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

TECHNICAL REPORT NO. 1085

February 1994

**TESTING FOR INDEPENDENCE IN
HEAVY TAILED AND POSITIVE
INNOVATION TIME SERIES**

by

Paul D. Feigin, Sidney I. Resnick
and Cătălin Stărică

Paul Feigin and Sidney Resnick were partially supported by the United States-Israel Binational Science Foundation (BSF) Grant No. 89-00323 and 92-000324. Sidney Resnick and Cătălin Stărică were also partially supported by NSF Grant DMS-9100027 at Cornell University. Sidney Resnick also received some support from NSA Grant 92G-116.

TESTING FOR INDEPENDENCE IN HEAVY TAILED AND POSITIVE INNOVATION TIME SERIES

PAUL D. FEIGIN, SIDNEY I. RESNICK AND CĂTĂLIN STĂRICĂ

Technion, Israel Institute of Technology and Cornell University

ABSTRACT. For time series with positive innovations, a test is given to distinguish between data coming from a stationary process where the variables are dependent versus the model being independent identically distributed random variables. The techniques are suitable either when the innovation distribution has heavy right tails or regularly varying left tails and is based on linear programming estimators given by Feigin and Resnick (1992, 1994). Examples using teletraffic data and the lynx data are discussed. The method has applications to model confirmation where the fit of a model is also examined by gauging whether the residuals are independent or not.

1. Introduction.

Consider time series data $x_1, \dots, x_n$ which is generated by a stationary autoregressive model with non-negative innovations. How can we decide if the data came from a model $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ where the random variables are dependent or if it came from random variables which are in fact independent, identically distributed (iid)? This problem of testing for independence is not only interesting in its own right but has applications to model selection and confirmation and goodness of fit criteria for time series models. This is due to the fact that a well fitted model should have estimated residuals which appear to be a realization of an iid sequence.

One type of data that we have in mind comes from teletraffic applications. It is typically non-negative and often heavy tailed and could represent holding times or interarrival times. Assuming the stationarity assumption is met, the heavy tailed assumption implies that for some $\alpha > 0$, $c > 0$ and some slowly varying function $L(x)$, for every t

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

or equivalently

$$\lim_{s \rightarrow \infty} \frac{P[X_t > sx]}{P[X_t > s]} = x^{-\alpha}.$$

In general, by a heavy tailed time series model, we mean a model that can be expressed in the form

$$(1.1) \quad X_t = \sum_{k=0}^{\infty} c_k Z_{t-k}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where for each t , $Z_t \geq 0$ and $\{c_k\}$ is a real sequence making (1.1) converge and $\{Z_t\}$ is iid satisfying

$$(1.2) \quad P[Z_t > x] \sim x^{-\alpha}L(x), \quad x \rightarrow \infty.$$

Key words and phrases. Poisson processes, linear programming, autoregressive processes, parameter estimation, weak convergence, consistency, time series analysis, estimation, independence.

Paul Feigin and Sidney Resnick were partially supported by the United States-Israel Binational Science Foundation (BSF) Grant No. 89-00323 and 92-000324. Sidney Resnick and Cătălin Stărică were also partially supported by NSF Grant DMS-9100027 at Cornell University. Sidney Resnick also received some support from NSA Grant 92G-116.

In this paper we will firstly concentrate on p th order autoregressive models (denoted $AR(p)$) of the form

$$(1.3) \quad X_t = \sum_{i=1}^p \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

where $\{Z_t\}$ is a positive iid sequence satisfying (1.2).

Heavy tailed models for teletraffic applications can arise in practice as when long times between computer transmissions initiated by a human operator are mingled with very short times between transmissions which are computer controlled (cf. Meier-Hellstern, et al, 1991). An application of this type is discussed in Section 4 where we analyze 783 interarrival times of packets to a host machine.

Our techniques apply in another circumstance as well, namely when the autoregressive model has positive innovations and the innovation distribution is regularly varying at zero. This means that the variables $\{Z_t\}$ in (1.3) satisfy

$$(1.4) \quad P[Z_1 \leq x] \sim x^\alpha L(x), \quad x \downarrow 0, \alpha > 0.$$

This is the case for example if the innovation distribution is exponential, Weibull or Gamma. The case of positive innovations has been considered by Li and McCleod (1988) who modelled, among other things, the lynx data using a Gamma distributed innovations model.

Precise statements of conditions and needed asymptotic theory are given in Section 2. The heavy tailed case is discussed under Condition R (R for **R**ight tail) and Condition L covers the left tail case of (1.4).

Our test for independence is based on estimators for the coefficients in (1.3) using a linear programming (lp) technique developed in Feigin and Resnick (1992, 1994). These estimators are defined to be

$$(1.5) \quad \hat{\phi}^{(n)} = \arg \max_{D_n} \sum_{i=1}^p \delta_i$$

where the feasible region D_n is

$$(1.6) \quad D_n := \{\delta \in \mathbb{R}^p : X_t - \sum_{i=1}^p \delta_i X_{t-i} \geq 0; \quad t = p+1, \dots, n\}.$$

Assuming reasonable conditions (Conditions L or R) hold, if the true model is $AR(p)$ given by (1.3) with true parameter vector ϕ_0 , then for suitably chosen q_n , there is a limit distribution

$$q_n(\hat{\phi}^{(n)} - \phi_0) \Rightarrow \mathbf{L}$$

for some limiting random variable $\mathbf{L}$. The limit distribution corresponding to $\mathbf{L}$ is quite complicated in general but if $\phi_0 = \mathbf{0}$, then under Condition R the limit distribution can be written explicitly and under Condition L can be approximated. The limit distribution depends on the (unknown) α but α can be estimated by the Hill estimator (Hill, 1975; Mason, 1982) or by graphical techniques so that percentiles of desired functionals of $\mathbf{L}$ can be readily estimated.

Techniques for identifying situations where heavy tailed models may be appropriate are given in Resnick and Stărică (1994) and center on the Hill estimator which may be applied to the original time series data directly. For heavy tailed time series models satisfying (1.1) and Condition R, the Hill estimator is a consistent estimate of α . For autoregressions satisfying Condition R one can also estimate the coefficients with any consistent estimator $\hat{\phi}^{(n)}$ and then apply the Hill estimator to the estimated residuals

$$\hat{Z}_t = X_t - \sum_{i=1}^p \hat{\phi}_i X_{t-i}, \quad t = p+1, \dots, n.$$

This also yields a consistent procedure for estimating α (Resnick and Stărică, 1994). However, in the left tail case, the marginal distribution of X_t is not regularly varying in the left tail (Davis and Resnick, 1991) and there seems no choice but to estimate α by first estimating the residuals and then applying Hill's estimator to the estimated residuals. This approach is justified in Section 6.

An important goal of this paper is to exhibit several techniques for analyzing heavy tailed time series models. These techniques have been discussed in the theoretical literature but not often applied to data. Thus, Section 4 treats a data set of interarrival times in some detail and displays a variety of techniques of analysis. Sections 2 and 3 discuss the mathematical background and the test of independence. Section 6 discusses estimation of the index of regular variation under Condition L for autoregressive data and Section 5 gives a data example of the use of the independence test in the left tail case. Section 7 shows how our techniques can be extended in order to test for when a particular component in the parameter vector $\phi = (\phi_1, \dots, \phi_p)$ can be set equal to zero. Explicit results are possible in the AR(2) case. Some concluding remarks are given in Section 8.

2. Asymptotic theory for the null case.

We now present the main theorem upon which the asymptotic theory of the tests for independence are based. This theorem is an improved version of Theorem 2.1 of Feigin and Resnick (1994) - the improvement is for the right tail case (Condition R), and the proof below will only relate to this improvement. We first state the main conditions of the theorem.

Condition M (model specification): The process $\{X_t : t = 0, \pm 1, \pm 2, \dots\}$ satisfies the equations

$$(2.1) \quad X_t = \sum_{i=1}^p \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

where $\{Z_t\}$ is an independent and identically distributed sequence of random variables with essential infimum (left endpoint) equal to 0 and common distribution function F .

Condition S (stationarity): The coefficients $\phi_1, \dots, \phi_p$ satisfy the stationarity condition that the autoregressive polynomial $\Phi(z) \equiv 1 - \sum_{i=1}^p \phi_i z^i$ has no roots in the unit disk $\{z : |z| \leq 1\}$. Furthermore we assume $\Phi(1) > 0$; i.e. we require

$$(2.2) \quad \sum_{i=1}^p \phi_i < 1.$$

Condition R (right tail): The distribution F of the innovations Z_t satisfies, for some $\alpha > 0$:

$$(2.3) \quad 1. \lim_{s \rightarrow \infty} \frac{1 - F(sx)}{1 - F(s)} = x^{-\alpha} \text{ for all } x > 0;$$

$$(2.4) \quad 2. E(Z_t^{-\beta}) = \int_0^\infty u^{-\beta} F(du) < \infty \text{ for some } \beta > \alpha.$$

Condition L (left tail): The distribution F of the innovations Z_t satisfies, for some $\alpha > 0$:

$$(2.5) \quad 1. \lim_{s \downarrow 0} \frac{F(sx)}{F(s)} = x^\alpha \text{ for all } x > 0;$$

$$(2.6) \quad 2. E(Z_t^\beta) = \int_0^\infty u^\beta F(du) < \infty \text{ for some } \beta > \alpha.$$

Our results have as hypotheses M, S, and either L or R. For additional remarks concerning these conditions see Feigin and Resnick (1992, 1994).

Theorem 2.1. Suppose conditions *M* and *S* hold and let $\hat{\phi}^{(n)}$ be the linear programming estimator based on $X_1, \dots, X_n$ given in (1.5) and (1.6). Let $\{E_j, j \geq 1\}$ be iid unit exponential random variables and define

$$\Gamma_k = E_1 + \dots + E_k, \quad k \geq 1,$$

so that $\{\Gamma_k\}$ are the points of a homogeneous Poisson process.

(a) If Condition *L* holds, define

$$a_n = F^{\leftarrow}\left(\frac{1}{n}\right).$$

Then

$$a_n^{-1}(\hat{\phi}^{(n)} - \phi^{(0)}) \Rightarrow \mathbf{L}$$

where $\mathbf{L}$ is non-degenerate and

$$(2.7) \quad \mathbf{L} \stackrel{d}{=} \arg \max_{\delta \in \Lambda} \delta' \mathbf{1}$$

where

$$(2.8) \quad \Lambda = \{\delta \in R^p : \delta' \mathbf{1} \geq -1, \delta' \mathbf{u}_k \leq 1, k \geq 1\}.$$

The points $\{\mathbf{u}_k\}$ are specified as follows. Let $\{\mathbf{Y}_j, j \geq 1\}$ be iid R^p valued random vectors with

$$\mathbf{Y}_1 \stackrel{d}{=} (X_p, \dots, X_1)$$

and suppose $\{\mathbf{Y}_j, j \geq 1\}$ are independent of $\{\Gamma_k\}$. Then

$$\mathbf{u}_k = \Gamma_k^{-1/\alpha} \mathbf{Y}_k.$$

(b) If Condition *R* holds, define b_n by

$$(2.9) \quad b_n = \left(\frac{1}{1-F}\right)^{\leftarrow}(n) = F^{\leftarrow}\left(1 - \frac{1}{n}\right).$$

Define for $|z| \leq 1$

$$(2.10) \quad C(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{\Phi(z)}.$$

Then

$$b_n(\hat{\phi}^{(n)} - \phi^{(0)}) = O_p(1)$$

so the rate of convergence of $\hat{\phi}^{(n)}$ to $\phi^{(0)}$ is b_n . Furthermore, if for any $p-1$ distinct indices $\{l_1, \dots, l_{p-1}\}$, for which the set of p vectors

$$(2.11) \quad \{\mathbf{1}, (c_{l_j}, c_{l_j-1}, \dots, c_{l_j-p+1}); 1 \leq j \leq p-1\}$$

does not contain the zero vector, the set is also linearly independent, then

$$b_n(\hat{\phi}^{(n)} - \phi^{(0)}) \Rightarrow \mathbf{L}$$

where $\mathbf{L}$ is non-degenerate,

$$(2.12) \quad \mathbf{L} \stackrel{d}{=} \arg \max_{\delta \in \Lambda} \delta' \mathbf{1},$$

and

$$(2.13) \quad \Lambda = \{\delta \in R^p : \delta' \mathbf{1} \geq -1, \delta' \mathbf{v}_k \leq 1, k \geq 1\}.$$

The points $\{\mathbf{v}_k\}$ are specified as follows. Let $\{Y_{kl}, k \geq 1, l \geq 0\}$ be a doubly infinite array of iid random variables which is independent of $\{\Gamma_k\}$ with the distribution F . Then

$$\mathbf{v}_l = \left(\bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} Y_{kl}^{-1} \right) (c_{l-1}, \dots, c_{l-p})' = V_l(c_{l-1}, \dots, c_{l-p}); \quad l = 1, 2, \dots$$

Proof. The proof of Theorem 2.1(a) is given in Feigin and Resnick (1994). The improvement in the statement of 2.1(b) is that in (2.11) only vectors of p successive c_j 's ($c_{l_j}, \dots, c_{l_j-p+1}$) that are not zero vectors need be investigated. The reason is that Lemma 3.6 of that paper tells us that the solution of the limiting linear program (2.83) is unique if in (2.11) we consider all points that are potentially extreme points of Λ . However $\mathbf{0} \in \text{int}(\Lambda)$ and so cannot be an extreme point. $\square$

The following corollaries specialize the theorem to the null case where the coefficients are all zero — $\phi_i = 0; i = 1, \dots, p$.

Corollary 2.2. Under conditions M, S and L and $\phi^{(0)} = \mathbf{0}$

$$a_n^{-1} \hat{\phi}^{(n)} \Rightarrow \arg \max \{\delta' \mathbf{1} : \delta' \mathbf{u}_k \leq 1; k \geq 1\} \equiv \mathbf{L}$$

where

$$\mathbf{u}_k = \Gamma_k^{-1/\alpha} (Z_{k1}, \dots, Z_{kp})$$

with Γ_k as in the theorem and $\{Z_{kj}; k \geq 1, j = 1, \dots, p\}$ iid with distribution F and independent of $\{\Gamma_k\}$.

Proof. Immediate application of the the result of (2.7) and (2.8) of Theorem 2.1(a) using the fact that when $\phi^{(0)} = \mathbf{0}$ then $(X_p, \dots, X_1) \stackrel{d}{=} (Z_p, \dots, Z_1)$. $\square$

Corollary 2.3. Under conditions M, S and R and $\phi^{(0)} = \mathbf{0}$

$$b_n \hat{\phi}^{(n)} \Rightarrow \mathbf{L} \equiv (V_1^{-1}, \dots, V_p^{-1})$$

where for $x_i \geq 0; i = 1, \dots, p$ we have that

$$(2.14) \quad P[V_i \leq x_i, i = 1, \dots, p] = \exp \left\{ - \int_{(y_1, \dots, y_p) \in [0, \infty)^p} \left(\bigwedge_{l=1}^p y_l x_l \right)^{-\alpha} F(dy_1) \dots F(dy_p) \right\}.$$

Proof. Applying $\phi^{(0)} = \mathbf{0}$ to the results of Theorem 2.1(b) we first note that $c_0 = 1$ and $c_j = 0; j \neq 0$ in (2.10). The only non-zero vectors $\mathbf{v}_l$ are those for $l = 1, \dots, p$, and then

$$\mathbf{v}_l = V_l \mathbf{e}_l$$

with $\mathbf{e}_l$ being the unit vector with 1 in the l -th place. Hence

$$\begin{aligned} \delta &= \arg \max \{\delta' \mathbf{1} : \delta' \mathbf{v}_l = \delta_l V_l \leq 1; l = 1, \dots, p\} \\ &= (V_1^{-1}, \dots, V_p^{-1}) = \mathbf{L}. \quad \square \end{aligned}$$

Note that in Corollaries 2.2 and 2.3 we have denoted the limit random vector by $\mathbf{L}$ but it will have a different distribution in each case. We do this in order to simplify the derivations in the next section.

3. Tests for independence.

If the null hypothesis is that the true $\text{AR}(p)$ parameter vector $\phi^{(0)}$ satisfies $\phi^{(0)} = \mathbf{0}$, then we would reject this hypothesis if the lp estimator $\hat{\phi}^{(n)}$ was outside a neighborhood of $\mathbf{0}$. The rejection region that we adopt is based on the largest component of the vector of lp estimators — $M^{(n)} = \bigvee_{i=1}^p \hat{\phi}_i^{(n)}$. By the continuous mapping theorem, under the null hypothesis we have that

$$q_n M^{(n)} \Rightarrow M^{(\infty)} \equiv \bigvee_{i=1}^p L_i$$

where $q_n = a_n^{-1}$ or $= b_n$ depending on which case we are in (Condition L or Condition R). So in order to determine a critical value for our statistic $M^{(n)}$, or in order to compute a p-value based on it, we need to approximate $P[M^{(n)} > x]$ based on the asymptotic theory. This approximation is done as follows:

$$(3.1) \quad P[M^{(n)} > x] \approx P[M^{(\infty)} > q_n x]$$

and therefore we seek expressions or an upper bound for the right hand side. In each case that we wish to apply the asymptotic theory we need to estimate both the value of α and that of q_n . The estimators we use are based on the Hill estimator for α and are discussed below. The distributions of $M^{(\infty)}$ to be used are given below for 3 cases.

Left Tail Case (Condition L and p=1) Here we directly use the limit result and conclude that

$$P[M^{(\infty)} > y] = \exp\{-y^\alpha c_\alpha\}$$

where $c_\alpha = E(X_1^\alpha)$. In order to apply the result the α moment of the X_k series is also required to be estimated — this can be done in the obvious way from the original time series.

Left Tail Case (Condition L and p=2) Here we note that $\{M^{(\infty)} > y\} \subset \{L_1 > y\} \cup \{L_2 > y\}$ and that from the definition of **L** in Corollary 2.2 $L_1 > y$ implies that $(y, -y)$ must be in the feasible region $\Lambda = \{\delta : \delta' \mathbf{u}_k \leq 1; k \geq 1\}$. The probability that a point (δ_1, δ_2) is in Λ is given by

$$P[\Gamma_k^{1/\alpha} > \delta_1 Y_{k2} + \delta_2 Y_{k1} \equiv W_k(\delta); k \geq 1] = \exp\{-E[(W_k(\delta))^\alpha I(W_k(\delta) > 0)]\}$$

and so the probability that $(y, -y)$ or $(-y, y)$ is feasible is given by:

$$(3.2) \quad \exp\{-y^\alpha E[(X_2 - X_1)^\alpha I(X_2 > X_1)]\} + \exp\{-y^\alpha E[(X_1 - X_2)^\alpha I(X_1 > X_2)]\}.$$

In the null case we have that $(X_1, X_2) \stackrel{d}{=} (Z_1, Z_2) \stackrel{d}{=} (X_2, X_1)$ so that we can conclude that

$$(3.3) \quad P[M^{(\infty)} > y] \leq 2 \exp\left\{-\frac{1}{2} y^\alpha E|X_2 - X_1|^\alpha\right\}.$$

In order to apply the results in (3.3) we need to estimate $E|X_2 - X_1|^\alpha$ which can again be done naturally from the original data. Alternatively we may resort to (3.2), and compute estimates of $E[(X_2 - X_1)^\alpha I(X_2 > X_1)]$ and $E[(X_1 - X_2)^\alpha I(X_1 > X_2)]$ from the original data.

Right Tail Case (Condition R) Applying the inclusion exclusion argument we have that

$$P[M^{(\infty)} > y] = \sum_{i=1}^p P[V_i < y^{-1}] - \sum_{1 \leq i < j \leq p} P[V_i < y^{-1}, V_j < y^{-1}] + \cdots + (-1)^{p+1} P[V_1 < y^{-1}, \dots, V_p < y^{-1}]$$

and by taking marginal distributions in (2.14) and replacing x_i by y^{-1} we get the following result:

$$\begin{aligned}
 P[M^{(\infty)} > y] &= p \exp\{-E[Z_1^{-\alpha}]y^\alpha\} \\
 &\quad - \binom{p}{2} \exp\{-E[(Z_1 \wedge Z_2)^{-\alpha}]y^\alpha\} \\
 &\quad + \binom{p}{3} \exp\{-E[(Z_1 \wedge Z_2 \wedge Z_3)^{-\alpha}]y^\alpha\} - \dots \\
 &\quad + (-1)^{p+1} \exp\{-E[(\bigwedge_{i=1}^p Z_i)^{-\alpha}]y^\alpha\}.
 \end{aligned}
 \tag{3.4}$$

In order to apply this result we would of course have to estimate expectations of the form $E[(\bigwedge_{i=1}^s Z_i)^{-\alpha}]$ and this would be done using the observed residuals $\hat{Z}_t; t = p+1, \dots, n$. Care must be taken when estimating these residual sample moments: in particular p (or $p+1$ if a left endpoint is estimated as well) of the estimated residuals are identically zero by definition of the lp estimation procedure.

To complete the above approach, we need to suggest reasonable estimators for α and for b_n or a_n . This is discussed in Sections 4, 5 and 6.

4. An example of analysis for heavy tails.

If our null hypothesis is that the true AR(p) parameter vector ϕ_0 satisfies $\phi_0 = \mathbf{0}$, then we reject this hypothesis if the lp estimator $\hat{\phi}^{(n)}$ is outside a neighborhood of $\mathbf{0}$. The rejection region for our test of level δ is

$$\bigcup_{i=1}^p [\hat{\phi}_i^{(n)} > K(\delta)],$$

and the probability of rejection is approximated by (3.4) with $y = b_n K(\delta)$. We may use a further approximation to the exact expression in (3.4), namely

$$P\left[\bigvee_{i=1}^p L_i > b_n K(\delta)\right] \leq p P[L_1 > b_n K(\delta)] = p e^{-c(b_n K(\delta))^\alpha},$$

where $c = E(Z_1^{-\alpha})$. For given δ , say $\delta = .05$, we choose $K(\delta)$ to achieve an approximate level $\delta = .05$ test by setting

$$p e^{-c(b_n K(\delta))^\alpha} = .05$$

or

$$K(.05) = \frac{\left(\frac{-\log(.05/p)}{c}\right)^{1/\alpha}}{b_n} = \frac{\left(\frac{\log(20p)}{c}\right)^{1/\alpha}}{b_n}.$$

Alternatively, (3.4) could be used to compute a p-value for a value of the test statistic $M^{(n)}$.

For applications, α , c and b_n have to be estimated and there are a variety of techniques to do this. One estimator for α^{-1} is the Hill estimator (Hill, 1975; Mason 1982). If $X_{(1)} < \dots < X_{(n)}$ are the order statistics of the observed sample, then the Hill estimator based on $k = k(n)$ upper order statistics is

$$H_{k,n} := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{(n-i)} - \log X_{(n-k)},
 \tag{4.1}$$

and

$$H_{k,n} \xrightarrow{P} \alpha^{-1},$$

as $n \rightarrow \infty$, $n/k(n) \rightarrow \infty$. In practice, one makes a plot of $(k, H_{k,n}^{-1})$, $1 \leq k \leq n-1$ and hopes to be able to infer the value of α from where the plot is stable. This is sometimes difficult since there is no reliable guide as to how many upper order statistics to choose. However judicious use of pp-plots can supplement the information provided by the Hill plots.

Once α has been estimated by an estimator $\hat{\alpha}$ an obvious estimator for $c = E(Z_1^{-\alpha})$ is

$$(4.2) \quad \hat{c} = n^{-1} \sum_{i=1}^n X_i^{-\hat{\alpha}},$$

and this leaves the problem of how to estimate b_n . There are various methods for choosing b_n but a simple procedure is to assume there exists $x_0 \geq x_l > 0$, $x_0 \geq 1$ such that

$$(4.3) \quad P[Z_1 > x] = 1 - F(x) = \left(\frac{x}{x_l}\right)^{-\alpha}, \quad x > x_0,$$

that is, assume F is Pareto beyond some point. If (4.3) is true, then of course for sufficiently large n

$$\left(\frac{b_n}{x_l}\right)^{-\alpha} = \frac{1}{n},$$

or

$$(4.4) \quad b_n = x_l n^{1/\alpha}.$$

A graphical technique for determining x_l and α is the pp-plot. Observe that (4.3) implies that for $y > \alpha \log(x_0/x_l)$

$$P[\alpha \log \left(\frac{Z_1}{x_l}\right) > y] = e^{-y} =: \bar{G}(y).$$

So G is a distribution function of a unit exponential random variable. Let $F_*(y) = P[\log Z_1 \leq y]$. Then for $y > \log x_0$

$$\bar{F}_*(y) = P\left[\frac{\log Z_1 - \log x_l}{\alpha^{-1}} > \frac{y - \log x_l}{\alpha^{-1}}\right] = \bar{G}\left(\frac{y - \log x_l}{\alpha^{-1}}\right)$$

and therefore for $x > 1 - x_0^{-1}$:

$$F_*^{\leftarrow}(x) = \alpha^{-1} G^{\leftarrow}(x) + \log x_l.$$

Note $G^{\leftarrow}(x) = -\log(1-x)$ for $x > 1 - x_0^{-1}$. Thus, if we plot

$$\left\{\left(-\log\left(1 - \frac{j}{n+1}\right), \log Z_{(j)}\right), 1 \leq j \leq n\right\},$$

where $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ are the order statistics from the distribution given by (4.3), we should get approximately a line whose slope is α^{-1} and whose intercept is $\log x_l$. For specific data, choosing the threshold x_0 is something of an art and some trial and error is frequently involved.

We now analyse a set of real data which exhibits large values and seems to be generated by a sequence of independent random variables. The data set represents the interarrival times between packets generated and sent to a host by a terminal during a logged-on session. 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 received by the host were recorded as our data. The total length of the data is 783.

The index of variation is first estimated using the Hill estimator and the p-p plot technique described earlier. Then the question of independence is addressed. The heavy tail sample autocorrelation function is first used on the sample. Then the independence test developed in Section 3 is applied to the data set. Figure 1 is a time series plot of our data set showing indications that a heavy tail time series model is appropriate.

Figure 1

The Hill estimator applied to the data set, shown on the left of Figure 2, indicates an index of variation around 0.8. The p-p plot technique is used to produce the graph on the right of Figure 2. This graph shows a rather good fit between our data and a Pareto distribution with an estimated α of 0.86. The left point x_0 (see (4.3)), beyond which the data behaves approximately like a Pareto, involves the last 650 order statistics of the data set and the plot yields $x_l = 0.11$. From (4.4), this technique also gives an estimate of b_n which in this case is 258.9.

The heavy tail sample autocorrelation function (HTSACF) is a useful diagnostic tool and is defined as

$$(4.5) \quad \hat{\rho}(h) := \frac{\sum_{t=h+1}^n X_t X_{t-h}}{\sum_{t=1}^n X_t^2}.$$

If X_t is a stationary independent sequence then, as $n \rightarrow \infty$

$$\hat{\rho}(h) \xrightarrow{P} 0$$

for all $h \geq 1$ (Davis and Resnick (1985, 1986)).

Figure 2

The left hand plot in Figure 3 is an evaluation of $\hat{\rho}(h)$, $0 \leq h \leq 20$ for our data set and gives strong preliminary indications of independence. The right side of Figure 3 shows the estimated coefficients of an AR(20) model fit to the data. The dotted line represents the boundary of the lower 95% confidence interval where we use an estimated value of $\hat{\alpha} = 0.86$ and $\hat{b}_n = 258.9$. All the values for the coefficients are well inside the interval. Similar plots with similar results have been made for $p = 10, 5, 4, 3, 2, 1$. Both the HTSACF and the independence test seem to point to the same conclusion that the hypothesis of independence should not be rejected.

Figure 3

5. Implementation for the left tail case — lynx data.

As we will see later in this section, the Canadian lynx data seems to provide an example of a time series for which the innovations have a regular varying left tail. In fact, in their analysis of this series, Li and McLeod (1988) fit a Gamma model to the innovations process for the log transformed data. We will compare our approach to theirs in what follows.

Before proceeding we need to explain how the theory that we have developed can be implemented for this case. There are three basic issues to examine: (a) estimating a possibly non-zero left endpoint for the innovation distribution; (b) estimating the value of α , the parameter of regular variation, from the estimated

residuals and (c) estimating the normalizing sequence a_n^{-1} . We discuss these now.

(a) **Estimating the left endpoint.** Let us assume that the left endpoint of the Z_t distribution is η so that the distribution F of $Z_t - \eta$ satisfies Condition L. Then using the maximum likelihood paradigm based on exponentially distributed innovations as in Feigin and Resnick (1994) the appropriate linear programming estimator for $\theta = (\eta, \phi_1, \dots, \phi_p)$ is

$$(5.1) \quad \begin{aligned} & \arg \max \left\{ (n-p)\eta + \phi_1 \sum_{t=p}^{n-1} X_t + \dots + \phi_p \sum_{t=1}^{n-p} X_t : X_t - \eta - \sum_{i=1}^p \phi_i X_{t-i} \geq 0; t = p+1, \dots, n \right\} \\ & \approx \arg \max \left\{ \frac{\eta}{\mu_{n-p}} + \phi_1 + \dots + \phi_p : X_t - \eta - \sum_{i=1}^p \phi_i X_{t-i} \geq 0; t = p+1, \dots, n \right\} \equiv \hat{\theta}^{(n)} \end{aligned}$$

where $\mu_n = \sum_{t=1}^n X_t/n$ is the sample mean. If we presume that $\mu = E(X_t) < \infty$ then we can establish a limit result for the asymptotic distribution of $\hat{\theta}^{(n)}$. In fact it is not hard to show that for the true parameter vector $\theta^{(0)}$:

$$a_n^{-1}(\hat{\theta}^{(n)} - \theta^{(0)}) = \arg \max \{ \delta_0 + \dots + \delta_p : \delta' \mathbf{u}_t^n \leq 1; t = p+1, \dots, n \}$$

where

$$\mathbf{u}_t^n = \frac{1}{a_n^{-1} Z_t} (\mu_{n-p}, X_{t-1}, \dots, X_{t-p}).$$

Just as in Feigin and Resnick (1992,1994), we can show that

$$\sum_{t=p+1}^n \epsilon_{\mathbf{u}_t^n} \Rightarrow \sum_{k=1}^{\infty} \epsilon_{\mathbf{u}_k}$$

where $\mathbf{u}_k = \Gamma_k^{-1/\alpha}(\mu, Y_{k1}, \dots, Y_{kp})$ and Γ_k and $\mathbf{Y}_k$ are as defined in Theorem 2.1(a). The complete proof of this result will be given elsewhere.

(b) **Estimation of α .** Once the estimated residuals $\hat{Z}_t$ are obtained, we can use them to estimate the value of α . These residuals are estimated as follows:

$$\hat{Z}_t = X_t - \hat{\eta} - \sum_{i=1}^p \hat{\phi}_i X_{t-i}; t = p+1, \dots, n$$

and we note that by definition they are non-negative and exactly $p+1$ of them will be identically zero if the linear program has a unique solution. These zeroes are removed from the set as they are artificial and will invalidate the estimation of α . This leaves us with $n-2p-1$ positive residuals which we denote by $R_t; t = 1, \dots, m \equiv n-2p-1$. Our estimator of α is based on the Hill estimator (of the reciprocals) of these residuals. For $1 \leq k \leq m-1$ let

$$H_{k,m} = \log(R_{(k+1)}) - \frac{1}{k} \sum_{j=1}^k \log(R_{(j)}).$$

Then, as we show in Section 6, for $k = k(m) \rightarrow \infty$ such that $k/m \rightarrow 0$ and $k^2/m \rightarrow \infty$ as $m \rightarrow \infty$, $H_{k,m}$ is consistent for $1/\alpha$. In practice, we need to choose k during the data analysis.

The approach we have used is to estimate α as the reciprocal of the average of $H_{k,m}$ for $h \equiv [m^{0.75}] \leq k \leq 2h$ as well as viewing a plot of the variability of

$$(5.2) \quad \alpha_{k,m} = 1/H_{k,m}$$

as a function of k . This approach has proved to be very successful in that it gives a numerical answer that is reasonable if the plot is showing signs of stability, and we have the opportunity to reject the model if the plot is unstable. In summary, we have,

$$\hat{\alpha} = \left[\frac{1}{h+1} \sum_{k=h}^{2h} H_{k,m} \right]^{-1}.$$

(c) **Estimation of a_n^{-1} .** If we assume that the distribution function is of the form $F(x) = cx^\alpha; x \leq x_l$, then we have that $a_m = (cm)^{-1/\alpha}$ and $a_m = k^{-1/\alpha} a_{m/k}$. Since we know from Proposition 6.2 that

$$\frac{R_{(k)}}{a_{m/k}} \xrightarrow{P} 1,$$

a plausible estimator for a_m^{-1} is

$$\hat{a}_m^{-1} = k^{1/\alpha} / R_{(k)}$$

where $k \rightarrow \infty$ and $k/m \rightarrow 0$ as $m \rightarrow \infty$ and $R_{(k)}$ is the k th smallest of $\{R_t; t = 1, \dots, m\}$.

In our application we estimate based on the residuals again averaging:

$$\hat{a}_m^{-1} = \exp \left[\frac{1}{h+1} \sum_{k=h}^{2h} \log(k^{1/\hat{\alpha}} / R_{(k)}) \right]$$

where the $\hat{\alpha}$ was defined above. We plot $a_{k,m}^{-1} \equiv k^{1/\hat{\alpha}} / R_{(k)}$ as a function of k in order to visually be comfortable that the number obtained corresponds to a stable estimate as k varies.

We applied the above analysis to the Canadian lynx data $(Y_1, \dots, Y_{114})$ as was also done by Li and McLeod (1988) for the 114 observations from 1821 through 1934. They were interested in fitting a Gamma innovation model to the $\log_{10}$ transformed data, and considered the AR(2) model applied to the data *centered* as follows: $X_t = \log_{10}(Y_t) - 1.591$. The shape parameter that they obtained was 2.658 which corresponds to the α parameter of regular variation for the Gamma distribution. Their complete fitted model is:

$$X_t = 1.4771X_{t-1} - 0.6210X_{t-2} + Z_t$$

or

$$(5.3) \quad \log_{10}(Y_t) = 0.2289 + 1.4771 \log_{10}(Y_{t-1}) - 0.6210 \log_{10}(Y_{t-2}) + \hat{Z}_t$$

where the $\hat{Z}_t$ are the estimated innovations with 0 for left endpoint. They also report a variance of residuals for this model of 0.0372. However, this value is evaluated from the fitted Gamma distribution to these residuals: the actual sample variance of these residuals is 0.0656 !

In order to compare these suggested analyses with that obtained using the lp estimator, we fitted an AR(2) with estimation of the left endpoint as explained in (5.1). The result was:

$$(5.4) \quad \log_{10}(Y_t) = 0.7543 + 1.2927 \log_{10}(Y_{t-1}) - 0.7343 \log_{10}(Y_{t-2}) + \hat{Z}_t$$

with a sample residual variance of 0.05416. So from the point of view of residual variance one can argue that the fit is substantially better than that obtained by Li and McLeod (1988).

Log10(Lynx) data - AR(2) with left endpoint

Histogram of Residuals Density of Residuals

Figure 4

The next step was to look at the residuals. Histograms and a smoothed density estimator of the residuals are presented in the top two plots of Figure 4. The value of α based on the residuals was 1.896, and should be compared with the shape parameter reported by Li and McLeod (1988) — 2.658. Looking at the plot of the estimates $\alpha_{k,m}$ as a function of k , see (5.2), we are quite happy that the estimated value between $k = h$ and $k = 2h$ as marked by the vertical reference lines is a stable estimate of the true α . We also see that estimating a_n^{-1} also seems quite satisfactory. At this point, we can be satisfied that the model fits well, and that the residuals do behave as predicted by the “left tail” theory.

A further step is to test whether there is no remaining dependence in this series of residuals. We test the null hypothesis of independence versus an AR(1) model using the results of Section 3. The estimate of ϕ_1 for these 109 residuals is 0.1221. There is no intercept fitted at this stage. The residuals from this fit (the residuals of the residuals) have an estimated α of 1.501 and an estimated a_n^{-1} of 28.63. The estimate of c_α is found as the appropriate sample mean (see Section 3) and equals 0.4267. Putting all this together into the formula for the p-value of the test statistic $\hat{\phi}_1$ we obtain a value of 0.0614. We may therefore be concerned that the residuals do still have some remaining dependence.

Log10(Lynx) data - AR(3) with left endpoint

Histogram of Residuals Density of Residuals

Figure 5

Going through the same procedure fitting an AR(3) model to the series $\{\log_{10}(Y_t)\}$ and testing the independence of the residuals we find that the resulting p-value is very large (0.99), and that therefore the fit should be considered more satisfactory since there is no evidence that the residuals are not independent. In this case the fitted model is

$$(5.5) \quad \log_{10}(Y_t) = 0.8268 + 1.1722 \log_{10}(Y_{t-1}) - 0.5059 \log_{10}(Y_{t-2}) - 0.1247 \log_{10}(Y_{t-3}) + \hat{Z}_t$$

with sample residual variance 0.05367. Again we are satisfied with the stability of the estimates of α (1.686) and of a_n^{-1} (19.36) as can be seen from Figure 5.

6. Left tail estimation of α from the residuals of AR processes.

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

$$(6.1) \quad X_t = \sum_{i=1}^p \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

and that Condition L holds. We suppose

$$(6.2) \quad \Phi(z) = 1 - \sum_{i=1}^p \phi_i z^i \neq 0, \quad |z| \leq 1$$

so that (Brockwell and Davis, (1991)) the autoregression (6.1) exists and has a stationary solution of the form

$$(6.3) \quad X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}, \quad -\infty < n < \infty$$

where

$$(6.4) \quad C(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{\Phi(z)}, \quad |z| \leq 1.$$

The coefficients $\{c_i\}$ are assumed to also satisfy

$$(6.5) \quad \sum_{j=0}^{\infty} |c_j|^{\beta \wedge 1} < \infty$$

where β is defined in Condition L. The goal of this section is to show that the left tail Hill estimator applied to the residuals is consistent under condition L. The main tool for proving this procedure is consistent is the left tail empirical measure associated to the estimated residuals.

Hill's estimator for the left tail of a sequence of positive random variables $\{\xi_i; i = 1, \dots, n\}$ is

$$(6.6) \quad H_{k,n} = H_{k,n}(\xi) := \frac{1}{k} \sum_{i=1}^k (\log \xi_{(k+1)} - \log \xi_{(i)}),$$

where $\xi_{(1)} \leq \xi_{(2)} \leq \dots \leq \xi_{(n)}$ are the order statistics. When $k \rightarrow \infty$ and $n/k \rightarrow \infty$ as $n \rightarrow \infty$, $H_{k,n}$ is consistent for α^{-1} in the iid case when Condition L holds.

As stated in Section 2, if Conditions M, S and L hold, then the linear programming estimators $\hat{\phi}^{(n)}$ satisfy

$$(6.8) \quad a(n)^{-1}(\hat{\phi}^{(n)} - \hat{\phi}) \Rightarrow \mathbf{L}$$

where $\mathbf{L}$ is non-degenerate and

$$a(n) = a_n = F^{\leftarrow}(1/n).$$

For this sequence of estimators the estimated residuals $\hat{Z}_t^{(n)}$ are defined as

$$(6.9) \quad \hat{Z}_t^{(n)} := X_t - \sum_{i=1}^p \hat{\phi}_i^{(n)} X_{t-i}$$

In order to prove consistency of $H_{k,n}(\hat{Z})$, we follow the outline in Section 4 in Resnick and Stărică (1994) and first prove that the left tail empirical measure

$$(6.10) \quad \mu_{\hat{Z},n} = \frac{1}{k} \sum_{i=1}^n \epsilon_{\hat{Z}_i^{(n)}/a(n/k)} \Rightarrow \mu_2$$

in $M_+([0, \infty))$, the Radon measures on $[0, \infty)$. Here $\mu_2([0, x]) = x^\alpha$, $x > 0$ and note that " $\Rightarrow$ " denotes weak convergence in $M_+([0, \infty))$; since the limit μ_2 is deterministic, weak convergence is the same as convergence in probability with respect to the vague metric. The convergence in (6.10) will be shown to imply the consistency of Hill's estimator. Here is the first step.

Proposition 6.1. *If Condition L holds, then as $n \rightarrow \infty$, $n/k \rightarrow \infty$ and $n/k^2 \rightarrow 0$,*

$$\mu_{\hat{Z},n} = \frac{1}{k} \sum_{i=1}^n \epsilon_{\hat{Z}_i^{(n)}/a(n/k)} \Rightarrow \mu_2$$

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

Proof. Since

$$(6.11) \quad \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/a(n/k)} \Rightarrow \mu_2$$

in $M_+([0, \infty))$ (Resnick (1986), Proposition 5.3), the proof reduces to showing

$$d\left(\frac{1}{k} \sum_{i=1}^n \epsilon_{\hat{Z}_i^{(n)}/a(n/k)}, \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/a(n/k)}\right) \xrightarrow{P} 0,$$

where d is the vague metric. From the definition of d (Resnick (1987), Kallenberg (1983)), it suffices to check that, for any $f \in C_K^+([0, \infty))$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{k} \sum_{i=1}^n f(\hat{Z}_i^{(n)}/a(n/k)) - \frac{1}{k} \sum_{i=1}^n f(Z_i/a(n/k))\right| > \varepsilon\right) = 0.$$

Since f is continuous with compact support, the support of f is contained in $[0, c]$, for some $c > 0$ and f is uniformly continuous. Therefore

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

as $\theta \rightarrow 0$. Observe that

$$\begin{aligned} \left| \frac{\hat{Z}_i^{(n)}}{a(n/k)} - \frac{Z_i}{a(n/k)} \right| &= \left| \sum_{j=1}^p (\hat{\phi}_j^{(n)} - \phi_j) \frac{X_{i-j}}{a(n/k)} \right| \\ &\leq \sum_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j| \frac{X_{i-j}}{a(n/k)} =: R_i. \end{aligned}$$

Let $0 < \delta < 1$ and suppose M is a large quantity to be specified later. Define the following sets:

$$\begin{aligned} A_i &= \{R_i \leq \delta, \frac{Z_i}{a(n/k)} \geq c + \delta\}, \\ B_i &= \{R_i \leq \delta, \frac{Z_i}{a(n/k)} < c + \delta\}, \\ C_i &= \{R_i > \delta\}, \\ D_i &= \left\{ \sum_{j=1}^p \frac{X_{i-j}}{a(n/k)} > \frac{\delta}{Ma(n)} \right\} \end{aligned}$$

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

$$\begin{aligned}
& P(|\frac{1}{k} \sum_{i=1}^n f(\hat{Z}_i^{(n)}/a(n/k)) - \frac{1}{k} \sum_{i=1}^n f(Z_i/a(n/k))| > \varepsilon) \\
& \leq P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{A_i} > \varepsilon/3) \\
& \quad + P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{B_i} > \varepsilon/3) \\
& \quad + P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{C_i} > \varepsilon/3, \frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} \leq M) \\
& \quad + P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{C_i} > \varepsilon/3, \frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} > M) \\
& \leq 0 + P(\omega_f(\delta) \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/a(n/k)}([0, c+1]) > \varepsilon/3) + P(\frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} > M) \\
& \quad + P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{D_i} > \varepsilon/3)
\end{aligned}$$

where the first term is 0 because neither $Z_i/a(n/k)$ nor $\hat{Z}_i^{(n)}/a(n/k)$ belongs to the support of f . Because of (6.11), if δ is sufficiently small,

$$P(\omega_f(\delta) \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/a(n/k)}([0, c+1]) > \varepsilon/3) \rightarrow 0$$

as $n \rightarrow \infty$. From (6.8), $\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|/a(n)$ is tight. Hence, given any $\zeta > 0$, there exists M such that, for all n

$$P(\frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} > M) \leq \zeta.$$

Finally, set $\|f\| = \sup f(x)$ and note that by Chebishev

$$\begin{aligned}
& P(\frac{1}{k} \sum_{i=1}^n |f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{D_i} > \varepsilon/3) \\
& \leq \frac{3}{k\varepsilon} \sum_{i=1}^n E|f(\hat{Z}_i^{(n)}/a(n/k)) - f(Z_i/a(n/k))| 1_{D_i} \\
& \leq \frac{6\|f\|n}{k\varepsilon} P(\sum_{j=1}^p X_{1-j} > \frac{\delta a(n/k)}{Ma(n)}).
\end{aligned}$$

Condition L implies that the β moment of X_i exists (see below) so applying Chebishev again we get the following bound

$$\frac{6\|f\|n}{\varepsilon k} E(\sum_{j=1}^p X_{1-j})^\beta / (\frac{\delta a(n/k)}{Ma(n)})^\beta = (\text{const}) \frac{n}{k} (\frac{a(n)}{a(n/k)})^\beta.$$

Since $a(\cdot) \in RV_{-1/\alpha}$ at 0, we get from Potter's bounds (Bingham, Goldie and Teugels (1987)) that, for any $\zeta > 0$ and n sufficiently large

$$\frac{a(n)}{a(n/k)} = \frac{a(n/k \cdot k)}{a(n/k)} \leq (1 + \zeta)k^{-1/\alpha + \zeta}.$$

Consequently,

$$\frac{n}{k} \left(\frac{a(n)}{a(n/k)} \right)^\beta \leq (\text{const}) \frac{n}{k} k^{-\beta/\alpha + \beta\zeta} = (\text{const}) \frac{n}{k^{1+\beta/\alpha - \zeta\beta}}.$$

If ζ is sufficiently small and $n/k^2 \rightarrow 0$ then $n/k^{1+\beta/\alpha - \zeta\beta} \rightarrow 0$ as desired.

It remains to show $EX_i^\beta < \infty$. For the case when $\beta \leq 1$ it is easy to see that

$$EX_i^\beta = E\left(\sum_{j=0}^{\infty} c_j Z_{i-j}\right)^\beta \leq \sum_{j=0}^{\infty} |c_j|^\beta E Z_{i-j}^\beta < \infty$$

by (6.5) and Condition L. When $\beta > 1$ the map $x \mapsto x^\beta$ is convex and therefore

$$EX_i^\beta \leq E\left(\sum_{j=0}^{\infty} |c_j| Z_{i-j}\right)^\beta \leq \left(\sum_{j=0}^{\infty} |c_j|\right)^{\beta-1} \sum_{j=0}^{\infty} |c_j| E Z_{i-j}^\beta < \infty$$

again by (6.5) and Condition L. $\square$

For $x \in \mathbb{R}$, x_+ denotes $\max(x, 0)$.

Proposition 6.2. *As $n \rightarrow \infty$, $n/k^2 \rightarrow 0$ and $k/n \rightarrow 0$,*

$$\begin{aligned} (i) \quad & \frac{\hat{Z}_{(k)}^{(n)}}{a(n/k)} \xrightarrow{P} 1 \text{ in } \mathbb{R}, \\ (ii) \quad & \frac{1}{k} \sum_{i=1}^n \epsilon_{\hat{Z}_i^{(n)}/\hat{Z}_{(k)}^{(n)}} \xrightarrow{P} \mu_2 \text{ in } M_+([0, \infty)) \\ (iii) \quad & H_{k,n}^+(\hat{Z}) := \frac{1}{k} \sum_{i=1}^n (\log a(n/k) - \log \hat{Z}_i^{(n)})_+ \xrightarrow{P} \frac{1}{\alpha}, \\ (iv) \quad & H_{k,n}(\hat{Z}) := \frac{1}{k} \sum_{i=1}^k (\log \hat{Z}_{(k+1)}^{(n)} - \log \hat{Z}_{(i)}^{(n)}) \xrightarrow{P} \frac{1}{\alpha}. \end{aligned}$$

Proof. Only a sketch of the proof is given due to its resemblance to Proposition 4.2 in Resnick and Stărică (1994). Parts (i) and (ii) have proofs similar to those of Proposition 2.1 and Proposition 2.2 of Resnick and Stărică (1994). Let us denote

$$(6.12) \quad \mu_{\hat{Z},n} := \frac{1}{k} \sum_{i=1}^n \epsilon_{\hat{Z}_i/a(n/k)},$$

so that

$$H_{k,n}^+(\hat{Z}) = \int_0^1 -\log(y) \mu_{\hat{Z},n}(dy) = \int_0^1 \mu_{\hat{Z},n}([0, u]) \frac{du}{u}.$$

Because of Proposition 6.1, to prove (iii) we need only show

$$\lim_{t \rightarrow 0} \limsup_{n \rightarrow \infty} P\left(\int_0^t -\log(y) \mu_{\hat{Z},n}(dy) > \varepsilon\right) = 0,$$

for which it is enough to show

$$\lim_{t \rightarrow 0} \limsup_{n \rightarrow \infty} P(|\int_0^t \mu_{\hat{Z},n}([0, u]) \frac{du}{u} - \int_0^t \mu_{Z,n}([0, u]) \frac{du}{u}| > \varepsilon) = 0.$$

This follows because it is relatively easy using Chebishev and Potter's inequalities to show that

$$(6.13) \quad \lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} P[\int_0^t \mu_{Z,n}([0, u]) \frac{du}{u} > \epsilon] = 0.$$

We have that:

$$\begin{aligned} & P(|\int_0^t \mu_{\hat{Z},n}([0, u]) \frac{du}{u} - \int_0^t \mu_{Z,n}([0, u]) \frac{du}{u}| > \varepsilon) \\ & \leq P(\int_0^t \frac{1}{k} \sum_{i=1}^n |\epsilon_{\hat{Z}_i^{(n)}/a(n/k)}([0, u]) - \epsilon_{Z_i/a(n/k)}([0, u])| \frac{du}{u} > \varepsilon) \\ & \leq P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{[\hat{Z}_i^{(n)}/a(n/k) > u, Z_i/a(n/k) \leq u]} \frac{du}{u} > \frac{\varepsilon}{2}) \\ & \quad + P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{[\hat{Z}_i^{(n)}/a(n/k) \leq u, Z_i/a(n/k) > u]} \frac{du}{u} > \frac{\varepsilon}{2}). \end{aligned}$$

Applying Chebishev's and Potter's inequalities to the first term proves it goes to 0. For the second term:

$$\begin{aligned} & P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{(\hat{Z}_i^{(n)}/a(n/k) \leq u, Z_i/a(n/k) > u)} \frac{du}{u} > \varepsilon) \\ & = P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{(\hat{Z}_i^{(n)}/a(n/k) \leq u, Z_i/a(n/k) > u(1+\varepsilon))} \frac{du}{u} > \frac{\varepsilon}{2}) \\ & \quad + P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{(\hat{Z}_i^{(n)}/a(n/k) \leq u, Z_i/a(n/k) \in (u, u(1+\varepsilon)])} \frac{du}{u} > \frac{\varepsilon}{2}). \end{aligned}$$

By the argument yielding (6.13), it follows that the second term in this sum has the desired behavior. Moreover, for the first term of this sum, using the main idea in the proof of Proposition 6.1, it is enough to look at

$$\begin{aligned} & P(\frac{1}{k} \sum_{i=1}^n \int_0^t 1_{[\hat{Z}_i^{(n)}/a(n/k) \leq u, Z_i/a(n/k) > u(1+\varepsilon)]} \frac{du}{u} > \frac{\varepsilon}{4}, \frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} \leq M) \\ & \leq P(\int_0^t \frac{1}{k} \sum_{i=1}^n 1_{[\sum_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j| X_{i-j}/a(n/k) > u\varepsilon]} \frac{du}{u} > \frac{\varepsilon}{4}, \frac{\bigvee_{j=1}^p |\hat{\phi}_j^{(n)} - \phi_j|}{a(n)} \leq M) \\ & \leq P(\int_0^t \frac{1}{k} \sum_{i=1}^n 1_{[\sum_{j=1}^p X_{i-j} > u\varepsilon/M a(n/k)/a(n)]} \frac{du}{u} > \frac{\varepsilon}{4}) \end{aligned}$$

where M is defined in the proof of Proposition 6.1. By Chebishev's inequality we get the following bound:

$$\int_0^t \frac{n}{k} P(\sum_{j=1}^p X_{i-j} > \frac{u\varepsilon}{M} \frac{a(n/k)}{a(n)}) \frac{du}{u}.$$

It has been previously proven that, for any given u

$$\frac{n}{k} P\left(\sum_{j=1}^p X_{i-j} > \frac{u\varepsilon}{M} \frac{a(n/k)}{a(n)}\right) \rightarrow 0$$

for $n \rightarrow \infty$. We conclude the proof of (iii) by noting that, for any fixed t , the dominated convergence theorem implies

$$\int_0^t \frac{n}{k} P\left(\sum_{j=1}^p X_{i-j} > \frac{u\varepsilon}{M} \frac{a(n/k)}{a(n)}\right) \frac{du}{u} \rightarrow 0$$

for $n \rightarrow \infty$. The proof for (iv) mimics exactly the proof of Proposition 2.4 in Resnick and Stărică (1993) to which we refer the reader. $\square$

7. Testing for a zero coefficient.

In this section we show how the techniques of Sections 3 and 4 can be extended to the problem of statistically deciding when a coefficient in an $AR(p)$ process can be set equal to zero. For this problem we only have results for $p = 2$ as the case $p > 2$ seems quite complex.

Suppose then that (1.3) holds for $p = 2$:

$$X_t = \sum_{i=1}^2 \phi_i X_{t-i} + Z_t$$

and that conditions M, S and R hold. We assume the order of the model p is known to be 2 and thus we assume $\phi_2 \neq 0$. We seek to test the hypothesis that $\phi_1 = 0$.

If indeed $\phi_1 = 0$ then the autoregressive polynomial is

$$\Phi(z) = 1 - \phi_2 z^2$$

and $|\phi_2| < 1$ and so

$$C(z) = \frac{1}{1 - \phi_2 z^2} = \sum_{k=0}^{\infty} \phi_2^k z^{2k},$$

so that for the $MA(\infty)$ representation we have

$$c_{2k} = \phi_2^k, \quad c_{2k+1} = 0.$$

This means that

$$b_n(\hat{\phi} - (0, \phi_2)) \Rightarrow \mathbf{L}$$

where

$$\mathbf{L} = \arg \max_{\Lambda} \delta' \mathbf{1}$$

and

$$\Lambda = \{\delta : \delta' \mathbf{1} \geq -1, \delta' \mathbf{v}_k \leq 1, k \geq 1\}.$$

Here

$$\mathbf{v}_l = V_l(c_{l-1}, c_{l-2})$$

where

$$V_l = \bigvee_{k=1}^{\infty} \frac{\Gamma_k^{-1/\alpha}}{Y_{kl}}$$

and $\{Y_{kl}\}$ are iid random variables with distribution F . Thus taking into account the specific form of the coefficients $\{c_k\}$ we get

$$\mathbf{v}_l = \begin{cases} V_l(\phi_2^{\frac{l-1}{2}}, 0)', & \text{if } l \text{ is odd,} \\ V_l(0, \phi_2^{\frac{l-2}{2}})', & \text{if } l \text{ is even.} \end{cases}$$

Therefore, we can explicitly represent Λ as

$$\Lambda = \{\delta : \delta' \mathbf{1} \geq -1, \delta_1 \bigvee_{l \text{ odd}} V_l \phi_2^{\frac{l-1}{2}} \leq 1, \delta_2 \bigvee_{l \text{ even}} V_l \phi_2^{\frac{l-2}{2}} \leq 1\}$$

so that

$$\mathbf{L} = \left(\frac{1}{\bigvee_{l \text{ odd}} V_l \phi_2^{\frac{l-1}{2}}}, \frac{1}{\bigvee_{l \text{ even}} V_l \phi_2^{\frac{l-2}{2}}} \right).$$

We would reject the null hypothesis that $\phi_1 = 0$ if $\hat{\phi}_1$ is big so we need the distribution of L_1 . Note that

$$L_1 \stackrel{d}{=} 1 / \bigvee_{l=0}^{\infty} V_l \phi_2^l$$

and that for $x > 0$

$$\begin{aligned} P\left[\bigvee_{l=0}^{\infty} V_l \phi_2^l \leq x\right] &= P\left[\bigvee_{l=0}^{\infty} \bigvee_{k=1}^{\infty} \frac{\Gamma_k^{-1/\alpha}}{Y_{kl}} \phi_2^l \leq x\right] \\ &= P\left[\bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \bigvee_{l=0}^{\infty} \frac{\phi_2^l}{Y_{kl}} \leq x\right]. \end{aligned}$$

Let

$$Q_k = \bigvee_{l=0}^{\infty} \frac{\phi_2^l}{Y_{kl}}$$

and note $Q_k < \infty$ a.s. since $Q_k < \infty$ iff $Q_k^\beta < \infty$ where $\beta > \alpha$ in Condition R and $EQ_k^\beta < \infty$ since $E(1/Y_{kl}^\beta) < \infty$. Also we have

$$G(x) := P[Q_k \leq x] = \prod_{l=0}^{\infty} \bar{F}\left(\frac{\phi_2^l}{x}\right).$$

Since the point process with points $\{(\Gamma_k^{-1/\alpha}, Q_k), k \geq 1\}$ is Poisson with mean measure $\alpha s^{-\alpha-1} ds \times G(dy)$ we have

$$\begin{aligned} P\left[\bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} Q_k \leq x\right] &= \exp \left\{ - \iint_{\{(s,y): sy > x\}} \alpha s^{-\alpha-1} ds G(dy) \right\} \\ &= \exp \left\{ - \int_y \left(\frac{x}{y}\right)^{-\alpha} G(dy) \right\} \\ &= \exp \left\{ -x^{-\alpha} \int_y y^\alpha G(dy) \right\} = e^{-cx^{-\alpha}}, \end{aligned}$$

where

$$c = \int_y y^\alpha G(dy).$$

Thus for $x > 0$

$$P[L_1 \leq x] = P\left[\bigvee_{l=0}^{\infty} V_l \phi_2^l > x^{-1}\right] = 1 - e^{-cx^\alpha},$$

which is a Weibull distribution.

The test is to reject the null hypothesis that $\phi_1 = 0$ if $\hat{\phi}_1 > K(\delta)$ and $K(\delta)$ can be chosen so that

$$\begin{aligned} \delta \geq P[\hat{\phi}_1 > K(\delta)] &= P[b_n \hat{\phi}_1 > b_n K(\delta)] \\ &\approx P[L_1 > b_n K(\delta)] = e^{-c(K(\delta)b_n)^\alpha}. \end{aligned}$$

So for a given δ , we choose $K(\delta)$ as

$$K(\delta) = \frac{1}{b_n} \left(\frac{-\log \delta}{c} \right)^{1/\alpha}.$$

The quantities c, b_n, α need to be estimated.

8. Concluding remarks.

As is always the case when using parameter estimates in asymptotic distributions, there is the chance that even if the model is correct, that the resulting critical values or p-values are quite different from the true ones. At this stage, we propose the above methods in order to give the practitioner some guidelines for determining significant coefficients. A simulation study will be conducted in order to verify the efficacy of the procedures. The results will be reported elsewhere.

REFERENCES

- Bingham, N., Goldie, C. and Teugels, J., *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, UK, 1987.
- Davis, R. and Resnick, S., *Limit theory for moving averages of random variables with regularly varying tail probabilities*, Ann. Probability **13** (1985), 179–195.
- 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., *Extremes of moving averages of random variables with finite endpoint*, Ann. Probability **19** (1991), 312–328.
- Feigin, P.D. and Resnick, S., *Estimation for autoregressive processes with positive innovations*, Stochastic Models **8** (1992), 479–498.
- Feigin, P. D. and Resnick, S., *Limit distributions for linear programming time series estimators*, To appear: Stochastic Processes and their Applications (1994).
- Hill, B., *A simple approach to inference about the tail of a distribution*, Ann. Statist. **3** (1975), 1163–1174.
- Kallenberg, O., *Random Measures*, Third edition, Akademie-Verlag, Berlin, 1983.
- Li, W.K. and McLeod, A.I., *ARMA modelling with non-Gaussian innovations*, J. Time Series Analysis **9** (1988), 155–168.
- Mason, D., *Laws of large numbers for sums of extreme values*, Ann. Probability **10** (1982), 754–764.
- Meier-Hellstern, K., Wirth, P., Yan, Y., Hoefflin, D., *Traffic models for ISDN data users: office automation application*, Preprint Bell Laboratories, Holmdel (1991).
- Neveu, J., *Processus Ponctuels*, Ecole d'Été de Probabilités de Saint-Flour VI. Lecture Notes in Mathematics, Number 598, Springer-Verlag, Berlin, 1976.
- Resnick, S., *Point processes, regular variation and weak convergence*, Advances Appl. Probability **18** (1986), 66–138.
- Resnick, S., *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York, 1987.
- Resnick, S. and Stărică, C., *Consistency of Hill's estimator for dependent data*, To appear: Journal of Applied Probability (1994).

PAUL D. FEIGIN, FACULTY OF INDUSTRIAL ENGINEERING AND MANAGEMENT, TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

E-mail: paulf@ie.technion.ac.il

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

E-mail: sid@orie.cornell.edu

CĂTĂLIN STĂRICĂ, CORNELL UNIVERSITY, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, ETC BUILDING, ITHACA, NY 14853 USA

E-mail: starica@orie.cornell.edu