PITFALLS OF FITTING AUTOREGRESSIVE MODELS FOR HEAVY–TAILED TIME SERIES

Paul D. Feigin and Sidney I. Resnick

Technion, Israel Institute of Technology and Cornell University

Abstract. We consider the analysis of time series data, with particular emphasis on series which have a heavy-tailed structure — that is, whose marginal distributions have a right tail which is regularly varying at infinity with index $-\alpha$. A natural model to attempt to fit to time series data is an autoregression of order p, where p itself is often determined from the data. Recently several methods of parameter estimation for heavy tailed series have been considered, including Yule-Walker estimation, linear programming estimators, and periodogram based estimators. We investigate the statistical pitfalls of the first two methods when the models are mis-specified — either completely or due to the presence of outliers. We illustrate the results of our considerations on both simulated and real data sets.

1. Introduction. Data analysis of time series has undergone some changes due to the availability of large data sets. A preliminary analysis of such series, especially those coming from telecommunications data, has revealed that the observed data series often have heavy-tailed marginal distributions. Recent examples from the telecommunications field include time series of file lengths, cpu times to complete jobs, call holding times, times between terminal transmissions, 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).

The preliminary analysis which confirms the presence of heavy tails is based on results that show that using estimators such as the Hill estimator (see Hill, 1975; Mason, 1982) — originally designed for independent and identically distributed (iid) observations — is legitimate for stationary time series. See Resnick and Stărică (1995) for linear time series and Rootzen, H., Leadbetter, M. and de Haan, L. (1990), Hsing (1991) and Resnick and Stărică (1997) for results under mixing conditions. Examples are given in Feigin, Resnick and Stărică, 1995; Resnick, 1995.

Having established the heavy-tailed nature of the series, the next issue often investigated is the nature of the model describing the dependence, if any, between successive observations. One common choice of model is an autoregression of order p (AR(p)):

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t \; ; \; t = 1, \dots , n \tag{1.1}$$

where $\{Z_t\}$ are an iid sequence of innovations. Questions that ensue include:

- "What is the nature of the innovation series?";
- "What is the value of p?"; and
- "How should the parameters $\phi_1, \dots , \phi_p$ be estimated?"

Key words and phrases. autoregressions, time series analysis, heavy tails, regular variation, linear programming estimation, Yule–Walker estimation, autocorrelation, robustness.

This research was partially supported by the United States-Israel Binational Science Foundation (BSF) Grant No. 92-00227. Sidney I. Resnick was also partially supported by NSF Grant DMS-9400535 at Cornell University.

We qualitatively address these questions now.

In the heavy-tail case, if the model (1.1) were to be appropriate, then the Z_t distribution presumably would also have to be heavy-tailed. Thus we should consider AR processes with heavy-tailed innovations $\{Z_t\}$.

Extensions of the classical Gaussian analysis for determining p have been shown to work for series satisfying (1.1) with heavy-tailed innovations and in fact a modified AIC criterion can be shown to be consistent — see Knight (1989), Bhansali (1988). In addition, one can investigate the behavior of both regular and heavy-tailed sample autocorrelation functions (acf) and sample partial autocorrelation functions (pacf). The heavy-tailed versions compute correlations based on *uncentered* data values. See Davis and Resnick (1985a,b; 1986) and also Resnick and Stărică (1995) who provide examples of this kind of analysis. Large values of partial autocorrelations would provide an indication of the order of the AR required for adequate representation of the data.

Estimation of the parameters ϕ_i has been examined recently by several authors — see Feigin and Resnick (1994) and Davis and Resnick (1985a) and Mikosch et al (1995). Feigin and Resnick propose using *linear programming* (LP) estimators and show that they are asymptotically consistent and determine their limiting distribution. Alternatively, one can show that the Yule-Walker (YW) estimators are also consistent and also have a limiting distribution. Mikosch et al (1985) have suggested the use of periodogram estimators for the case of *symmetric α-stable* (SαS) innovations. We will not refer to these periodogram based estimators in the sequel as the positive time series we are concerned with have a single heavy right tail. They therefore cannot have SαS innovations. See also Davis, Knight and Liu (1991) who use least γ–deviation estimation techniques.

The aim of this current investigation and analysis is to evaluate what happens to these model identification and fitting strategies when the AR(p) model is not an accurate description of the structure of the time series. We consider alternatives such as

- a model with contamination by (large) outliers
- a non-linear times series model
- a moving average (MA(q))

Before we investigate the effects of these mis-specifications (in Section 5) we present some background information on the LP (Section 2) and YW (Section 3) estimation procedures. This information will lead us to consider ways of robustifying the LP estimators so as to make their statistical implementation more practical. A further issue that we consider in Section 4 is the development of diagnostic tests to verify model adequacy. This and various tests for independence are discussed in Section 4. Section 6 concludes with a discussion of several data examples.

2. Linear programming estimators. The LP estimators were derived to estimate the parameters of the AR(p) model (1.1) when the innovation distribution has either a heavy right tail or a "heavy" (regularly varying) left tail at some finite left endpoint a. See Feigin and Resnick (1992, 1994, 1995) and Feigin, Resnick and Stărică (1995). When we refer to heavy left tail, we refer to the regular variation of the innovation distribution at a. In both the right and left tail cases, we assume the innovation distribution has a *finite* left endpoint which will often be taken to be 0.

The LP estimators are defined as follows:

$$\hat{\boldsymbol{\phi}}(n) = \arg\max_{\boldsymbol{\phi}}\{\sum_{i=1}^{p}\phi_i \ : \ \sum_{i=1}^{p}\phi_i X_{t-i} \le X_t;\ t=p+1,\dots,n\}. \tag{2.1}$$

These estimators were shown to have good properties when model (1.1) holds, under conditions of stationarity and regular variation of the right *or left* tail of the Z_t-distribution — see Feigin and Resnick (1994). A small adaptation of the AR model (1.1), for the case of an unknown non-zero left endpoint a of the Z_t-distribution, leads to the following alternative linear program which is particularly relevant for the heavy left tail case:

$$(\hat{\boldsymbol{\phi}}(n), \hat{a}) = \arg\max_{(\boldsymbol{\phi},a)}\{\sum_i^p \phi_i + a/\bar{X} \ : \ \sum_i^p \phi_i X_{t-i} + a \leq X_t;\ t = p+1, \dots, n\}. \tag{2.2}$$

See Feigin, Resnick and Stărică (1995) for more details.

2.1 Understanding linear programming estimators.

In the heavy right tail case, the asymptotic theory of LP estimators reveals that the estimated value $\hat{\boldsymbol{\phi}}(n)$ of $\boldsymbol{\phi}$ is basically determined by a very large X, say X_l. This large value will appear on the left hand side in p constraints

$$\sum_i^p \phi_i X_{t-i} \leq X_t;\ t = l+1, \dots, l+p\ , \tag{2.3}$$

which will most likely be the active constraints at the optimum of the LP problem. For large data sets which do follow model (1.1), there will be several large values (due to the heavy right tail), and so removing the p active constraints should not cause a drastic change in the parameter estimates.

However, if model (1.1) is not exactly true, say due to a very few additive outliers, then removing active constraints could have a dramatic effect on the parameter estimates and on the objective function. We will use this fact in an attempt to diagnose a contaminated AR(p) model and also in order to propose a robustification of the LP estimators in the next subsection. Another possibility is to consider removing those p constraints associated with the k largest X's. In the heavy right tail case, these large X's may be legitimate, uncontaminated, values and so removing their constraints may not reveal the effect of additive outliers at other locations.

We note here that in the regularly varying left tail case, the asymptotics of the LP estimators indicate that the value of $\hat{\boldsymbol{\phi}}(n)$ from (2.1) (or of $(\hat{\boldsymbol{\phi}}(n), \hat{a})$ from (2.2)) is determined by those X_t's for which the corresponding Z_t's are close to zero (or to a). The relevant constraints are those with the corresponding X_t on the *right hand side*. Unfortunately, by observing only the X's, we cannot know which are likely to be the active constraints at the optimum. However, the stability of the estimate $\hat{\boldsymbol{\phi}}(n)$ can still be investigated as indicated above, by removing the p active constraints at the optimum. Note also that in the heavy *left* tail case, the inclusion of an intercept a in the model may be very important as far as the estimation of the coefficients $\boldsymbol{\phi}$ is concerned. We also note that when the X_t's are both positive and negative, then the LP (2.1) may be infeasible if the data do not follow (1.1) exactly. However, there will always be a feasible solution to (2.2). Therefore, when considering estimation of possibly contaminated autoregressive time series data taking both negative and positive values — see the *testar* data set in Section 3 — we use the LP estimator with "left endpoint" to be estimated.

2.2. Robustifying linear programming estimation.

Based on the above observations we propose the following strategies when applying the LP estimators to data.

As is usual in the time series literature we refer to additive outliers as those perturbations of the underlying process performed by the adding of (large) values to the basic AR(p) series. These are distinguished from *innovation outliers*, which involve aberrant innovations whose effect continues for a sequence of observations due to the recursive dynamics of the AR system.

A strategy designed to overcome the problem of additive outliers is to remove active constraints sequentially. The first step is to remove the p active constraints and then to solve the new LP problem. Next, remove a further p active constraints so that $2p$ constraints have been removed altogether. Repeat this process of solving and removing active constraints k times. The number k should reflect the number of contaminated data points we might expect to have in the data set. (Note that each aberrant X will give rise to p *bad* constraints. However these sets of constraints may not be disjoint.)

As the number of constraints removed increases, we seek a levelling off in the objective function and in the value of $(\hat{\boldsymbol{\phi}}(n), \hat{a})$. Suppose such a change is observed after removing j sets of p constraints and from then on the parameter estimates remain relatively stable as further constraints are removed. We would then conclude that there are approximately j aberrant observations and that the LP estimates after removing the related constraints are the robust estimates of the parameters $\boldsymbol{\phi}$. Note that these aberrant data values *need not be the largest values of the series.*

3. AR model selection and estimation based on sample correlations.

We review several model fitting and selection techniques which are based on the sample autocorrelation function. In each case we illustrate the technique by applying it to a sample of size 5,000 from the autoregression

$$X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t$$

where $\{Z_t\}$ are iid with standard Pareto distribution: $P[Z_1 > x] = 1/x,\ x \geq 1$. We refer to the sample as *testar*.

3.1. The heavy tailed acf. Mature statistical computer packages have built in routines to graph the classical sample autocorrelation function (acf) of the data given by

$$\hat{\rho}(h) = \frac{\sum_{i=1}^{n-h}(X_i - \bar{X})(X_{i+h} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}. \tag{3.1}$$

In the classical L_2 case where the variance of the marginal distribution is finite and correlations exist, the sample correlation $\hat{\rho}(h)$ estimates the mathematical correlation $\rho(h)$ and in fact

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

In the heavy tailed case where variances and even means may be infinite, there is no point to the centering by $\bar{X}$ and the following modification called the heavy tailed sample acf is more appropriate:

$$\hat{\rho}_H(h) = \frac{\sum_{i=1}^{n-h} X_i X_{i+h}}{\sum_{i=1}^{n} X_i^2}. \tag{3.2}$$

If we have the heavy tailed model given by the MA(∞) process

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \tag{3.3}$$

where $\{Z_t\}$ are non-negative, iid, heavy tailed with regularly varying parameter $-\alpha$, then the mathematical correlations do not exist if $\alpha < 2$. However $\hat{\rho}_H(h)$ still converges to a limiting constant (Davis and Resnick, 1985a,b; 1986)

$$\hat{\rho}_H(h) \xrightarrow{P} \frac{\sum_{j=0}^{\infty} c_j c_{j+h}}{\sum_{j=0}^{\infty} c_j^2} := \rho(h). \tag{3.4}$$

The mean corrected function given in (3.1) also converges in probability to the same limit. The limit law for $\hat{\rho}_H(h)$ is complicated and is established in Davis and Resnick (1986, 1985b). The most tractable cases are when either $\alpha < 1$, or when the distribution of Z_1 is either symmetric or has mean zero; the latter two cases are not particularly relevant to our assumption that innovations are positive. In the tractable cases for every $l \geq 1$

$$\left(\tilde{b}(n)^{-1} b(n)^2 (\hat{\rho}_H(h) - \rho(h)), 1 \leq h \leq l\right) \tag{3.5}$$

has a limit distribution in $\mathbb{R}^l$ where the normalizing constants are given by

$$b(n) = \left(\frac{1}{1-F}\right)^{\leftarrow}(n), \quad \tilde{b}(n) = \left(\frac{1}{P[Z_1 Z_2 > \cdot]}\right)^{\leftarrow}(n). \tag{3.6}$$

The rate of convergence is inferior to the LP estimator.

Even in the simple cases, the limit distribution of $\hat{\rho}_H$ is complicated and depends on the value of α. The percentiles of the distribution are hard to obtain and usually must be calculated by simulation. See Section 6 for an example. Nonetheless, $\hat{\rho}_H$ can be used as an exploratory tool to make preliminary investigations of dependence. Note that if the MA(∞) process $\{X_t\}$ is iid, so that $X_t = Z_t$, then $c_j = 0$ for $j \geq 1$ and for $h \geq 1$

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

This provides an exploratory tool for discovering independence: If on graphing the sample heavy tailed acf one finds only small values, then it may be possible to model the data as iid. Similarly, if the sample acf is small beyond lag q, then there is some evidence that MA(q) may be an appropriate model. Of course, without firm knowledge of the quantiles of the limit distribution of $\hat{\rho}_H(h)$, it is impossible to say with precision what *small* means and this emphasizes the importance of at least being able to simulate the quantiles.

Figure 3.1 displays on the left the classical acf given by (3.1) applied to *testar*; on the right is the comparable heavy tailed version given by (3.2).

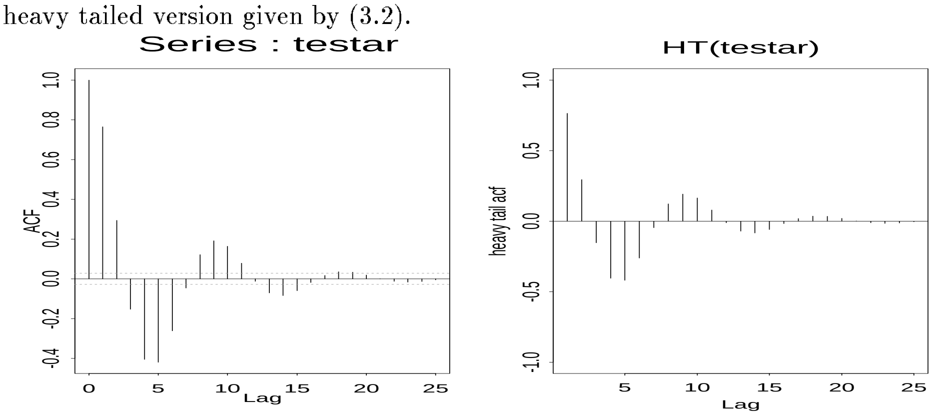

Figure 3.1. Sample acf and heavy tailed acf for testar.

3.2 Yule–Walker estimation in the heavy tailed case. In classical time series, the Yule–Walker estimators are moment based estimators for the coefficients in an autoregression. The method can easily be adapted to the heavy tailed case. Suppose $0 < \alpha < 2$ and that $\{X_t\}$ is a stationary, invertible autoregressive process of order p of the form

$$X_n = \sum_{i=1}^{p} \phi_i X_{n-i} + Z_n, \quad n = 0, \pm 1, \pm 2, \ldots \tag{3.7}$$

which can be inverted and written as the MA(∞) process

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}.$$

Write $c_j = 0$ if $j < 0$ so that

$$\rho(-i) = \frac{\sum_{k=i}^{\infty} c_k c_{k-i}}{\sum_{k=0}^{\infty} c_k^2} = \rho(i).$$

Set

$$\mathbf{R} = (R_{ij})_{i,j=1}^{p} = (\rho(i-j))_{i,j=1}^{p}, \quad \boldsymbol{\rho} = (\rho(1), \dots, \rho(p))'$$

and we have the Yule–Walker equation

$$\mathbf{R}\boldsymbol{\phi} = \boldsymbol{\rho} \tag{3.8}$$

where of course $\boldsymbol{\phi}$ is the p-vector of autoregressive coefficients in (3.7). Furthermore, for every m, the matrix

$$\mathbf{R}_m = (\rho(i-j))_{i,j=1}^{m} \tag{3.9}$$

is non-singular, provided $\sum_k c_k^2 > 0$. The heavy tailed Yule–Walker estimator $\hat{\boldsymbol{\phi}}^{YW}$ of $\boldsymbol{\phi}$ satisfies

$$\hat{\mathbf{R}}\hat{\boldsymbol{\phi}}^{YW} = \hat{\boldsymbol{\rho}_H} \tag{3.10}$$

where

$$\hat{\mathbf{R}} = (\hat{\rho}_H(i-j))_{i,j=1}^{p}, \quad \hat{\boldsymbol{\rho}}_H = (\hat{\rho}_H(1), \dots, \hat{\rho}_H(p))'.$$

Since $\hat{\boldsymbol{\rho}}_H \xrightarrow{P} \boldsymbol{\rho}$ (Davis and Resnick, 1985a) and $\hat{\mathbf{R}} \xrightarrow{P} \mathbf{R}$ as $n \to \infty$, the consistency of the heavy tailed Yule–Walker estimators follows.

Furthermore, in nice cases (eg, when $\alpha < 1$)

$$\tilde{b}(n)^{-1} b(n)^2 (\hat{\boldsymbol{\phi}}^{YW} - \boldsymbol{\phi})$$

has a limit distribution which is a function of the limit distribution obtained for the sample correlation function (Davis and Resnick, 1986). Again, the rate of convergence to this limit distribution is inferior to that of the linear programming estimators.

Applying the heavy tailed Yule–Walker procedure to *testar* produced correct answers of 1.3, -0.7. Since the order was known, this is not too surprising. This success was duplicated when the sample size was reduced from the original 5000 to 1000 and then to 500. For a sample of size 200, the procedure produces 1.1354 and -0.5555. Fitting Yule–Walker with the wrong order of $p = 4$ produces answers of 1.3, -0.7, $2 \times 10^{-9}, -3 \times 10^{-9}$.

3.3. The heavy tailed partial autocorrelation function. Continue to suppose (3.1), (3.4), (3.7) and now define

$$\boldsymbol{\phi}_m^* = \begin{cases} (\phi_1, \dots, \phi_m), & \text{if } m \le p, \\ (\phi_1, \dots, \phi_p, 0, \dots, 0), & \text{if } m \ge p \end{cases}$$

and

$$\boldsymbol{\rho}_m = (\rho(1), \dots, \rho(m))'.$$

For $m > p$ we have

$$\mathbf{R}_m \boldsymbol{\phi}_m^* = \boldsymbol{\rho}_m$$

and so

$$\boldsymbol{\phi}_m^* = \mathbf{R}_m^{-1} \boldsymbol{\rho}_m.$$

Recall that in classical time series analysis, the m-th component on the right would be the partial autocorrelation at lag m (Brockwell and Davis, 1991, page 102) and we call the m-th component of the m-vector

$$\hat{\boldsymbol{\phi}}_m^* = \hat{\mathbf{R}}_m^{-1} \hat{\boldsymbol{\rho}}_{H,m} \tag{3.8}$$

the sample heavy tailed partial autocorrelation function (pacf) at lag m. For $m > p$ we have

$$\hat{\boldsymbol{\phi}}^*_m \xrightarrow{P} \boldsymbol{\phi}^*_m$$

in $\mathbb{R}^m$ so that for the m-th component we have when $m > p$

$$\hat{\phi}^*_{m,m} \xrightarrow{P} 0.$$

Again, in simple cases such as when $\alpha < 1$, we have that for $m > p$

$$\tilde{b}_n b_n^2 (\hat{\boldsymbol{\phi}}^*_m - \boldsymbol{\phi}^*_m)$$

has a limit distribution depending on the limit achieved for the sample acf in (3.5). This result of course means that the m-th component $\tilde{b}_n b_n^2 \hat{\phi}^*_{m,m}$ has a limit distribution.

This result also yields an exploratory technique for diagnosing when an autoregression might be a suitable candidate model for a data set. Graph the heavy tailed sample pacf and if it dies after p lags, try fitting an AR(p). Again, because of the complexity of the limit distribution, there is difficulty in deciding at what lag the graph has died.

Figure 3.1 graphs both the classical pacf (left) and the sample heavy tailed pacf (right) applied to the *testar* data.

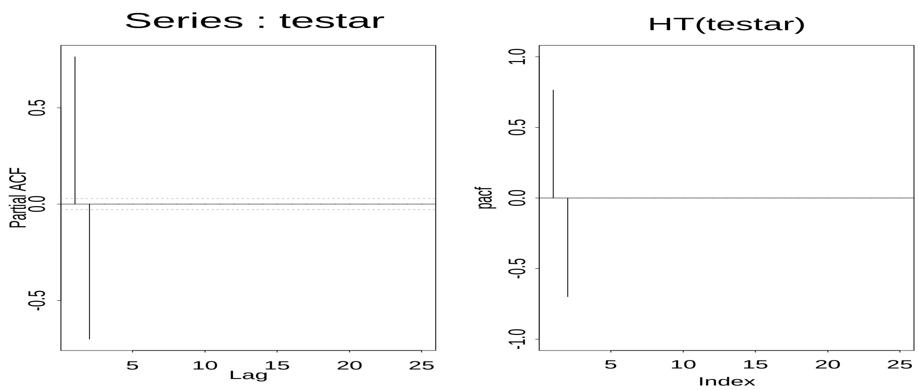

Figure 3.2. Pacf and heavy tailed pacf for testar.

3.4. Order selection and the AIC. The AIC criterion for order selection is not consistent in the classical L_2 case (Brockwell and Davis, 1991) but is consistent for heavy tails (Knight, 1989; Bhansali, 1988). Define recursively

$$\hat{\sigma}^2(0) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\hat{\sigma}^2(m) = \hat{\sigma}^2(0)\prod_{j=1}^{m}(1 - \hat{\phi}^*_{j,j})^2), \quad m \geq 1.$$

The heavy tailed AIC function is defined by

$$AIC(k) = n \log \hat{\sigma}^2(k) + 2k,$$

and the estimate of p is obtained by minimizing this function:

$$\hat{p} = \text{argmin}_{k \leq K} AIC(k)$$

where K is an upper bound which is magically known to exist. As $n \to \infty$, $\hat{p} \xrightarrow{P} p$, the true order.

Graphing $\{AIC(k), k \geq 1\}$ helps in determining the order. However, for real data, it may be the case that even though $AIC(k)$ has a nice minimum, residual analysis of the fitted model reveals a lack of iid structure in the estimated residuals, throwing into doubt the goodness of fit.

Figure 3.3 displays a plot of $\{(k, AIC(k), 1 \leq k \leq 10\}$. Observe the occurrence of the minimum.

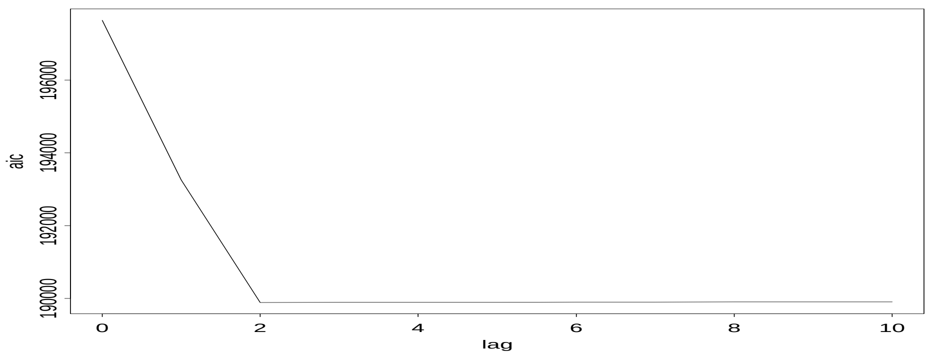

Figure 3.3. AIC plot for testar.

4. Tests for independence and AR model adequacy.

We outline several tests for independence which can help distinguish when the iid model is adequate from the situation that a model incorporating dependencies is necessary. Our tests are motivated by the situation in which we must choose between the alternatives of iid modeling versus autoregressive modeling.

Any test for independence can also be used as a diagnostic to examine model adequacy when using autoregressive modeling by considering the following steps:

(1) Determine the order p of the autoregressive model to be used.

(2) Fit the p autoregressive coefficients using an effective method such as the linear programming estimators. Call the vector of parameter estimates $\hat{\boldsymbol{\phi}}$.

(3) Compute the estimated residuals. If the model is

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

then the estimated residuals are

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

Disregard the p zero (0) residuals that arise from the use of linear programming estimators.

(4) Test the estimated residuals for independence.

We now outline several methods which can be used to examine independence. Some of these are based on asymptotic methods using heavy tailed analysis and the rest are standard time series tests of homogeneity.

4.1. Method based on sample acf. An exploratory, informal method for testing for independence can be based on the *heavy tailed sample autocorrelation function* $\hat{\rho}_H(h)$ where for h any positive integer

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

This was discussed in the previous section. Here, note that (Davis and Resnick, 1985a) if $\{X_t\}$ are iid with regularly varying tail probabilities, then

$$\lim_{n\to\infty} \hat{\rho}_H(h) = \begin{cases} 1, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0. \end{cases}$$

Thus, if upon graphing $\hat{\rho}_H(h), h = 0, \dots, n-h$ we get only small values for $h \neq 0$ there is no evidence against independence. The limit distribution of $\hat{\rho}_H(h), h = 1, \dots, q$ is known but it is difficult to work with and the percentiles must be calculated by simulation.

4.2. Tests based on asymptotic theory. The LP estimator can be used to fashion a test for independence against autoregressive alternatives: Test if

$$\phi_1 = \dots = \phi_p = 0$$

by rejecting when

$$\bigvee_{i=1}^{p} |\hat{\phi}_i(n)|$$

is too large (Feigin, Resnick and Stărică, 1996).

It would not be possible to fix the size of the test if the limit distribution of the LP estimator did not considerably simplify. Fortunately it does and under the null hypothesis of $\boldsymbol{\phi}(0) = \mathbf{0}$

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

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

$$P[V_i \leq x_i, i = 1, \dots, p] = \exp\{-\int_{(y_1,\dots,y_p)\in[0,\infty)^p} \left(\bigwedge_{l=1}^{p} y_l x_l\right)^{-\alpha} F(dy_1)\dots F(dy_p)\}. \tag{4.12}$$

This means that if we want a 0.05 level rejection region, we should reject when

$$P[\bigvee_{i=1}^{p} |\hat{\phi}_i(n)| > K(.05)] = .05$$

and to find an approximate value of $K(.05)$ we write

$$P[\bigvee_{i=1}^{p} |\hat{\phi}_i(n)| > K(.05) \approx P[\bigvee_{i=1}^{p} L_i > b_n K(.05)] \leq pP[L_1 > b_n K(.05)] = pe^{-c(b_n K(.05))^\alpha}, \tag{4.13}$$

where $c = E(Z_1^{-\alpha})$. This yields

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

We need to estimate α, c and b_n. One way to do this is to use the QQ-plot (Feigin, Resnick and Stărică, 1996; Kratz and Resnick (1996)) which yields both $\hat{b}_n$ (as the intercept of the fitted line) and $\hat{\alpha}$ (as the reciprocal of the slope of the fitted line) and then we can get

$$\hat{c} = n^{-1} \sum_{i=1}^{n} X_i^{-\hat{\alpha}}.$$

4.3. Standard tests of randomness. There are several standard time series tests of randomness (Brockwell and Davis, 1991, Section 9.4; Kendall and Stewart, 1976) which are nonparametric and can be employed in the present context. We give some examples below. We use the notation

$$\chi_n \sim AN(\mu_n, \sigma_n^2)$$

as shorthand to mean that

$$(\chi_n - \mu_n)/\sigma_n \Rightarrow N(0,1).$$

(1) Turning point test. If T is the number of turning points among $X_1, \dots, X_n$ then under the null hypothesis that the random variables are iid we have

$$T \sim AN(2(n-2)/3, (16n-29)/90)$$

and this can be used as the basis of a test.

(2) Difference-sign test. Let S be the number of $i = 2, \dots, n$ that $X_i - X_{i-1}$ is positive. Under the null hypothesis that the random variables $X_1, \dots, X_n$ are iid we have

$$S \sim AN(\frac{1}{2}(n-1), (n+1)/12).$$

(3) Rank test. Let P be the number of pairs (i,j) such that $X_j > X_i$ for $j > i$ and $i = 1, \dots, n-1$. Under the null hypothesis that the random variables $X_1, \dots, X_n$ are iid we have

$$P \sim AN(\frac{1}{4}n(n-1), n(n-1)(2n+5)/8).$$

We would reject the iid hypothesis at the 0.05 level if any of these standardized variables had an absolute value greater than 1.96. All of these tests are implemented in the Brockwell and Davis (1991) package ITSM. Residuals can easily be imported into their program and tested within the package for randomness.

4.4. Stability testing on subsets of the data. An informal but useful technique is to take a statistic, such as the sample acf, and compute it relative to different subsets of the sample. If the data is iid, the values of the statistic should be similar across different subsets.

For the sample acf, if the graphs of $\hat{\rho}_H(h), h = 1, \dots, q$ look different for different subsets, then one should be skeptical of the correctness of the iid assumption. Often it is enough to split the sample into halves or thirds to generate some skepticism.

4.5 Permutation tests for independence. Another approach to testing for independence in time series analysis is based on permutation tests. Here we can use any desired statistic that is designed to measure some form of dependence between successive data. This statistic might be a maximum autocorrelation or partial autocorrelation, or it may be a maximal autoregressive coefficient estimated by the linear programming paradigm. Of course many other candidates are available, including robust measures such as biserial autocorrelations.

The permutation test is based on comparing the observed value of the statistic with the permutation distribution of that statistic — that is with the distribution of values of the statistic under all the possible permutations of the time series data. If there is no dependence structure in the data, then the observed value should be a typical value for this reference permutation distribution. If there is some dependence of the type to which the statistic is sensitive, then the observed value should be extreme with respect to this reference distribution.

This approach allows one to perform tests without relying on the asymptotic theory for the particular statistic. As we have seen earlier, the asymptotic distribution for

$$\bigvee_{i=1}^{p} |\hat{\phi}_i(n)|$$

involves various parameters that have to be estimated. Moreover, the fact that we are not sure of the rate of convergence to the asymptotic distribution, also suggests the precautionary tactic of using a permutation test.

In the implementation we use below, we approximate the *p-value* of the actually observed statistic. This is achieved by generating 99 permutations of the time series, computing the statistic for each one, and counting the number (C) of these that are greater than or equal to the actually observed statistic. The p-value is approximated by $(1+C)\%$. The statistics considered are the maximum absolute autocorrelation (macf), the maximum absolute partial autocorrelation (mpacf), and the maximum absolute linear programming coefficient estimate (mphi). In each case, one must specify the value of p, the maximal order over which the maximum is taken.

5. Effects of mis–specified models on AR estimation.

In this section we sound several cautionary warnings about the potential for mischief when the model is mis–specified. What if we try to fit a heavy tailed autoregression when the model is really bilinear or a finite order moving average?

5.1. Non–linear time series: a bilinear example. Evidence is strong that the presence of non-linearities will dramatically change behavior of heavy tailed time series estimators such as the LP and Yule Walker estimators. Davis and Resnick (1996) show that for a simple heavy tailed bilinear process, the heavy tailed sample acf

$$(\hat{\rho}_H(1), \dots, \hat{\rho}_H(h))$$

will converge in distribution in $\mathbb{R}^h$ to a non-degenerate random limit. This is in sharp distinction to the behavior of the heavy tailed sample acf of an MA(∞), where convergence is to a limiting constant as in (3.4). The theory behind such a result also shows that Yule-Walker estimators applied to bilinear data will converge in distribution to non-degenerate random limits rather than to constants. Furthermore, it is clear that the LP estimators, at least for $p = 1$, will converge to constants and under certain circumstances this can lead to confusion between independence and nonlinearity.

To illustrate the effect of non-linearities on estimation procedures for autoregressive processes, we simulated three independent samples ($test_i,\ i = 1, 2, 3$) of size 5000 from the bilinear process

$$X_t = .1Z_{t-1}X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots, \tag{5.1}$$

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

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

A stationary solution for (5.1) exists and is of the form

$$X_t = Z_t + \sum_{k=1}^{\infty} (.1)^k \left(\prod_{j=1}^{k-1} Z_{t-j} \right) Z_{t-k}^2,$$

so that the tail behavior of X_t is governed by $\{Z_t^2\}$. Since Z_t has a distribution tail which is regularly varying with index -1, $P[X_t > x]$ should be regularly varying with index $-1/2$. Figure 5.1 presents a time series plot of $test_1$ with an accompanying Hill plot which estimates the α and shows it is in a neighborhood of .5 as theory dictates. A Hill plot is a plot of $\{(k, H_{k,n}^{-1}), 1 \leq k \leq n\}$ where $H_{k,n}$ is the Hill estimator

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

which is constructed from the k largest order statistics

$$X_{(1)} \geq X_{(2)} \geq \cdots$$

of a sample of size n. The Hill plot is a popular method of estimating the parameter of regular variation.

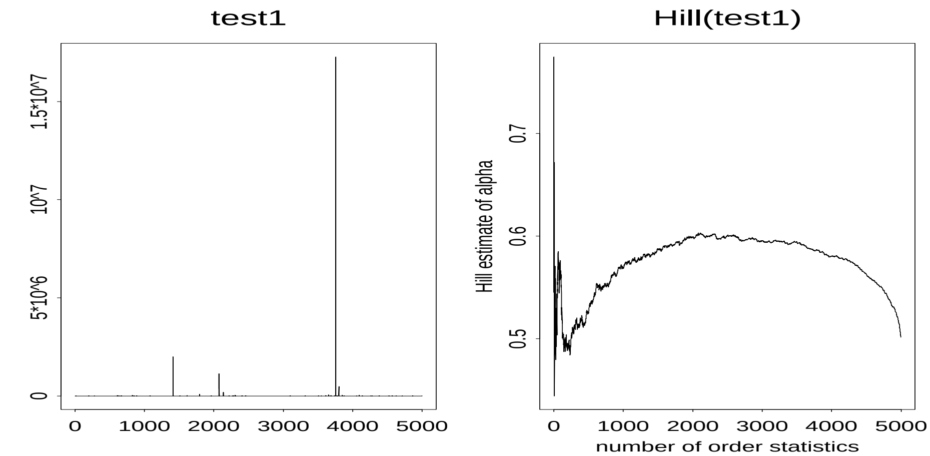

Figure 5.1. Test1: tsplot and Hill plot.

The erratic nature of the behavior of $\hat{\rho}_H$ is illustrated in Figure 5.2 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.

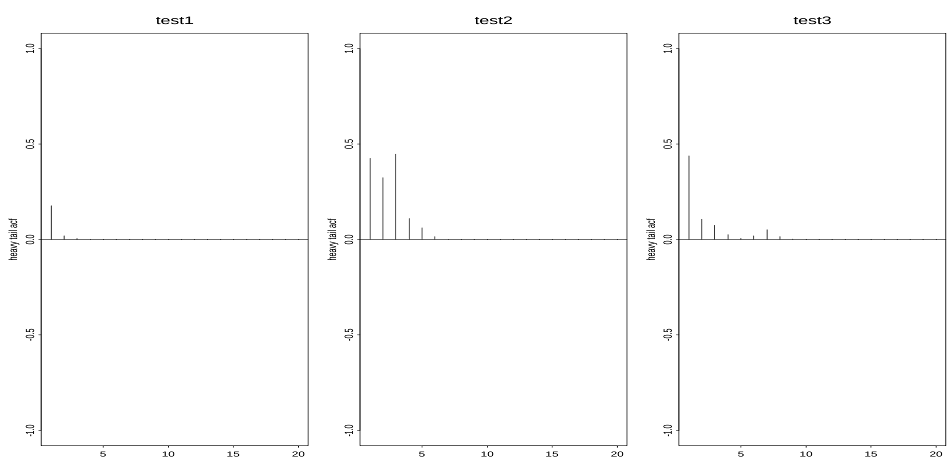

Figure 5.2. Heavy tailed acf for 3 bilinear samples.

In contrast, we present in Figure 5.3 comparable heavy tailed acf plots for three samples of testar of size 1500. Here, the pictures look identical.

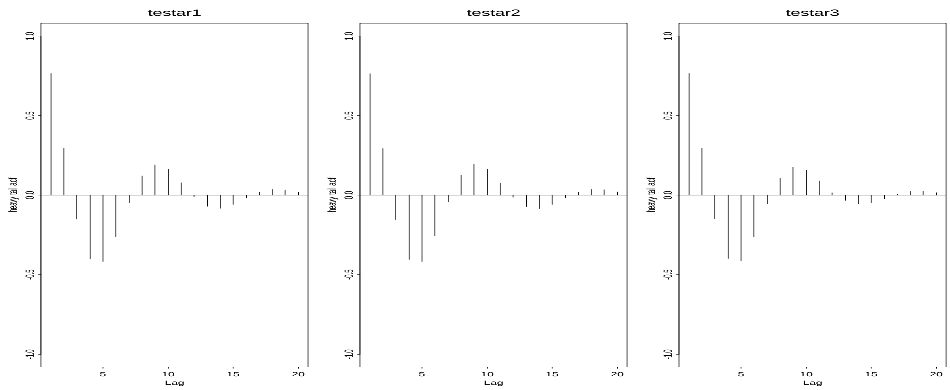

Figure 5.3. Heavy tailed acf for 3 autoregressive samples.

The erratic behavior of the heavy tailed sample acf of course affects the behavior of the Yule-Walker estimators. We attempted to fit an AR(2) to the three nonlinear *test* sets using Yule-Walker and obtained (ϕ_1, ϕ_2) values of (0.17863100, -0.01221288), (0.3506413, 0.1749626) and (0.4849370, -0.1066039) which are rather different from each other and reflect the fact that for this bilinear process, the Yule–Walker estimators converge as $n \to \infty$ to non-degenerate random variables.

Figure 5.4 gives the heavy tailed pacf plots for the three *test* data sets and Figure 5.5 gives the AIC plots. Again, one observes tremendous variability in the plots. Observing only one of the plots would be very misleading and would provide no insight into the non-linearity. In particular, the first AIC plot on the left of Figure 5.5 would suggest a low order AR model.

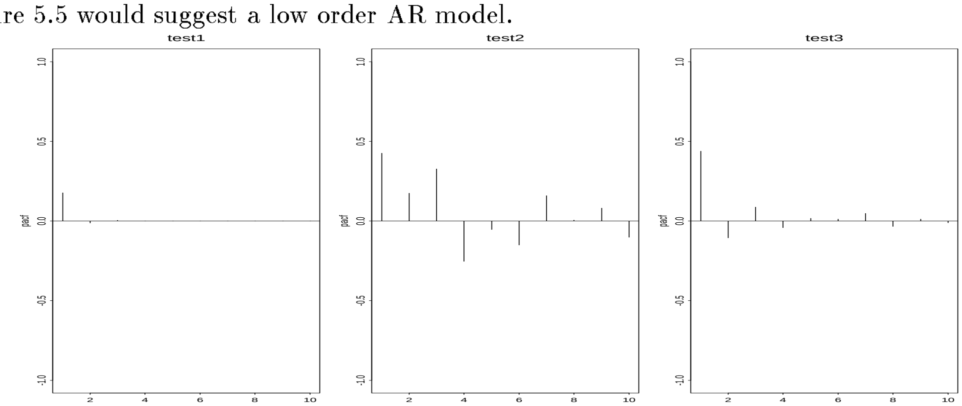

Figure 5.4. Heavy tailed pacf for 3 bilinear samples.

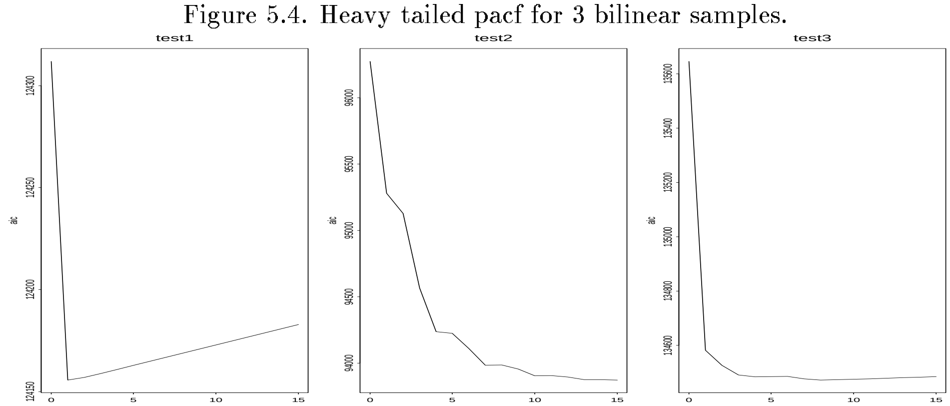

Figure 5.5. Heavy tailed AIC for 3 bilinear samples.

The limit theory for this bilinear example predicts that the LP estimator applied to data from the model

$$X_t = cZ_{t-1}X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots \tag{5.2}$$

will yield an estimate for $p = 1$ with the property that

$$\hat{\phi}_1 \xrightarrow{P} c \bigwedge_{j=1}^{\infty} Z_j$$

(see Resnick, 1996). For the example in (5.1), $\hat{\phi}_1 = .1$ for all three *test* data sets, as it should. For $p = 2$ we obtained for the LP estimation (ϕ_1, ϕ_2) values of (0.1021767, -0.0006200289), (0.1055141, -0.001255001), (0.1019985, 0.0001410015). Theory dictates that for $p = 2$ the LP estimators applied to the process in (5.2) will satisfy

$$(\hat{\phi}_1, \hat{\phi}_2) \xrightarrow{P} (c, 0)$$

where c is the constant in (5.2). In our example, $c = 0.1$.

An important point is that if the $\{Z_t\}$ variables in the construction of $\{X_t\}$ were centered to have an infimum close to 0 instead of 1, then the non-linearity in the model produces LP estimates which are close to 0. This has the potential to raise confusion between non-linearity and independence, since an independent model would also yield LP estimates which were close to 0.

5.2. Moving average processes. We considered what happens if the true underlying time series is a heavy tailed moving average. We created three independent sample data sets, called $testma_i$, $i = 1, 2, 3$, of length 3000 each from the process

$$X_t = Z_t + .3Z_{t-1} + .5Z_{t-2}, \quad t = 1, \dots, 3000.$$

For $p = 1$, the LP estimators applied to the three *testma* sets yielded estimates of 0.003263236, 0.003972783 and 0.000755287. As $n \to \infty$, back of the envelope calculations indicate that the theoretical limit in probability of the LP estimator should be 0. For $p = 2$ the LP estimates for (ϕ_1, ϕ_2) were (0.001206104, 0.003434608), (0.001226002, 0.00457661) and (0.0002060519, 0.0009152887). As $n \to \infty$ preliminary theoretical calculations show that the LP estimator for $p = 2$ should converge to $(0, 0)$.

For $p = 1$ the Yule–Walker method applied to the three *test* sets yielded the estimates 0.3622376, 0.3537322 and 0.3389519, while for $p = 2$ the Yule–Walker estimates were (0.2492008, 0.3120514), (0.2431817, 0.3125259) and (0.2380277, 0.2977536). For comparison purposes we note that the AR(∞) representation of this MA(2) process is

$$X_t - 0.3X_{t-1} - 0.41X_{t-2} + 0.2730000X_{t-3} + 0.1231X_{t-4} - 0.17343X_{t-5} + \cdots = Z_t \ .$$

For $p = 1$,

$$\hat{\phi}_1^{YW} = \hat{\rho}_H(1) \xrightarrow{P} \rho(1) := \frac{\sum_{i=0}^{\infty} c_i c_{i+1}}{\sum_{i=0}^{\infty} c_i^2}$$

and for our case $c_0 = 1, c_1 = .3, c_2 = .5$ so the value of the limit is .33582. For $p = 2$, we have $\rho(1) = .33582$, $\rho(2) = .3731$ and $(\hat{\phi}_1^{YW}, \hat{\phi}_2^{YW})$ converges in probability as $n \to \infty$ to the solution $\mathbf{x}$ of

$$\begin{pmatrix} 1 & .33582 \\ .335821 & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} .33582 \\ .3731 \end{pmatrix}.$$

The solution is $\mathbf{x} = (.2373, .2934)'$.

Despite the model being badly misspecified, the Yule–Walker method gives seemingly reasonable but misleading answers and offers no warnings that something is amiss. The LP estimators give values close to zero. Since examining the residuals quickly generates skepticism that the residuals are independent, a warning is sounded that something is wrong. So for this case of model misspecification, the LP estimation technique offers a significant advantage.

5.3. Effect of outliers.

We now turn to the situation when the underlying series $\{X_t\}$ does follow the AR(p) model (1.1), but the observed series is contaminated by (additive) outliers:

$$Y_t = X_t + \delta_t E_t$$

where $\delta_t = 0$ with probability $1 - \alpha$; and $\delta_t = 1$ with probability α; and E_t comes from some contamination distribution.

In order to illustrate the robust estimation strategy outlined in Section 2.2, we generated a contaminated series $\{Y_t\}$ as follows: taking as the $\{X_t\}$ series the first 500 observations of *testar* (say, *testar5*); choosing $\alpha = 2\%$ contamination; and letting $E_t = 100 Z_t^*$ where $\{Z_t^*\}$ are iid standard Pareto independent of $\{Z_t\}$.

The resulting $\{Y_t\}$ series had 14 observations contaminated (out of 500), and the four largest values of $\{Y_t\}$ were *not* the contaminated ones.

Applying the LP estimator (with left endpoint) for $p = 2$ to *testar5* produces estimates:

$$\hat{\phi}_1 = 1.300 \quad \hat{\phi}_2 = -0.700 \quad \hat{a} = 1.010$$

whereas applying it to the contaminated $\{Y_t\}$ data we obtain:

$$\hat{\phi}_1 = 0.069 \quad \hat{\phi}_2 = 0.019 \quad \hat{a} = -17862.83.$$

Clearly the contamination has destroyed the structure that the ordinary LP estimator is designed to detect. However, by applying the robust approach of Section 2.2 we obtain the following table.

Parameter	phi 1	phi 2	intercept	objective
Complete	0.06874115	0.01869025	-17862.832492	-15.1904368
Remove 1p	0.61776376	-0.39346766	-8491.514521	-7.0383953
Remove 2p	0.54820333	-0.22656304	-6165.655764	-4.9517718
Remove 3p	0.56094262	-0.23389676	-5807.283135	-4.6398543
Remove 4p	0.84151922	-0.43806125	-3135.604547	-2.2783873
Remove 5p	0.88760179	-0.46876247	-2499.568627	-1.7190120
Remove 6p	0.85136500	-0.45508233	-1934.530351	-1.2582981
Remove 7p	1.08747261	-0.54213733	-1226.596428	-0.5037581
Remove 8p	1.00948201	-0.52423239	-948.564440	-0.3260463
Remove 9p	1.18667011	-0.60409521	-398.300346	0.2419134
Remove 10p	1.25498991	-0.66743968	-135.376219	0.4717646
Remove 11p	1.30000716	-0.69999965	1.001867	0.6008644
Remove 12p	1.30000867	-0.69999613	1.038484	0.6009007
Remove 13p	1.30000436	-0.69998809	1.036626	0.6009029
Remove 14p	1.30000928	-0.69998891	1.035657	0.6009062
Remove 15p	1.30001042	-0.69997061	1.087074	0.6009696
Remove 16p	1.30003501	-0.69998171	1.094979	0.6009898
Remove 17p	1.30004242	-0.69998661	1.123277	0.6010165
Remove 18p	1.30006577	-0.69999983	1.130155	0.6010325
Remove 19p	1.30005763	-0.69998747	1.145117	0.6010496
Remove 20p	1.30006141	-0.69998558	1.153146	0.6010621

It is clear that the robust estimates of the parameters are given after 11 sets of constraints (22 constraints in all) have been removed, and that they correspond to values very close to the ones obtained from the uncontaminated series.

Comparing this analysis with what is possible from the Yule-Walker estimation procedure, we first note that the (heavy tailed) YW estimates of (ϕ_1, ϕ_2) for the contaminated $\{Y_t\}$ series are (0.900,-0.335): these are very different to those for the uncontaminated *testar5* series — (1.288,-0.689). However, we have no clearcut way of detecting the source of the discrepancy when only observing the contaminated series! The YW estimates, based on the autocorrelations does not provide us with a straightforward robust alternative.

6. Data example. This section analyzes several real telecommunications data sets and shows some of the limitations of autoregressive and sample acf based methods. Our strategy for analysis of the data is as follows.

- Plot the data and confirm that heavy tailed analysis is appropriate through qq and Hill plots.
- Check for independence of the data using techniques described in Section 4. If we are doing acf analysis, we use permutation tests based on the heavy tailed sample acf and also inspect the sample acf plot. If we are doing analysis based on the lp estimator, we use the permutation test based on the maximum estimated autoregressive coefficient and also the asymptotic test of independence given in Section 4.2.
- Assuming the data is not independent, we try to model the data as an autoregression. The order p must be determined. If we do acf analysis, we may use the pacf plot and the AIC plot to fix the order p and we may also try to fit successively larger models using Yule-Walker estimation until we find a model whose residuals pass the independence tests of Section 4. If we are doing lp analysis, we fit AR models using lp estimation with positive innovations (or perhaps with intercepts or with estimation of the left endpoint) until estimated residuals pass independence tests.
- If the result of this model fitting exercise (with p ranging say from 1 to 15) produces an acceptable model then we are done and autoregressive modelling has provisionally been successful. However, stability over subsets of various statistics like the sample acf and the lp coefficient estimates should be examined as an additional diagnostic to ensure no undetected anomalies remain.
- If the AR model fitting fails, other models have to be considered. Nonlinear models, mixture models and hidden Markov chain models are some possibilities.

6.1. SILENCE. Consider a time series of length 1026 shown in Figure 6.1 which represents the off periods between transmission of packets generated by a terminal during a logged-on session. (See also Resnick and Stărică (1995), Resnick (1996).) The left graph in Figure 6.1 is the time series plot and the right graph is the static qq-plot of the logarithm of the data using 500 upper order statistics giving ample evidence of heavy tails. The estimate of α given by this plot is 0.67.

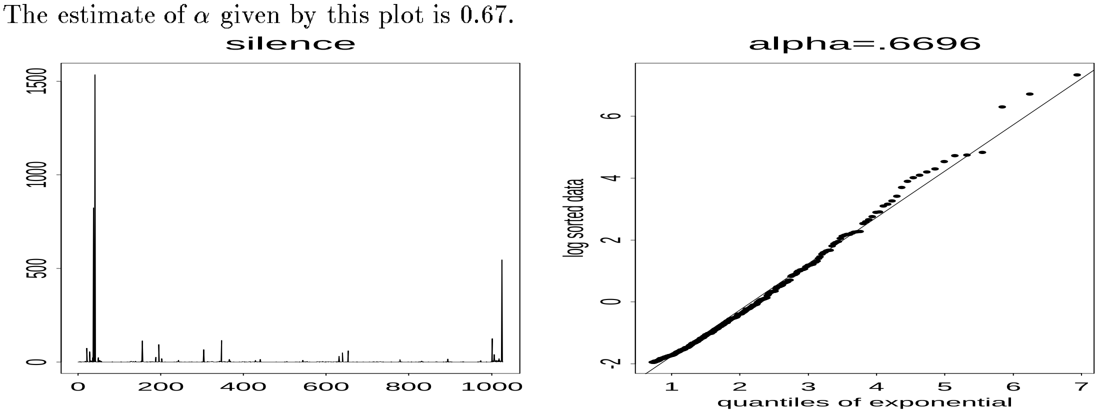

Figure 6.1. Tsplot and static qq-plot of SILENCE.

The Hill plots in Figure 6.2 give a similar estimate of α, namely $\alpha = .64$ and this is the estimate we adopt. The upper left plot is the ordinary Hill plot of $\{(k, H_{k,n}^{-1}), 1 \le k \le n\}$ and the upper right plot is the Hill plot in *alt* scale $\{(\theta, H_{\lceil n^\theta \rceil, n}^{-1}), 0 \le \theta \le 1\}$. The lower left graph provides smoothing in *alt* scale and the lower right plot overlays two plots. See Resnick and Stărică (1996) for more information on these plots.

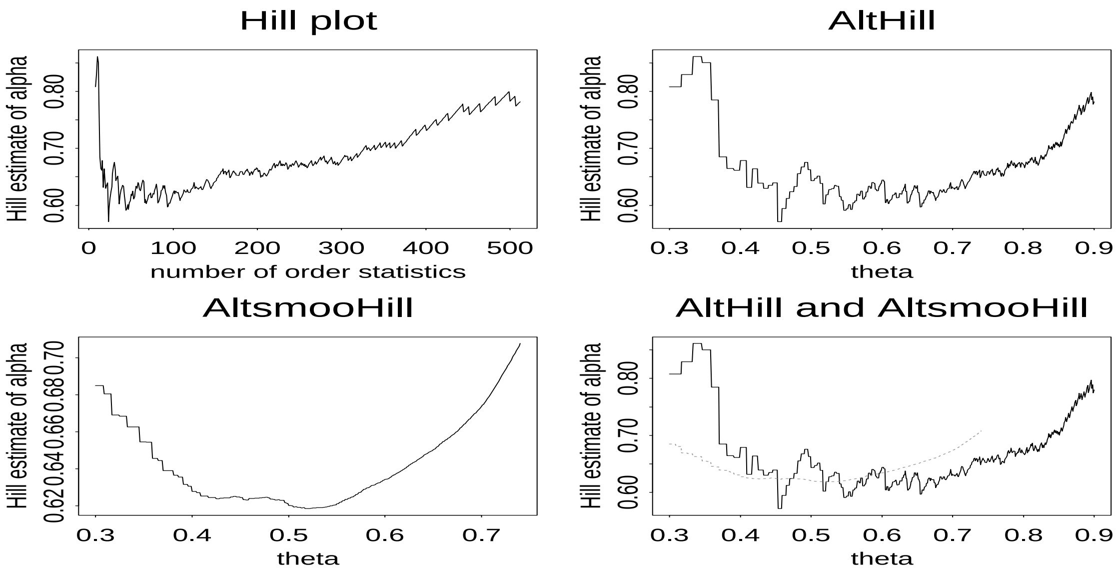

Figure 6.2. Hill plots of SILENCE.

6.1.1. ACF analysis. We now apply sample autocorrelation based analysis to try to model the data. First we applied the permutation test based on the heavy tailed sample acf (up to lags of 25) and this rejected independence with a p-value of 0.03. Successive runs of this test supplied p-values of 0.02, 0.01 0.03. So it seems the data cannot be modeled as iid. Some indication of trouble when using sample autocorrelation based methods shows up quickly when one applies the permutation test to SILENCE with the first 100 data excluded. The p-value is now 0.33 and thus at a reasonable level one would not reject independence for these data. This anomaly indicates that the very large values clustering between serial indices 30 and 50 are inordinately influencing the sample acf. The acf also exhibits lack of stability over subsets as is clear from plotting the acf for the data split in halves or thirds (Resnick, 1996).

Despite these reservations, we proceed with sample acf analysis. Figure 6.1.1 displays the heavy tailed acf and pacf plots. From the pacf, an AR(9) model is an attractive possibility and this choice is confirmed by the heavy tailed AIC plot in Figure 6.1.2 which has its minimum at lag 9.

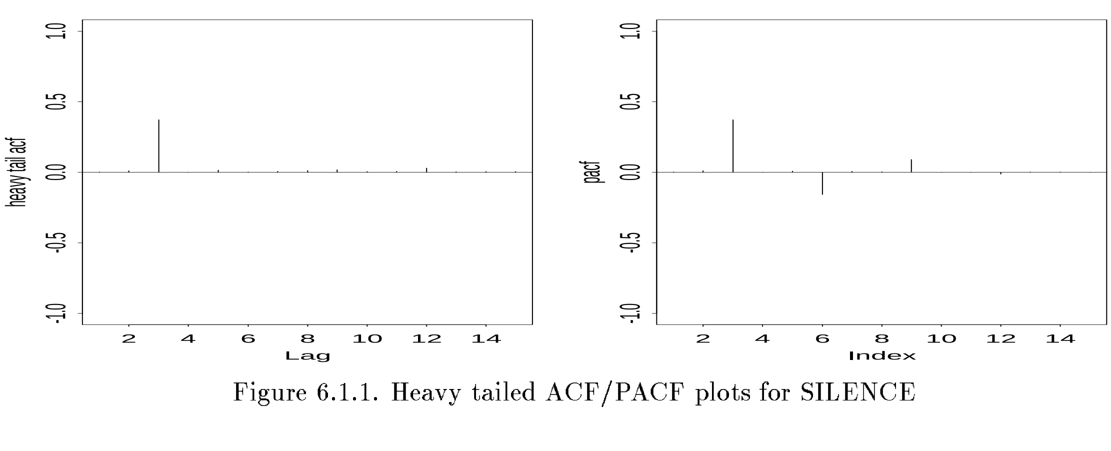

Figure 6.1.1. Heavy tailed ACF/PACF plots for SILENCE

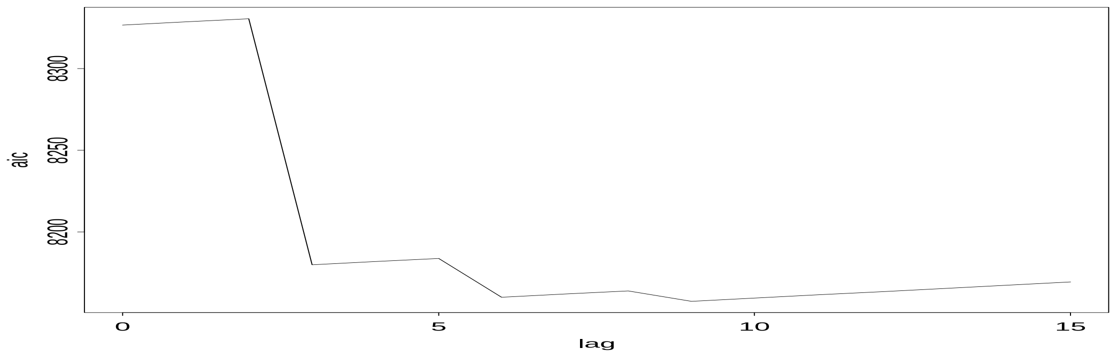

Figure 6.1.2. AIC plot for SILENCE

Fitting an AR(9) to the data using the Yule-Walker estimators produces the following autoregressive coefficients

$$\hat{\phi}_{YW} = \begin{pmatrix} -0.0004836027 \\ 0.0069079981 \\ 0.4442829274 \\ -0.0007584241 \\ 0.0064452867 \\ -0.1961397316 \\ 0.0058430780 \\ 0.0046091042 \\ 0.0901514058 \end{pmatrix}.$$

The permutation test based on the sample acf and using 25 lags when applied to the residuals has a p–values 0.36 and hence fails to reject the hypothesis of independence. Successive runs of this test produced p-values of 0.24, 0.31 and 0.34. A plot of the residuals is given in Figure 6.1.3.

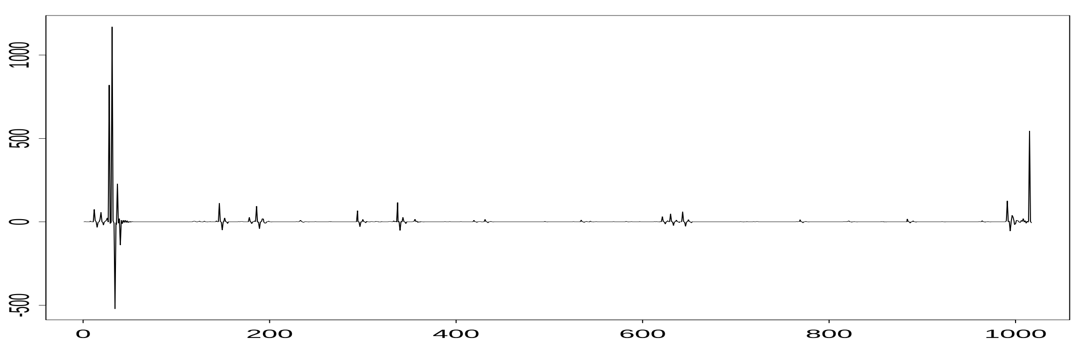

Figure 6.1.3. Tsplot of the SILENCE residuals

Yule-Walker analysis is designed to make the sample acf of the residuals look good and it is no surprise that the plot (not shown) of the sample acf and pacf of the residuals reveals no obvious problems. However, one's satisfaction is tempered by the lack of stability of the heavy tailed sample acf of the residuals across subsets; the acf plot of the first 400 residuals and the acf plot for residuals with serial index 600 to 1000 are quite different and are pictured in Figure 6.1.4.

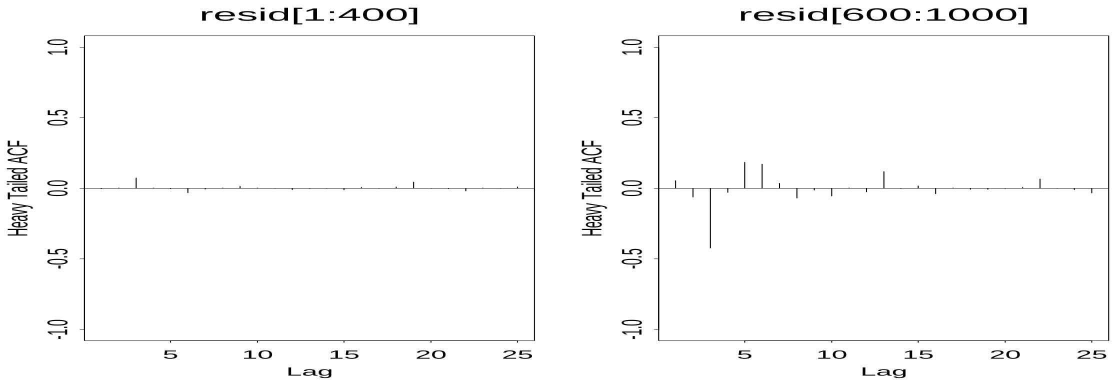

Figure 6.1.4. Lack of stability in acf of residuals

Somewhat suspicious, we decided to investigate further whether the residuals really are independent. Call the residuals $\xi_1, \dots, \xi_{1017}$. If the residuals were independent, then according to Theorem 3.3 of Davis and Resnick (1986), we would have

$$\lim_{n\to\infty} P[\tilde{b}_n^{-1} b_n^2 \frac{\sum_{t=1}^{n-h} \xi_t \xi_{t+h}}{\sum_{t=1}^{n} \xi_t^2} \le x] = P[U/V \le x] \tag{6.1}$$

where U is a stable random variable with index $\alpha = .64$ and V is a positive stable random variable with index $\alpha/2 = .32$ and b_n is the solution to

$$P[|\xi_1| > x] = 1/n$$

and $\tilde{b}_n$ is the solution to

$$P[|\xi_1 \xi_2| > x] = 1/n.$$

Thus an approximate symmetric 95% confidence window for the heavy tailed sample correlations would be placed at $\pm l\tilde{b}_n/b_n^2$ where l satisfies

$$P[|U/V| \le l] = .95.$$

We estimated the 95%-quantile of $|U/V|$ by simulation obtaining 163.48 and assuming the distribution of ξ_i's to be Pareto we find

$$l\frac{\tilde{b}_n}{b_n^2} = l\alpha^{1/\alpha} \frac{n^{-1/\alpha}}{\log n}$$

which when computed yields the value 0.000235 The values of the first 15 heavy tailed sample acf's for the residuals are

$$\begin{pmatrix} \hat{\rho}_H(1) \\ \vdots \\ \hat{\rho}_H(15) \end{pmatrix} = \begin{pmatrix} -0.0003398865 \\ -0.0002235024 \\ 0.0599596189 \\ 0.0002053652 \\ 0.0001518907 \\ -0.0271219926 \\ -0.0017466411 \\ 0.0010538758 \\ 0.0133693048 \\ 0.0015380144 \\ -0.0004159059 \\ -0.0084004526 \\ 0.0007366633 \\ 0.0021101667 \\ -0.0128681139 \end{pmatrix}$$

making the hypothesis that the residuals are independent implausible.

A note on the simulation of the quantile of U/V. In the notation of Samorodnitsky and Taqqu (1994)

$$U = \Big(\cos(\frac{\pi\alpha}{2})\Gamma(1-\alpha)\Big)^{1/\alpha} S_\alpha(1,1,0)$$
$$V = \Big(\cos(\frac{\pi\alpha}{4})\Gamma(1-\frac{\alpha}{2})\Big)^{2/\alpha} S_{\alpha/2}(1,1,0).$$

As pointed out to us by Gennady Samorodnitsky and John Nolan, care must be taken when simulating stable random variables with Splus since Splus uses a different parameterization which produces simulations of random variables with symmetry parameter 1 which may be negative. The connection is that for the Splus function rstab

$$\text{rstab}(\alpha,1) + \tan(\pi\frac{\alpha}{2}) = S_\alpha(1,1,0).$$

We also tested the residuals for independence using the standard tests of randomness outlined in Section 4.3 using Brockwell and Davis' ITSM. The results were as follows:

Turning points	837	AN(676.67, 13.43^2)
Difference sign	507	AN(508, 9.21^2)
Rank test	260158	AN(258318,16228.22^2).

The turning points test was significant and rejected independence.

6.1.2. LP analysis. We now run an analysis parallel to what was provided by the sample acf. First we run the asymptotic test (Section 4.2) of independence at the 0.05 level on the data. Figure 6.1.5 shows the results of applying this test which rejects independence.

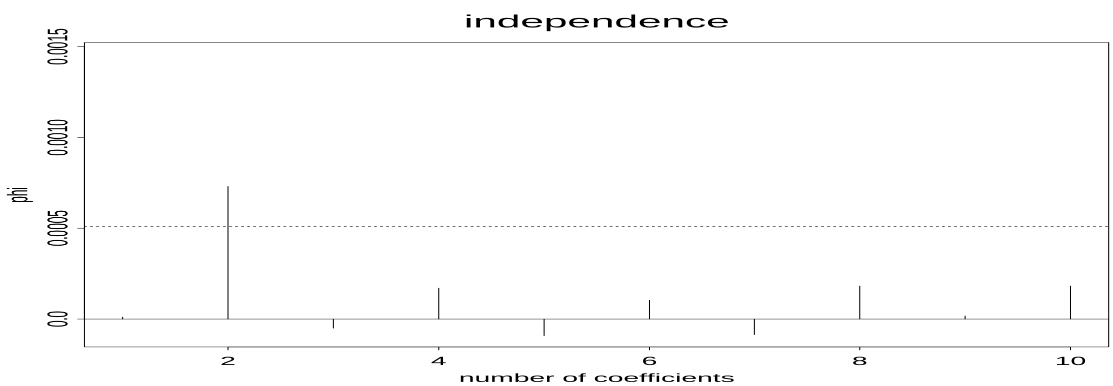

Figure 6.1.5. LP independence test of SILENCE

Trying to fit an AR(9) using the linear programming estimates requires assuming positive innovations and yields rather small estimates for the coefficients:

$$\hat{\phi}_{LP} = \begin{pmatrix} 0.0000227 \\ 0.0007290 \\ -0.0000504 \\ 0.0001177 \\ -0.0000881 \\ 0.0001042 \\ 0.0000116 \\ 0.0001804 \\ 0.0000166 \end{pmatrix}.$$

Care must be taken when applying the asymptotic test of independence to estimated residuals since $\hat{c} = n^{-1}\sum_{i=1}^{n} X_i^{-\hat{\alpha}}$ will become infinite due to the p values of the estimated residuals which are approximately 0. Consequently, the test was performed using only the nonzero residuals. It rejected at level 0.05 as shown in Figure 6.1.6.

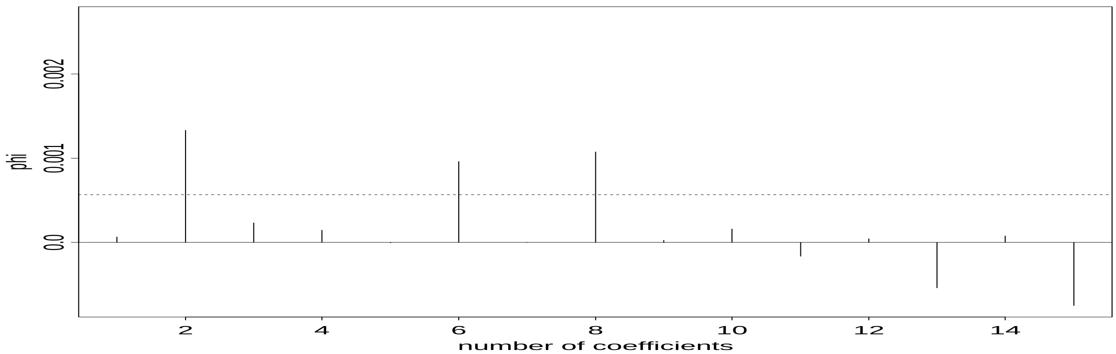

Figure 6.1.6. LP independence test of SILENCE residuals

We also ran the permutation test based on lp estimation discussed in Section 4.5 using 10 coefficients which produced insignificant p-values of 0.11, 0.16 and 0.19. The permutation tests do not reject independence of the residuals conclusively even though the asymptotic test does. The heavy tailed acf plot (not shown) has big spikes, particularly at lag 3 which is of size 0.37, so asymptotic theory based on the acf when applied to the lp residuals would certainly reject the hypothesis of independence of these residuals.

This analysis was repeated using the linear programming estimates with an intercept to fit an AR(9). The intercept was quite small, namely 0.008653932. With a small intercept we would not expect a change in the conclusions obtained previously without an intercept and this was indeed the case.

We next investigated whether robustifying the procedure would produce better results. We considered whether to remove large observations on the basis that they could be contaminated outliers but based on our experience with acf analysis we suspected that removing large observations would leave data that could be modelled as iid and this turned out to be the case. For example, considering the 1015 observations which are less than 50 produces a new data set, called SHORTSILENCE, which passes the asymptotic independence test at level .05 as shown in Figure 6.1.7.

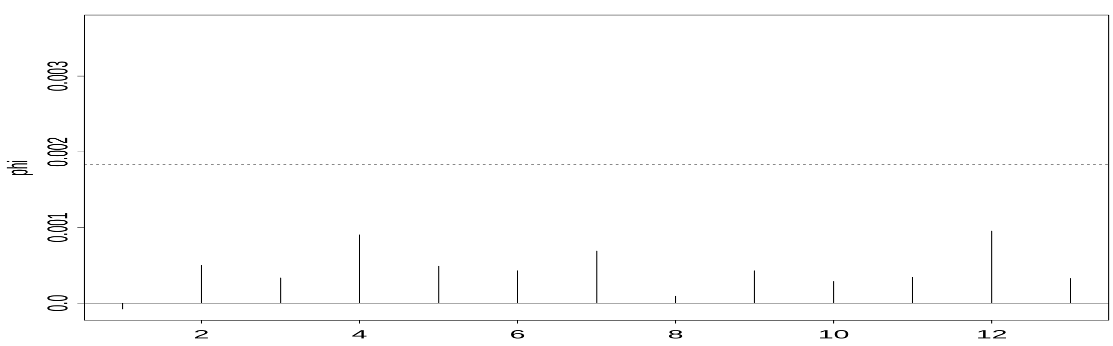

Figure 6.1.7. LP independence test of SHORTSILENCE

We also tried removing successive sets of p (in this case $p = 9$) active constraints and checking the stability which would have made us confident that contaminated data had been removed. However, this stability is hard to discern. We display the results of removing $0p, 1p, \ldots, 5p$ sets of constraints in the next table; the bottom row gives the value of the objective function. Based on our experience with SHORTSILENCE, we did not proceed further than removing $5p$ constraints.

	Remove $0p$	Remove $1p$	Remove $2p$	Remove $3p$	Remove $4p$	Remove $5p$
ϕ_1	0.000022	0.000096	0.000135	−0.000205	0.000226	0.000288
ϕ_2	0.000729	0.001937	0.002190	0.002287	0.002884	0.002965
ϕ_3	−0.000050	0.000165	0.000091	0.000194	0.000505	0.000641
ϕ_4	0.000117	0.000315	0.000226	0.000976	0.001236	0.001692
ϕ_5	−0.000088	0.000357	0.000381	0.000402	0.000454	0.000550
ϕ_6	0.000104	0.000358	0.000444	0.000872	0.001010	0.001391
ϕ_7	0.000011	0.000322	0.000570	0.000915	0.001304	0.001387
ϕ_8	0.000180	0.000221	0.001256	0.001341	0.001379	0.001652
ϕ_9	0.000016	0.000390	0.000505	0.000426	0.000666	0.000962
object	0.001044	0.004165	0.005803	0.007211	0.009668	0.011531

It is a difficult to feel comfortable with the fit of the AR model. Based on our experience with removing the first 100 observations and also with SHORTSILENCE, it seems likely that a mixture model is a better candidate for fitting SILENCE. This might also conform to the physical reality of a terminal operator continuously generating traffic except during interruptions such as when the operator is called from the room or engaged by a distraction preventing operation of the terminal.

References

Bhansali, R., *Consistent order determination for processes with infinite variance*, JRSS B **50** (1988), 46–60.

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

Brockwell, P. and Davis, R., *ITSM: An Interactive Time Series Modelling Package for the PC*, 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, C., 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., Knight, K. and Liu, J., *M estimation for autoregressions with infinite variance*, Stoch. Proc and their Appl. (1991).

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.ps.Z at http:www.orie.cornell.edu/trlist/trlist.html (1995), (to appear)Annals of Applied Probability.

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*, Advances in Applied Probability; forthcoming (1995).

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

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.

Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics, Vol. 3*, Griffin, London, 1976.

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

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

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.

Mikosch,T., Gadrich, T., Klüppelberg, C., and Adler, R., *Parameter estimation for ARMA models with infinite variance innovations*, Ann. Statist. **23** (1995), 305–326.

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

Resnick, S., *Heavy tail modelling and teletraffic data* Available as TR1134.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1995), (to appear)Ann. Statist..

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

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

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

Resnick, S. and Stărică, Cătălin, *Tail Index estimation for dependent data*, Available as TR1174.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html, Preprint (1997).

Rootzen, H., Leadbetter, M. and de Haan, L., *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 (1990).

Samorodnitsky, G. and Taqqu, M., *Stable Non–Gaussian Random Processes*, Chapman and Hall, New York, 1994.

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).

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, Rhodes Hall 223 Ithaca, NY 14853 USA

E-mail: sid@orie.cornell.edu