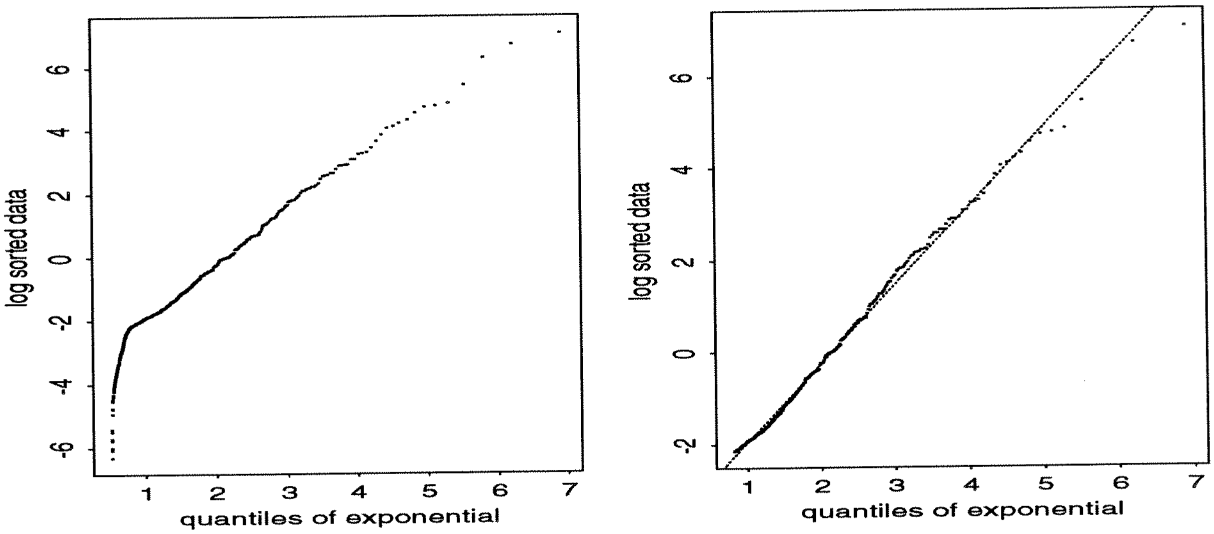

Figure 11

Removing the first 577 order statistics and fitting a line to the graph, the estimate for α is 0.5886 and for x_0 is 0.026. The right hand plot of Figure 11 shows the fit of a straight line to the upper part of the estimated residuals. The same procedure applied to X_1, gives an estimate of 0.6335 for α.

As a conclusion, all these methods indicate a value of α somewhere between 0.58 and 0.68.

References

[1] Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

[2] Bingham, N., Goldie, C. and Teugels, J. (1987). Regular Variation. Cambridge University Press, Cambridge, UK.

[3] Brockwell, P. and Davis, R (1991). Time Series: Theory and Methods, 2nd edition, Springer-Verlag, New York.

[4] Brockwell, P. and Davis, R (1991). ITSM, an Interactive Time Series Modelling Package for the PC. Springer-Verlag, New York.

[5] Cline, D. (1983). Infinite series of random variables with regularly varying tails. Tech. Report 83-24, Institute of Applied Mathematics and Statistics, University of British Columbia.

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

[7] Davis, R. and Resnick, S. (1986). Limit theory for sample covariance and correlation functions. Ann.Statist., 14, 533–558.

[8] Deheuvels, P. and Mason, D. M. (1991). A tail empirical process approach to some non-standard laws of the iterated logarithm. J. Theor. Probab., 4, 53-85.

[9] Feigin, P. D. and Resnick, S. (1992). Estimation for autoregressive processes with positive innovations. Stochastic Models 8, 479–498.

[10] Feigin, P. D. and Resnick, S. (1993). Limit distributions for linear programming time series estimators. To appear: Stochastic Processes and their Applications.

[11] Hill, B. (1975). A simple approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.

[12] Hsing, T. (1991). On tail estimation using dependent data. Ann. Statist. 19, 1547–1569.

[13] Kallenberg, O. (1983). Random Measures, 3rd edition, Akademie-Verlag Berlin.

[14] Knight, K. (1989). Order selection for autoregressions. Ann. Statist. 17, 824–840.

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

[16] Mason, D. (1988). A strong invariance theorem for the tail empirical process. Ann. Inst. Henri Poincaré 24, 491-506.

[17] Mikosch,T., Gadrich, T., Klüppelberg, C., and Adler, R. (1993) Estimation for infinite variance ARMA models. Submitted to Ann. Statist.

[18] Resnick, S. (1986). Point processes, regular variation and weak convergence. Adv. Appl. Probab. 18, 66–138.

[19] Resnick, S. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag, Berlin.

[20] Rootzen, H. , Leadbetter, M. and De Haan, L. (1990). Tail and quantile estimation for strongly mixing stationary sequences. Preprint, Econometric Institute.