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

TECHNICAL REPORT NO. 1157

May 1996

# WHY NON-LINEARITIES CAN RUIN THE HEAVY TAILED MODELER'S DAY

by

Sidney I. Resnick

Sidney Resnick was partially supported by the National Science Foundation grant DMS-9400535 at Cornell University and also received some support from NSA Grant MDA904-95-H-1036.

# WHY NON–LINEARITIES CAN RUIN THE HEAVY TAILED MODELER'S DAY

Sidney I. Resnick

Cornell University

Abstract. A heavy tailed time series that can be expressed as an infinite order moving average has the property that the sample autocorrelation function (acf) at lag $h$, converges in probability to a constant $\rho(h)$ despite the fact that the mathematical correlation typically does not exist. A simple bilinear model considered by Davis and Resnick (1996) has the property that the sample autocorrelation function at lag $h$ converges in distribution to a non-degenerate random variable. Examination of various data sets exhibiting heavy tailed behavior reveals that the sample correlation function typically does not behave like a constant. Usually, the sample acf of the first half of the data set looks considerably different than the sample acf of the second half. A possible explanation for this acf behavior is the presence of nonlinear components in the underlying model and this seems to imply that infinite order moving average models and in particular ARMA models do not adequately capture dependency structure in the presence of heavy tails. Some additional results about the simple nonlinear model are discussed and in particular we consider how to estimate coefficients.

**1. Introduction.** There are now numerous data sets from the fields of telecommunications, finance and economics which appear to be compatible with the assumption of heavy-tailed marginal distributions. Examples include file lengths, cpu time to complete a job, call holding times, inter-arrival times between packets in a network and lengths of on/off cycles (Duffy, et al 1993, 1994; Meier–Hellstern et al, 1991; Willinger, Taqqu, Sherman and Wilson, 1995; Crovella and Bestavros, 1995; Cunha, Bestavros and Crovella, 1995).

A key question of course is how to fit models to data which require heavy tailed marginal distributions. 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. For such time series, the autocorrelation function can be well approximated by that of an finite order ARMA$(p,q)$ model. In particular, one can choose an autoregressive model of order $p$ (AR$(p)$) such that the acf of the two models agree for lags $1,\dots,p$ (see Brockwell and Davis (1991), p. 240). So when finite variance models are considered from a second order point of view, linear models are sufficient for data analysis. In the infinite variance case, we have no such confidence that linear models are sufficiently flexible and rich enough for modeling purposes. Yet theoretical attempts to date to study heavy tailed time series models have concentrated effort on ARMA models or infinite order moving averages despite little evidence that such models would actually fit heavy tailed data. Understandably, these attempts were motivated by the desire to see how well classical ARMA models perform in the heavy tailed world. However, the point which this paper emphasizes is that the class of infinite order moving averages is unlikely to provide a sufficiently broad class which is capable of accurately capturing the dependency structure of a variety of heavy tailed data.

Some theoretical perspective on this issue is provided in the interesting work of Rosiński (1995) who decomposes a general symmetric $\alpha$-stable process $\{X(t), t \geq 0\}$ into an independent sum of 3 processes

$$X(t) = X_1(t) + X_2(t) + X_3(t)$$

*Key words and phrases.* heavy tails, regular variation, Hill estimator, Poisson processes, linear programming, autoregressive processes, parameter estimation, weak convergence, consistency, time series analysis, estimation, independence.

Sidney Resnick was partially supported by NSF Grant DMS-9400535 at Cornell University and also received some support from NSA Grant MDA904-95-H-1036.

where $X_1$ is a superposition of moving average processes, $X_2$ is a Harmonizable $\alpha$-stable process and $X_3$ is a process of 'third' type. There is no reason to suppose the moving average processes are in any way dense within the class of symmetric $\alpha$-stable processes and thus no reason to suspect that moving averages should be given a prominent role in data analysis.

The challenge, of course, is to find flexible parametric families of heavy tailed models. One possible family is the class of bilinear models which has received attention in the finite variance world (Gabr and Subba Rao, 1984). Davis and Resnick (1995) study a simple bilinear model and show that the behavior of the sample correlation function is strikingly different for such models than for linear heavy tailed time series. This has the important implication that any heavy tailed inference such as Yule-Walker estimation (Brockwell and Davis, 1991; Resnick, 1995) based on the sample acf, will be dramatically misleading if the analyst fails to adequately account for nonlinearities. This is discussed further in Section 2. Section 3 gives examples of several data sets whose sample correlation function behaves like that of the bilinear process and it is doubtful if such data can be fit by linear heavy tailed models. Further consideration is given in Section 4 to linear programming estimators (Feigin and Resnick, 1992, 1994, 1996; Feigin, Resnick, Stărică, 1995; Feigin, Kratz, Resnick, 1994; Davis and McCormick, 1989) applied to a simple bilinear process. Generalizations of this simple process will be necessary to achieve the flexibility needed of a desirable parametric family.

**2. Sample correlations of linear and bilinear processes.** The sample correlation function is a basic tool in classical time series for not only assessing dependence but also for estimation purposes since, for example, the Yule-Walker estimators of autoregressive coefficients in an autoregressive model depend on sample correlations. (See Brockwell and Davis, 1991; Resnick, 1995.) For a stationary sequence $\{X_n, n = 0, \pm 1, \pm 2, \dots\}$ the classical definition of the sample correlation function at lag $h$ is $(h = 0, 1, \dots)$

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}.$$

When heavy tails are present, and especially when the data is positive as is frequently the case, it makes little sense to center at $\bar{X}$ and the following heavy tailed version is used:

$$\hat{\rho}_H(h) = \frac{\sum_{t=1}^{n-h} X_t X_{t+h}}{\sum_{t=1}^{n} X_t^2}.$$

Consider an infinite order moving average

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{2.1}$$

where $\{Z_t\}$ is an iid sequence of heavy tailed random variables satisfying

$$\begin{aligned} P[|Z_1| > x] &= x^{-\alpha} L(x), \quad (x \to \infty), \\ \frac{P[Z_1 > x]}{P[|Z_1| > x]} &\to p, \quad (x \to \infty), \end{aligned} \tag{2.2}$$

and $\alpha > 0$, $L$ is slowly varying and $0 \le p \le 1$ and the $\psi$'s satisfy mild summability conditions. Note that if $Z_1 \ge 0$, then $p = 1$. If $\alpha < 2$, there is no finite variance and hence the mathematical correlations of the $Z$'s and presumably the $X$'s do not exist. However, Davis and Resnick (1985a,b; 1986) proved that $(\hat{\rho}_H(h), h = 1, \dots, q)$ still has nice asymptotic properties which can be used for assessing dependence and for Yule-Walker estimation. Define

$$\rho(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} / \sum_{j=0} \psi_j^2$$

and then for the linear model (2.1) we have the consistency result

$$\hat{\rho}_H(h) \xrightarrow{P} \rho(h)$$

and a reasonably fast rate of convergence also ensues. This leads to consistency of Yule-Walker estimates of autoregressive coefficients in an AR(p) model and allows for computation of a limit distribution for these estimates.

For contrasting behavior we consider a simple bilinear process satisfying the recursion

$$X_t = cX_{t-1}Z_{t-1} + Z_t, \tag{2.3}$$

where the $Z$'s satisfy (2.2) and

$$|c|^{\alpha/2} E|Z_1|^{\alpha/2} < 1. \tag{2.4}$$

Under this condition (see Liu (1989)), there exists a unique stationary solution to the equations (2.3) given by

$$X_t = \sum_{j=0}^{\infty} c^j Y_t^{(j)},$$

where

$$Y_t^{(j)} = \begin{cases} Z_t, & \text{if } j = 0, \\ \left(\prod_{i=1}^{j-1} Z_{t-i}\right) Z_{t-j}^2, & \text{if } j \geq 1. \end{cases} \tag{2.5}$$

Define $b_n$ to be the $1 - n^{-1}$ quantile of $|Z_1|$, i.e.

$$b_n = \inf\{x : P[|Z_1| > x] < n^{-1}\}. \tag{2.6}$$

In order to describe the basic result about acf's of bilinear processes and also for later work on estimation, we review rapidly some notation and concepts about point processes. For a locally compact Hausdorff topological space $\mathbb{E}$, we let $M_p(\mathbb{E})$ be the space of Radon point measures on $\mathbb{E}$. This means $m \in M_p(\mathbb{E})$ is of the form

$$m = \sum_{i=1}^{\infty} \epsilon_{x_i},$$

where $x_i \in \mathbb{E}$ are the locations of the point masses of $m$ and $\epsilon_{x_i}$ denotes the point measure defined by

$$\epsilon_{x_i}(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

We emphasize that we assume that all measures in $M_p(\mathbb{E})$ are Radon which means that for any $m \in M_p(\mathbb{E})$ and any compact $K \subset \mathbb{E}$, $m(K) < \infty$. On the space $M_p(\mathbb{E})$ we use the vague metric $\rho(\cdot,\cdot)$. Its properties are discussed for example in Resnick (1987, Section 3.4) and Kallenberg (1983). Note that a sequence of measures $m_n \in M_p(\mathbb{E})$ converge vaguely to $m_0 \in M_p(\mathbb{E})$ if for any continuous function $f : \mathbb{E} \mapsto [0,\infty)$ with compact support we have $m_n(f) \to m_0(f)$ where $m_n(f) = \int_{\mathbb{E}} f \, dm_n$. The non-negative continuous functions with compact support will be denoted by $C_K^+(\mathbb{E})$.

A Poisson process on $\mathbb{E}$ with mean measure $\mu$ will be denoted by PRM($\mu$). The primary example of interest in our applications is the case when $\mathbb{E}_m = [-\infty,\infty]^m \setminus \{\mathbf{0}\}$, where compact sets are closed subsets of $[-\infty,\infty]^m$ which are bounded away from $\mathbf{0}$.

Here is the result from Davis and Resnick (1995) describing the behavior of the simple bilinear process.

**Theorem 2.1.** *Suppose $\{X_t\}$ is the bilinear process (2.3) where the marginal distribution $F$ of the iid noise $\{Z_t\}$ satisfies (2.2), the constant $c$ satisfies (2.4) and $b_n$ is given by (2.6). Suppose further that $\sum_{s=1}^{\infty} \epsilon_{j_s}$ is PRM($\mu$) with $\mu$ given by*

$$\mu(dx) = p\alpha x^{-\alpha-1} dx 1_{(0,\infty)}(x) + q|x|^{-\alpha-1} dx 1_{(-\infty,0)}(x),$$

*and $\{U_{s,k},\ s \geq 1, k \geq 1\}$ are iid with distribution $F$.*

*(i) In $M_p(\mathbb{E}_1)$,*

$$\sum_{t=1}^{n} \epsilon_{b_n^{-2} X_t} \Rightarrow \sum_{k=1}^{\infty} \sum_{s=1}^{\infty} \epsilon_{j_s^2 c^k W_{s,k}},$$

*where*

$$W_{s,k} = \begin{cases} \prod_{i=1}^{k-1} U_{s,i}, & \text{if } k > 1, \\ 1, & \text{if } k = 1, \\ 0, & \text{if } k < 1. \end{cases}$$

*(ii) In $M_p(\mathbb{E}_{h+1})$,*

$$\sum_{t=1}^{n} \epsilon_{b_n^{-2}(X_t, X_{t-1}, \dots, X_{t-h})} \Rightarrow \sum_{k=1}^{\infty} \sum_{s=1}^{\infty} \epsilon_{j_s^2 (c^k W_{s,k}, c^{k-1} W_{s,k-1}, \dots, c^{k-h} W_{s,k-h})}.$$

*Furthermore, if $0 < \alpha < 4$, we have for any $h = 1, 2, \dots$ that*

$$(\hat{\rho}_H(l), l = 1, \dots, h) \Rightarrow (L_i, i = 1, \dots, h)$$

*in $\mathbb{R}^h$, where*

$$L_i = \frac{\sum_{s=1}^{\infty} \sum_{k=1}^{\infty} j_s^4 c^{2k-i} W_{s,k} W_{s,k-i}}{\sum_{s=1}^{\infty} \sum_{k=1}^{\infty} j_s^4 c^{2k} W_{s,k}^2}, \quad i = 1, \dots, h.$$

Contrast the random limits for $(\hat{\rho}_H(l), l = 1, \dots, h)$ when $\{X_t\}$ is the bilinear process (2.3) with the nonrandom limits obtained when $\{X_t\}$ is the linear process (2.1). This difference can lead to dramatic errors if one models heavy tailed nonlinear data with a linear model. This contrast is demonstrated clearly with simulated data. In Section 3 we present simple analyses of several real heavy tailed data sets to illustrate the likelihood that linear models are unsuitable.

We simulated three independent samples ($test_i,\ i = 1, 2, 3$) of size 5000 from the bilinear process

$$X_t = .1 Z_{t-1} X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots, \tag{1.1}$$

where $\{Z_t\}$ are iid Pareto random variables,

$$P[Z_1 > x] = 1/x, \quad x > 1.$$

For contrast, we also simulated three independent samples of size 1500 of AR(2) data. The AR(2) is

$$X_t = 1.3 X_{t-1} - 0.7 X_{t-2} + Z_t, t = 0, \pm 1, \pm 2, \dots$$

and the innovations have a Pareto distribution as for the bilinear example. The AR data sets were called $testar_i,\ i = 1, 2, 3$.

The erratic nature of the behavior of $\hat{\rho}_H$ for the bilinear model is illustrated in Figure 2.1 which graphs the heavy tail acf for $test_i,\ i = 1, 2, 3$. The graphs look rather different reflecting the fact that we are basically sampling independently three times from the non-degenerate limit distribution of the heavy tailed acf. If one were not aware of the non-linearity in the data, one would be tempted to model with a low order moving average based for example on the left hand plot. Furthermore, partial autocorrelation plots and plots of the AIC statistic as a function of the order of the model all show similar erratic behavior as one moves from independent sample to independent sample. So failure to account for non-linearity means there is great potential to be misled in the sorts of models one tries to fit.

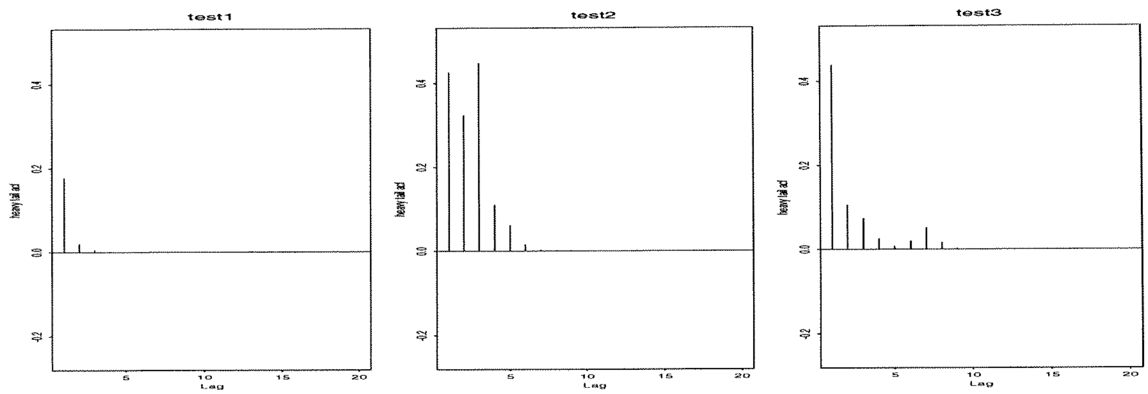


Figure 2.1. Heavy tailed ACF for 3 bilinear samples.

As a contrast, Figure 1.2 presents the comparable heavy tailed acf plots for the three independent AR samples. Here, the pictures look identical reflecting the fact that we are sampling from an essentially degenerate distribution.

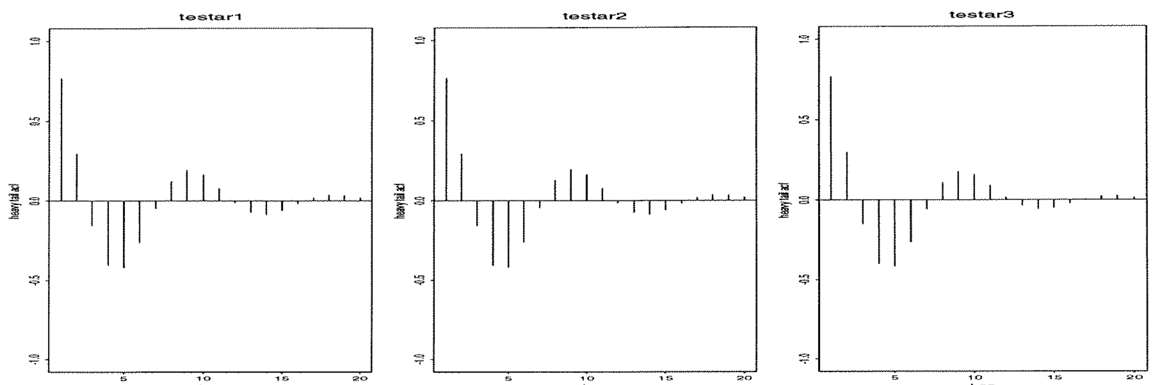


Figure 2.2. Heavy tailed ACF for 3 autoregressive samples.

**3. Analysis of heavy tailed data sets.**

We now present several examples of real data and make the argument that it is unlikely that the data can be modelled as a linear model of the form (2.1). For each data set we note why we believe a heavy tailed model is appropriate and why any sort of infinite order moving average is likely to be an inadequate model.

Given a particular data set, there are various methods of checking that a heavy tailed model is appropriate. Such methods are reviewed in Resnick (1995). Suppose $\{X_n, n \geq 1\}$ is a stationary sequence and that

$$P[X_1 > x] = x^{-\alpha} L(x), \quad x \to \infty \tag{3.1}$$

where $L$ is slowly varying and $\alpha > 0$. Consider the following techniques:

*(1) The Hill plot.* Let

$$X_{(1)} > X_{(2)} > \cdots > X_{(n)}$$

be the order statistics of the sample $X_1, \ldots, X_n$. We pick $k < n$ and define the Hill estimator (Hill, 1975) to be

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

Note $k$ is the number of upper order statistics used in the estimation. The Hill plot is the plot of

$$(k, H_{k,n}^{-1}), 1 \le k < n)$$

and if the process is linear or satisfies mixing conditions then since $H_{k,n} \xrightarrow{P} \alpha^{-1}$ as $n \to \infty$, $k/n \to 0$ the Hill plot should have a stable regime sitting at height roughly $\alpha$. See Mason (1982), Hsing (1991), Resnick and Stărică (1995a), Rootzen et. al (1990), Rootzen (1996). In the iid case, under a second order regular variation condition, $H_{k,n}$ is asymptotically normal with asymptotic variance $1/\alpha^2$. (Cf. de Haan and Resnick, 1996).

*(2) The smooHill plot.* The Hill plot often exhibits extreme volatility which makes finding a stable regime in the plot more guesswork than science and to counteract this, Resnick and Stărică (1995b) developed a smoothing technique yielding the smooHill plot: Pick an integer $u$ (usually 2 or 3) and define

$$smooH_{k,n} = \frac{1}{(u-1)k} \sum_{j=k+1}^{uk} H_{j,n}.$$

In the iid case when a second order regular variation condition holds, the asymptotic variance of $smooH_{k,n}$ is less than than of the Hill estimator, namely:

$$\frac{1}{\alpha^2} \frac{2}{u} (1 - \frac{\log u}{u}).$$

*(3) Alt plotting; Changing the scale.* As an alternative to the Hill plot, it is sometimes useful to display the information provided by the Hill or smooHill estimation as

$$\{(\theta, H^{-1}_{\lceil n^\theta \rceil, n}), 0 \le \theta \le 1, \}$$

and similarly for the smooHill plot where we write $\lceil y \rceil$ for the smallest integer greater or equal to $y \ge 0$. We call such plots the *alternative Hill plot* abbreviated AltHill and the *alternative smoothed Hill plot* abbreviated AltsmooHill. The alternative display is sometimes revealing since the initial order statistics get shown more clearly and cover a bigger portion of the displayed space.

*(4) Dynamic and static qq plots.* As we did for the Hill plots, pick $k$ upper order statistics

$$X_{(1)} > X_{(2)} > \ldots X_{(k)}$$

and neglect the rest. Plot

$$\{(-\log(1 - \frac{j}{k+1}), \log X_{(j)}), 1 \le j \le k\}. \tag{3.2}$$

If the data is approximately Pareto or even if the marginal tail is only regularly varying, this should be approximately a straight line with slope=$1/\alpha$. The slope of the least squares line through the points is an estimator called the qq-estimator (Kratz and Resnick, 1995). Computing the slope we find that the qq–estimator is given by

$$\widehat{\alpha^{-1}}_{k,n} = \frac{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))\log(\frac{X_{(i)}}{X_{(k+1)}}) - \frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))H_{k,n}}{\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2 - (\frac{1}{k}\sum_{i=1}^{k}(-\log(\frac{i}{k+1}))^2} \tag{3.4}$$

There are two different plots one can make based on the qq–estimator. There is the dynamic qq–plot obtained from plotting $\{(k, 1/\widehat{\alpha^{-1}}_{k,n}, 1 \le k \le n\}$ which is similar to the Hill plot. Another plot, the static qq–plot, is obtained by choosing and fixing $k$, plotting the points in (3.2) and putting the least squares line through the points while computing the slope as the estimate of $\alpha^{-1}$ .

The qq-estimator is consistent for the iid model if $k \to \infty$ and $k/n \to 0$ and under a second order regular variation condition and further restriction on $k(n)$, it is asymptotically normal with asymptotic variance $2/\alpha^2$. This is larger than the asymptotic variance of the Hill estimator The volatility of the qq–plot always seems to be less than that of the Hill estimator.

We now consider several data sets and illustrate some features and describe problems encountered when trying to fit MA($\infty$) models.

*(i) ISDN2.* This dataset consists of 4868 interarrival times of ISDN D-channel packets. The time series plot and qq–plot giving evidence of heavy tails are shown in Figure 3.1.

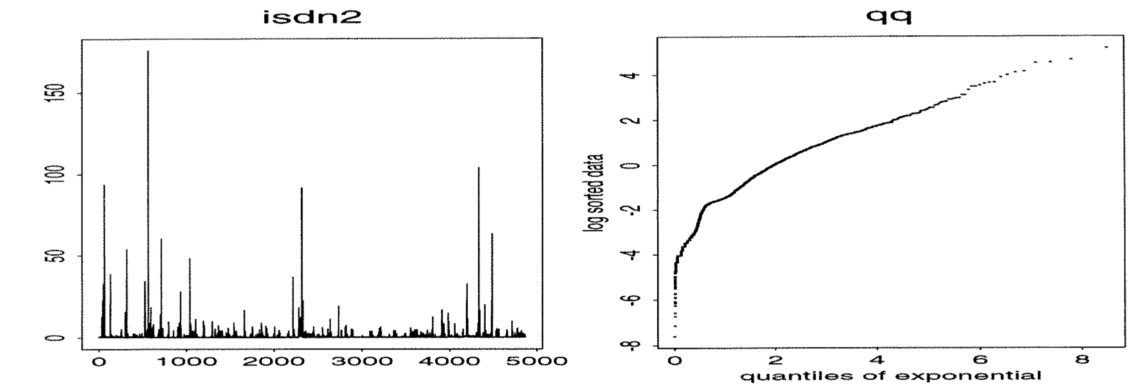


Figure 3.1. Tsplot of ISDN2 and qq-plot.

Hill plots given in Figure 3.2 indicate an $\alpha$ in the neighborhood of 1.2. The static qq-plot based on 1000 upper order statistics given in Figure 3.3 yields an $\alpha$ value of 1.136.

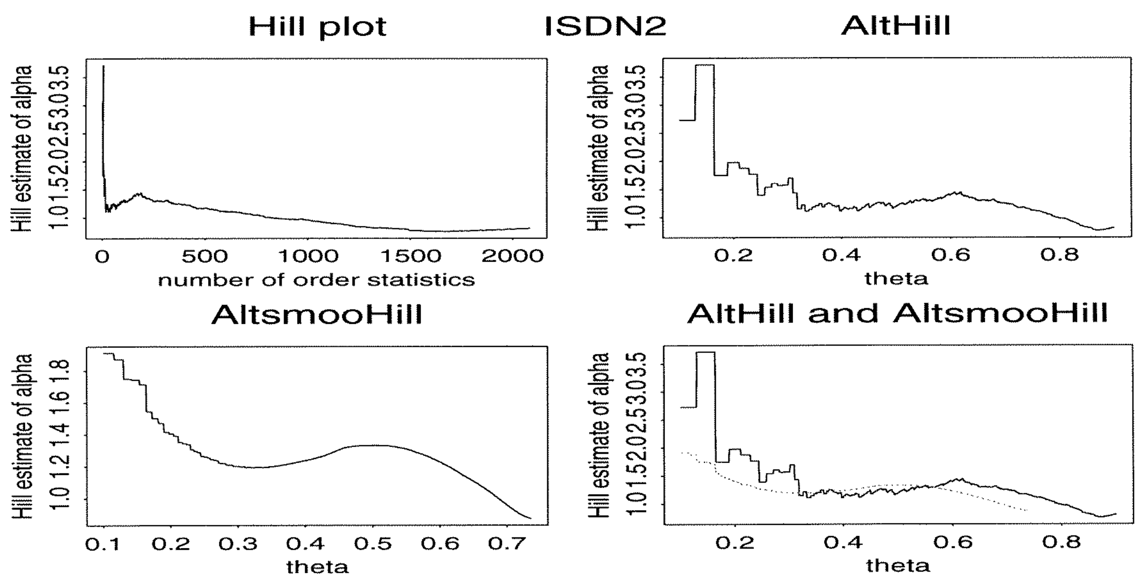


Figure 3.2. Hill plots of ISDN2.

The last two plots in Figure 3.3 give the acf of the first 1500 observations and the acf of the last 1500 observations in the data set. Note the two graphs are quite different which seems to rule out any sort of ARMA model as a potential candidate to be fit to the data.

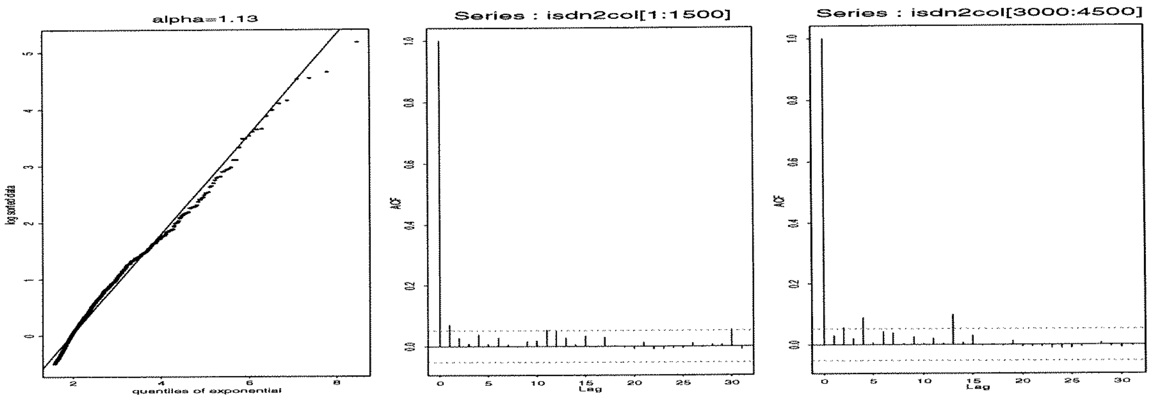


Figure 3.3. qq-plot and acf plots of ISDN2.

*(ii) Interar.* This data set represents 176834 interarrivals of externally generated TCP packets to a server. The recording period was one hour. Figure 3.4 give the time series plot and the qq-plot. Both show clear evidence of heavy tails.

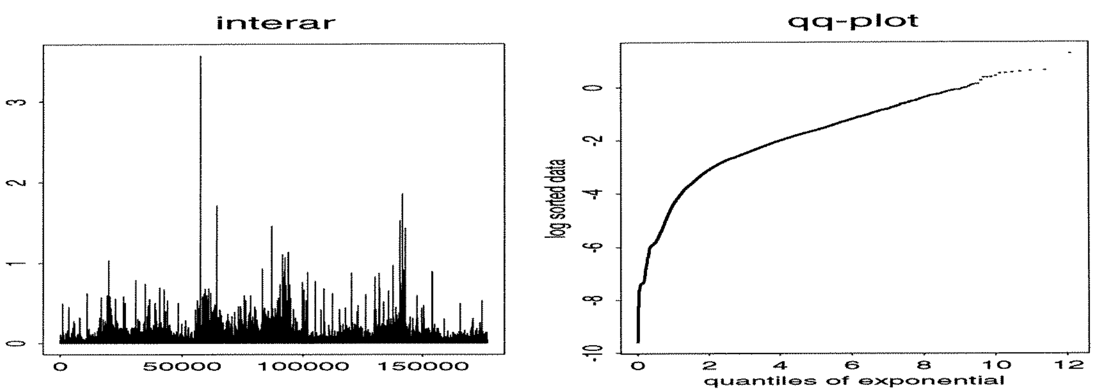


Figure 3.4. Tsplot and qq-plot for interar.

The Hill plots given in Figure 3.5 show a value of $\alpha$ in the neighborhood of 2.5 and the static qq-plot in Figure 3.6 gives a value of $\alpha = 2.28$.

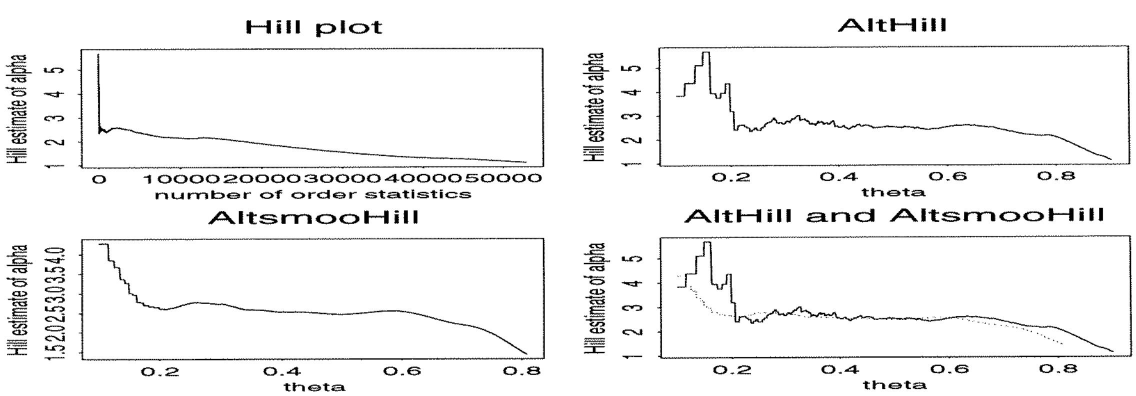

Figure 3.5. Hill plots for interar.

The acf of the first 10,000 values does not remotely resemble the plot for 10,000 values taken in the middle of the time series. These are the last two plots in figure 3.6.

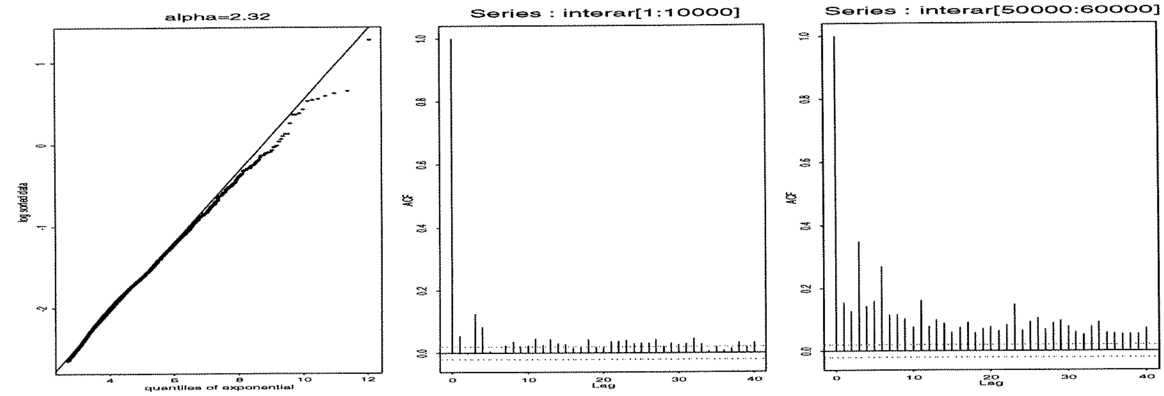


Figure 3.6. Static qq-plot and acf plots for interar.

*(iii) SILENCE.* Consider a time series of length 1027 shown in Figure 3.7 which represents the off periods between transmission of packets generated by a terminal during a logged-on session. The left graph in Figure 3.7 is the time series plot and the right graph is the static qq-plot using 500 upper order statistics giving ample evidence of heavy tails. The estimate of $\alpha$ given by this plot is 0.6696.

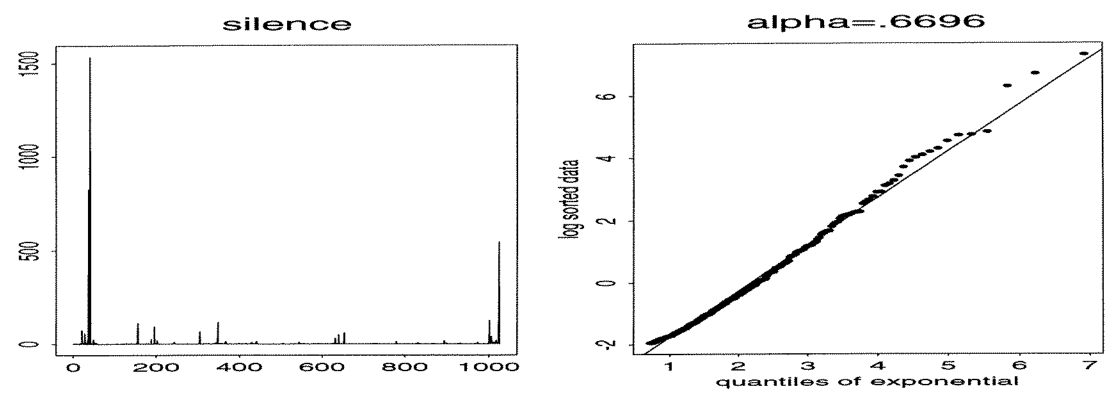


Figure 3.7. Tsplot and static qq-plot of SILENCE.

The Hill plots confirm the estimate of $\alpha$ given by the static qq-plot and give $\alpha \approx .64$.

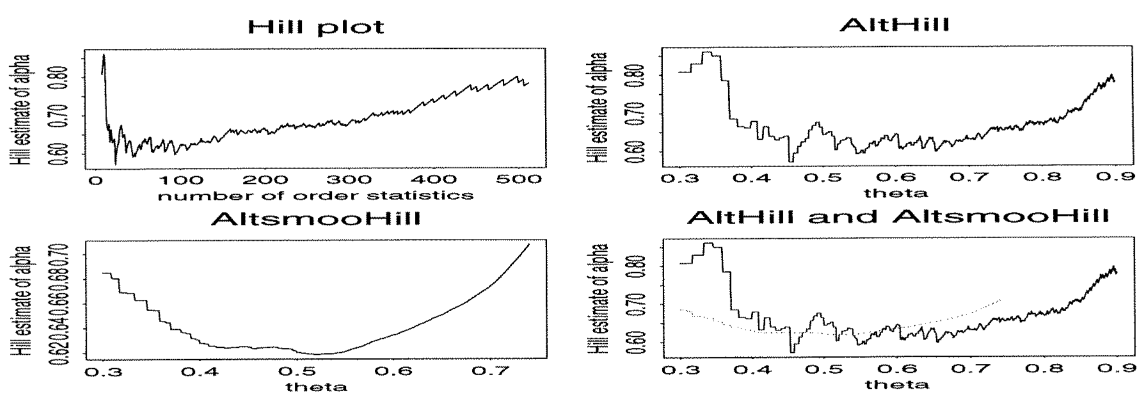

Figure 3.8. Hill plots of SILENCE.

For the last graph, we split the data set into thirds and graphed the acf for each piece separately. The pictures are obviously quite different.

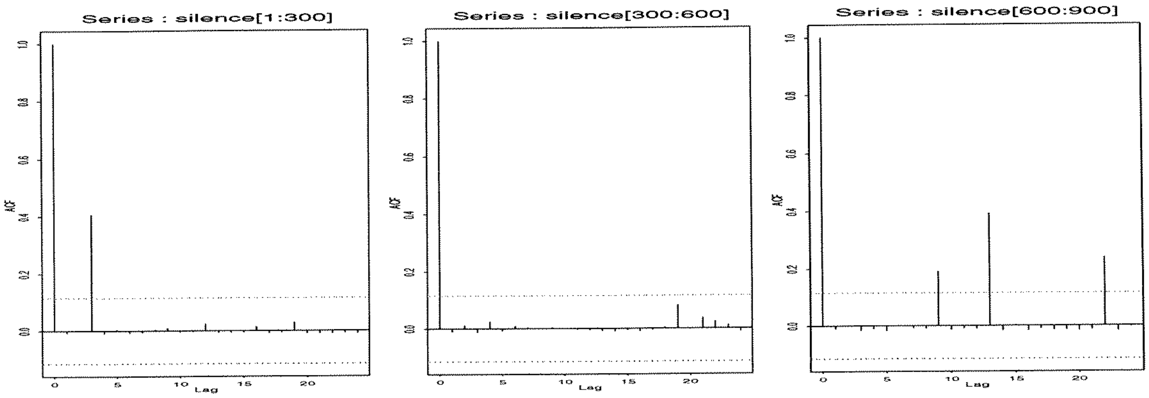


Figure 3.9. Acf plots of SILENCE.

**4. Estimation for the simple bilinear process.** We now consider the simple bilinear process given in (2.3)

$$X_t = cX_{t-1}Z_{t-1} + Z_t \tag{4.1}$$

where in this section we assume $Z_t \geq 0$ and the distribution of $Z_t$ has left endpoint

$$l = \inf\{x > 0 : P[Z_t \leq x] > 0. \tag{4.2}$$

Further, we assume $c > 0$ and (2.4) assumes the form

$$c^{\alpha/2} E Z_1^{\alpha/2} < 1. \tag{4.3}$$

In particular, this implies $c^{\alpha/2} l^{\alpha/2} < 1$ or $cl < 1$.

The next proposition gives consistent estimators for $(c, l)$.

**Proposition 4.1.** *Suppose $\{X_t\}$ is the bilinear process given in (4.1) and that $(c, l)$ satisfy (4.2) and (4.3). Suppose in addition that $l > 0$ and $0 < \alpha < 4$. Define*

$$\hat{m} = \bigwedge_{t=1}^{t} X_t \text{ and } \hat{r} = \bigwedge_{t=2}^{n} \frac{X_t}{X_{t-1}}.$$

*Then consistent estimators of $(c, l)$ are given by*

$$\hat{l} = \hat{m}(1 - \hat{r}), \quad \hat{c} = \frac{\hat{r}}{\hat{m}(1 - \hat{r})}.$$

Note that $\hat{r}$ is the linear programming estimator studied by Davis and McCormick (1989) and Feigin and Resnick (1992, 1994) for linear autoregressions.

*Proof.* Since $Z_t \geq l$ with probability 1, we have from (2.5) that a.s. for each $t$

$$\begin{aligned} X_t \geq & l + \sum_{j=1}^{\infty} c^j l^{j-1} l^2 \\ = & l + \frac{cl^2}{1 - cl} \\ = & \frac{l}{1 - cl} := l^{\#}. \end{aligned}$$

Now we apply the Davis and Resnick (1995) result quoted as Theorem 2.1 in this paper. We have in $M_p[l^{\#},\infty]^2$ (and therefore by Theorem 3.3 of Feigin, Kratz and Resnick (1994) in $M_p[l^{\#},\infty)^2$) that

$$\sum_{t=2}^{n} \epsilon_{b_n^{-2}(X_t,X_{t-1})} \Rightarrow \sum_{k=1}^{\infty}\sum_{s=1}^{\infty} \epsilon_{j_s^2(c^k W_{s,k},c^{k-1}W_{s,k-1})}. \tag{4.4}$$

We seek to apply a map

$$(x,y)\mapsto x/y$$

but in order to do this we must compactify the state space of the converging point processes. From (4.4) we get for any large $K>0$

$$\sum_{t=2}^{n} 1_{[l^{\#}\le b_n^{-2}X_{t-i}\le K,\ i=0,1]}\epsilon_{b_n^{-2}(X_t,X_{t-1})} \Rightarrow \sum_{k=1}^{\infty}\sum_{s=1}^{\infty} 1_{[l^{\#}\le j_s^2(c^{k-i}W_{s,k-i}\le K,\ i=0,1]}\epsilon_{j_s^2(c^k W_{s,k},c^{k-1}W_{s,k-1})}. \tag{4.5}$$

and now applying the division map we get in $M_p[l^{\#},\infty)$

$$\sum_{t=2}^{n} 1_{[l^{\#}\le b_n^{-2}X_{t-i}\le K,\ i=0,1]}\epsilon_{X_t/X_{t-1}} \Rightarrow \sum_{k=1}^{\infty}\sum_{s=1}^{\infty} 1_{l^{\#}\le j_s^2(c^{k-i}W_{s,k-i}\le K,\ i=0,1]}\epsilon_{cU_{s,k-1}}. \tag{4.6}$$

We now argue we can remove the truncation level $K$. In order to do this we must show by Billingsley, 1968, Theorem 4.2 that for any $\eta>0$ and any continuous function $f$ with compact support in $[0,\infty)$

$$\lim_{K\to\infty}\limsup_{n\to\infty} P[|\sum_{t=2}^{n} 1_{[l^{\#}\le b_n^{-2}X_{t-i}\le K,\ i=0,1]} f(\frac{X_t}{X_{t-1}}) - \sum_{t=2}^{n} f(\frac{X_t}{X_{t-1}})| > \eta] = 0. \tag{4.7}$$

Suppose the compact support of $f$ is contained in $[0,\theta]$. The probability in (4.7) is bounded by

$$\begin{aligned} & nP[\frac{X_t}{b_n^2}>K, \frac{X_t}{X_{t-1}}\le\theta] + nP[\frac{X_{t-1}}{b_n^2}>K, \frac{X_t}{X_{t-1}}\le\theta] \\ \le & 2nP[\frac{X_t}{b_n^2}>K] \\ \to & (const)K^{-\alpha}, \end{aligned}$$

as $n\to\infty$ by Corollary 2.4 of Davis and Resnick (1995) and as $K\to\infty$ the above

$$\to 0.$$

This shows that (4.6) holds with the truncation level $K$ replaced by $\infty$.

We thus conclude that in $M_p[0,\infty)$

$$\sum_{t=2}^{n}\epsilon_{X_t/X_{t-1}} \Rightarrow \sum_{k=1}^{\infty}\sum_{s=1}^{\infty}\epsilon_{cU_{s,k-1}}$$

and applying the a.s. continuous function that maps the point measure into the minimum of the points yields

$$\hat r = \bigwedge_{t=2}^{n}\frac{X_t}{X_{t-1}} \Rightarrow c\bigwedge_{s,k} U_{s,k} = cl.$$

Since it is also clear that

$$\hat m \Rightarrow l^{\#},$$

the desired consistency result follows. □

In a simulation experiment, we simulated a time series of length 5000 from the simple bilinear process given in (4.1) with $c=.3$ and $l=.1$. Our estimators yielded values of

$$(\hat c,\hat l) = (.317023, .1011468).$$

## 5. Concluding remarks.

The estimator $\hat{c}$ proposed in Proposition 4.1 for the simple bilinear process given in (4.1) is consistent but it is not at all clear that there is an asymptotic distribution or even what the rate of convergence might be. It seems likely that if an asymptotic distribution exists, it will depend on the unknown parameters $c$ and $\alpha$. Other estimators need to be explored.

However the obvious priority must be to find a flexible parametric family which is large enough to fit the abundance of heavy tailed data that exists but is tractable enough to yield excellent model selection and estimation techniques. The point of emphasis of this paper is that any parametric family of stationary processes which can be expressed as infinite order moving averages is not likely to satisfy the requirements of adequately fitting existing heavy tailed data. The general bilinear model is one possible family of processes that merits exploration.

The Hill estimator has been proven to be consistent for observations coming from a process which is iid (Mason, 1982) or MA($\infty$) (Resnick and Stărică, 1995a) or which satisfy mixing conditions (Rootzen, et al, 1990; Rootzen, 1995). The Hill estimator appears to work just fine for the simple bilinear process in (4.1). Figure 5.1 displays the Hill plots for 5000 observations coming from (4.1) with $c = .1$ and $\alpha = 1$. Since the tail of $X_t$ satisfies

$$P[X_t > x] \sim (const)P[Z_1^2 > x]$$

the correct answer that the Hill plots seek is 0.5.

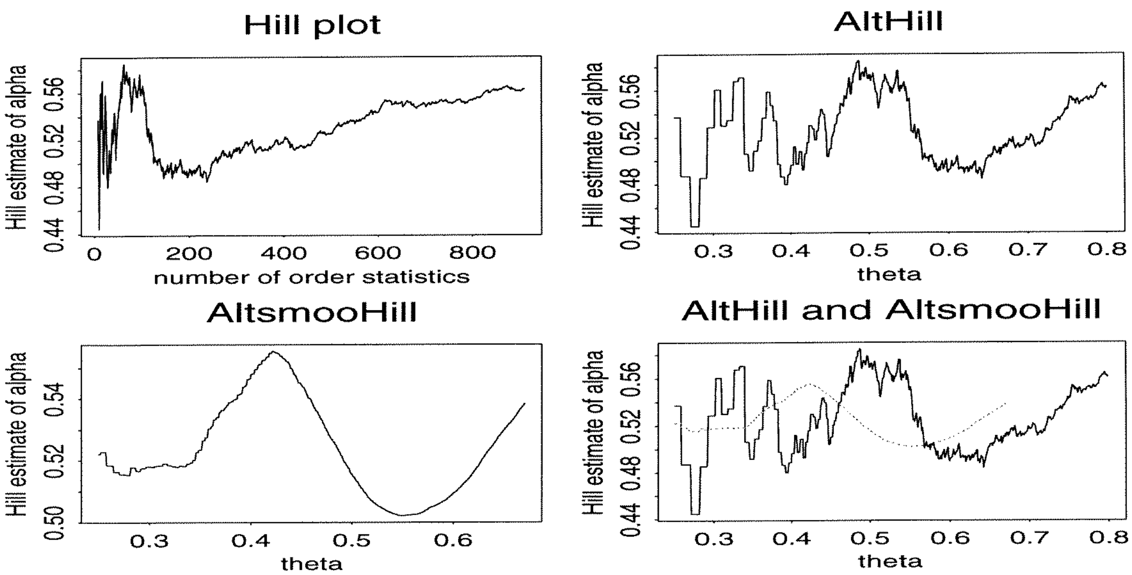


Figure 5.1. Hill plots for the simple bilinear process.

We intend to give some thought to showing directly that the Hill estimator can be applied successfully to estimating the shape parameter when the underlying model is nonlinear.

## References


Billingsley, P., *Convergence of Probability Measures*, Wiley, New York, 1968.

Brockwell, P. and Davis, R., *Time Series: Theory and Methods, 2nd edition*, Springer–Verlag, New York, 1991.

Crovella, M and Bestavros, A., *Explaining world wide web traffic self-similarity*, Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu (1995).

Cunha, Bestavros, A. and Crovella, M, *Characteristics of www client-based traces*, Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu.

Davis, R. and McCormick, W., *Estimation for first-order autoregressive processes with positive or bounded innovations*, Stochastic Processes and their Applic. **31** (1989), 237–250.

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

Davis, R. and Resnick, S., *More limit theory for the sample correlation function of moving averages*, Stochastic Processes and their Applications **20** (1985b), 257–279.

Davis, R. and Resnick, S., *Limit theory for the sample covariance and correlation functions of moving averages*, Ann. Statist. **14** (1986), 533-558.

Davis, R. and Resnick, S., *Limit theory for bilinear processes with heavy tailed noise*, Available as TR1140 at ftp.orie.cornell.edu after changing directory to /ftp/pub/techreps. Also at http:www.orie.cornell.edu/trlist/trlist.html as TR1140.ps.Z (1995).

Duffy, D., McIntosh, A., Rosenstein, M., and Willinger, W., *Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks*, IEEE Journal on Selected Areas in Communications **12** (1994), 544–551.

Duffy, D., McIntosh, A., Rosenstein, M., and Willinger, W., *Analyzing telecommunications traffic data from working common channel signaling subnetworks*, Proceedings of the 25th Interface, San Diego Ca (1993).

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

Feigin, P. and Resnick, S., *Limit distributions for linear programming time series estimators*, Stochastic Processes and their Applications **51** (1994), 135–166.

Feigin, P. and Resnick, S., *Linear programming estimators and bootstrapping for heavy tailed phenomena*, Forthcoming: Advances in Applied Probability (1996).

Feigin, P., Kratz, M. and Resnick, S., *Parameter estimation for moving averages with positive innovations*, Technical Report (1994).

Feigin, P. Resnick, S. and Stărică, Cătălin, *Testing for independence in heavy tailed and positive innovation time series*, Stochastic Models **11** (1995), 587–612.

Gabr, M. and Subba Rao, T., *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture notes in statistics v. 24, Springer-Verlag, New York :, 1984.

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

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

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

Kallenberg, O., *Random Measures*, Third edition, Akademie-Verlag, Berlin, 1983.

Kratz, M. and Resnick, S., *The qq–estimator and heavy tails*, Available as TR 1122.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1995).

Liu, J., *On the existence of a general multiple bilinear time series*, J. Time Series Analysis **10** (1989), 341–355.

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

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

Resnick, Sidney, *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York, 1987.

Resnick, S., *Heavy tail modelling and teletraffic data*, Available by ftp from ftp.orie.cornell.edu as TR1134.ps.Z in directory /ftp/pub/techreps or at http://www.orie.cornell.edu/trlist/trlist.html, Preprint (1995).

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

Resnick, S. and Stărică, Cătălin, *Smoothing the Hill estimator*, Technical report (1995b).

Rootzen, H., Leadbetter, M. and de Haan, L., *Tail and quantile estimation for strongly mixing stationary sequences*, Preprint, Econometric Institute, EUR, PO Box 1738, 3000 DR, Rotterdam, Holland.

Rootzen, H., *The tail empirical process for stationary sequences*, Preprint 1995:9 ISSN 1100-2255, Studies in Statistical Quality Control and Reliability, Chalmers University of Technology (1995).

Rosiński, Jan, *On the structure of stationary stable processes*, Ann. Probab. **23** (1995), 1163–1187.

Willinger, W., Taqqu, M., Sherman, R. and Wilson, D., *Self–similarity through high–variability: Statistical analysis of ethernet LAN traffic at the source level*, Preprint (1995).

SIDNEY I. RESNICK, CORNELL UNIVERSITY, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, ETC BUILDING, ITHACA, NY 14853 USA

*E-mail*: sid@orie.cornell.edu