HEAVY TAIL MODELING AND TELETRAFFIC DATA

Sidney I. Resnick

Cornell University

Abstract. Huge data sets from the teletraffic industry exhibit many non-standard characteristics such as heavy tails and long range dependence. Various estimation methods for heavy tailed time series with positive innovations are reviewed. These include parameter estimation and model identification methods for autoregressions and moving averages. Parameter estimation methods include those of Yule-Walker and the linear programming estimators of Feigin and Resnick as well estimators for tail heaviness such as the Hill estimator and the qq-estimator. Examples are given using call holding data and inter-arrivals between packet transmissions on a computer network. The limit theory makes heavy use of point process techniques and random set theory.

1. Introduction

Classical queuing and network stochastic models contain simplifying assumptions guaranteeing the Markov property and insuring analytical tractability. Frequently inter-arrivals and service times are assumed to be iid and typically underlying distributions are derived from operations on exponential distributions. At a minimum, underlying distributions are usually assumed nice enough that moments are finite.

Increasing instrumentation of teletraffic networks has made possible the acquisition of large amounts of data. Analysis of this data is disturbing since there is strong evidence that the classical queuing assumptions of thin tails and independence are inappropriate for this data. Video conference data and packet counts per unit time in Ethernet traffic appear to exhibit long range dependence and self similarity (Beran, et al, 1995; Beran, 1995; Willinger, Taqqu, Leland and Wilson, 1995) while such phenomena as file lengths, cpu time to complete a job, call holding times, inter-arrival times between packets in a network and lengths of on/off cycles appear to be generated by distributions which have heavy tails (Duffy, et al 1993, 1994; Meier–Hellstern et al, 1991; Willinger, Taqqu, Sherman and Wilson, 1995).

Although we will not dwell heavily on long range dependence in this survey, it is instructive to quickly view in Figure 1.1 a video conferencing data set in excess of 48,000 data. The top graph is the time series plot and the bottom plot is the sample autocorrelation function (acf), graphed out to 700 lags. Usual short range dependent data sets would show a sample correlation function dying after only a few lags and then persisting within the magic 95% confidence window determined by Bartlett's formula (see for example Brockwell and Davis (1991). However the video-conference data has an acf showing significant correlations for hundreds of lags. This data also exhibits self–similarity; see Beran et al (1995), Garrett and Willinger (1994).

It is logical and natural to question whether departures from classical queuing assumptions matter that much. There is persistent skepticism about the importance of correctly detecting heavy tails or long range dependence which is founded on unverified faith in the robustness of classical assumptions and perhaps, understandably, a bit of inertia. Also, mathematical analysts have been slow to embrace these features

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

Sidney Resnick was partially supported by NSF Grant DMS-9400535 at Cornell University and also received some support from NSA Grant MDA904-95-H-1036. This paper is partly based on the Statistics Society of Canada Presidential address given in Montreal, Canada in July of 1995 and special thanks to President R. James Tomkins and the SSC for the invitation to speak and for the encouragement to organize my thoughts on the subject of heavy tailed modeling .

since queues and networks with such complex inputs are difficult to analyze mathematically. However there is simulation evidence (for example see Livny, Melamed and Tsiolis, 1993) that correctly accounting for dependence and correctly modeling tail heaviness is crucial. Resnick and Samorodnitsky (1995) verify that for a simple G/M/1 queue, a stationary input with long range dependence can induce heavy tails for the waiting time distributions and for the distribution of the number in the system. The inputs in the Resnick and Samorodnitsky model are of the form of inter-arrivals $\{T_n\}$ of customers where $\{T_n\}$ is stationary with a form of long range dependence. The dramatic degradation of system performance exemplified by heavy tailed outputs compared with what would be predicted by a model satisfying classical assumptions indicates that the phenomena of long range dependence and heavy tails cannot be ignored if estimation of network capacity is the goal. This point will undoubtedly receive more systematic study in the near future and it is expected that long range dependence and heavy tails will be of critical importance for even crude first order solutions of and guidelines for a wide range of network engineering problems.

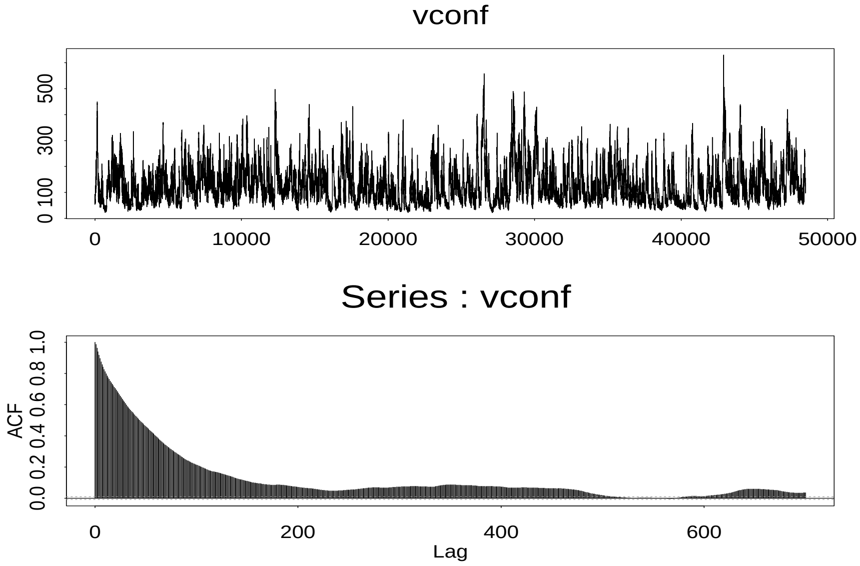

Figure 1.1

It is becoming increasingly clear that there are strong connections between long range dependence and heavy tails but these connections have not yet been systematically explored. The Resnick and Samorodnitsky (1995) paper suggests that long range dependent inputs to a queuing system can induce heavy tailed outputs. It is also clear that heavy tails can induce long range dependence. Consider the following example of an on/off model for packet transmission of a source and destination pair fashioned after one described in Willinger, Taqqu, Sherman, Wilson (1995) and discussed in Heath, Resnick, Samorodnitsky (1996): We have a stationary alternating renewal process (Resnick, 1992) with counting function $\{N(t), t \geq 0\}$ and renewal times $\{S_n, n \geq 0\}$. A renewal interval consists of an *on* period which has distribution F_{on} and an *off* period with distribution F_{off}. Let the means of F_{on} and F_{off} be μ_{on} and μ_{off} respectively. Set $\mu = \mu_{\text{on}} + \mu_{\text{off}}$. The process $\{N(t)\}$ can be constructed to be stationary by flipping a coin to decide if the process starts in an *on* or *off* period, and then choosing the right residual distribution. The coin chooses *on* with probability $\mu_{\text{on}}/(\mu_{\text{on}} + \mu_{\text{off}})$. If the coin indicates for example *on*, then after the initial residual period, we alternate independent *off* periods with independent *on* periods. $N(t)$ is the number of points corresponding to termination of *off* periods in $[0, t]$. Next, we may define the stationary process $\{Z_t, t \geq 0\}$ so that $Z_t = 1$ iff t is in an *on* period. If

$$1 - F_{\text{on}}(t) = t^{-\alpha} L(t), \ t \to \infty, 1 < \alpha < 2$$

$$1 - F_{\text{off}}(t) = o\Big(1 - F_{\text{on}}(t)\Big), \ t \to \infty,$$

then

$$\mathrm{Cov}(Z_0, Z_t) \sim \frac{\mu_{\mathrm{off}}^2}{(\alpha-1)\mu^3} t^{-(\alpha-1)} L(t), \; t \to \infty.$$

The slow rate of decay of the covariance function is characteristic of long range dependence.

The effects of heavy tails in the inputs to an associated reservoir model are quite dramatic. Suppose the alternating renewal process controls inputs to the reservoir; water flows into the reservoir at unit rate in *on* period and evaporates at rate $r > \mu_{\mathrm{on}}/\mu$ during *off* periods. Let $X(t)$ be the contents of the reservoir at time t. Then because of the regenerative nature of $\{X(t)\}$ we have

$$X(t) \Rightarrow X(\infty)$$

and it follows that the limit distribution of $X(\infty)$ is heavy tailed

$$P[X(\infty) > x] \sim (const) x^{-(\alpha-1)} L(x).$$

These results follow by standard results in the theory of regularly varying functions; see Resnick (1987), de Haan (1970), Bingham et al (1987), Geluk and de Haan (1987).

This model is a simplified idealization but effectively illustrates the simple way in which heavy tails can induce long range dependence. Traffic on an Ethernet system can be considered as a superposition of traffic from many source-destination pairs. Willinger et al (1995) show how superimposing independent copies of the Z_t process and taking limits in a suitable manner yields fractional Brownian Motion (Samorodnitsky and Taqqu, 1994) as a limit. This offers one plausible explanation for the observed self–similar nature of Ethernet traffic.

This example serves as a reminder that for modeling the data the following choices must be faced:

- Is it better to use structural models such as on/off models where specific physical features are accounted for? Structural models which successfully capture relevant features have immediate connections to applications. Here the engineering comfort level may be high and there may be some possibility of analytical work succeeding if such physical models are used as model inputs.
- Or should we use black box models from time series analysis? Such models ignore physical structure but have long tradition in data analysis. They may be easier for analysts who do not have the close ties to engineers required to intelligently construct structural models. Furthermore, such models may be applicable across a broad range of disciplines.

If we use structural models there is no guarantee we have correctly summarized relevant features. Statistical verification of goodness of fit can be challenging. The probability analysis may be easier but the statistics may be harder. Time series models have an advantage of being applicable across a broad range of disciplines. However if we use time series models we have to decide whether to use linear or non–linear models. Linear models are the simplest but there is no guarantee that in the heavy tailed world they form a large flexible class. In traditional, finite variance time series modeling based on Hilbert space methodology, the Yule–Walker estimators guarantee that any correlation structure may be mimicked out to any fixed number of lags by autoregressions of a suitable order (see Brockwell and Davis (1991), page 240) and in this very limited sense linear models are sufficient for data analysis. For infinite variance models, we have no such confidence that linear models are adequately flexible and in fact the theoretical perspective offered by Rosinski's (1995) work as well as some data experience (Resnick, 1996; Feigin and Resnick, 1996) make this unlikely.

This is an important point because evidence is emerging that for heavy tailed modeling there are significant differences in the behavior of key summary statistics depending on whether the model is linear or non–linear. For example consider the sample autocorrelation function (acf). In the heavy tailed case it is best to define it as follows: If the stationary time series is $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ then define for $h = 0, 1, \dots$

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

This substitutes for the usual correlation function in finite variance time series modeling. If the model is linear, the sample acf at lag h converges in probability to a constant depending on h (Davis and Resnick (1985a,b, 1986)) but if the model is non–linear, Davis and Resnick (1995) show that the sample acf at lag h may converge in distribution to a non–degenerate random variable depending on h.

Sets of data displaying characteristics of heavy tails are encountered in diverse fields other than teletraffic engineering, for example in hydrology (Gumbel, 1958; Castillo, 1988), economics and finance (Koedijk, Schafgans and de Vries 1990; Janson and de Vries, 1991) reliability and structural engineering (Grigoriu,1995). Because of the diversity of applications and because heavy tails are one of the causes of long range dependence, we expect interest in the detection and modeling of heavy tailed phenomena to grow and hope the survey which follows will contribute to understanding the uses and limitations of the growing body of techniques.

Section 2 discusses some mathematical background of regular variation and specifies the type of heavy tailed models we will study. Section 3 focuses on the obvious steps in heavy tailed modeling: detecting heavy tails and detecting dependencies. We describe various graphical techniques of an exploratory nature which can be helpful but point out some limitations. Heavy tailed autoregressive models have been successfully studied and relevant model selection and estimation methods are summarized in Section 4 which also contains some brief remarks on the bootstrap. An example is provided in Section 5. Section 6 contains some closing remarks.

2. Background for heavy tailed models

2.1. Regular variation. What is a heavy tail? The bulk of statistical work deals with light tails which decay exponentially fast as typified by the normal distribution tail, since by Mills' ration, as $x \to \infty$

$$P[N > x] \sim \frac{n(x)}{x} = \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x} \to 0.$$

Contrast this with a typical heavy tail such as possessed by the Pareto distribution: X has a Pareto tail with index $\alpha > 0$ if for $x > 0$

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

More generally, we say X has a heavy tailed distribution F if

$$P[X > x] = x^{-\alpha} L(x) \tag{2.1}$$

where L is slowly varying; that is for $x > 0$

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1. \tag{2.2}$$

Note that

$$\begin{aligned} E(X^\beta) &< \infty, \quad \beta < \alpha, \\ E(X^\beta) &= \infty, \quad \beta > \alpha. \end{aligned}$$

Typical slowly varying functions include the following examples

$$\begin{aligned} L(x) =& c + o(1), \quad x > 0, \\ =& \log x, \quad x > 1, \\ =& \log(\log x) \quad x \text{ large}, \\ =& 1/\log x, \quad x > 1. \end{aligned}$$

In the first example of L where $L(x) = c + o(1)$ the term $o(1)$ may look innocent and harmless but can wreak havoc and there can be a big difference between detecting Pareto tails and detecting, say, tails of stable distributions (Samorodnitsky and Taqqu, 1994) where

$$1 - F(x) \sim x^{-\alpha}, \quad x \to \infty.$$

This point is illustrated by examples in Section 3.1.

Another way to express (2.1) is to say that $1 - F$ is regularly varying with index $-\alpha$: A distribution F concentrating on $[0, \infty)$ has a tail $1 - F(x)$ which is regularly varying with index $-\alpha$, $\alpha > 0$ (written $1 - F \in RV_{-\alpha}$) if

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \quad x > 0. \tag{2.3}$$

The distribution tail $1 - F$ is second order regularly varying with first order parameter $-\alpha$ and second order parameter ρ (written $1 - F \in 2RV(-\alpha, \rho)$) if there exists a function $A(t) \to 0$, $t \to \infty$ which ultimately has constant sign such that the following refinement of (2.3) holds:

$$\lim_{t \to \infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)} = H(x) := cx^{-\alpha} \int_1^x u^{\rho - 1} du, \quad x > 0, \tag{2.4}$$

for $c \neq 0$. Note that for $x > 0$

$$H(x) = \begin{cases} cx^{-\alpha} \log x, & \text{if } \rho = 0, \\ cx^{-\alpha} \frac{x^\rho - 1}{\rho}, & \text{if } \rho < 0. \end{cases}$$

It follows that $|A| \in RV_\rho$ and no other choices of ρ are consistent with $A(t) \to 0$. See de Haan and Stadtmüller (1995), Geluk and de Haan (1987).

Regular variation is the basic analytic theory underlying extreme value theory and stable processes (de Haan, 1970; Geluk and de Haan, 1987; Bingham et al, 1987; Resnick, 1987; Samorodnitsky and Taqqu, 1994). Second order regular variation has proven very useful and natural for establishing asymptotic normality of extreme value statistics (see the discussion in Section 3 on the Hill estimator) and also for the study of rates of convergence to extreme value and stable distributions (de Haan and Resnick, 1995; de Haan and Peng, 1995a,b,c; Smith, 1982).

There is a well developed technique for integrating regularly varying functions called Karamata's theorem which roughly says that when one integrates a regularly varying function, one may treat the slowly varying function as a constant.

2.2. Point Processes and Random Measures. Mathematical and asymptotic properties of heavy tailed models are analyzed with heavy reliance on weak convergence theory and point process techniques. Occasionally mildly exotic items like weak convergence of random closed sets are necessary (Feigin and Resnick, 1994). The central role of point processes for the mathematical analysis of heavy tailed phenomena is well documented (Resnick, 1986, 1987, 1991) and will not be emphasized here beyond some reminders about notation and one central result. In what follows $M_+(E)$ is the set of positive Radon measures on a nice locally compact space E; $M_p(E)$ is the set of point measures. $M_+(E)$ is metrized by the vague metric (cf. Kallenberg, 1983; Resnick, 1987; Neveu, 1976). We denote weak convergence of random elements or probability measures by "$\Rightarrow$" and "$\xrightarrow{v}$" denotes vague convergence of measures in $M_+(E)$. For $x \in E$ and $A \subset E$ define

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

A Radon point measure with points in E is denoted $\sum_i \epsilon_{x_i}$. The collection of all such point measures is $M_p(E)$. The following two results giving weak convergence results for special random measures provides the necessary theoretical background.

Proposition 2.1. *Let $m = m(n)$ be a sequence satisfying $m \to \infty$, $m/n \to 0$ as $n \to \infty$.*

(a) Suppose for each n that $\{Z_j(n), j \ge 1\}$ are iid random elements of the space E. The sequence of point processes

$$\sum_{j=1}^{n} \epsilon_{Z_j(n)}, \quad n = 1, 2, \dots$$

converges weakly to a limiting Poisson process on E with mean measure ν iff

$$E\left(\sum_{j=1}^{n} \epsilon_{Z_j(n)}(\cdot)\right) = nP[Z_1(n) \in \cdot] \xrightarrow{v} \nu(\cdot).$$

(b) Suppose for each n that $\{Y_j(n), j \ge 1\}$ are iid random elements of the space E. The sequence of random measures

$$\frac{m}{n}\sum_{j=1}^{n} \epsilon_{Y_j(n)}, \quad n = 1, 2, \dots$$

converges weakly to a limiting non-random measure $\nu \in M_+(E)$, iff

$$E\left(\frac{m}{n}\sum_{j=1}^{n} \epsilon_{Y_j(n)}(\cdot)\right) = mP[Y_1(n) \in \cdot] \xrightarrow{v} \nu.$$

Proof. Part (a) is Proposition 3.21, page 154 in Resnick (1987) and part (b) is 3.57, page 161, Resnick (1987) and is also given in Resnick (1986).

An application of Proposition 2.1 of immediate interest to analysis of heavy tailed phenomena is as follows: Let $\{Z_j, j \ge 1\}$ be iid and non-negative with common distribution F. Set $E = (0, \infty]$, $b(t) = F^{\leftarrow}(1 - 1/t)$ where $F^{\leftarrow}$ is the left continuous inverse of the monotone function F, $Z_j(n) = Z_j/b(n)$, $Y_j(n) = Z_j/b(m)$ and $m = m(n)$ is a sequence satisfying $m \to \infty$, $m/n \to 0$ as $n \to \infty$. Also, the measure ν is given by $\nu((x, \infty]) = x^{-\alpha}$. We then have the following equivalences to the regular variation condition given in (2.3):

(1) We have in $M_+(E)$

$$nP[\frac{Z_1}{b(n)} \in \cdot] \xrightarrow{v} \nu.$$

(2) In $M_+(E)$

$$N_n := \sum_{j=1}^{n} \epsilon_{Z_j/b(n)} \Rightarrow N_\infty,$$

where N_∞ is a Poisson process on E with mean measure ν.

(3) In $M_+(E)$

$$\nu_n := \frac{m}{n}\sum_{j=1}^{n} \epsilon_{Z_j/b(m)} \Rightarrow \nu.$$

2.3. Heavy Tailed Time Series Models. We will try to model data with linear time series models but keep in mind that although this class of models is relatively simple, there is no guarantee that it is adequate. In particular this choice of class excludes non-linear models and such random coefficient models as ARCH and GARCH which are dearly beloved in economics. Linear time series models is familiar territory but one often encounters the fact that many statistical techniques are tried in the heavy tailed domain because they work in the finite variance arena and this is fairly weak justification.

We will deal with moving average processes of order infinity, written MA(∞). These are specified as follows: Let $\{Z_t, -\infty < t < \infty\}$ be iid and non-negative with common distribution F satisfying (2.3). For suitable constants $\{c_j\}$ define

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \quad t = 0, \pm 1, \dots. \tag{2.5}$$

Summability conditions must be assumed for the $\{c_j\}$ to guarantee that the random series in (2.5) converges; the following is typically assumed:

$$\sum_{j=0}^{\infty} |c_j|^{\delta} < \infty, \text{ for some } 0 < \delta < \alpha \wedge 1.$$

Note we assume $Z_j \geq 0$ but only assume $c_j \in \mathbb{R}$. It may be mildly puzzling and even controversial to assume positive Z's but the sort of data that we have in mind to model is inherently positive and if the support of Z_i included negative values it would be difficult to preclude an X_t from taking negative values with positive probability. Furthermore, there is an important practical reason for the assumption: Without this assumption limit distributions of important statistics are a mess and the only practical alternative to assuming positivity is to assume the distribution of the Z's is symmetric which seems more harsh than assuming positivity.

Of course the usual classical models are special cases of MA(∞) (cf. Brockwell and Davis, 1991):

(1) Autoregressions of order p, abbreviated AR(p):

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + Z_t, \quad t = 1, \dots, n.$$

(2) Finite order moving averages of order q, written MA(q):

$$X_t = \sum_{j=0}^{q} \theta_j Z_{t-j}, \quad t = 1, \dots, n.$$

(3) Autoregressive moving average processes with orders p, q written ARMA(p,q):

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=0}^{q} \theta_j Z_{t-j}, \quad t = 1, \dots, n.$$

In each case, the process can be written as an $MA(\infty)$.

Issues in a classical context when fitting time series models to data must also be dealt with in the heavy tailed case. These include:

(1) Model selection: Is ARMA a good choice of a parametric family and if so, which ARMA should we choose; i.e. how do we choose p and q?
(2) Parameter estimation: After p and q are chosen, how do you estimate parameters $\boldsymbol{\phi}, \boldsymbol{\theta}$, α, etc ?
(3) Confirmation: After model selection and estimation have been accomplished, how does one confirm that the fitted model is an adequate description of what produced the data?
(4) Given a well selected, estimated and confirmed model, what can you do with it? Predict? There does not seem to be any practical prediction theory for heavy tailed phenomena but if a good fit to the data has been achieved, synthetic data can be fed into a network model to try to estimate things like the distribution of the time to buffer overflow.

3. First steps for heavy tailed modeling

Suppose one is inclined to try fitting a heavy tailed model to a data set. The following seems like a rational set of exploratory first steps:

- Decide if the data could plausibly be explained by a heavy tailed model.
- Try to assess if there is dependency in the data.

We take up these two steps in turn.

3.1 Are heavy tails present? To help with assessing whether heavy tails are present and to estimate the index α in (2.3), various exploratory plotting techniques are available. These are based on the Hill estimator and the qq–plot. We begin with the Hill estimator which is widely used.

3.1.1. The Hill estimator. Suppose $X_1, \dots, X_n$ are iid from a distribution F. Let

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

be the order statistics. If F has an exact Pareto distribution,

$$1 - F(x) = x^{-\alpha}, \quad x > 1,$$

then taking logarithms $\log X_1, \dots, \log X_n$ yields a sample from an exponential density with parameter α. Since the mean is α^{-1} the maximum likelihood estimator (mle) of α^{-1} is the sample mean and thus

$$H_n = \frac{1}{n} \sum_{i=1}^{n} \log X_{(i)}$$

is the mle of α^{-1}. If instead of assuming a Pareto distribution, we only assume

$$1 - F(x) = x^{-\alpha} L(x), \quad x \to \infty,$$

then we may pick $k < n$ and define the Hill estimator (Hill, 1975) to be

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

Note k is the number of upper order statistics used in the estimation.

The rough idea behind using only k upper order statistics is that you should only sample from that part of the distribution which looks most Pareto–like. A more precise explanation is that conditional on $X_{(k+1)}$, the sample

$$\frac{X_{(1)}}{X_{(k+1)}}, \dots, \frac{X_{(k)}}{X_{(k+1)}}$$

is distributed like the order statistics from a sample of size k from the distribution with tail

$$\frac{1 - F(xX_{(k+1)})}{1 - F(X_{(k+1)})}, \quad x \geq 1.$$

Because of (2.3), if $X_{(k+1)}$ is large, the regular variation condition implies that

$$\frac{1 - F(xX_{(k+1)})}{1 - F(X_{(k+1)})} \approx x^{-\alpha}$$

and thus it seems sensible to do what we did in the Pareto case.

Here are the theoretical properties of the Hill estimator.

Proposition 3.1. *Suppose $\{X_t\}$ is stationary and the marginal distribution satisfies*

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

If either
(a) $\{X_t\}$ is iid (Mason, 1982) or
(b) $\{X_t\}$ is weakly dependent (Rootzen et al, 1990; Hsing, 1991) or
(c) $\{X_t\}$ is an MA(∞) process (Resnick and Stărică, 1995, 1996c)
then if $n \to \infty$ and $k \to \infty$ but $k/n \to 0$, we have

$$H_{k,n} \xrightarrow{P} \alpha^{-1} \tag{3.1}$$

and usually (one needs an extra unverifiable assumption on the distribution like second order regular variation and a further restriction on k) Hill's estimator is asymptotically normal. For the iid case we have

$$\sqrt{k}(H_{k,n} - \alpha^{-1}) \Rightarrow N(0, \alpha^{-2}) \tag{3.2}$$

Our purposes would not be served by worrying about the precision of the statement of Proposition 3.1 but note the form of the asymptotics: We need to let $k \to \infty$ but $k/n \to 0$. So k, the number of upper order statistics used in the estimation is considered a function of n, the sample size. We note that regular variation is equivalent to consistency of the Hill estimator in a manner made precise in Mason (1982) and second order regular variation is equivalent to asymptotic normality of Hill's estimator in a manner made precise in Geluk et al (1995). The condition of second order regular variation controls the bias $EH_{k,n} - \alpha^{-1}$. See also Csorgo and Mason (1985), Davis and Resnick (1984), Dekkers and de Haan (1989), Hall (1982), Mason (1988), Mason and Turova (1994), Resnick and Stărică (1996a).

Proofs of the facts given in Proposition 3.1 can be based on the following observation. Suppose the Hill estimator is based on the stationary observations $\{X_1, \dots, X_n\}$, which have marginal distribution $G(x) = P[X_1 \le x]$, and quantile function

$$b(t) = G^{\leftarrow}(1 - t^{-1}).$$

For this sequence of X's, define the tail empirical process as in Section 2.2 by

$$\nu_{X,n}(\cdot) = \frac{1}{k} \sum_{i=1}^{n} \epsilon_{X_i/b(n/k)}(\cdot).$$

Suppose

$$\nu_{X,n}(\cdot) \xrightarrow{P} \nu(\cdot)$$

where $\nu(x, \infty] = x^{-\alpha}$, $x > 0$. Simple inversion and scaling arguments show that it then follows that the same limit relation follows with $b(n/k)$ replaced by $X_{(k)}$, the kth largest order statistic. Since

$$H_{k,n} = \int_1^{\infty} \log y \; \frac{1}{k} \sum_{i=1}^{n} \epsilon_{X_i/X_{(k)}}(dy)$$

it appears that the convergence of the random measures can drag with it the convergence of the integral functional $H_{k,n}$. This argument is the one given in Resnick and Stărică (1995). The philosophy can also be adapted to verify asymptotic normality at least when the sequence $\{X_n\}$ is iid.

In practice, the Hill estimator is used as follows: We graph

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

and hope the graph looks stable so you can pick out a value of α.

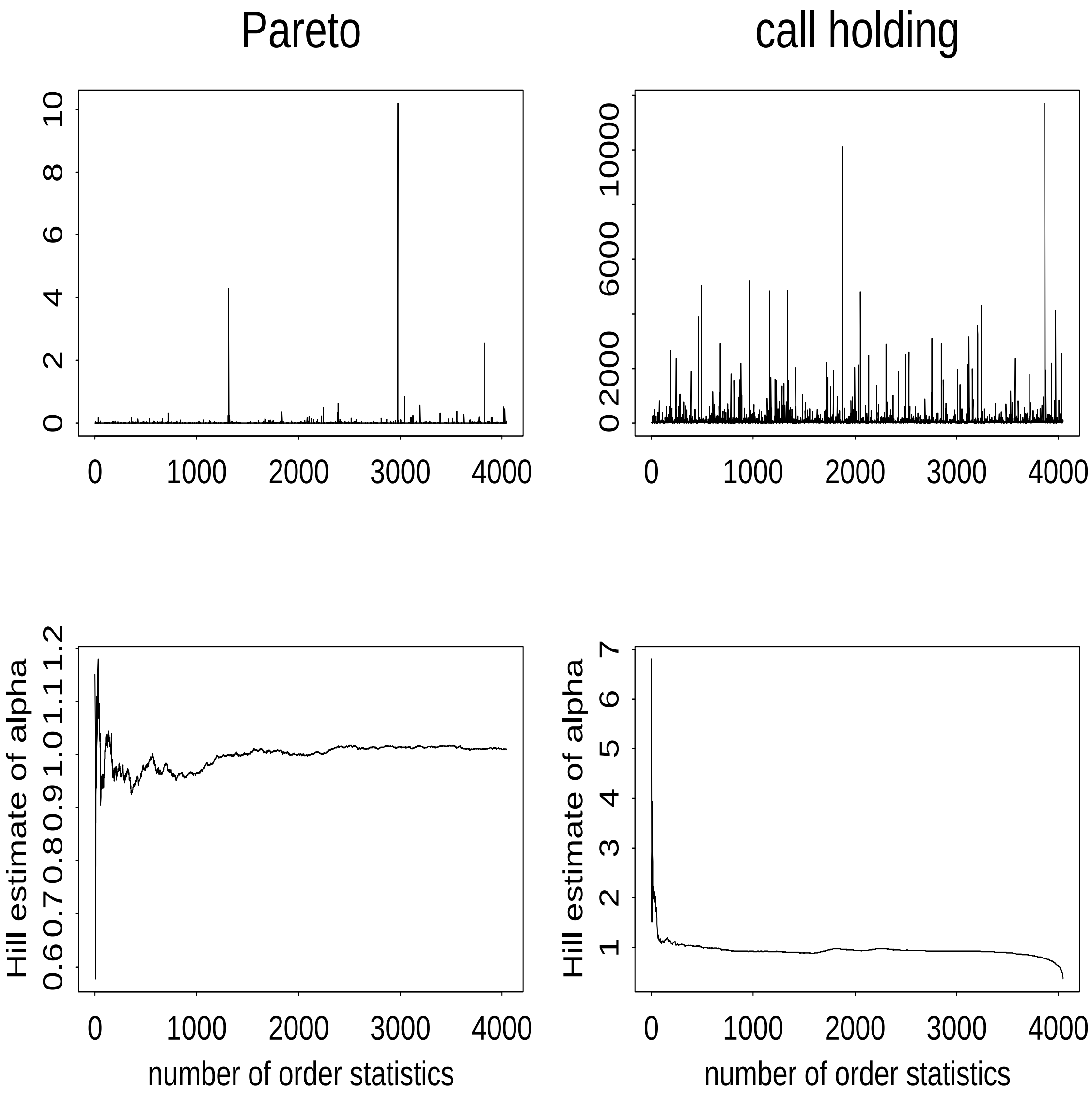

Figure 3.1. Time series and Hill plots for Pareto (left) and call holding (right) data.

Sometimes this works beautifully but sometimes there are problems and it pays to be on good terms with a higher power. Consider Figure 3.1 which shows two cases where the procedure is heart-warming. The top row are time series plots. The top left plot is 4045 simulated observations from a Pareto distribution with $\alpha = 1$ and the top right plot is 4045 telephone call holding times indexed according to the time of initiation of the call. Both plots are scaled by division by 1000. The range of the Pareto data is (1.0001, 10206.477) and the range of the call holding data is (2288,11714735). Readers with teenagers living at home or who use a modem to dial into remote computers will not be surprised that call holding times can be heavy tailed. The bottom two plots are Hill plots $\{(k, H_{k,n}^{-1}), 1 \leq k \leq 4045\}$, the bottom left plot being for the Pareto sample and the bottom right plot for the call holding times. Both Hill plots are gratifyingly stable after settling down and are in a tight neighborhood. The Hill plot for the Pareto seems to nail $\alpha = 1$ correctly and the estimate in the call holding example seems to be between .9 and 1. (So in this case, not only does the variance not exist but the mean appears to be infinite as well.) The Hill plots could be modified to include a confidence interval based on the asymptotic normality of the Hill estimator but we have not done this.

The Hill plot is not always so revealing. Consider Figure 3.2, which has come to be known as the Hill Horror Plot. The left plot is for a simulation of size 10,000 from a symmetric α–stable distribution with $\alpha = 1.7$. One would be hard pressed to discern the correct answer of 1.7 from the plot. The middle plot is for a sample of size 10,000 called *perturb* from the distribution tail

$$1 - F(x) \sim x^{-1}(\log x)^{10}, \quad x \to \infty,$$

so that $\alpha = 1$. The plot exhibits extreme bias and comes nowhere close to indicating the correct answer of 1. The problem of course is that the Hill estimator is designed for the Pareto distribution and thus does not know how to interpret information correctly from the factor $(\log x)^{10}$ and merely readjusts its estimate of α based on this factor rather than identifying the logarithmic perturbation. The third plot is 783 real data called *packet* representing inter-arrival times of packets to a server in a network. The problem here is that the graph is volatile and it is not easy to decide what the estimate should be.

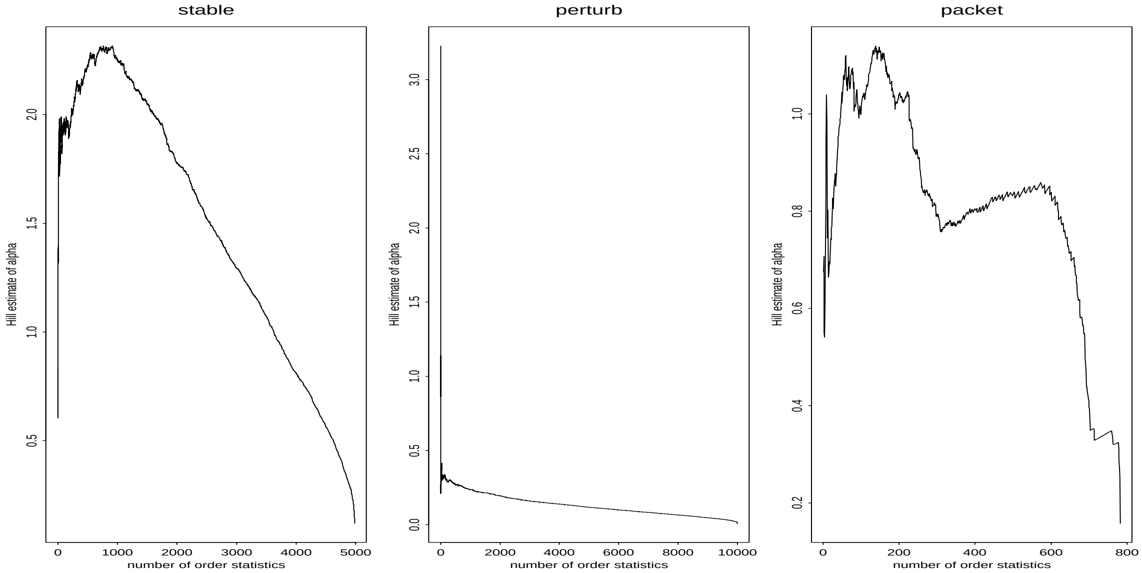

Figure 3.2. The Hill Horror Plot.

Here is a summary of difficulties when using the Hill estimator:

(1) How do you get a point estimate from a graph? What value of k do you use?
(2) The graph may exhibit considerable volatility and/or the true answer may be hidden in the graph.
(3) The Hill estimate has optimality properties only when the underlying distribution is close to Pareto. If the distribution is far from Pareto, there may be outrageous bias even for sample sizes like 1,000,000.

For point 1, several previous studies advocate choosing k to minimize the asymptotic mean squared error of Hill's estimator (Hall (1982), Dekkers and de Haan (1991), de Haan and Peng (1995d)). There are several problems with such formulas. First, they require one to know the distribution rather explicitly and thus, although interesting and welcome, are not a practical solution although there is a possibility of adaptively modifying the procedure which would improve on practicality. Second, the formulas are frequently only asymptotic formulas and asymptotic equivalence is often not helpful for finite samples. If $k^* = k^*(n)$ is the choice of k which minimizes the asymptotic mse, then an equally acceptable asymptotic solution is

$$k_1^* = (1 + \frac{10^{96}}{n})k^*.$$

Even if one accepts a value of k^* for finite n from a displayed formula, this does not always work well in practice.

For point 2, there are simple smoothing techniques which always help to overcome the volatility of the plot and plotting on a different scale frequently overcomes the difficulty associated with the stable example. These techniques are discussed in the next paragraph. For the bias problem, there is no completely satisfactory resolution yet. Two possibilities under investigation are the bootstrap and fitting within a smaller parametric family.

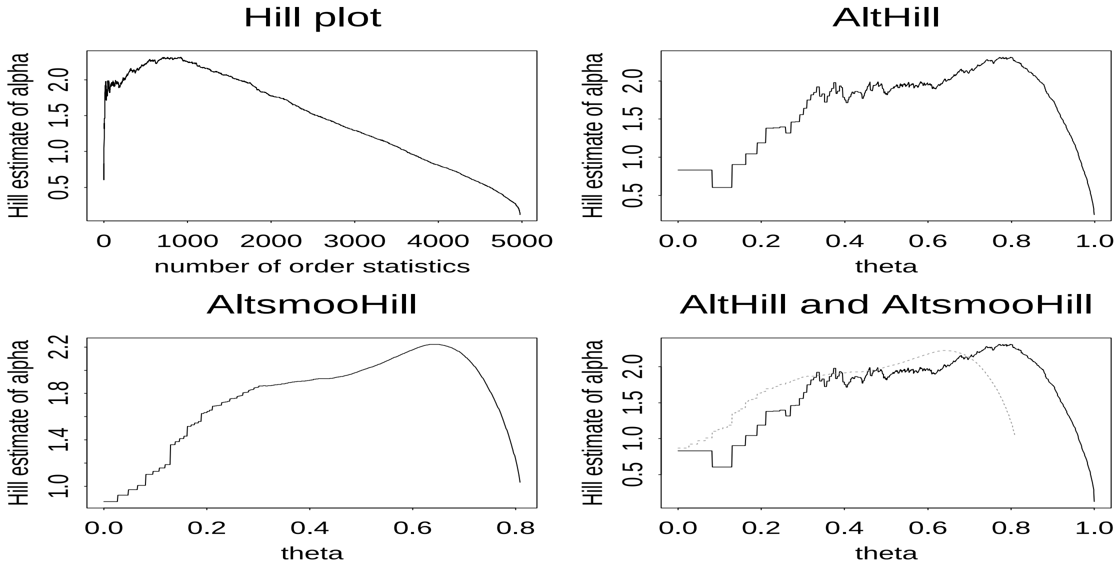

Figure 3.3. Stable, $\alpha = 1.7$

3.1.2. SmooHill: Smoothing the Hill estimator. A simple smoothing technique (Resnick and Stărică, 1996a), although powerless to correct bias, reduces the volatility of the plot and the uncertainty about how to pick out an estimate of α. Pick integer u (usually 2 or 3) and define the SmooHill estimator:

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

The asymptotic variance of the Hill estimator $H_{k,n}$ is $1/\alpha^2$. The asymptotic variance of $\mathrm{Smoo}H_{k,n}$ is less, namely:

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

The bigger the u, the more the asymptotic variance is reduced. However there is a tradeoff between variance reduction and the fact that for large u fewer points are plotted in SmooHill. Usually we pick u between $n^{0.1}$ and $n^{0.2}$ where n is the sample size.

3.1.3. Alt plotting; Changing the scale. As an alternative to the Hill plot, it is sometimes useful to display the information provided by the Hill estimation as

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

where we write $\lceil y \rceil$ for the smallest integer greater or equal to $y \ge 0$. We call such a plot the *alternative Hill plot* abbreviated AltHill. The alternative display is sometimes revealing since the initial order statistics get shown more clearly and cover a bigger portion of the displayed space. The part of the graph corresponding order statistics from the left tail gets rescaled. This method was suggested by C. Stărică and efforts are currently underway to quantify the improvement.

We return now to the examples given in Figure 3.2. Figure 3.3 gives four views of the Hill plot of the stable data where $\alpha = 1.7$. The traditional Hill plot offers little hope of correctly discerning the answer but the alternate scale of AltHill in the upper right seems to reveal the answer fairly clearly. The bottom left plot is the SmooHill plot in alt scale and the bottom right presents both Hill and SmooHill in alt scale together.

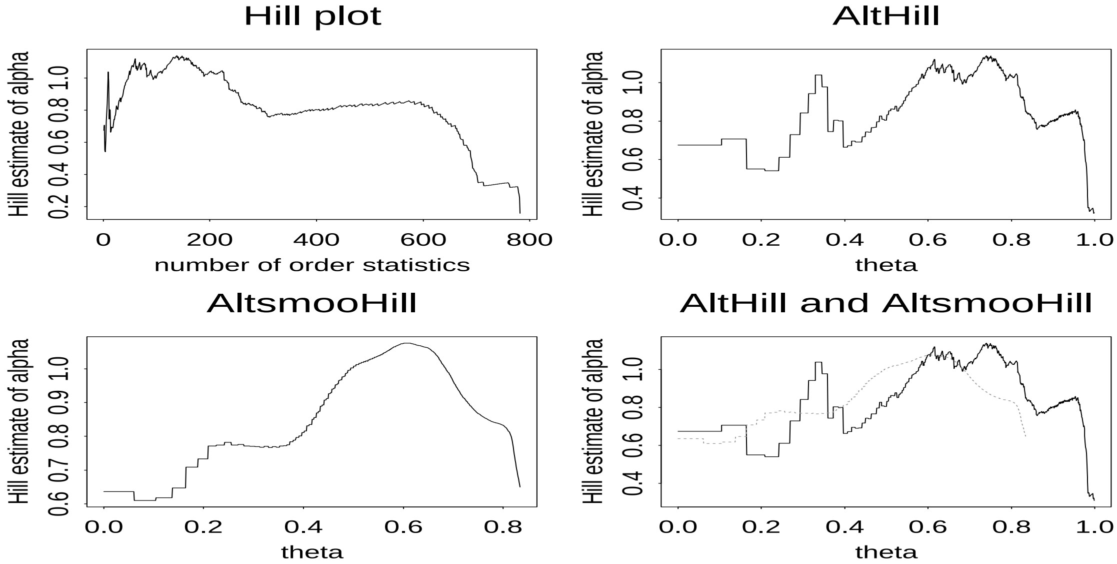

Figure 3.4. Hill plots for packet inter-arrivals.

Figure 3.4 analyzes the packet inter-arrival data introduced for Figure 3.2 and displays the Hill analysis in a manner parallel to the previous Figure 3.3. The Hill plot might lead one to guess a value of α of about .8 until one notices with unease that this occurs around $k = 400$ which means $\theta = \log(400)/\log(783) = 0.899$. Picking k midway between 1 and 783 or picking θ so close to 1 seem unwise. Examining AltSmooHill makes $\alpha = 1.1$ a more likely choice. The sample size of 783 seems too small to provide much assurance of a correct estimate.

3.1.4. Alternative estimators: qq-plotting. The following statement, while not precise, is suggestive: The closer the underlying distribution F is to Pareto, the better the Hill estimator seems to do. The qq-plot can help assess this. This graphical technique is a commonly used method of visually assessing goodness of fit and of estimating location and scale parameters. See for example Rice (1988) and Castillo (1988). It can be adapted to the problem of detecting heavy tails and for estimating the α. It rests on the simple observation that for a sample of size n uniformly distributed on $(0, 1)$, since the spacings are identically distributed, plotting $i/(n + 1)$ vs the ith largest in the sample should yield approximately a straight line of slope 1.

Suppose $\{X_1, \ldots, X_n\}$ are iid with distribution F. Pick k upper order statistics

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

and neglect the rest. Plot

$$\{(-\log(1 - \frac{j}{k+1}), \log X_{(j)}), 1 \leq j \leq k\}. \tag{3.3}$$

If the data is approximately Pareto or even if $1 - F$ is only regularly varying and satisfies (2.3), this should be approximately a straight line with slope$=1/\alpha$. The slope of the least squares line through the points is an estimator called the qq-estimator (Kratz and Resnick, 1995). Computing the slope we find that the qq–estimator is given by

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

There are two different plots one can make based on the qq–estimator. There is the dynamic qq–plot obtained from plotting $\{(k, 1/\widehat{\alpha^{-1}}_{k,n}, 1 \leq k \leq n\}$ which is similar to the Hill plot. Another plot, the static qq–plot, is obtained by choosing and fixing k, plotting the points in (3.3) and putting the least squares line through the points while computing the slope as the estimate of α^{-1} .

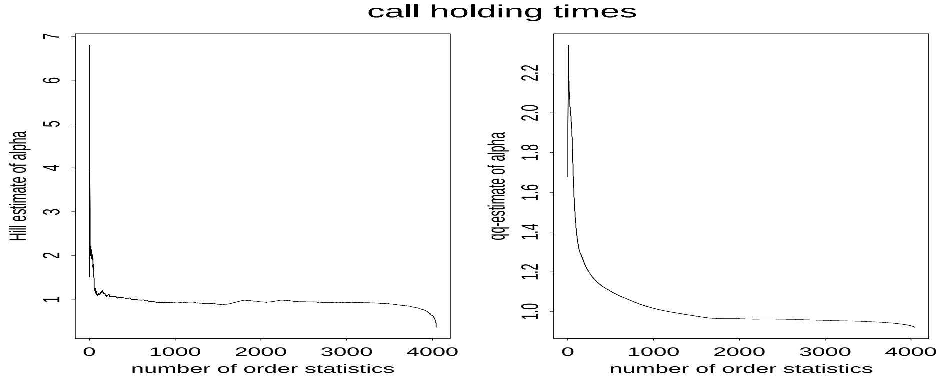

Figure 3.5. Hill and qq-plots for call holding times.

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

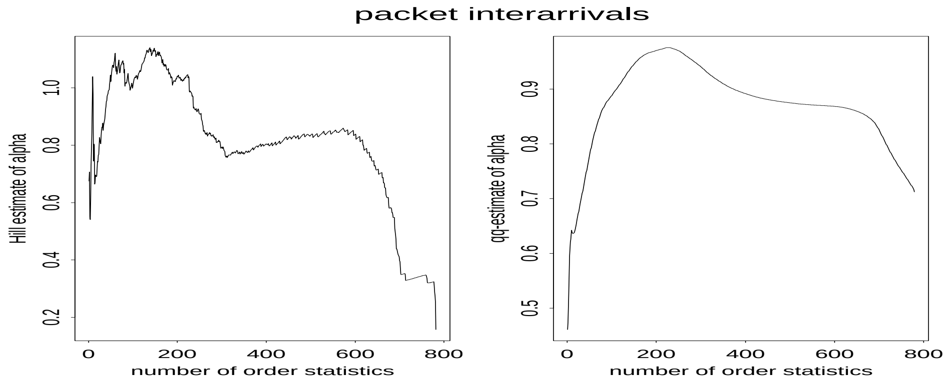

Figure 3.6. Hill and qq-plots for packet inter-arrivals.

Figure 3.5 compares the Hill plot with the dynamic qq-plot for the call holding data and Figure 3.6 does the same thing for the packet inter-arrival data. Figure 3.7 gives two static qq–plots for the call holding data, one using $k = 3500$ and the other using $k = 1500$ yielding estimates of α of 0.95 and 0.977 respectively. For this data set, the estimators are unusually insensitive to the choice of k; the Hill plots and the dynamic qq–plots are quite stable and the static qq–plots do not change much as k varies. Figure 3.8 gives two static qq–plots for the packet inter-arrival data. The data set is only 783 in length and now there is some sensitivity to k.

3.1.5. De Haan's moment estimator. The extreme value distributions (Resnick, 1987; de Haan, 1970; Leadbetter et al, 1983; Castillo, 1988) can be parameterized as a one parameter family

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-\gamma^{-1}}\}, \quad \gamma \in \mathbb{R},\ 1+\gamma x > 0.$$

When $\gamma = 0$, we interpret G_0 as the Gumbel distribution

$$G_0(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

A distribution whose sample maxima when properly centered and scaled converges in distribution to G_γ is said to be in the *domain of attraction* of G_γ, which is written $F \in D(G_\gamma)$. If $\gamma > 0$ and $F \in D(G_\gamma)$ then $1 - F \in RV_{-1/\gamma}$. De Haan's moment estimator $\hat{\gamma}$ (Dekker's, Einmahl, de Haan, 1989; de Haan, 1991, Dekkers and de Haan, 1991; Resnick and Stărică , 1996b) is designed to estimate γ for a random sample in the domain of attraction of G_γ. When $\gamma > 0$, this is the same thing as estimating $\gamma = 1/\alpha$. Since the exponential, normal, gamma densities and many others are in the $D(G_0)$, the domain of attraction of the Gumbel distribution, this provides another method of deciding when a distribution is heavy tailed or not. If $\hat{\gamma}$ is negative or very close to zero, there is considerable doubt that heavy tailed analysis should be applied.

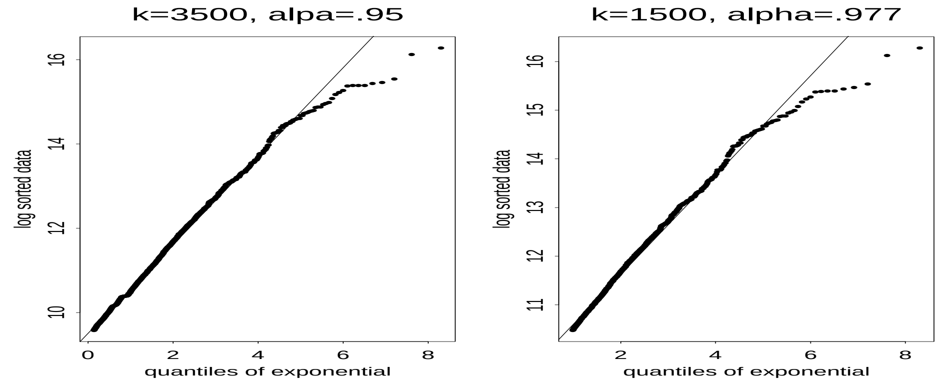

Figure 3.7. Static qq-plots for call holding times.

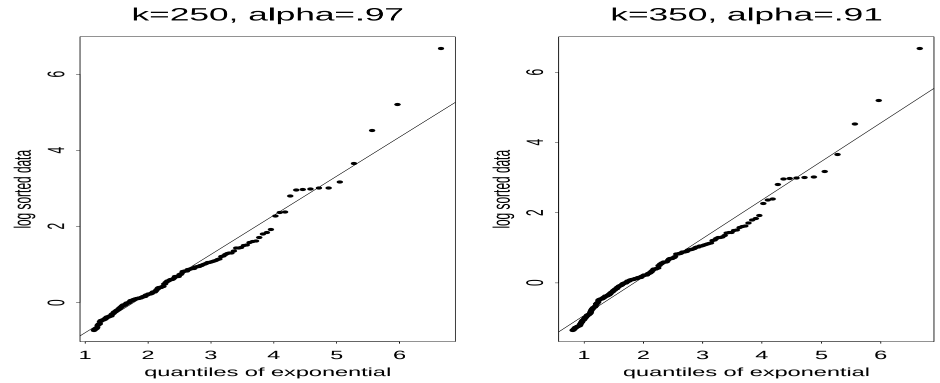

Figure 3.8. Static qq-plots for packet inter-arrivals.

The moment estimator is defined as follows: Let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ be the order statistics from a random sample of size n. Define for $r = 1, 2$

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

so that $H^{(1)}_{k,n}$ is the Hill estimator. Define

$$\hat{\gamma}_n = H^{(1)}_{k,n} + 1 - \frac{1/2}{1 - \frac{(H^{(1)}_{k,n})^2}{H^{(2)}_{k,n}}}. \tag{3.5}$$

Then assuming $F \in D(G_\gamma)$, we have consistency

$$\hat{\gamma}_n \xrightarrow{P} \gamma,$$

as $n \to \infty$ and $k/n \to 0$. Furthermore under a second order condition and a further restriction on k,

$$\sqrt{k}(\hat{\gamma} - \gamma) \Rightarrow N,$$

where N is a normal random variable with 0 mean and variance

$$\sigma(\gamma) = \begin{cases} 1+\gamma^2, & \text{if } \gamma \geq 0, \\ (1-\gamma)^2(1-2\gamma)\left(4 - 8\frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \text{if } \gamma < 0. \end{cases}$$

The asymptotic variance of the moment estimator exceeds that of the Hill estimator when $\gamma > 0$ so from the point of view of asymptotic variance, there is no reason to prefer it. However, the moment estimator discerns a light tail more effectively than the Hill estimator and thus it is often useful to apply the moment estimator to see if $\gamma \leq 0$ which would rule out heavy tail analysis.

Figure 3.9 compares the effectiveness of the moment estimator for discerning a light tail ($\gamma = 0$) with that of of the Hill estimator $\alpha = \infty$. The Hill estimator does not seem very reliable for this purpose. Both estimators are applied to the same random sample of size 1000 taken from an exponential distribution with unit mean.

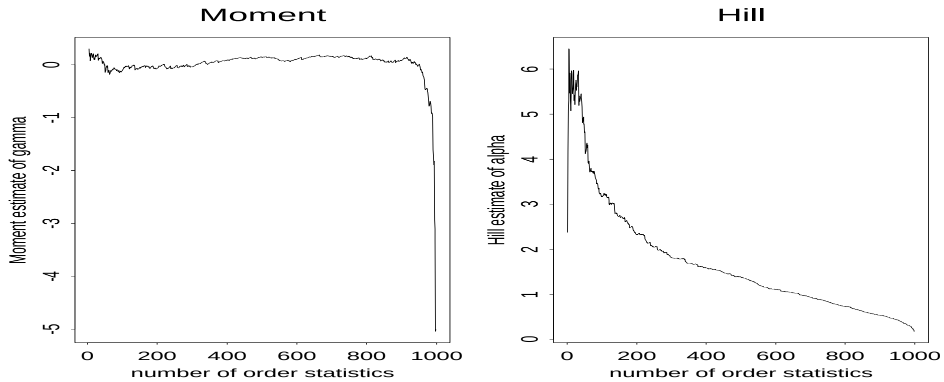

Figure 3.9. Unit exponential data.

Figure 3.10 shows the moment estimator applied to the call holding data on the left and the packet inter-arrival data on the right. Keep in mind when comparing these graphs with previous graphs that $\gamma = 1/\alpha$.

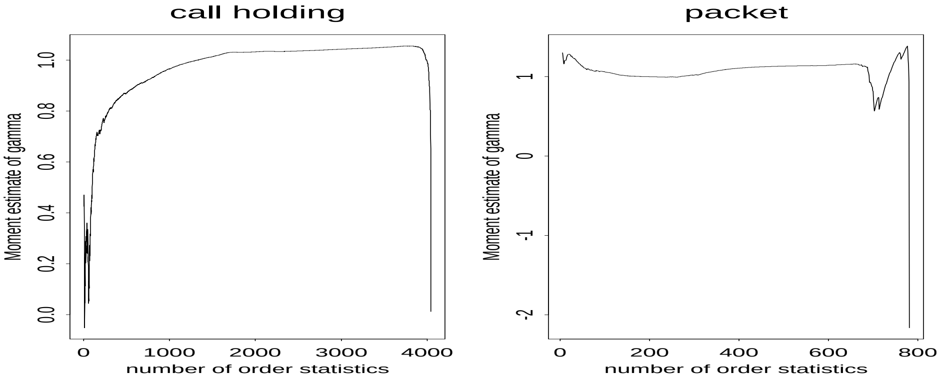

Figure 3.10. Moment estimator applied to call holding and packet data.

3.2 Are dependencies present? Mature statistical computer packages have built in routines to graph the classical sample autocorrelation function (acf) of the data given by

$$\hat{\rho}(h) = \frac{\sum_{i=1}^{n-|h|}(X_i - \bar{X})(X_{i+h} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}. \tag{3.6}$$

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 heavy tailed modification is more appropriate:

$$\hat{\rho}_H(h) = \frac{\sum_{i=1}^{n-|h|} X_i X_{i+h}}{\sum_{i=1}^{n} X_i^2}. \tag{3.7}$$

If we have the heavy tailed model given by (2.5)

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

where $\{Z_t\}$ are non-negative, iid, heavy tailed, 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)

$$\hat{\rho}_H(h) \xrightarrow{P} \frac{\sum_{j=0}^{\infty} c_j c_{j+h}}{\sum_{j=1}^{\infty} c_j^2} := \rho(h). \tag{3.8}$$

The mean corrected function given in (3.6) also converges in probability to the same limit. The limit law for $\hat{\rho}_H(h)$ is complex and is established in Davis and Resnick (1986, 1985b). The most tractable case is for $\alpha < 1$ and is given as follows:

Proposition 3.2. *Suppose $\alpha < 1$ and for some $\delta < \alpha$*

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

Define the quantile functions

$$b(n) = \left(\frac{1}{1-F}\right)^{\leftarrow}(n), \quad \tilde{b}(n) = \left(\frac{1}{P[Z_1Z_2 > \cdot]}\right)^{\leftarrow}(n).$$

Then for any $l \geq 1$

$$\left(\tilde{b}(n)^{-1}b(n)^2(\hat{\rho}_H(h) - \rho(h), 1 \leq h \leq l\right) \Rightarrow (Y_1, \dots, Y_l),$$

in $\mathbb{R}^l$ *where*

$$Y_h = \sum_{j=1}^{\infty}(\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h))\frac{S_j}{S_0}.$$

Here $S_0, S_1, \dots$ *are independent and* S_0 *is one sided stable of index* $\alpha/2$ *and* $S_1, S_2, \dots$ *are iid, one sided stable of index* α*.*

Even in this simplest case, the limit distribution is quite complex. (Compare this with the classical Bartlett formula where the sample acf has a limiting normal distribution. See Brockwell and Davis (1991), page 221ff, page 538ff.) In distribution Y_h can be re-expressed as

$$\left(\sum_{j=1}^{\infty}|\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h)|^{\alpha}\right)^{1/\alpha}\frac{U}{V}$$

where U and V are independent, non-negative stable random variables and the index of V is $\alpha/2$ and the index of U is α. So even if the value of α were known, the percentiles of the distribution of Y_h are not easy to obtain and certainly impossible to obtain analytically.

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 give an exploratory indication of 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.

Figure 3.11 shows the classical acf for the call holding data side by side with the heavy tailed modification. The right graph of the heavy tailed sample acf has a dotted line drawn at height $h = 0.0035$ and the interval $[0, h]$ is a 95% confidence window analogous to the one given by Bartlett's formula (Brockwell and Davis, 1991) in classical time series. The confidence window is drawn based on the assumption that the data is independent and has Pareto tails. According to Theorem 3.3 of Davis and Resnick (1986), h is given by

$$h = l\alpha^{1/\alpha}\frac{n^{-1/\alpha}}{\log n}$$

where l is the quantile satisfying

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

for U, V independent positive stable random variables with indices α and $\alpha/2$. In the case of the call holding data, we used the estimated value of $\alpha = .97$. The quantile l was estimated by simulation. The position of h relative to the heights of the sample heavy tailed acf values casts serious doubts on the assumption of independence. Figure 3.12 exhibits comparable graphs for the packet inter-arrival data. The right heavy tailed graph does not offer evidence against the hypothesis of independence.

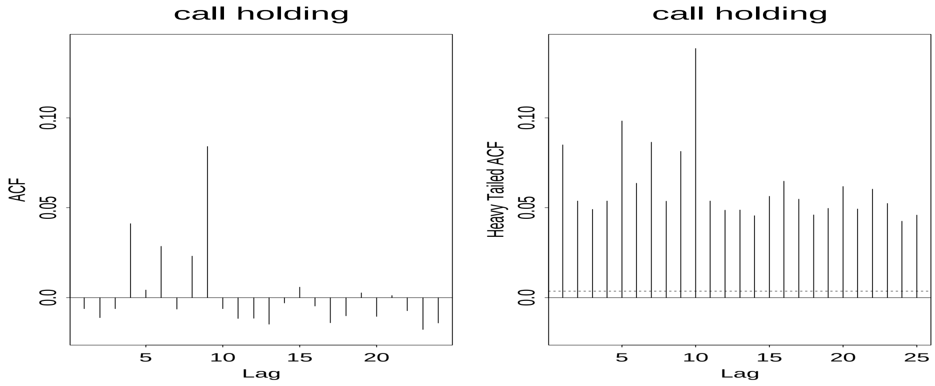

Figure 3.11. Call holding times

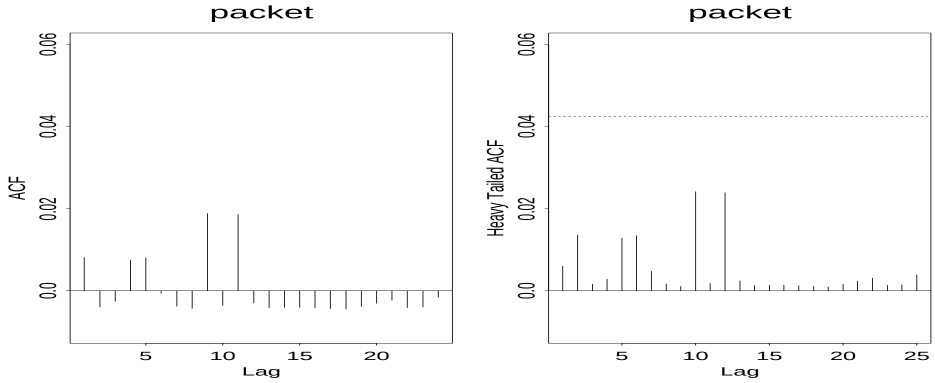

Figure 3.12. Packet inter-arrivals

4. Modeling dependent data

Modeling data which is not a realization of an iid model presents great challenges. There is no guarantee that the simple linear ARMA models form a large enough subclass of the class of heavy tailed models that hunting within this class will yield a model with an acceptable fit to the data.

From the theoretical point of view, there has been considerable success in analyzing autoregressive models of order p, abbreviated AR(p). These are models of the form

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

where $\{Z_t\}$ is iid and heavy tailed.

Several methods have been proposed for estimating the coefficients $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$. These methods are all consistent and asymptotic distributions have been worked out. They include:

- Yule–Walker estimators; after slight modification, these work for heavy tailed autoregressions (Davis and Resnick, 1985a,b;1986).
- Spectral density estimators (Mikosch, Gadrich, Klüppelberg, Adler, 1995).
- Least gamma deviation estimators (Davis, Knight, Liu, 1991).
- Linear programming (LP) estimators (Feigin and Resnick, 1992, 1994a,b; Feigin, Resnick, Stărică, 1995; Feigin, Kratz, Resnick, 1994).

The LP and least gamma deviation estimators have the best rate of convergence. We will only deal with the Yule–Walker and LP methods. For both methods we review model fitting and estimation.

4.1. The Yule–Walker method. The classical Yule–Walker method is based on sample correlations. In the heavy tailed context, the heavy tailed sample correlation function is the basis for fitting and estimation.

4.1.1 Yule–Walker estimation. Recall the definition of $\hat\rho_H(h)$ and $\rho(h)$ from (3.7) and (3.8). Suppose $0 < \alpha < 2$ and that $\{X_t\}$ is a stationary, invertible autoregressive process of order p of the form given by (4.1) which can be inverted and written as the MA(∞) process

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}. \tag{4.2}$$

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

$$\mathbf{R}\boldsymbol{\phi} = \boldsymbol{\rho} \tag{4.3}$$

where recall $\boldsymbol{\phi}$ is the p-vector of autoregressive coefficients in (4.1). Furthermore, for every m, the matrix

$$\mathbf{R}_m = (\rho(i-j))_{i,j=1}^{m} \tag{4.4}$$

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

$$\hat{\mathbf{R}}\hat{\boldsymbol{\phi}}^{YW} = \hat{\boldsymbol{\rho}_H} \tag{4.5}$$

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, 1986a) 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). The rate of convergence of this limit distribution as measured by $\tilde b(n)^{-1} b(n)^2$ is inferior to that of the linear programming estimators.

4.1.2. Model selection based on sample correlations: pacf and AIC.

Standard techniques for model selection based on sample correlations are available for heavy tailed analysis after minor modification. 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

$$\hat{\boldsymbol{\phi}}^*_m = \hat{\mathbf{R}}_m^{-1} \hat{\boldsymbol{\rho}}_m \tag{4.6}$$

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. This of course means that the m-th component $\tilde{b}_n b_n^2 \hat{\phi}^*_{m,m}$ has a limit distribution.

This 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.

The classical AIC criterion for order selection is not consistent for selecting finite variance models since it tends to overestimate the order. Brockwell and Davis (1991) have a nice example where they simulate an AR(1) process and then apply the AIC criterion which insists the simulated process is an AR(2). However, the AIC criterion is consistent for heavy tails (Knight, 1989a; Bhansali, 1988). Define

$$\begin{aligned}
\hat{\sigma}^2(0) &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\
\hat{\sigma}^2(m) &= \hat{\sigma}^2(0) \prod_{j=1}^{m} (1 - \hat{\phi}^{*}_{j,j})^2), \quad m \geq 1.
\end{aligned}$$

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:

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

where K is an upper bound which is assumed 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. Below in Figure 4.1 we display the heavy tailed pacf plot and the AIC plot $\{(k, AIC(k)), 1 \leq k \leq 15\}$ for 500 simulated data from the heavy tailed AR(2) process

$$X_t = 1.3X_{t-1} - 0.7X_{t-2} + Z_t \tag{4.7}$$

where $\{Z_t\}$ are iid with standard Pareto distribution: $P[Z_1 > x] = 1/x,\ x \geq 1$. Both techniques are seen to work quite well. However, for real data, it may be the case that even though $AIC(k)$ has a nice minimum, residual analysis of the model fitted by the AIC criterion reveals a lack of iid structure in the estimated residuals, throwing into doubt the goodness of the fit.

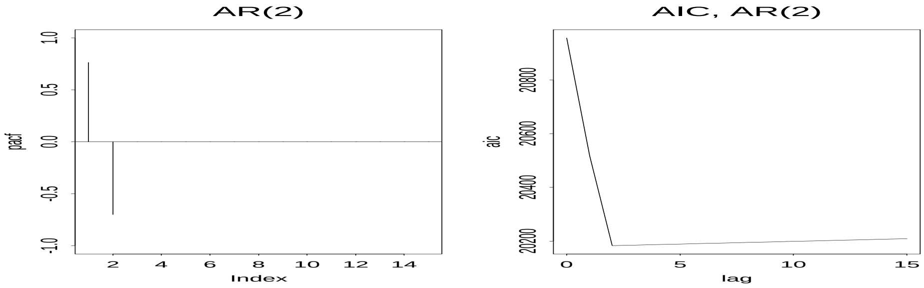

Figure 4.1 Heavy tailed pacf and AIC for simulated AR(2).

4.2. Linear programming estimators. The linear programming (LP) estimators were devised by Feigin and Resnick (1992, 1994a) to capitalize on the assumption that the residuals $\{Z_t\}$ in (4.1) are non-negative. This work built on ideas of Davis and McCormick (1989) and Andel (1989).

4.2.1. Linear programming estimation. Assume the innovations $\{Z_t\}$ in (4.1) are non-negative with a common distribution whose left endpoint is 0. The LP estimators of the AR(p) coefficients $\boldsymbol{\phi}$ are

$$\hat{\phi}(n) = \arg\max_{\boldsymbol{\delta} \in D_n} \boldsymbol{\delta}'\mathbf{1} \tag{4.8}$$

where $\mathbf{1}' = (1, \ldots, 1)$ and where the feasible region D_n is defined as

$$D_n = \bigcap_{t=1}^{n} \{\delta \in R^p : X_t - \sum_{i=1}^{p} \delta_i X_{t-i} \geq 0\}. \tag{4.9}$$

Here is a rapid review of one derivation of this estimator: Suppose temporarily that the common distribution of the Z's is unit exponential so that

$$P[Z_1 > x] = e^{-x}, \quad x > 0.$$

In this case, conditionally on $\{X_0 = x_0, X_{-1} = x_{-1}, \ldots, X_{-p+1} = x_{-p+1}\}$, the likelihood is proportional to (using $I(\cdot)$ for the indicator function):

$$I\left(\bigwedge_{t=1}^{n} (X_t - \sum_{i=1}^{p} \phi_i X_{t-i}) \geq 0\right) \exp\left\{\phi_1 \sum_1^n X_{t-1} + \phi_2 \sum_1^n X_{t-2} + \cdots + \phi_n \sum_1^n X_{t-p}\right\}$$
$$\approx I\left(\bigwedge_{t=1}^{n} (X_t - \sum_{i=1}^{p} \phi_i X_{t-i}) \geq 0\right) \exp\left\{\sum_{i=1}^{p} \phi_i \sum_{t=1}^{n} X_t\right\}.$$

Assuming that $\sum_{t=1}^n X_t$ is ultimately positive, the corresponding maximum likelihood estimator will thus be approximately determined by solving the linear program (LP)

$$\max(\sum_{i=1}^p \phi_i)$$

subject to

$$X_t \geq \sum_{i=1}^p \phi_i X_{t-i}; \quad t = 1, \dots, n.$$

The simplified approximate form of the objective function is justified by the fact that $\sum_1^n X_{t-1} / \sum_1^n X_{t-i} \approx 1$.

Now drop the assumption that the density of the Z's is exponential. The LP estimator still gives an estimation procedure with good properties which is applicable to the heavy tailed case.

The consistency and asymptotic distribution for the LP estimator was established in Feigin and Resnick (1992, 1994a), Feigin, Resnick and Stărică (1995).

Theorem 4.1. *Suppose $\{X_t\}$ is a stationary autoregression given by (4.1) where $\{Z_t\}$ are iid non-negative innovation variables with common distribution F which has a regularly varying tail of index $-\alpha$. Suppose also that for some $\beta > \alpha$*

$$EZ_1^{-\beta} = \int_0^\infty u^{-\beta} F(du) < \infty.$$

Let $\hat{\boldsymbol{\phi}}(n)$ be the linear programming estimator based on $X_1, \dots, X_n$ given in (4.8) and (4.9). Let $\{E_j, j \geq 1\}$ be iid unit exponential random variables and define

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

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

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

and for $|z| \leq 1$ set

$$C(z) = \sum_{j=0}^\infty c_j z^j = \frac{1}{\Phi(z)}, \quad \Phi(z) = 1 - \sum_{i=1}^p \phi_i(0) z^i,$$

where $\{\phi_i(0), i = 1, \dots, p\}$ are the true autoregressive coefficients. Then

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

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

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

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

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

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

$$\mathbf{L} \stackrel{d}{=} \arg\max_{\boldsymbol{\delta} \in \Lambda} \delta' \mathbf{1}, \tag{4.10}$$

and

$$\Lambda = \{\boldsymbol{\delta} \in R^p : \boldsymbol{\delta}'\mathbf{1} \geq -1, \boldsymbol{\delta}'\mathbf{v}_k \leq 1, \ k \geq 1\}. \tag{4.11}$$

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

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

Note that the asymptotic distribution is complicated and depends on a limiting Poisson process and the unknown distribution of the autoregression. In one case, however, namely if $\boldsymbol{\phi}(0) = \mathbf{0}$ making $X_t = Z_t$ the limit distribution considerably simplifies and this is the basis of a test for independence discussed in the next subsection.

We applied the LP estimator and the Yule–Walker estimator to a sample of size 100 from the AR(2) described in (4.7). The LP estimator yielded (ϕ_1, ϕ_2) values of $(1.30202, -0.7004143)$, which is quite good performance in view of the small sample size. The corresponding Yule–Walker were not nearly so accurate and were $(0.9084571, -0.4122331)$.

4.2.2. Model selection and confirmation. 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 large. (Feigin, Resnick and Stărică, 1995). This also provides a model selection tool since a well fitted model should have the property that the estimated residuals

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

are approximately iid and we may test this using the independence test.

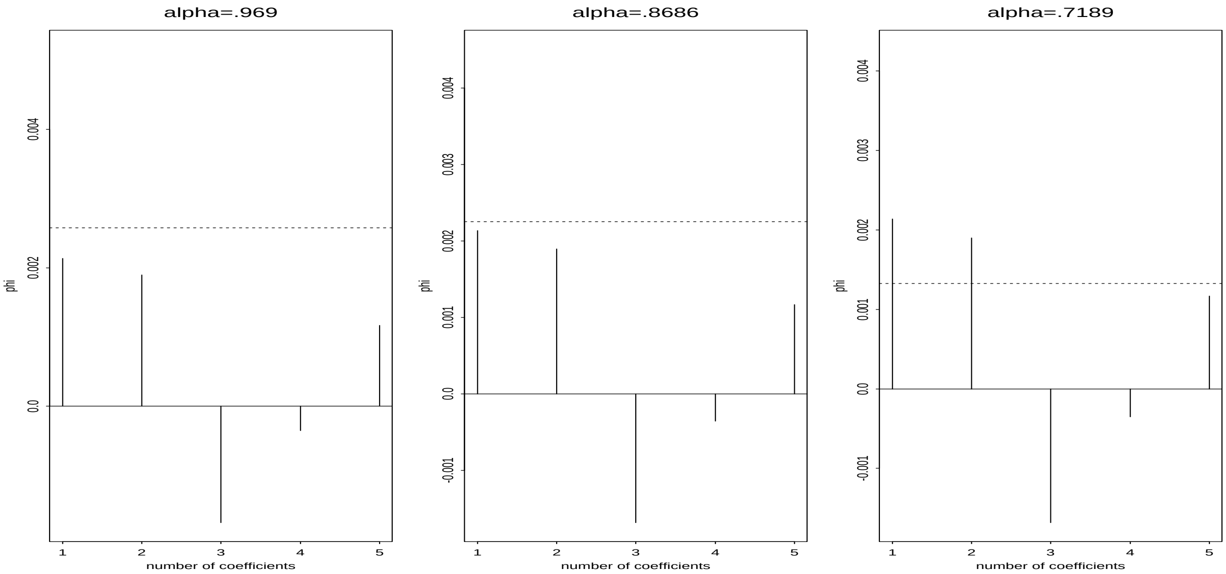

Figure 4.2 Tests for independence for packet inter-arrivals.

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 $\boldsymbol{\phi}(0) = \mathbf{0}$

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

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

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

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. The qq-plot yields both $\hat{b}_n$ and $\hat{\alpha}$ and then we can get

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

Figure 4.2 shows the implementation of the independence test for the data set of packet inter-arrivals. The earlier discussion of the estimate of α was not entirely conclusive so the test was conducted three times with various values of α which are shown in the figure. We tested against the alternative for 5 autoregressive coefficients. Because the subadditive approximation does more damage for larger values of p, we limited ourselves to $p = 5$. The test fails only for the improbably small estimate of $\alpha \approx .7$ and this coupled with the heavy tailed acf and pacf plots indicates strong evidence for independence. In the graphs, the dotted horizontal line represents $K(.05)$ and the vertical lines represent $\{\hat{\phi}_i^{(n)}, i = 1, \ldots, 5\}$.

4.2.3. Left tail analysis. The linear programming estimators and test for independence are designed to work under the assumption that either the right tail is regularly varying as in (2.3) or that the left tail is regularly varying (Feigin and Resnick, 1994a). For the left tail case, the precise assumptions that ensure consistency and an asymptotic distribution for the LP estimator is as follows:

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

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

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.

(2) **Condition S** (stationarity): The coefficients $\phi_1, \dots, \phi_p$ satisfy the stationarity condition that the autoregressive polynomial $\Phi(z) \equiv 1 - \sum_1^p \phi_i z^i$ has no roots in the unit disk $\{z : |z| \leq 1\}$. Furthermore we assume $\Phi(1) > 0$; i.e. we require

$$\sum_{i=1}^{p} \phi_i < 1.$$

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

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

$$\text{2. } E(Z_t^{\beta}) = \int_0^{\infty} u^{\beta} F(du) < \infty \text{ for some } \beta > \alpha. \tag{4.14}$$

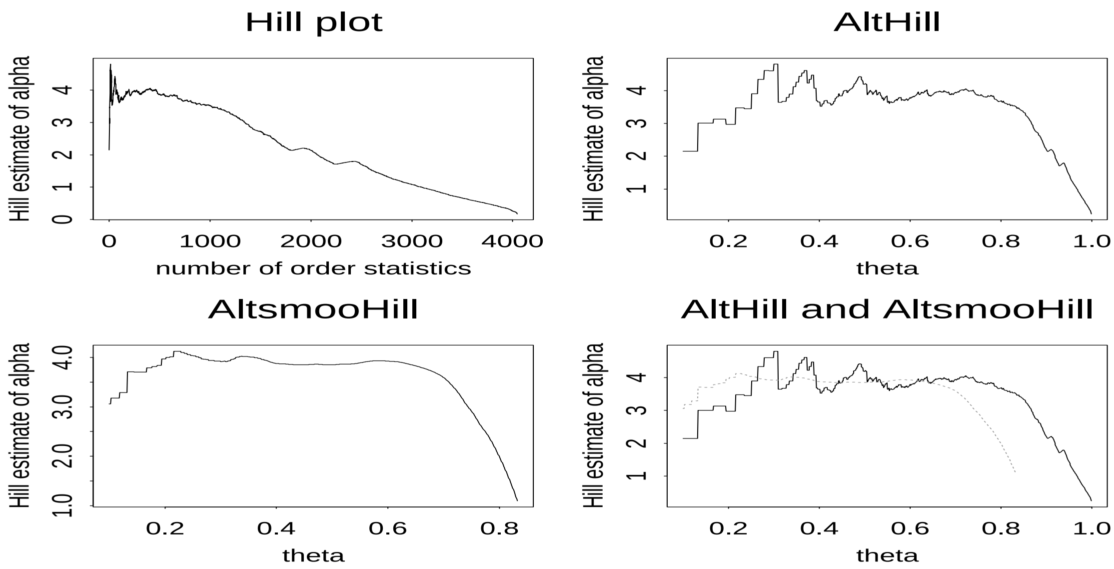

Figure 4.3. Hill plots for left tail of call holding data.

A notable success in the left tail case was given in Feigin, Resnick and Stărică (1995) where the LP estimator was used to fit an autoregression to the lynx data and the test for independence was used to fine tune the fit and perform model confirmation.

Under regular variation of either the left or right tail of Z_1, the rate of convergence of the LP estimator is of the order of n^q where q is the reciprocal of the index of variation. For some phenomena, the distribution of Z_1 may have both tails regularly varying and in this case best results are obtained by working with the heavier tail, that is, the tail with the smallest index. This achieves the best rate of convergence.

Figure 4.3 displays four views of the Hill plot for 1/callholding giving strong indication that the *left* tail of the call holding data is regularly varying with an index in the neighborhood of $\alpha = 4$. Figure 4.4 gives a qq-plot yielding an estimate of 3.52. Since the left tail parameter is so much bigger than the right tail index, there is little temptation to switch to left tail analysis.

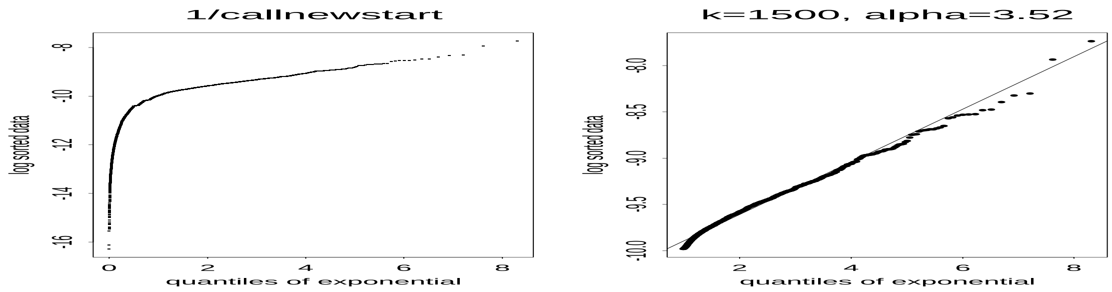

Figure 4.4. QQ- plots for left tail of call holding data.

4.2.4. The bootstrap and LP estimation.
A review of Theorem 4.1 affirms that the limit distribution for the LP estimators cannot be calculated explicitly except in the case of independence. To overcome this difficulty, a bootstrap procedure can be devised (Feigin and Resnick, 1995; Datta and McCormick, 1995). Caution: It remains to be seen how practical such procedures will be and right now there is a sizable gap between theory and practice.

The proof (Feigin and Resnick, 1995) of the validity of the bootstrap depends heavily on a stochastic version of Karamata's Theorem where $1-F$ is replaced by $\nu_n(x,\infty]$ where

$$\nu_n(x,\infty] = \frac{m}{n}\sum_{t=1}^{n} \epsilon_{Z_t/b(m)}$$

and also by a version of the Karamata Theorem where $1-F$ is replaced by by $\hat{\nu}_n(x,\infty]$ where

$$\hat{\nu}_n = \frac{m}{n}\sum_{t=1}^{n} \epsilon_{\hat{Z}_t(n)/b(m)}$$

and $\hat{Z}_t(n), t=1,2,\ldots$ are the estimated residuals.

The necessity for the bootstrap sample size to be $m=o(n)$ makes use of the bootstrap difficult in practice. Just as we had difficulty picking k when using the Hill estimator, for the bootstrap we must pick m without reliable guidelines. In connection with bootstrapping extremes and heavy tailed phenomena, many authors have noticed that if the original sample is of size n, in order for the bootstrap asymptotics to work as desired the bootstrap sample should be of size m where m is a function of n and $m/n \to 0$ as $n \to \infty$. See for example Athreya (1987), Gine and Zinn (1989), Hall (1990), Kinateder (1992), Knight (1989b), Lepage (1992), and Deheuvels, Mason and Shorack (1993). Some perspective on this necessity to reduce the bootstrap sample size to something of smaller order than the observed sample size is provided in the discussion of the behavior of random measures in Proposition 2.1. See Feigin and Resnick (1995) for a discussion.

4.2.5 Estimating α for autoregressions. Suppose you observe $X_1,\ldots,X_n$ from a heavy tailed autoregression

$$X_t = \sum_{i=1}^{p} \phi_j X_{t-j} + Z_t, t = 0, \pm 1, \pm 2, \ldots \tag{4.15}$$

where $Z_t \geq 0$ and

$$P[Z_1 > x] = x^{-\alpha}L(x). \tag{4.16}$$

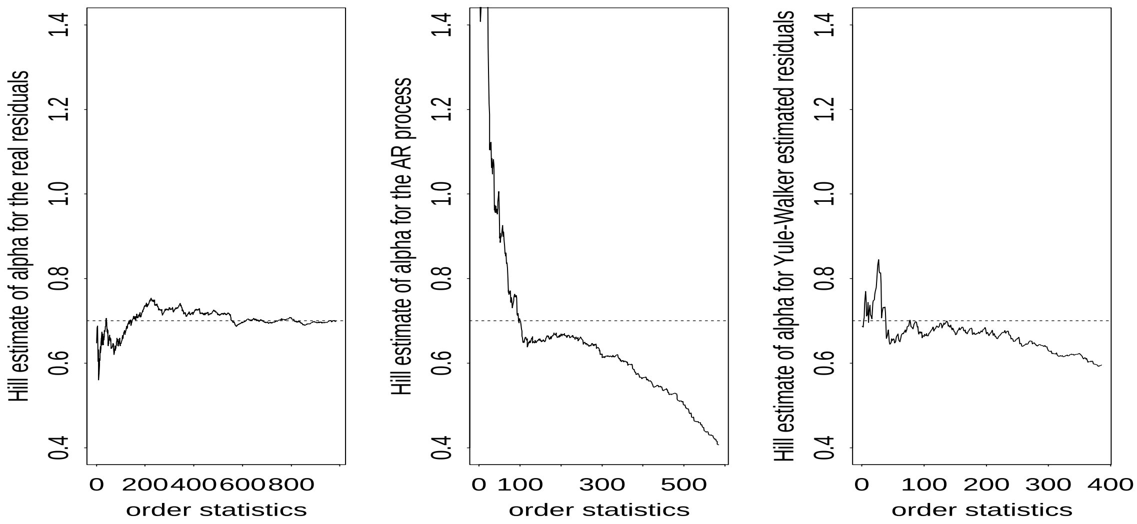

Figure 4.5. Hill plots for an autoregression.

There are two possible ways to estimate α:

• Apply the Hill estimator directly to $X_1, \dots, X_n$, ie

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

This seems a sensible thing to do since the tail of X_1 contains the same information as the tail of Z_1 by a result of Cline (1983) which says that

$$P[X_1 > x] \sim (const)P[Z_1 > x].$$

We know this works from Proposition 3.1.

• Alternatively we could estimate autoregressive coefficients $\boldsymbol{\phi}$ with a consistent estimator $\hat{\boldsymbol{\phi}}^{(n)}$ and then estimate the residuals

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

These estimate $Z_1, \dots, Z_n$. If we apply Hill's estimator to the estimated residuals we should get a sensible procedure:

$$H_{k,n}(\hat{Z}) = \frac{1}{k}\sum_{i=1}^{k}(\log \hat{Z}_{(i)}^{(n)} - \log Z_{(k+1)}^{(n)}.$$

Both are consistent estimators of α^{-1} (Resnick and Stărică, 1995). Experience and asymptotic variance calculations (Resnick and Stărică, 1996c) indicate that the second is surely a better procedure.

To compare these procedures we present 3 Hill plots plots. An autoregression of length 1000 was simulated with $\phi_1 = 1.3$, $\phi_2 = -0.7$ using iid Pareto random variables with $\alpha = 0.7$. The left hand Hill plot is for the actual Pareto residuals used. The middle graph is a Hill plot for $\{X_t\}$ and the right hand Hill plot is based on the estimated residuals. The middle plot seems to be considerably worse than the right plot based on estimated residuals. The dotted horizontal line in each case represents the true value of α.

5. A data example

We consider a data set consisting of 3802 inter-arrival times of ISDN D-channel packets. Figure 5.1 gives the time series plot and Figure 5.2 gives the acf and pacf plots in both classical and heavy tailed form. Independence does not seem likely based on these graphs.

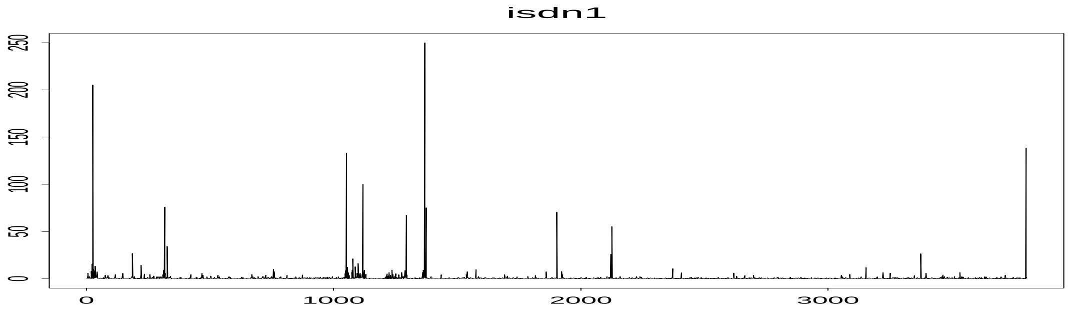

Figure 5.1. Time series plot of isdn1.

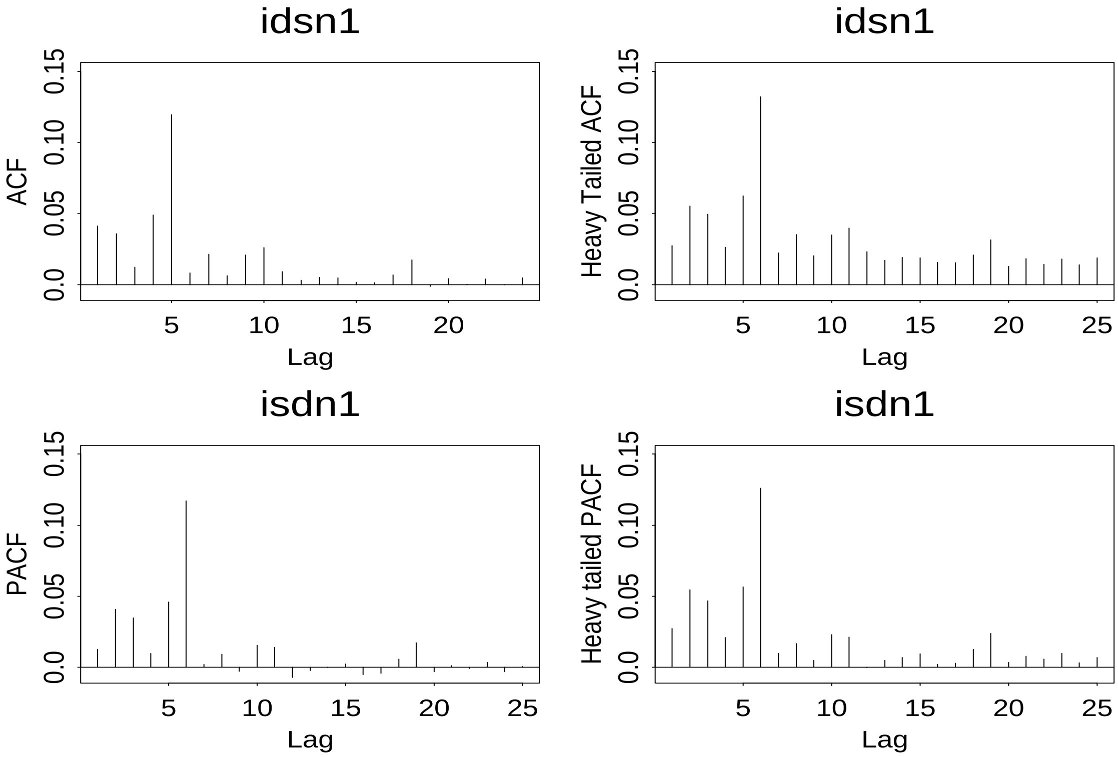

Figure 5.2. Acf and pacf, heavy and classical, for isdn1.

We next checked the heavy tailed nature of the data. The Hill estimator was exceptionally stable and so were the qq-plots. Based on these diagnostics a value of $\alpha = 1.06$ was estimated. The plots are given in Figure 5.3.

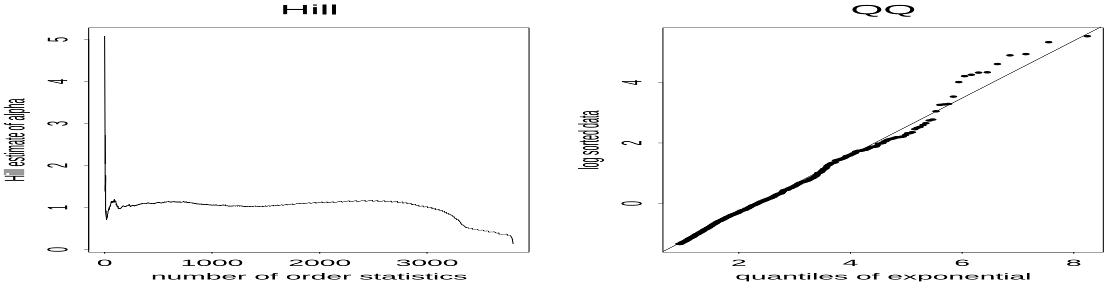

Figure 5.3. Hill and qq-plots for isdn1.

Based on the acf/pacf plots, it is unlikely the data can be modeled by independence, but to confirm this we applied the independence test with $p = 6$. The independence hypothesis is rejected at the 5% level as shown in the next plot Figure 5.4.

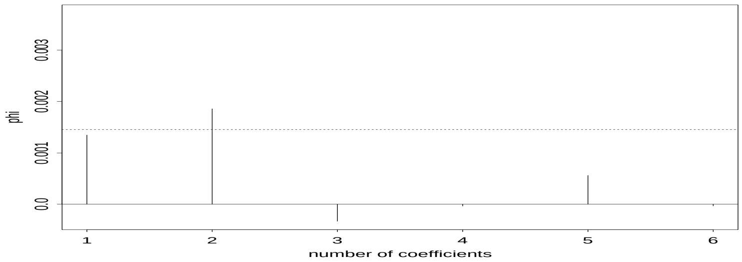

Figure 5.4. Independence test on isdn1.

If we wish to try modeling the data using a heavy tailed autoregression, we have to decide on the order. The pacf plot indicates $p = 6$ is likely and this is confirmed by the AIC plot given in Figure 5.5.

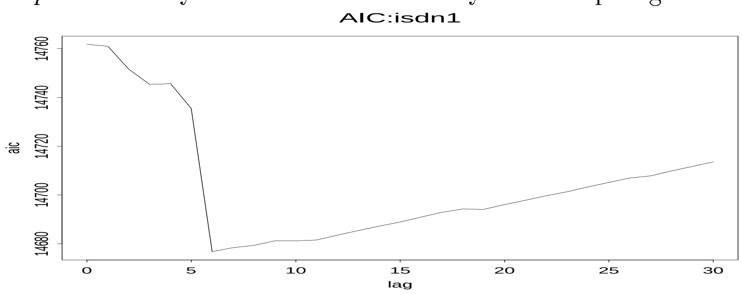

Figure 5.5. AIC plot for isdn1.

We then estimated autoregressive coefficients using the LP estimator and obtained coefficients

$$(\hat{\phi}_1^{(n)}, \ldots, \hat{\phi}_6^{(n)}) = (0.00135, 0.00186, -0.00033, -0.00003, 0.00056, -0.00003).$$

The estimated coefficients are rather small and it does not appear that the autoregressive structure changes the data very much. However, applying the independence test to the residuals as before with $p = 6$ yields better results and the independence hypothesis cannot now be rejected.

The satisfaction one feels at apparently finding a suitable fit for the data is tempered by the acf plots in Figure 5.7 which splits the data into 3 successive parts of length 1000 each and computes an acf plot for each. The plots do not look very similar which could be an indication of lack of stationarity or, more disturbing, could be an indication of non-linearity in the data. Tools need to be developed to cope with this possibility. This is discussed in greater detail in Feigin and Resnick (1996), Resnick (1996) and Davis and Resnick (1995). In any events, tests on estimated residuals should always be interpreted cautiously since the estimation method used usually tries to whiten residuals.

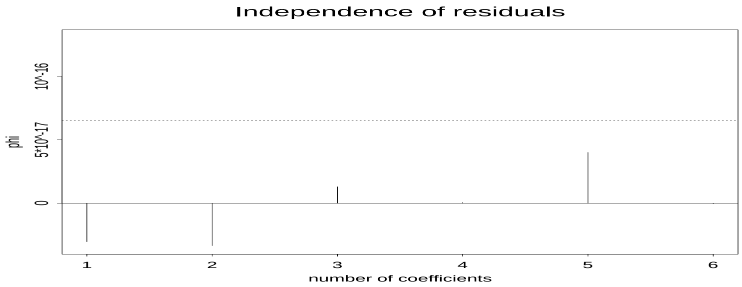

Figure 5.6. Independence test for residuals of isdn1.

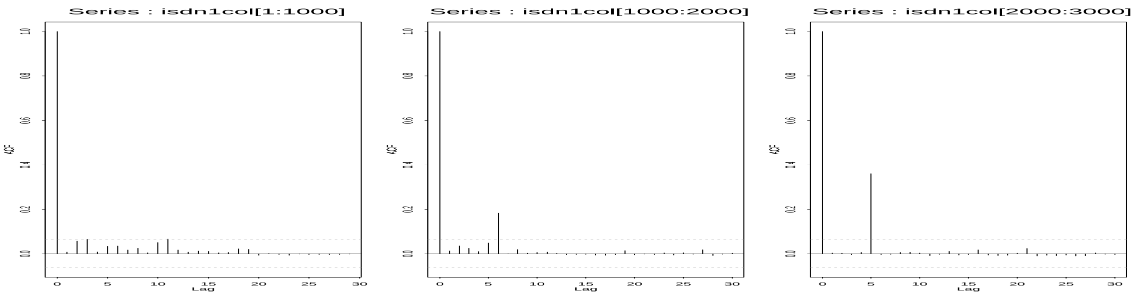

Figure 5.6. Acf of partitioned data.

6. Closing remarks

At this point in the development of the subject, it is clear that there is an abundance of data which seems to need heavy tailed modeling but there is not an abundance of heavy tailed models with a rich set of accompanying data fitting techniques. It is difficult to get autoregressions to fit given data sets and although this class offers attractive features for analysis, it is probably too small a class within the heavy tailed universe. There is a crying need to develop tools which will work with a larger class. The class of non-linear models and the class of hidden Markov models are two possible classes worthy of investigation. See Meier-Hellstern et al (1991) for a closely related example.

References

Andel, J., *Non-negative autoregressive processes*, J. Time Series Analysis **10** (1989), 1–11.

Athreya, K., *Bootstrap of the mean in the infinite variance case*, Ann. Statist. **15** (1987), 724–731.

Beran, Jan, *Statistics for Long-Memory Processes*, Monographs on Statistics and Applied Probability, Number 61, Chapman and Hall, New York, 1994.

Bhansali, R., *Consistent order determination for processes with infinite variance*, JRSS B **50, no. 1** (1988), 46–60.

Beran, J., Sherman, R., Taqqu, M., and Willinger, W., *Long–range dependence in Variable–bit rate video traffic*, To appear, IEEE Trans. Comm. (1995).

Bingham,N., Goldie, C. and Teugels, J., *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, UK, 1987.

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

Castillo, Enrique, *Extreme value theory in engineering*, Academic Press, San Diego, California, 1988.

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

Csorgo, S and Mason, D., *Central limit theorems for sums of extreme values*, Math. Proc. Camb. Phil. Soc. **98** (1985), 547–558.

Datta, S. and McCormick, W., *Bootstrap inference for a first order autoregression with positive innovations*, To appear December, 1995, JASA (1995).

Davis, R., Knight, K. and Liu, J., *M estimation for autoregressions with infinite variance*, Stochastic Processes and their Applications (1991).

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

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

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

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

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

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

Deheuvels, P., Mason, D. and Shorack, G., *Some results on the influence of extremes on the bootstrap*, Ann. Inst. Henri Poincare **29** (1993), 83–103.

Dekkers, A. and Haan, L. de, *On the estimation of the extreme value index and large quantile estimation*, Ann. Statist. **17** (1989), 1795–1832.

Dekkers, A. and de Haan, L., *Optimal choice of sample fraction in extreme value estimation*, Thesis, Erasmus University, Rotterdam (1991).

Dekkers, A., Einmahl, J., and Haan, L. de, *A moment estimator for the index of an extreme value distribution*, Ann. Statist. **17** (1989), 1833–1855.

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** (1994a), 135–166.

Feigin, P. and Resnick, S., *Linear programming estimators and bootstrapping for heavy tailed phenomena*, To appear: Advances in Applied Probability, Available as TR1124.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1994b).

Feigin, P. and Resnick, S., *Pitfalls of fitting autoregressive models for heavy-tailed time series*, To appear (1996).

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

Feigin, P., Kratz, M. and Resnick, S., *Parameter estimation for moving averages with positive innovations*, To appear: Annals of Applied Probability, Available as TR1111.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1994).

Garrett, M. and Willinger, W., *Analysis, modeling and generation of self-similar VBR video traffic*, Proceedings of ACM SigComm, London (1994).

Geluk, J. and Haan, L. de, *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Center for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands, 1987.

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

Gine, E. and Zinn, J., *Necessary conditions for the bootstrap of the mean*, Ann. Statist. **17** (1989), 684–691.

Grigoriu, M., *Applied Non–Gaussian Processes*, Prentice Hall, Englewood Cliffs, NY 07632, 1995.

Gumbel, E., *Statistics of Extremes*, Columbia University Press, New York, 1958.

Haan, L. de, *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Mathematical Centre, Amsterdam, Holland, 1970.

Haan, L. de, *Extreme Value Statistics*, Lecture Notes, Econometric Institute, Erasmus University, Rotterdam (1991).

Haan, L. de and Peng, L., *Rate of convergence for bivariate extremes (total variation metric)*, Report 9465/A, Econometric Institute, Erasmus University Rotterdam (1995a).

Haan, L. de and Peng, L., *Rate of convergence for bivariate extremes (uniform metric)*, Report 9466/A, Econometric Institute, Erasmus University Rotterdam (1995b).

Haan, L. de and Peng, L., *Exact rates of convergence to a symmetric stable law*, Report 9467/A, Econometric Institute,Erasmus University Rotterdam (1995c).

Haan, L. de and Peng, L., *Comparison of tail index estimators*, Report 9464/A, Econometric Institute,Erasmus University Rotterdam (1995).

Haan, L. de and Resnick, S., *Second–order regular variation and rates of convergence in extreme–value theory*, To appear, Ann. Probability (1995).

Haan, L. de and Stadtmüller, U., *Generalized regular variation of second order*, Preprint, Econometric Institute, Erasmus University Rotterdam (1995).

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

Hall, P., *Asymptotic properties of the bootstrap for heavy-tailed distributions*, Ann. Probability **18** (1990), 1342–1360.

Heath, D., Resnick, S. and Samorodnitsky, G., *Heavy tails and long range dependence in on/off processes and associated fluid models*, Available as TR1144.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1996).

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

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

Jansen, D. and de Vries, C., *On the frequency of large stock returns: putting booms and busts into perspective*, The Review of Economics and Statistics **LXXIII, No 1** (1991), 18–24.

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

Kinateder, J., *An invariance principle applicable to the bootstrap*, Exploring the Limits of Bootstrap (R. LePage and Lynne Billard, ed.), John Wiley and Sons, New York, 1992.

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

Knight, K., *On the bootstrap of the sample mean in the infinite variance case*, Ann. Statist. **17** (1989b), 1168–1175.

Koedijk, K., Schafgans, M. and de Vries, C., *The tail index of exchange rate returns*, Journal of International Economics **29** (1990), 93–108.

Kratz, M. and Resnick, S., *The qq–estimator and heavy tails*, Available as TR1122.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html (1995).

Leadbetter, M., Lindgren, G. and Rootzen, H., *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, 1983.

LePage, R., *Bootstrapping signs*, Exploring the Limits of Bootstrap (R. LePage and Lynne Billard, ed.), John Wiley and Sons, New York, 1992.

Livny, M., Melamed, B. and Tsiolis, A.K., *The impact of autocorrelations on queueing systems*, Management Sciences **39** (1993), 322-339.

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

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

Mason, D. and Turova, T., *Weak convergence of the Hill estimator process*, Extreme Value Theory and Applications (Galambos, J., Lechner, J. and Simiu, E., ed.), Kluwer Academic Publishers, Dordrecht, Holland, 1994.

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.

Neveu, Jacques, *Processus Ponctuels*, Ecole d'Eté de Probabilités de Saint-Flour VI. Lecture Notes in Mathematics, 598, Springer-Verlag, Berlin, 1976.

Resnick, Sidney, *Point processes, regular variation and weak convergence*, Advances Appl. Probability **18** (1986), 66-138.

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

Resnick, S., *Point processes and Tauberian theory*, The Mathematical Scientist (1991), 83–106.

Resnick, S., *Adventures in Stochastic Processes*, Birkhauser, Boston, 1992.

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

Resnick, S. and Samorodnitsky, G., *Performance Decay in a Single Server Exponential Queuing Model with Long Range Dependence*, To appear, Operations Research (1995).

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 Proability (1996a).

Resnick, S. and Stărică, Cătălin, *Smoothing the moment estimator of the extreme value parameter*, Available by ftp from ftp.orie.cornell.edu as TR1158.ps.Z in directory /ftp/pub/techreps or at http://www.orie.cornell.edu/trlist/trlist.html, Preprint (1996b).

Resnick, S. and Stărică, Cătălin, *On the asymptotic behavior of Hill's estimator for dependent data*, Forthcoming (1996c).

Rice, J., *Mathematical Statistics and Data Analysis*, Brooks/Cole Publishing, Pacific Grove, California, 1988.

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).

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

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

Smith, R., *Uniform rates of convergence in extreme value theory*, Advances in Appl. Probability **l4** (1982), 543–565.

Willinger, W., Taqqu, M., Leland, W. and Wilson, D., *Self–similarity in high–speed packet traffic: analysis and modeling of ethernet traffic measurements*, Statistical Science **10** (1995), 67–85.

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

Sidney I. Resnick, Cornell University, School of Operations Research and Industrial Engineering, ETC Building, Ithaca, NY 14853 USA

E-mail: sid@orie.cornell.edu