Tail Index Estimation for Dependent Data

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

September 5, 1996

Abstract

A popular estimator of the index of regular variation in heavy tailed models is Hill's estimator. We discuss consistency of Hill's estimator when it is applied to certain classes of heavy tailed stationary processes. One class of processes discussed consists of processes which can be appropriately approximated by sequences of m-dependent random variables and special cases of our results show the consistency of Hill's estimator for (i) infinite moving averages with heavy tail innovations, (ii) a simple stationary bilinear model driven by heavy tail noise variables, (iii) solutions of stochastic difference equations of the form

$$Y_t = A_t Y_{t-1} + Z_t, \quad -\infty < t < \infty$$

where $\{(A_n, Z_n), -\infty < n < \infty\}$ are iid and the Z's have regularly varying tail probabilities. Another class of problems where our methods work successfully are solutions of stochastic difference equations such as the ARCH process where the process cannot be successfully approximated by m-dependent random variables. A final class of models where Hill estimator consistency is proven by our tail empirical process methods is the class of hidden semi-Markov models.

1 Introduction.

This paper discusses how to estimate the Pareto index or the index of regular variation for stationary dependent sequences. If $\{X_t, -\infty < t < \infty\}$ is a stationary time series with the property that

$$P[X_t > x] \sim x^{-\alpha} L(x), \quad x \to \infty,$$

L being a slowly varying function, then a key question in tail estimation is how to estimate the index α. A popular estimator which arose in the iid context as a conditional maximum likelihood estimator is Hill's estimator (Hill (1975)) which is defined as follows: For $1 \le i \le n$, write $X_{(i)}$ for the i-th largest value of $X_1, X_2, \ldots, X_n$. Hill's estimator based on the observations $X_1, \ldots, X_n$ is

$$H^X_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}. \tag{1.1}$$

*Partially supported by NSF Grant DMS-9400535 at Cornell University. Some support was also received from NSA Grant MDA904-95-H-1036.

†Supported by NSF Grant DMS-9400535 at Cornell University.

This estimator has been well studied when $\{X_n\}$ is iid (Hall (1982), Mason (1982, 1988), Mason and Turova (1994), de Haan and Resnick (1996), Geluk et al (1996), Davis and Resnick (1984), Hausler and Teugels (1985), Resnick and Stărică (1996a)) and our goal here is to better understand its behavior when it is applied to stationary dependent sequences. Related papers which study Hill's estimator in the dependent case are Hsing (1991), Rootzen et al. (1990), Rootzen (1995).

A great deal of time series analysis has been based on the assumption that the structure of the series can be described by linear models. In the traditional setting of a stationary time series with finite variance every purely non-deterministic process can be expressed as a linear process driven by an uncorrelated input sequence. From a second order point of view, linear models are sufficient for data analysis. The situation is totally different when the stationary series has heavy tails and perhaps infinite variance. In this case we have no such confidence that heavy tailed linear models are sufficiently flexible and rich enough for modeling purposes and in any case, for heavy tailed infinite order moving averages it is already known (Resnick and Stărică (1995) and see also Section 3) that Hill's estimator is consistent. Thus in this paper we concentrate on non-linear models.

Linear models do not seem to describe adequately the underlying random mechanism when heavy tails are present (Davis and Resnick (1995), Resnick (1996)). Insistence upon modeling heavy tailed data with linear time series can be quite misleading (Feigin and Resnick (1996)). A popular non-linear alternative to the linear model is the bilinear process introduced by Mohler (1973) and considered by Granger and Andersen (1978). To date, little use has been made of bilinear models in heavy tailed data analysis though Davis and Resnick (1995) present some evidence for their relevance. Other worthy non-linear models which we consider are two classes of random coeficient models, one of which includes the important example of the ARCH process (Engle (1982)) and hidden semi-Markov models or random variables defined on a semi-Markov chain. Such models have recently been used to fit times between packet transmissions at a terminal in the stimulating paper by Meier-Hellstern et al (1991).

Section 2 presents two general theorems which can be applied to prove consistency of Hill's estimator for heavy tailed stationary sequences. Section 3 applies one of the theorems to the case of processes which can be approximated by m-dependent sequences. Among the examples considered are infinite order moving averages, simple bilinear processes and solutions of certain random coefficient autoregressions. Section 4 applies the other theorem from Section 2 to a class of random coefficient autoregressions which includes the first order ARCH process. This result yields not only an estimator for the Pareto index of the ARCH process but also an estimator of one of the scaling parameters. Section 5 deals directly with hidden semi-Markov models using Laplace functional methods.

The tail empirical measure plays a central role in our approach to proving consistency of Hill's estiamtor. This method was also used in Resnick and Stărică (1995). For using this method, we need the following notation. Let $\mathbb{E} := (0, \infty]$ be the one point uncompactification of $[0, \infty]$ so that the compact sets of $\mathbb{E}$ are of the form U^c, where $0 \in U$ and U is an open set in $[0, \infty)$. Suppose $\mathcal{E}$ is the Borel σ-field on $\mathbb{E}$. Let $M_+(\mathbb{E})$ be the space of positive Radon measures on $\mathbb{E}$ endowed with the vague topology (Resnick (1987), Kallenberg (1983)). Let $C_K^+(\mathbb{E})$ be the space of continuous, non-negative functions on $\mathbb{E} = (0, \infty]$ with compact support. The vague topology on $M_+(\mathbb{E})$ can be generated by a countable family of semi-norms

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

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

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

2 General consistency results.

We now prove two general Hill estimator consistency results for heavy tailed stationary sequences. The first, Proposition 2.1, is designed to be easily specialized for processes which can be approximated by m-dependent sequences and this specialization comes in Proposition 2.2. Proposition 2.3 is similar to Proposition 2.1 but is better suited for application to the ARCH model (cf. Section 4). The proofs of Propositions 2.1 and 2.3 use the standard big block–little block technique explained carefully and exploited in Leadbetter, Lindgren and Rootzen (1988). See also Hsing, Husler and Leadbetter (1988) as well as Davis and Resnick (1988) where a parallel result for Poisson convergence is given.

Proposition 2.1 *Suppose for each $n = 1, 2, \ldots$ that $\{X_{n,i}, i \geq 1\}$ is a stationary sequence of random elements of $\mathbb{E}$. Let $\{k = k(n)\}$ be a sequence such that $k \to \infty$, $n/k \to \infty$. Suppose $\{X_{n,i}\}$ satisfies the following two conditions:*

1. *For any $f \in C_K^+(\mathbb{E})$,*

$$\lim_{n\to\infty} \frac{n}{k^2} \sum_{j=2}^{k} E\left(f(X_{n,1}) f(X_{n,j})\right) = 0. \tag{2.1}$$

2. *For any sequence $\{l_n\}$ such that $l_n \to \infty$ and*

$$l_n/k \to 0 \tag{2.2}$$

and intervals

$$I_1 = [1, k - l_n], I_2 = [k+1, 2k - l_n], \ldots, I_{[n/k]} = [([n/k]-1)k, [n/k]k - l_n] \tag{2.3}$$

we have for $f \in C_K^+(\mathbb{E})$

$$\lim_{n\to\infty} E\left(\prod_{j=1}^{[n/k]} \exp\{-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\}\right) - \prod_{j=1}^{[n/k]} E\left(\exp\{-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\}\right) = 0. \tag{2.4}$$

Assume also that

$$\frac{n}{k} P(X_{n,1} \in \cdot) \xrightarrow{v} \nu. \tag{2.5}$$

where $\nu(\{x\}) = 0$ for any $x \in (0, \infty]$. Then

$$\nu_n := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{X_{n,i}} \Rightarrow \nu \tag{2.6}$$

in $M_+(\mathbb{E})$. Moreover if $X_{n,i} = X_i/b_n$, $i = 1, \ldots, n$, where $\{X_n, n \geq 1\}$ is a sequence of stationary random variables and $b_n \to \infty$ and if ν satisfies $\int_1^\infty \log(u)\nu(du) < \infty$, it also follows that

$$H_{k,n}^X := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} \xrightarrow{P} \int_1^\infty \log(u)\nu(du). \tag{2.7}$$

Remark: Condition 2.1 is implied by the condition that for any $x > 0$,

$$\lim_{n\to\infty} \frac{n}{k^2} \sum_{j=2}^{k} P[X_{n,1} > x, X_{n,j} > x] = 0. \tag{2.8}$$

This follows since if $f \in C_K^+(\mathbb{E})$ and we set $[c, \infty]$ for the support of f and set $\|f\| = \sup_{\mathbb{E}} f(x)$, then

$$f \leq \|f\| 1_{[c,\infty]}$$

and

$$E\left(f(X_{n,1})f(X_{n,j})\right) \leq \|f\|^2 P[X_{n,1} > c, X_{n,j} > c].$$

Proof. Suppose $f \in C_K^+(\mathbb{E})$. To show (2.6), it suffices to show (Kallenberg (1983), Resnick (1987))

$$\lim_{n\to\infty} E\exp\{-\frac{1}{k}\sum_{i=1}^{n} f(X_{n,i})\} = e^{-\nu(f)}. \tag{2.9}$$

For typographical ease, we write $f_i = f(X_{n,i})$ and $p = [n/k]$. Then

$$I_j = \{(j-1)k_n + 1, \ldots, jk_n - l_n\}, \quad I_j^* = \{jk_n - l_n + 1, \ldots, jk_n\}, \quad j = 1, \ldots, p-1 \tag{2.10}$$

and

$$I_p = \{(p-1)k_n + 1, \ldots, pk_n - l_n\}, \quad I_p^* = \{pk_n - l_n + 1, \ldots, n\}. \tag{2.11}$$

We have

$$\begin{aligned}
&|E\exp\{-\frac{1}{k}\sum_{i=1}^{n} f_i\} - \exp\{-\nu(f)\}| \\
&\leq \left| E\exp\{-\frac{1}{k}\sum_{i=1}^{n} f_i\} - E\exp\{-\frac{1}{k}\sum_{j=1}^{p}\sum_{i\in I_j} f_i\} \right| \\
&\quad + \left| E\exp\{-\frac{1}{k}\sum_{j=1}^{p}\sum_{i\in I_j} f_i\} - \left(E\exp\{-\frac{1}{k}\sum_{i\in I_j} f_i\}\right)^p \right| \\
&\quad + \left| \left(E\exp\{-\frac{1}{k}\sum_{i\in I_j} f_i\}\right)^p - \left(E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\}\right)^p \right| \\
&\quad + \left| \left(E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\}\right)^p - \exp\{-\nu(f)\} \right| \\
&= I + II + III + IV.
\end{aligned}$$

Let us look at the individual terms in turn.

We have

$$
\begin{aligned}
I &\le \left| E\exp\{-\frac{1}{k}\sum_{j=1}^{p}\left(\sum_{i\in I_j} f_i + \sum_{i\in I_j^*} f_i\right)\}\right| \\
&\le E|1-\exp\{-\frac{1}{k}\sum_{j=1}^{p}\sum_{i\in I_j^*} f_i\}| \\
&\le \sum_{j=1}^{p}\sum_{i\in I_j^*}\frac{1}{k}Ef_i \\
&\le p l_n \frac{1}{k}Ef_1 \sim \frac{l_n}{k}\frac{n}{k}Ef_1 \\
&\sim \frac{l_n}{k}\nu(f) \to 0,
\end{aligned}
$$

as $n\to\infty$ from (2.2) and (2.5). Term III is handled very similarly:

$$
\begin{aligned}
III &\le p|E\exp\{-\frac{1}{k}\sum_{i\in I_1} f_i\} - E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\}| \\
&\le pE|1-\exp\{-\frac{1}{k}\sum_{i\in I_1^*} f_i\}| \\
&\le p\frac{l_n}{k}Ef_1 \to 0.
\end{aligned}
$$

Term II goes to 0 because of condition (2.4).

For IV we set $y_i = 1-\exp\{-\frac{1}{k}f_i\}$ and observe

$$
\begin{aligned}
E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\} &= E\prod_{i=1}^{k}(1-y_i) \\
&\le 1 - E\sum_{i=1}^{k} y_i + E\sum_{1\le i<j\le k} y_i y_j \\
&\le 1 - kEy_1 + k\sum_{l=2}^{k} Ey_1 y_l
\end{aligned}
$$

and thus

$$
\left(E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\}\right)^p \le \left(1-\frac{kp(Ey_1-\sum_{l=2}^{k}Ey_1y_l)}{p}\right)^p .
$$

Now

$$
\begin{aligned}
kp\sum_{l=2}^{k} Ey_1y_l &\sim n\sum_{l=2}^{k} E(1-\exp\{-\frac{1}{k}f_1)(1-\exp\{-\frac{1}{k}f_l) \\
&\le \frac{n}{k^2}\sum_{l=2}^{k} Ef_1f_l \to 0
\end{aligned}
$$

by (2.1). Also

$$kpEy_1 \sim nE(1 - \exp\{-\frac{1}{k}f_1\} \leq \frac{n}{k}Ef_1 \to \nu(f)$$

and

$$nE(1 - \exp\{-\frac{1}{k}f_1\} \geq nE(\frac{f_1}{k} - \frac{f_1^2}{2k^2}) \sim \nu(f) - \frac{1}{2k}\nu(f^2) \to \nu(f).$$

Thus we conclude

$$\limsup_{n\to\infty} \left(E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\} \right)^p \leq e^{-\nu(f)}.$$

A slightly simpler argument gives

$$\liminf_{n\to\infty} \left(E\exp\{-\frac{1}{k}\sum_{i=1}^{k} f_i\} \right)^p \geq e^{-\nu(f)}$$

and this completes the proof of (2.9).

To prove (2.7) we make use of Proposition 2.4 of Resnick and Stărică (1995) which shows that the convergence of the tail measure implies the consistency of Hill's estimator.□

Proposition 2.1 will be applied primarily to proving consistency of Hill's estimator for stationary processes which can be approximated by truncated versions which are m-dependent. In order for this approximation strategy to be successful, the truncated m-dependent approximation must carry enough information about the tail behavior of the marginal distribution of the original process $\{X_t\}$. This is true of the processes considered in the examples Section 3 and false for certain random coefficient models such as the ARCH process considered in Section 4. The adaptation of Proposition 2.1 to processes which can be successfully approximated by m-dependent processes is given next.

Proposition 2.2 *Suppose for each $n \geq 1$, $m \geq 1$, $\{X_{n,i}^{(m)}, i \geq 1\}$ is a stationary sequence of m-dependent random elements of $\mathbb{E}$ and for each $n \geq 1$, $\{X_{n,i}, i \geq 1\}$ is a stationary sequence of random elements of $\mathbb{E}$. Suppose there exist Radon measures $\nu^{(m)}$ on $\mathbb{E}$ and a sequence $k = k(n)$, $k \to \infty$ and $n/k \to \infty$ such that, for any fixed $m \geq 1$*

$$\frac{n}{k}P(X_{n,i}^{(m)} \in \cdot) \xrightarrow{v} \nu^{(m)} \tag{2.12}$$

as $n \to \infty$. Suppose further that

$$\nu^{(m)} \xrightarrow{v} \nu \tag{2.13}$$

as $m \to \infty$. Finally, assume that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{n}{k}P(|X_{n,1}^{(m)} - X_{n,1}| > \epsilon) = 0, \tag{2.14}$$

for all $\epsilon > 0$. Then

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_{n,i}} \Rightarrow \nu \tag{2.15}$$

in $M_+(\mathbb{E})$.

Moreover if $X_{n,i} = X_i/b_n$, $i = 1,\ldots,n$, *where* $\{X_n, n \geq 1\}$ *is a stationary sequence,* $b_n \to \infty$ *and* ν *satisfies* $\int_1^\infty \log(u)\nu(du) < \infty$, *then*

$$H^X_{k,n} := \frac{1}{k}\sum_{i=1}^{k} \log\frac{X_{(i)}}{X_{(k+1)}} \xrightarrow{P} \int_1^\infty \log(u)\nu(du). \tag{2.16}$$

Proof. We first show that for any fixed m

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{X^{(m)}_{i,n}} \Rightarrow \nu^{(m)} \tag{2.17}$$

by checking that the hypotheses of Proposition 2.1 hold for $\{X^{(m)}_{n,i}, i \geq 1\}$. Since (2.12) holds, we need only check condition (2.1) and condition (2.4). Condition (2.4) holds trivially since for $l(n) > m$, $\{\sum_{i\in I_j} f(X^{(m)}_{i,n}), j = 1,\ldots,p\}$ are independent random variables. To check condition (2.1) note that

$$\begin{aligned}
&\frac{n}{k^2}\sum_{j=2}^{k} P\left(X^{(m)}_{n,1} > x, X^{(m)}_{n,j} > y\right) \\
&\leq \frac{n}{k^2}\left(\sum_{j=2}^{m} P(X^{(m)}_{n,1} > x, X^{(m)}_{n,j} > y) + \sum_{j=m+1}^{k} P(X^{(m)}_{n,1} > x, X^{(m)}_{n,j} > y)\right) \\
&\leq \frac{n}{k^2}\sum_{j=2}^{m} P(X^{(m)}_{n,j} > y) + \frac{n}{k}P(X^{(m)}_{n,1} > x)P(X^{(m)}_{n,1} > y) \\
&= \frac{(m-1)n}{k^2}P(X^{(m)}_{n,1} > y) + \frac{k}{n}\frac{n^2}{k^2}P(X^{(m)}_{n,1} > x)P(X^{(m)}_{n,1} > y) \\
&= \frac{m-1}{k}\left(\nu^{(m)}((y,\infty]) + o(1)\right) + \frac{k}{n}\left(\nu^{(m)}((x,\infty])\nu^{(m)}((y,\infty]) + o(1)\right).
\end{aligned}$$

Therefore

$$\lim_{n\to\infty}\frac{n}{k^2}\sum_{j=2}^{k} P(X^{(m)}_{n,1} > x, X^{(m)}_{n,j} > y) \to 0$$

which completes the proof of (2.17).

The proof of (2.15) follows a converging together argument similar to the proof of Proposition 3.3 in Resnick and Stărică (1995). The conclusion (2.16) follows from Proposition 2.4 of Resnick and Stărică (1995) which shows that the convergence of the tail measure implies the consistency of Hill's estimator. □

For dealing with the ARCH process in Section 4, it is better to have a version of Proposition 2.1 adapted for use with sets rather than $C_K^+(\mathbb{E})$ functions. This is given next.

Proposition 2.3 *Suppose all the assumptions of Proposition 2.1 hold except that in place of condition (2.4) we assume*

$$\lim_{n\to\infty} |E\prod_{j=1}^{[n/k]}(1 - \frac{1}{k}\sum_{i\in I_j} f(X_{n.i})) - \prod_{j=1}^{[n/k]} E(1 - \frac{1}{k}\sum_{i\in I_j} f(X_{n.i}))| = 0 \tag{2.18}$$

for any function f of the form $f = \sum_{h=1}^{s} \beta_h 1_{[(x_h,\infty]}$ where $\beta_h > 0$, $h = 1,\ldots,s$, and $x_h > 0$, $h = 1,\ldots,s$ Then the conclusions of Proposition 2.1 hold.

Proof. We will use the fact that $\nu_n \Rightarrow \nu$ in $M_+(\mathbb{E})$ provided

$$(\nu_n(I_1),\ldots\nu_n(I_s)) \Rightarrow (\nu(I_1),\ldots\nu(I_s)), \quad (n\to\infty) \tag{2.19}$$

in $\mathbb{R}^s$, for any s and any intervals $I_i = (x_i,\infty]$, $i = 1,2,\ldots,s$ (Kallenberg (1983)). Using multivariate Laplace transforms we must show that for any positive $\beta_1,\ldots,\beta_s$ and $f = \sum_{h=1}^{s}\beta_h 1_{(x_h,\infty]}$

$$E\exp\left(-\frac{1}{k}\sum_{i=1}^{n} f(X_{n,i})\right) \to \exp(-\nu(f)). \tag{2.20}$$

Define blocks I_j, I_j^* as in Proposition 2.1 and decompose

$$\begin{aligned}
&\left|E\exp\left(-\frac{1}{k}\sum_{i=1}^{n} f(X_{n,i}))\right) - \exp(-\nu(f))\right| \\
&\le \left|E\exp\left(-\frac{1}{k}\sum_{i=1}^{n} f(X_{n,i}))\right) - E\exp\left(-\frac{1}{k}\sum_{j=1}^{p}\sum_{i\in I_j} f(X_{n,i}))\right)\right| \\
&\quad + \left|E\exp\left(-\frac{1}{k}\sum_{j=1}^{p}\sum_{i\in I_j} f(X_{n,i}))\right) - E\prod_{j=1}^{p}\left(1-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\right)\right| \\
&\quad + \left|E\prod_{j=1}^{p}\left(1-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\right) - \prod_{j=1}^{p}E\left(1-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\right)\right| \\
&\quad + \left|\left(1-\frac{1}{p}\frac{p(k-l_n)}{k}Ef(X_{n,1})\right)^p - \exp(-\nu(f))\right| \\
&= I + II + III + IV.
\end{aligned}$$

I is controlled as in Proposition 2.1. For II, denote $Q := \max_h \beta_h = \sup_{\mathbb{E}} f(x)$ and we have

$$\begin{aligned}
II &\le E\sum_{j=1}^{p}\left|\exp\left(-\frac{1}{k}\sum_{i\in I_j} f(X_{n,i}))\right) - 1 + \frac{1}{k}\sum_{i\in I_j} f(X_{n,i})\right| \\
&\le \frac{p}{2k^2}E(\sum_{i=1}^{k-l_n} f(X_{n,i})))^2 \\
&\le \frac{n}{2k^2}Ef^2(X_{n,1}) + \frac{n}{k^2}\sum_{j=2}^{k} Ef(X_{n,1})f(X_{n,j}) \\
&\le \frac{n}{2k^2}Ef^2(X_{n,1}) + \sum_{h=1}^{s}\sum_{g=1}^{s} Q^2\frac{n}{k^2}\sum_{j=2}^{k} P(X_{n,1} > x_h, X_{n,j} > x_g).
\end{aligned}$$

By condition (2.1) and (2.5) it follows that $\lim_{n\to\infty} II = 0$. Condition (2.18) is equivalent to $\lim_{n\to\infty} III = 0$. By (2.5) and (2.2)

$$\left(1-\frac{1}{p}\frac{p(k-l_n)}{k}Ef(X_{n,1})\right)^p \to \exp(-\nu(f)),$$

the conclusion (2.6) of the Proposition follows. The rest is the same as in Proposition 2.1.□

3 Examples.

We now consider three examples of heavy-tailed dependent, stationary processes which have m-dependent approximations and in each case we apply Proposition 2.2 to demonstrate the consistency of Hill's estimator. The three classes of processes are

- infinite order moving averages of iid heavy tailed random variables,
- the bilinear processes driven by heavy tailed innovations and
- processes satisfying a simple stochastic difference equation with random coefficients.

The first two processes are constructed using a sequence $\{Z_t, -\infty < t < \infty\}$ of iid random variables which for simplicity we take to be positive. These random variables have regularly varying tail probabilities; that is, for $x > 0$,

$$P[Z_1 > x] =: 1 - F(x) =: \bar{F}(x) = x^{-\alpha} L(x), \quad \alpha > 0, \tag{3.1}$$

where L is a slowly varying function at ∞.

3.1 Infinite order moving averages.

Suppose, that the sequence $\{c_i, i \geq 0\} \in \mathbb{R}^\infty$ contains at least one positive number and satisfies

$$0 < \sum_{j=0}^{\infty} |c_j|^\delta < \infty \tag{3.2}$$

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

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

and

$$\lim_{x \to \infty} \frac{P(\sum_{j=0}^{\infty} c_j Z_j > x)}{P(Z_1 > x)} = \sum_{\substack{j=0 \\ c_j > 0}}^{\infty} c_j^\alpha \tag{3.3}$$

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

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \quad -\infty < t < \infty. \tag{3.4}$$

Causal ARMA processes can be represented in the form (3.4) (Brockwell and Davis (1991), Chapter 3). The consistency of the Hill estimator for MA(∞) processes was considered in detail in Resnick and Stărică (1995). See also Resnick and Stărică (1996b).

3.2 The simple bilinear model.

Let X_t be the stationary bilinear model

$$X_t = cX_{t-1}Z_{t-1} + Z_t, \quad -\infty < t < \infty \tag{3.5}$$

where $c > 0$ is a positive constant satisfying

$$c^{\alpha/2} E Z_1^{\alpha/2} < 1. \tag{3.6}$$

Using the bilinear recursion formula (3.5), X_t can be written as an infinite series whose convergence is guaranteed by (3.6) (see Davis and Resnick (1995))

$$X_t = \sum_{j=0}^{\infty} c^j X_t^{(j)} \tag{3.7}$$

where

$$X_t^{(0)} = Z_t, \quad X_t^{(j)} = \left(\prod_{i=1}^{j-1} Z_{t-i} \right) Z_{t-j}^2, \quad j \geq 1.$$

Corollaries 2.3 and 2.4 of Davis and Resnick (1995) show that

$$\lim_{x\to\infty} \frac{P(\sum_{j=0}^m c^j X_t^{(j)} > x)}{P(Z_1^2 > x)} = \sum_{j=1}^{m} c^{j\alpha/2} \left(E Z_1^{\alpha/2} \right)^{j-1} \tag{3.8}$$

and

$$\lim_{x\to\infty} \frac{P(\sum_{j=0}^\infty c^j X_t^{(j)} > x)}{P(Z_1^2 > x)} = \sum_{j=1}^{\infty} c^{j\alpha/2} \left(E Z_1^{\alpha/2} \right)^{j-1} = \frac{c^{\alpha/2}}{1 - c^{\alpha/2} E Z_1^{\alpha/2}}. \tag{3.9}$$

3.3 Solutions of stochastic difference equations.

Let $\{Y_t, -\infty < t < \infty\}$ be a process which satisfies the stochastic difference equation

$$Y_t = A_t Y_{t-1} + Z_t, \quad -\infty < t < \infty, \tag{3.10}$$

where $\{(A_n, Z_n), -\infty < n < \infty\}$ are iid $\mathbb{R}_+^2$-valued random pairs (cf. Vervaat (1979), Grincevicius (1975)). For the case which we consider here, Z_1 will have regularly varying tail probabilities and the tail of Z_1 is heavier than that of A_1. We assume the pair (A_0, Z_0) satisfies

$$E A_0^\alpha < 1, \quad E A_0^\beta < \infty \tag{3.11}$$

for some $0 < \alpha < \beta$ and as usual

$$P(Z_0 > x) = x^{-\alpha} L(x), \tag{3.12}$$

where L is a slowly varying function at infinity. By iterating (3.10) we find for $t \geq 1$,

$$Y_t = \sum_{j=0}^{\infty} \left(\prod_{i=t-j+1}^{t} A_i \right) Z_{t-j} := \sum_{j=0}^{\infty} Y_t^{(j)} \tag{3.13}$$

(where $\prod_{i=t+1}^{t} A_i = 1$). It is suggestive to also write

$$Y_t = \sum_{t=0}^{\infty} C_{t,j} Z_{t-j}, \quad t \geq 1$$

where $C_j = \prod_{i=t-j+1}^{t} A_i$ so that the process is a random coefficient MA(∞) process. Furthermore (Resnick and Willekens, 1990, Theorem 2.1 and Grincevicius, 1975)

$$\frac{P(\sum_{j=0}^{m} Y_t^{(j)} > x)}{P(Z_0 > x)} = \sum_{j=1}^{m} (EA_0^\alpha)^{j-1} \tag{3.14}$$

and

$$\frac{P(\sum_{j=0}^{\infty} Y_t^{(j)} > x)}{P(Z_0 > x)} = \sum_{j=1}^{\infty} (EA_0^\alpha)^{j-1} = \frac{1}{1 - EA_0^\alpha}. \tag{3.15}$$

We now state the result which applies Proposition 2.2 and yields weak consistency of Hill's estimator for these three processes.

Corollary 3.1 *Suppose $\{Z_t\}$ are iid positive random variables satisfying (3.1). The Hill estimator is consistent for α^{-1} when applied either to the MA(∞) process of Section 3.1 or the solution of the random coefficient difference equation described in Section 3.3. For the simple bilinear process described in Section 3.2, the Hill estimator is consistent for $2/\alpha$.*

Proof. We apply Proposition 2.2. The key in each case is that each process can be approximated by an m-dependent sequence.

To prove the assertion for the simple bilinear process, let $k \to \infty$, $n/k \to \infty$ and define b_n such that

$$\frac{n}{k} P(X_1 > b_n) \to 1 \quad (n \to \infty). \tag{3.16}$$

For $m \geq 1$, let $X_{n,i}^{(m)} := \sum_{j=0}^{m} c^j X_i^{(j)}/b_n$. Define $X_{n,i} := \sum_{j=0}^{\infty} c^j X_i^{(j)}/b_n$. Since by (3.8) and (3.9) we have for $x > 0$,

$$\frac{n}{k} P(\sum_{j=0}^{m} c^j X_1^{(j)}/b_n > x) \to \frac{\sum_{j=1}^{m} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}}{\sum_{j=1}^{\infty} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}} x^{-\alpha/2},$$

we may define the measures $\nu^{(m)}$ of Proposition 2.2 by

$$\nu^{(m)}((x,\infty]) := \frac{\sum_{j=1}^{m} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}}{\sum_{j=1}^{\infty} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}} x^{-\alpha/2}.$$

Note that $\nu^{(m)} \xrightarrow{v} \nu$, where $\nu((x,\infty]) = x^{-\alpha/2}$. Since

$$\begin{aligned}\lim_{m\to\infty} \lim_{n\to\infty} \frac{n}{k} P(|X_{n,1}^{(m)} - X_{n,1}| > \epsilon) &= \lim_{m\to\infty} \lim_{n\to\infty} \frac{n}{k} P(\sum_{j=m+1}^{\infty} c^j X_1^{(j)}/b_n > \epsilon) \\ &= \lim_{m\to\infty} \frac{\sum_{j=m+1}^{\infty} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}}{\sum_{j=1}^{\infty} c^{j\alpha/2} \left(EZ_1^{\alpha/2}\right)^{j-1}} \epsilon^{-\alpha/2} = 0\end{aligned}$$

the condition (2.14) of Proposition 2.2 is also verified which proves consistency.

The proofs of the results for the MA(∞) process and the solution of the stochastic difference equation are very similar. $\square$

We simulated the bilinear process to get a sample of size 5000 using Pareto distributed Z's satisfying

$$P[Z_1 > x] = x^{-1}, \quad x > 1.$$

In Figure 3.1 we show a Hill plot of $\{(k, H_{k,n}^{-1}), 1 \leq k \leq 5000\}$. The graph hovers between 0.5 and 0.6. The correct answer is 0.5.

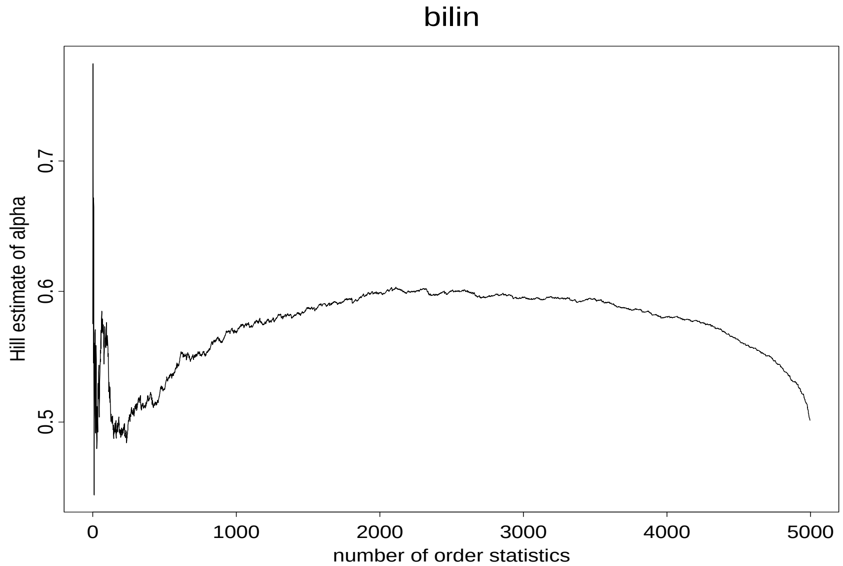

Figure 3.1

4 Tail estimation for solutions of stochastic difference equations and ARCH processes.

In this section we consider tail estimation for the process $\{Y_t, -\infty < t < \infty\}$ which satisfies the stochastic difference equation

$$Y_t = A_t Y_{t-1} + B_t, \quad -\infty < t < \infty, \tag{4.1}$$

where $\{(A_n, B_n), -\infty < n < \infty\}$ are iid $\mathbb{R}_+^2$-valued random pairs. In contrast to Section 3, we will now make different assumptions on the tail behavior of the pair (A_n, B_n) which preclude truncating the series solution of (4.1). Solutions to (4.1) include as a particular case the first order autoregressive conditional heteroschedastic (ARCH) process introduced by Engle (1982). The first order ARCH process is defined by

$$\xi_t = X_t(\beta + \lambda \xi_{t-1}^2)^{1/2}, \quad -\infty < t < \infty, \tag{4.2}$$

where $\{X_t\}$ are iid $N(0,1)$ random variables, $\beta > 0$, $0 < \lambda < 1$. Thus $\{\xi_t^2\}$ satisfies (4.1) with $A_t = \lambda X_t^2$, $B_t = \beta X_t^2$. (Higher order ARCH processes would satisfy higher order versions of (4.1) but these are not considered here.)

It is known (Kesten (1973), Vervaat (1979), Goldie (1991)) that if there exists $\alpha > 0$ with

$$EA_0^\alpha = 1, \quad EA_0^\alpha \log^+ A_0 < \infty, \quad 0 < EB_0^\alpha < \infty, \tag{4.3}$$

if $B_0/(1 - A_0)$ is non-degenerate and if the conditional distribution of $\log A_0$ given $A_0 \neq 0$ is nonlattice, then there exists a constant $c > 0$ such that as $x \to \infty$,

$$P(Y_t > x) \sim cx^{-\alpha}. \tag{4.4}$$

Furthermore, (cf. de Haan, Resnick, Rootzen, de Vries (1989), page 220) under the assumptions (4.3) there exists a γ such that $0 < \gamma < \alpha$ and $0 < c_0 < 1$ such that

$$EA_0^\gamma = c_0 < 1. \tag{4.5}$$

By iterating (4.1) we find for $t \geq 1$,

$$Y_t = \sum_{j=0}^{\infty} \left(\prod_{i=t-j+1}^{t} A_i \right) B_{t-j} := \sum_{j=0}^{\infty} Y_t^{(j)} \tag{4.6}$$

(where $\prod_{i=t+1}^t A_i = 1$). If we iterate (4.1) $t - s$ times for $s < t$ we get

$$Y_t = \sum_{j=0}^{t-s-1} Y_t^{(j)} + \left(\prod_{k=s+1}^{t} A_k \right) Y_s := Y_t^{s,t} + \Pi_{s+1}^t Y_s \tag{4.7}$$

where

$$\Pi_{s+1}^t = A_t A_{t-1} \cdots A_{s+1} \tag{4.8}$$

and

$$Y_t^{s,t} = B_t + A_t B_{t-1} + A_t A_{t-1} B_{t-2} + \ldots + A_t A_{t-1} \cdots A_{s+2} B_{s+1}. \tag{4.9}$$

Observe that $Y_t^{s,t}$ and Y_s are independent random variables as are Π_{s+1}^t and Y_s.

We begin with a lemma designed to help us check conditions (2.1) and (2.4) for the solution of the stochastic difference equation (4.1).

Lemma 4.1 *Assume (4.3) holds and that $\epsilon > 0$ is given. Suppose $i_1 < i_2 < \ldots < i_s$ and $x_i > 0$ for $i = 1, \ldots, s$. Recall the definition of γ and c_0 from (4.5).*

(a) We have that

$$\begin{aligned} &|P(Y_{i_1} > x_1, \ldots, Y_{i_s} > x_s) - \prod_{l=1}^{s} P(Y_{i_l} > x_l)| \\ &\leq \sum_{q=1}^{s-1} \left(\prod_{j=1}^{s-q} P(Y_0 > x_j) P(Y_0 \in (x_{s-q+1} - \epsilon, x_{s-q+1} + \epsilon]) \prod_{j=s-q+2}^{s} P(Y_0 > x_j - \epsilon) \right) \\ &\quad + \sum_{j=2}^{s} P(\Pi_1^{i_j - i_{j-1}} Y_0 > \epsilon). \end{aligned} \tag{4.10}$$

(b) There exists $M = M(x_1, x_2, \ldots, x_s)$ and $K = K(x_1, \ldots, x_s)$ such that, for n large enough,

$$
\begin{aligned}
&|P(Y_{i_1} > (n/k)^{1/\alpha} x_1, \ldots, Y_{i_s} > (n/k)^{1/\alpha} x_s) - P(Y_{i_1} > (n/k)^{1/\alpha} x_1) \ldots P(Y_{i_s} > (n/k)^{1/\alpha} x_s)| \\
&\qquad \le K\epsilon(s-1)M^{s-1}(k/n)^s + \epsilon^{-\gamma} E Y_0^\gamma (k/n)^{\gamma/\alpha} \sum_{j=2}^{s} c_0^{i_j - i_{j-1}}. \qquad (4.11)
\end{aligned}
$$

(c) There exists $C < \infty$ such that

$$
P(Y_1 > (n/k)^{1/\alpha} x, Y_t > (n/k)^{1/\alpha} y) \le P(Y_0 > (n/k)^{1/\alpha} x) P(Y_0 > (n/k)^{1/\alpha}(y-\epsilon)) + C\frac{k}{n} c_0^{t-1}. \quad (4.12)
$$

Proof. The conclusion of (a) follows from an induction argument. To keep the notation simple we prove (a) for $s = 2$ and then derive the result for $s = 3$. The basic ingredient of the proof is the observation in (4.7). We have for $s < t$, $x > 0$, $y > 0$

$$
\begin{aligned}
P(Y_s > x, Y_t > y) &= P(Y_s > x, Y_t^{s,t} + \Pi_{s+1}^t Y_s > y) \\
&\le P(Y_s > x, Y_t^{s,t} + \Pi_{s+1}^t Y_s > y, \Pi_{s+1}^t Y_s \le \epsilon) + P(Y_s > x, \Pi_{s+1}^t Y_s > \epsilon) \\
&\le P(Y_s > x) P(Y_t^{s,t} > y - \epsilon) + P(Y_s > x, \Pi_{s+1}^t Y_s > \epsilon) \qquad (4.13) \\
&\le P(Y_s > x) P(Y_t > y - \epsilon) + P(Y_s > x, \Pi_{s+1}^t Y_s > \epsilon) \qquad (4.14) \\
&= P(Y_s > x) P(Y_t > y) + P(Y_0 > x) P(y - \epsilon < Y_0 \le y) + P(\Pi_1^{t-s} Y_0 > \epsilon).
\end{aligned}
$$

This shows that

$$
P(Y_s > x, Y_t > y) - P(Y_s > x) P(Y_t > y) \le P(Y_0 > x) P(Y_0 \in (y - \epsilon, y]) + P(\Pi_1^{t-s} Y_0 > \epsilon).
$$

From (4.14) we also get

$$
P(Y_s > x, Y_t > y) \le P(Y_0 > x) P(Y_0 > y - \epsilon) + P(Y_0 > x, \Pi_1^{t-s} Y_0 > \epsilon). \qquad (4.15)
$$

We will use (4.15) in the proof of (c).

The other half of the inequality in (a) is derived as follows:

$$
\begin{aligned}
P(Y_s > x) P(Y_t > y + \epsilon) &\le P(Y_s > x) P(Y_t^{s,t} + \Pi_{s+1}^t Y_s > y + \epsilon, \Pi_{s+1}^t Y_s \le \epsilon) + P(\Pi_{s+1}^t Y_s > \epsilon) \\
&\le P(Y_s > x, Y_t^{s,t} > y) + P(\Pi_{s+1}^t Y_s > \epsilon) \\
&\le P(Y_s > x, Y_t > y) + P(\Pi_1^{t-s} Y_0 > \epsilon). \qquad (4.16)
\end{aligned}
$$

Hence

$$
\begin{aligned}
-P(Y_0 > x)\, P\,(y < Y_0 < y + \epsilon) - P(\Pi_1^{t-s} Y_0 > \epsilon) \\
\le P(Y_s > x, Y_t > y) - P(Y_s > x) P(Y_t > y) \\
\le P(Y_0 > x) P(y - \epsilon < Y_0 \le y) + P(\Pi_1^{t-s} Y_0 > \epsilon).
\end{aligned}
$$

The conclusion of (a) for $s=2$ follows. Based on the case $s=2$ we will now prove the inequality for $s=3$. For $s<t<u$ and $x>0$, $y>0$ and $z>0$ we have

$$\begin{aligned}
P(Y_s > x\ ,\ Y_t > y, Y_u > z) &\le P(Y_s > x, Y_t > y)P(Y_u^{t,u} > z-\epsilon) + P(\Pi_{t+1}^{u} Y_t > \epsilon) \\
&\le P(Y_s > x)P(Y_t > y)P(Y_u^{t,u} > z-\epsilon) \\
&\qquad + P(Y_0 > x)P(y-\epsilon < Y_0 \le y)P(Y_u^{t,u} > z-\epsilon) + P(\Pi_1^{t-s} Y_0 > \epsilon) \\
&\qquad + P(\Pi_1^{u-t} Y_0 > \epsilon) \\
&\le P(Y_s > x)P(Y_t > y)P(Y_u > z) \\
&\qquad + P(Y_0 > x)P(Y_0 > y)P(Y_0 \in (z-\epsilon, z]) \\
&\qquad + P(Y_0 > x)P(y-\epsilon < Y_0 \le y)P(Y_0 > z-\epsilon) \\
&\qquad + P(\Pi_1^{t-s} Y_0 > \epsilon) + P(\Pi_1^{u-t} Y_0 > \epsilon).
\end{aligned}$$

For the other half of the inequality, use (4.16) and independence and we get

$$\begin{aligned}
P(Y_s > x)\,P(Y_t > y)P(Y_u > z+\epsilon) &\le P(Y_s > x)P(Y_t > y)P(Y_u^{t,u} > z) + P(\Pi_{t+1}^{u} Y_t > \epsilon) \\
&\le P(Y_s > x, Y_t > y)P(Y_u^{t,u} > z) \\
&\qquad + P(Y_s > x)P(Y_t \in (y-\epsilon, y])P(Y_u^{t,u} > z) \\
&\qquad + P(\Pi_1^{t-s} Y_0 > \epsilon) + P(\Pi_1^{u-t} Y_0 > \epsilon) \\
&\le P(Y_s > x, Y_t > y, Y_u > z) \\
&\qquad + P(Y_0 > x)P(Y_0 \in (y-\epsilon, y])P(Y_0 > z-\epsilon) \\
&\qquad + P(\Pi_1^{t-s} Y_0 > \epsilon) + P(\Pi_1^{u-t} Y_t > \epsilon).
\end{aligned}$$

Therefore

$$\begin{aligned}
-P(Y_0 > x\ )\ & P(Y_0 > y)P(Y_0 \in (z, z+\epsilon]) - P(Y_0 > x)P(Y_0 \in (y-\epsilon, y])P(Y_0 > z-\epsilon) \\
&\qquad - P(\Pi_1^{t-s} Y_0 > \epsilon) - P(\Pi_1^{u-t} Y_t > \epsilon) \\
&\le P(Y_s > x, Y_t > y, Y_u > z) - P(Y_s > x)P(Y_t > y)P(Y_u > z) \\
&\le P(Y_0 > x)P(Y_0 > y)P(Y_0 \in (z-\epsilon, z]) + P(Y_0 > x)P(Y_0 \in (y-\epsilon, y])P(Y_0 > z-\epsilon) \\
&\qquad + P(\Pi_1^{t-s} Y_0 > \epsilon) + P(\Pi_1^{u-t} Y_0 > \epsilon).
\end{aligned}$$

To prove (b), we note that since the inequality in (a) holds whenever $x_i > 0$, $i=1,\ldots,s$ and $\epsilon > 0$ (provided $x_i - \epsilon > 0$, $i=1,\ldots,s$), we may replace x_i by $(n/k)^{1/\alpha} x_i$ and ϵ by $(n/k)^{1/\alpha}\epsilon$ to get a valid inequality. The inequality in (b) then results from the one in (a) by using $c_0 = EA_0^\gamma < 1$, $P[Y_0 > x] \sim cx^{-\alpha}$, $x \to \infty$ and Markov's inequality. To see this, note that the upper bound becomes

$$\sum_{q=1}^{s-1}\left(\prod_{j=1}^{s-q} P[\frac{Y_0}{(\frac{n}{k})^{\frac{1}{\alpha}}} > x_j] P[\frac{Y_0}{(\frac{n}{k})^{\frac{1}{\alpha}}} \in (x_{s-q+1}-\epsilon, x_{s-q+1}+\epsilon]] \prod_{j=s-q+2}^{s} P[\frac{Y_0}{(\frac{n}{k})^{\frac{1}{\alpha}1}} > x_j - \epsilon]\right) \quad (4.17)$$
$$+ \sum_{j=2}^{s} P[\Pi_1^{i_j - i_{j-1}} Y_0 > (n/k)^{1/\alpha}\epsilon].$$

Note that by Markov's inequality

$$\sum_{j=2}^{s} P\,[\Pi_1^{i_j-i_{j-1}} Y_0 > (n/k)^{1/\alpha}]$$
$$\leq (\frac{k}{n})^{\gamma/\alpha}\epsilon^{-\gamma}\sum_{j=2}^{s} E\left(\Pi_1^{i_j-i_{j-1}} Y_0\right)^{\gamma}$$
$$=(\frac{k}{n})^{\gamma/\alpha}\epsilon^{-\gamma}\sum_{j=2}^{s} c_0^{i_j-i_{j-1}} EY_0^{\gamma}.$$

Furthermore, for n sufficiently large and some constant $K = K(x_1, \ldots, x_s)$

$$\frac{n}{k}P\,[Y_0 > (n/k)^{1/\alpha}(x_j-\epsilon)] \leq M, \quad j=1,\ldots,s$$
$$\frac{n}{k}P\,[Y_0 \in (\frac{n}{k})^{1/\alpha}(x_{s-q+1}-\epsilon, x_{s-q+1}+\epsilon]] \leq \epsilon, \quad j=1,\ldots,s$$

and therefore, the first summation in (4.17) is bounded by

$$\sum_{q=1}^{s-1}\prod_{j=1}^{s-q}\tfrac{k}{n}\,M\epsilon K\frac{k}{n}\prod_{j=s-q+2}^{s}\frac{k}{n}M$$
$$=\sum_{q=1}^{s-1}(\frac{k}{n})^{s-q}M^{s-q}\epsilon K\frac{k}{n}(\frac{k}{n})^{q-1}M^{q-1}$$
$$=K\epsilon(s-1)M^{s-1}(\frac{k}{n})^{s}$$

which verifies (4.11).

To prove (c), substitute in (4.15) $s=1$ and replace ϵ, x and y by $(n/k)^{1/\alpha}\epsilon$, $(n/k)^{1/\alpha}x$ and $(n/k)^{1/\alpha}y$. The desired result is shown if we prove

$$P[Y_0 > (\frac{n}{k})^{1/\alpha}x, \Pi_1^{t-1}Y_0 > (\frac{n}{k})^{1/\alpha}\epsilon] \leq c\frac{k}{n}c_0^{t-1}.$$

The probability on the left is

$$\int_x^{\infty} P\,[\Pi_1^{t-1} > \epsilon u^{-1}]P[\frac{Y_0}{(n/k)^{1/\alpha}} \in du]$$
$$\leq c_0^{t-1}\epsilon^{-\gamma}\int_x^{\infty} y^{\gamma}P[\frac{Y_0}{(n/k)^{1/\alpha}} \in du]$$
$$=c_0^{t-1}\epsilon^{-\gamma}E\left(\frac{Y_0}{(n/k)^{1/\alpha}}\right)^{\gamma}1_{[\frac{Y_0}{(n/k)^{1/\alpha}}>x]}$$
$$=c_0^{t-1}\epsilon^{-\gamma}\frac{k}{n}\left(\frac{n}{k}E\left(\frac{Y_0}{(n/k)^{1/\alpha}}\right)\right)$$
$$\leq C\,c_0^{t-1}\frac{k}{n}$$

by Karamata's theorem. □

Lemma 4.2 *Assume (4.3) holds and let* $\{Y_t\}$ *be the solution of (4.1). Then condition (2.1) or (2.8) holds for the array* $\{Y_t/(n/k)^{1/\alpha}\}$; *that is*

$$\lim_{n\to\infty}\frac{n}{k^2}\sum_{j=2}^{k}P(Y_1>(n/k)^{1/\alpha}x, Y_j>(n/k)^{1/\alpha}y)=0 \tag{4.18}$$

for any $x>0$, $y>0$. *If in addition one chooses* l_n *such that* $l_n/k\to 0$ *and*

$$\frac{n}{k}=o(l_n) \tag{4.19}$$

then condition (2.18) also holds; that is

$$\lim_{n\to\infty}|E\prod_{j=1}^{p}(1-\frac{1}{k}\sum_{i\in I_j}f(Y_i/(n/k)^{1/\alpha}))-\prod_{j=1}^{p}E(1-\frac{1}{k}\sum_{i\in I_j}f(Y_i/(n/k)^{1/\alpha}))|=0. \tag{4.20}$$

where $p=[n/k]$, I_j, $j=1,\ldots,p$ *are defined in (2.3) and the function* f *is of the form given in Proposition 2.3.*

Proof. To check condition (2.1) use (c) of Lemma 4.2:

$$\begin{aligned}&\frac{n}{k^2}\sum_{j=2}^{k}P(Y_1>(n/k)^{1/\alpha}x, Y_{j+1}>(n/k)^{1/\alpha}y)\\&\quad\le\frac{n(k-1)}{k^2}P(Y_0>(n/k)^{1/\alpha}x)P(Y_0>(n/k)^{1/\alpha}(y-\epsilon))+C\frac{1}{k}\sum_{j=2}^{k}c_0^{j-1}\to 0\end{aligned}$$

as $n\to\infty$.

To prove condition (2.18) holds, note that

$$\begin{aligned}&E\prod_{j=1}^{p}(1-\frac{1}{k}\sum_{i\in I_j}f(Y_i/(n/k)^{1/\alpha}))-\prod_{j=1}^{p}E(1-\frac{1}{k}\sum_{i\in I_j}f(Y_i/(n/k)^{1/\alpha})) \qquad (4.21)\\&\le\sum_{u=2}^{p}\frac{1}{k^u}\sum_{1\le j_1<j_2<\ldots<j_u\le p}\ \sum_{i_1\in I_{j_1}}\sum_{i_2\in I_{j_2}}\cdots\sum_{i_u\in I_{j_u}}|E(\prod_{v=1}^{u}f(Y_{i_v}/(n/k)^{1/\alpha}))-\prod_{v=1}^{u}Ef(Y_{i_v}/(n/k)^{1/\alpha})|.\end{aligned}$$

Also due to the definition of f one has

$$\begin{aligned}&E(\prod_{v=1}^{u}f(Y_{i_v}/(n/k)^{1/\alpha}))-\prod_{v=1}^{u}Ef(Y_{i_v}/(n/k)^{1/\alpha})\\&=\sum_{h_1=1}^{s}\cdots\sum_{h_u=1}^{s}\beta_{h_1}\ldots\beta_{h_u}|P(Y_{i_1}/(n/k)^{1/\alpha}>x_{h_1},\ldots,Y_{i_u}/(n/k)^{1/\alpha}>x_{h_u}) \qquad (4.22)\\&\quad-P(Y_{i_1}/(n/k)^{1/\alpha}>x_{h_1})\ldots P(Y_{i_u}/(n/k)^{1/\alpha}>x_{h_u})|.\end{aligned}$$

From (4.11) it follows that

$$\begin{aligned}&|P(Y_{i_1}/(n/k)^{1/\alpha}>x_{h_1},\ldots,Y_{i_u}/(n/k)^{1/\alpha}>x_{h_u})-\prod_{j=1}^{u}P(Y_{i_j}/(n/k)^{1/\alpha}>x_{h_1})|\\&\qquad\le K\epsilon(u-1)M^u(k/n)^u+\epsilon^{-\gamma}EY_0^{\gamma}(k/n)^{\gamma/\alpha}(u-1)c_0^{l_n}.\end{aligned}$$

If we denote $Q := \max\{\beta_h : h = 1, \ldots, s\}$ then one can bound (4.21) by

$$\begin{aligned}
&\sum_{u=2}^{p} \frac{1}{k^u} \sum_{1 \le j_1 < j_2 < \ldots < j_u \le p} \sum_{i_1 \in I_{j_1}} \sum_{i_2 \in I_{j_2}} \cdots \sum_{i_u \in I_{j_u}} (sQ)^u \left(\epsilon (u-1)(M\frac{k}{n})^u + \epsilon^{-\gamma} E Y_0^\gamma (\frac{k}{n})^{\gamma/\alpha} (u-1) c_0^{l_n} \right) \\
&\le \sum_{u=2}^{p} \frac{1}{k^u} \binom{p}{u} (k - l_n)^u (sQ)^u \left(\epsilon K (u-1) M^u (k/n)^u + \epsilon^{-\gamma} E Y_0^\gamma (k/n)^{\gamma/\alpha} (u-1) c_0^{l_n} \right) \\
&\le K\epsilon \sum_{u=2}^{p} \binom{p}{u} (u-1)(sQM)^u (k/n)^u + \epsilon^{-\gamma} E Y_0^\gamma (k/n)^{\gamma/\alpha} c_0^{l_n} \sum_{u=2}^{p} \binom{p}{u} (u-1) \\
&= K\epsilon \left(\frac{sQMpk}{n} \left(1 + \frac{sQMk}{n} \right)^{p-1} - \left(1 + \frac{sQMk}{n} \right)^p + 1 \right) \\
&\quad + \epsilon^{-\gamma} E Y_0^\gamma (k/n)^{\gamma/\alpha} c_0^{l_n} (\frac{n}{k} 2^{n/k-1} - 2^{n/k} + 1) \\
&= A + B.
\end{aligned}$$

When $n \to \infty$, B goes to 0, due to (4.19) and $A \to \epsilon((sQM - 1)\exp(sQM) + 1)$. Letting $\epsilon \to 0$ ends the proof for condition (4.20). □

Proposition 4.1 *Assume (4.3) holds and let $\{Y_t\}$ be the solution of (3.10). Choose $k(n)$ such that*

$$n = o(k^{3/2}). \tag{4.23}$$

Then the Hill estimator applied to the sequence Y_t is consistent, i.e.

$$\frac{1}{k} \sum_{i=1}^{k} \log \frac{Y_{(i)}}{Y_{(k+1)}} \xrightarrow{P} \frac{1}{\alpha}. \tag{4.24}$$

Proof. Due to (4.23) it is possible to choose l_n such that $n/k << l_n << k^2/n$. This choice makes sure that (2.2) and (4.19) hold. The conclusion then follows from Proposition 2.1. □

For the ARCH process $\{\xi_t\}$ given by (4.2), we have

$$P[\xi_t^2 > x] \sim c x^{-\alpha}, \quad x \to \infty$$

where α satisfies

$$E(\lambda X_t^2)^\alpha = 1$$

with $\{X_t\}$ being iid $N(0,1)$ random variables. Equivalently, α satisfies

$$\Gamma(\alpha + \frac{1}{2}) = \sqrt{\pi}(2\lambda)^{-\alpha}.$$

Thus the Hill estimator applied to $\{\xi_1^2, \ldots, \xi_t^2\}$ is consistent for α^{-1} and a consistent estimator for λ is obtained from solving

$$\Gamma(\hat\alpha + \frac{1}{2}) = \sqrt{\pi}(2\hat\lambda)^{-\hat\alpha}$$

for $\hat\lambda$, where $\hat\alpha$ is the estimate of α given by the reciprocal of the Hill estimator.

We simulated 7000 data from the ARCH(1) model using $\beta = 1$ and $\lambda = 0.5$. In this case, the true value of α for $\{\xi_t^2\}$ is $\alpha \approx 2.365$. Figure 4.1 displays the Hill plots which indicate an estimate of α in the neighborhood of 2.1 or 2.2. The AltHill plot in the display is $\{(\theta, H^{-1}_{[n^\theta],n}, 0 \le \theta \le 1\}$ and the AltsmooHill plot smooths the AltHill plot. See Resnick and Starica (1996a) for a discussion of such plots.

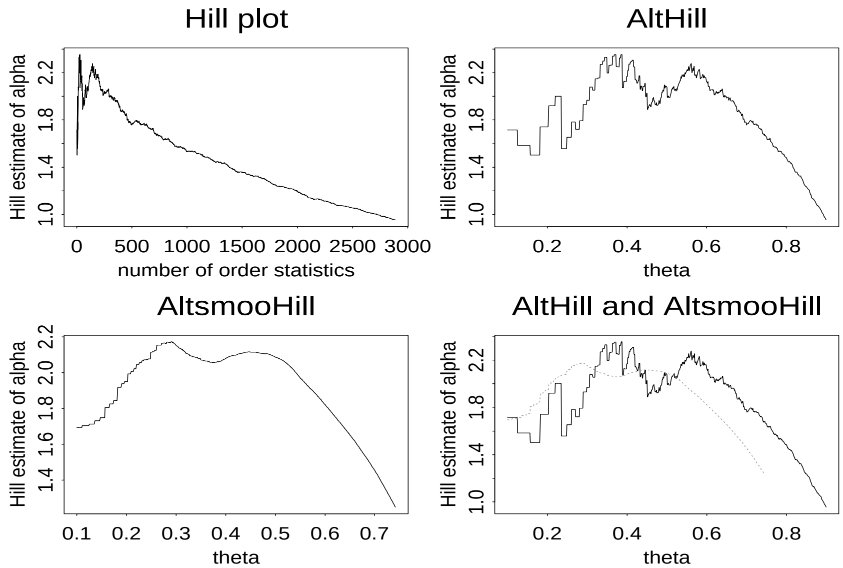

Figure 4.1. Hill plots of ARCH^2 when $\lambda = 0.5$.

5 Hidden Markov models.

A heavy tailed hidden Markov model is proposed in Meier-Hellstern et al (1991) to model the times between transmission of packets at a source. We show the Hill estimator is consistent when applied to such models.

The model has the following ingredients. Let $\{J_n, n \ge 0\}$ be an ergodic, m-state Markov chain on the state space $\{1, 2, \dots, m\}$. Suppose the transition probability matrix of this chain is $P = \{p_{ij}, 1 \le i, j \le m\}$ and that the stationary distribution is $\pi' = (\pi_1, \dots, \pi_m)$. Now suppose for $i = 1, \dots, m$ we are given holding time distributions $\{q_n^{(i)}, n \ge 1\}$ concentrating on $\{1, 2, \dots\}$ and that for $i = 1, \dots, m$, $\{D_n^{(i)}, n \ge 0\}$ are iid with common distribution $\{q_n^{(i)}\}$. Define $\{V_n, n \ge 0\}$ by

$$\begin{aligned} V_j &= J_0, \text{ if } 0 \le j < D_0^{(J_0)}, \\ &= J_1, \text{ if } D_0^{(J_0)} \le j < D_0^{(J_0)} + D_1^{(J_1)}, \\ &= J_2, \text{ if } D_0^{(J_0)} + D_1^{(J_1)} \le j < D_0^{(J_0)} + D_1^{(J_1)} + D_2^{(J_2)}, \\ \vdots \quad & \quad \vdots \end{aligned}$$

Thus

$$V_j = \sum_{k=0}^{\infty} J_k 1_{[\sum_{l=0}^{k-1} D_l^{(J_l)} \le j < \sum_{l=0}^{k} D_l^{(J_l)}]}.$$

The next ingredient we need are distributions $F_1, \ldots, F_m$ on $\mathbb{R}_+$ and iid uniform random variables with support $[0,1]$ which we call $\{U_n, n \geq 0\}$. Define for $n \geq 0$

$$X_n = F_{V_n}^{\leftarrow}(U_n) \tag{5.1}$$

and assume $\{U_n\}$, $\{J_n\}$, $\{D_n^{(i)}, n \geq 0, 1 \leq i \leq m\}$ are all independent.

So changes of state follow the Markov chain $\{J_n\}$ and a transition from i to j occurs with probability p_{ij}. Having entered state i, the system stays in state i for k time units with probability $q_k^{(i)}$. While in state i, random variables which we think of as interarrivals are generated from distribution F_i.

Proposition 5.1 *Suppose $\{J_n\}$ is a stationary, ergodic Markov chain and that*

$$ED_n^{(i)} < \infty, \quad i = 1, \ldots, m. \tag{5.2}$$

Suppose for $\alpha > 0$

$$\bar{F}_1(x) \sim x^{-\alpha} L(x), \quad x \to \infty \tag{5.3}$$

and

$$\lim_{x \to \infty} \frac{\bar{F}_j(x)}{\bar{F}_1(x)} = 0, \quad j = 2, \ldots, m. \tag{5.4}$$

Define the quantile function

$$b(t) = \left(\frac{1}{1 - F_1}\right)^{\leftarrow}(t).$$

If $k \to \infty$, $n/k \to \infty$ then

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

where

$$\nu((x, \infty] = \theta_1 x^{-\alpha}$$

and for $k = 1, \ldots, m$

$$\theta_k = \frac{ED_1^{(k)} \pi_k}{E \sum_{j=1}^{m} D_1^{(j)} \pi_j}.$$

Furthermore, the Hill estimator applied to $\{X_t\}$ is consistent for α^{-1}.

Proof. The proof uses Laplace functionals. For $f \in C_K^+(\mathbb{E})$, we need to show

$$\Psi_n(f) := E \exp\{-\frac{1}{k} \sum_{j=1}^{n} f(X_j / b(\frac{n}{k}))\} \to e^{-\nu(f)} = \exp\{-\int_{\mathbb{E}} f(x) \nu(dx)\}.$$

Define for $n \geq 0$

$$N_n^{(j)} = \sum_{l=0}^{n} 1_{[V_l = j]}, \quad j = 1, \ldots, m,$$

$$\mu^{(j)}(n) = \sum_{l=0}^{n} 1_{[J_l = j]}, \quad j = 1, \ldots, m.$$

Because $\{X_n\}$ is conditionally independent given $\{V_n\}$ (see (5.1)) we have

$$\Psi_n(f) = E\left(E\left(\exp\{-\frac{1}{k}\sum_{j=1}^{n} f(X_j/b(\frac{n}{k}))\}|V_0,\ldots,V_n\right)\right)$$
$$= E\prod_{j=1}^{m}\left(\int_{\mathbb{E}} e^{-f(x)/k}F_j(b(\frac{n}{k})dx)\right)^{N_n^{(j)}}. \tag{5.5}$$

We now study the behavior of $N_n^{(j)}$ and we will prove that as $n \to \infty$

$$\frac{N_n^{(n)}}{n} \xrightarrow{P} \theta_j, \quad j = 1,\ldots,m. \tag{5.6}$$

The semi-Markov process $\{V_j\}$ changes states at times $\{S_n\}$ where

$$S_n = \sum_{q=0}^{n} D_q^{(J_q)}$$

and as $n \to \infty$ we have

$$\begin{aligned}\frac{S_n}{n} &\stackrel{d}{=} \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{\mu^{(k)}(n)} D_i^{(k)}\\ &= \sum_{k=1}^{m}\frac{\sum_{i=1}^{\mu^{(k)}(n)} D_i^{(k)}}{\mu^{(k)}(n)}\frac{\mu^{(k)}(n)}{n}\\ &\to \sum_{k=1}^{m} ED_1^{(k)}\pi_k.\end{aligned}$$

Now we define the process inverse to $\{S_n\}$ as

$$M(t) = \sup\{n : S_n \le t\}$$

so that, as $t \to \infty$,

$$\frac{M(t)}{t} \to \frac{1}{\sum_{k=1}^{m} ED_1^{(k)}\pi_k}.$$

The relevance of $\{S_n\}$ and $\{M(t)\}$ is that

$$\begin{aligned}\frac{N_n^{(k)}}{n} &\le \frac{1}{n}\sum_{q=1}^{M(n)+1}(S_q - S_{q-1})1_{[J_{q-1}=k]}\\ &= \frac{1}{n}\sum_{q=1}^{M(n)+1} D_q^{(k)}1_{[J_{q-1}=k]}\\ &\stackrel{d}{=} \frac{1}{n}\sum_{q=0}^{\mu^{(k)}(M(n)+1)} D_q^{(k)}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{q=0}^{\mu^{(k)}(M(n)+1)} D_q^{(k)}}{\mu^{(k)}(M(n)+1)} \frac{\mu^{(k)}(M(n)+1)}{M(n)+1} \frac{M(n)+1}{n} \\
&\to \frac{ED_1^{(k)}\pi_k}{\sum_{j=1}^m ED_1^{(j)}\pi_j}.
\end{aligned}
$$

A lower bound is obtained similarly and this proves (5.6).

Note that because of (5.4) we have for $x > 0$ that

$$
\frac{n}{k}\bar{F}_j(b(\frac{n}{k})x) \to 0, \quad j = 2, \ldots, m.
$$

Thus for $2 \leq j \leq m$,

$$
\begin{aligned}
1 &\geq \left(\int_{\mathbb{E}} e^{-f(x)/k} F_j(b(\frac{n}{k})dx) \right)^{N_n^{(j)}} \\
&= \left(1 - \frac{\int_{\mathbb{E}} (1 - e^{-f(x)/k}) F_j(b(\frac{n}{k})dx)}{n} \right)^{N_n^{(j)}} \\
&\geq \left(1 - \frac{\int_{\mathbb{E}} f(x) \frac{n}{k} F_j(b(\frac{n}{k})dx)}{n} \right)^{n(N_n^{(j)})/n} \\
&\to e^{-0} = 1,
\end{aligned}
$$

since $\int_{\mathbb{E}} f(x)\frac{n}{k}F_j(b(\frac{n}{k})dx) \to 0$ if $f \in C_K^+(\mathbb{E})$. For $j = 1$ we claim

$$
\int_{\mathbb{E}} (1 - e^{-f(x)/k}) n F_1(b(\frac{n}{k})dx) \to \nu(f) \tag{5.7}
$$

and assuming this is true we get

$$
\begin{aligned}
\left(\int_{\mathbb{E}} e^{-f(x)/k} F_1(b(\frac{n}{k})dx \right)^{N_n^{(1)}} &= \left(1 - \frac{\int_{\mathbb{E}} (1 - e^{-f(x)/k}) n F_1(b(\frac{n}{k})dx)}{n} \right)^{n(N_n^{(1)}/n)} \\
&\to \exp\{-\nu(f)\}.
\end{aligned} \tag{5.8}
$$

To verify (5.7) observe that

$$
\int_{\mathbb{E}} (1 - e^{-f(x)/k}) n F_1(b(\frac{n}{k})dx) \leq \int_{\mathbb{E}} f(x) \frac{n}{k} F_1(b(\frac{n}{k})dx) \to \nu(f)
$$

and

$$
\begin{aligned}
\int_{\mathbb{E}} (1 - e^{-f(x)/k}) n F_1(b(\frac{n}{k})dx) &\geq \int_{\mathbb{E}} \left(\frac{f(x)}{k} - \frac{f^2(x)}{k^2} \right) n F_1(b(\frac{n}{k})dx) \\
&= \nu(f) + o(1) - \frac{1}{k} \int_{\mathbb{E}} f^2(x) \frac{n}{k} F_1(b(\frac{n}{k})dx) \\
&= \nu(f) + o(1) + \frac{1}{k} O(1) \\
&\to \nu(f).
\end{aligned}
$$

This proves (5.7).

Thus, the factors in $\Psi_n(f)$ in (5.5) not corresponding to state 1 converge to 1 while the factor from state 1 converges to the correct limit. The desired result follows from dominated convergence after taking expectations. □

References

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

[2] Cline, D. (1983) Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Thesis, Dept of Statistics, Colorado State University, Ft. Collins CO 80521 USA.

[3] Davis, R. and Resnick, S. (1984) Tail estimates motivated by extreme value theory. Ann. Statist., 12, 1467–1487.

[4] Davis, R. and Resnick, S. (1988) Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. Stochastic Processes and their Applications, 30, 41–68.

[5] Davis, R. and Resnick, S. (1995) Limit theory for bilinear processes with heavy tailed noise. Available at http://www.orie.cornell.edu/trlist/trlist.html as TR1140.ps.Z. To appear: Annals of Applied Probability.

[6] Engle, R. (1982) Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987–1007.

[7] Feigin, P. and Resnick, S. (1996) Pitfalls of fitting autoregressive models for heavy-tailed time series. Available at http:www.orie.cornell.edu/trlist/trlist.html as TR1163.ps.Z.

[8] Geluk, J and Haan, L. de, Resnick, S. and Stărică, C. (1995) Second order regular variation, convolution and the central limit theorem. Available as TR1133 at http://www.orie.cornell.edu/trlist/trlist.html.

[9] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. Annals of Applied Probability, 1, 126–166.

[10] Granger, C. W. J., Andersen, A. (1978) Non-linear time series modelling. In: Applied Time Series Analysis (Proc. First Sympos., Tulsa, Okla., 1976), 25–38, Academic Press, New York.

[11] Grincevicius, A. (1975) One limit distribution for a random walk on the line. Lithuanian Math. J. 15, 580–589.

[12] Haan, L. de and Resnick, S. I. (1996) On asymptotic normality of the Hill estimator. Preprint available as TR1155.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html.

[13] Haan, L. de, Resnick, S.I., Rootzen, H. and Vries, C. de (1989) Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. Stochastic Processes and their Applications, 32, 213–224.

[14] Hall, P. (1982) On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. Ser. B 44, 37–42.

[15] Häusler, E. and Teugels, J. (1985) On the asymptotic normality of Hill's estimator for the exponent of regular variation. Ann. Statist. 13, 743–756.

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

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

[18] Hsing, T., Husler, J. and Leadbetter, M.R. (1988) On the exceedance point process for a stationary sequence. Probab. Theory and related Fields, 78, 97–112.

[19] Kallenberg, O. (1983) Random Measures. Third edition. Akademie-Verlag, Berlin.

[20] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. Acta. Math., 131, 207–248.

[21] Leadbetter, M., Lindgren, G. and Rootzen, H. (1983) Extremes and Related Properties of Random Sequences and Processes Springer-Verlag, New York.

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

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

[24] Mason, D. and Turova, T. (1994). Weak convergence of the hill estimator process. In J. Galambos, J. Lechner, and Simiu, editors, *Extreme Value Theory and Applications*. Kluwer Academic Publishers, Dordrecht, Holland.

[25] Meier–Hellstern,K.S., Wirth, P.E., Yan,Y.L. and Hoeflin, D.A. (1991) Traffic models for ISDN data users: office automation application. In: Teletraffic and Datatraffic in a Period of Change. Proceedings of the 13th ITC, editors: A. Jensen and V.B. Iversen, 167–192, North Holland, Amsterdam, The Netherlands.

[26] Mohler, Ronald R. (1973) Bilinear control processes with applications to engineering, ecology, and medicine. Mathematics in Science and Engineering, 106. Academic Press, New York.

[27] Resnick, Sidney, (1987) Extreme Values, Regular Variation, and Point Processes. Springer-Verlag, New York.

[28] Resnick, S. (1996) Why non-linearities can ruin the heavy tailed modeler's day. Available as TR1157.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html.

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

[30] Resnick, S. and Stărică, Cătălin (1996a) Smoothing the Hill estimator. To appear: J. Applied Probability.

[31] Resnick, S. and Stărică, Cătălin (1996b) Behavior of Hill's estimator for autoregressive data. Available as TR1165.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html.

[32] Resnick, S. and Willekens, E. (1990) Moving averages with random coefficients and random coefficient autoregressive models. Stochastic Models, 7, 511–526.

[33] Rootzen, H., Leadbetter, M. and de Haan, L. (1990) Tail and quantile estimation for strongly mixing stationary sequences. Technical Report 292, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260.

[34] Vervaat, W. (1979) On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. Adv. Appl. Prob., 11, 750–783.