

THE INFINITE SOURCE MODEL FOR INTERNET TRAFFIC: STATISTICAL ANALYSIS AND LIMIT THEOREMS *

WALTER A. ROSENKRANTZ[†] AND J. HOROWITZ[‡]

1. Introduction. The discovery by Willinger et. al. that the input process to an Ethernet cable exhibits, at least empirically, *long range dependence* (LRD) and self-similarity has produced many models of this phenomenon, along with many, sometimes ad-hoc, statistical methods to analyze them ([1],[5],[14],[17]). It has been observed, for example, that detecting long range dependence (LRD) in internet traffic by fitting a straight line to a variance time (VT) plot can give misleading results; in particular, this method appears to be biased towards LRD even when the underlying process is known to be short range dependent (SRD) ([7],[8],[15]). In this paper we consider VT plots in the context of a readily-interpretable model for internet traffic that avoids the use of Gaussian processes, such as fractional Brownian motion (fBm) or fractional Gaussian noise (fGn), which take on negative values, to model a process that does not (Riedi et al. [14]). This model, which has been studied (e.g., in Guerin et al. [5]) under the rubric “infinite source Poisson data traffic model”, can capture the essential features of LRD and asymptotic self-similarity, but is simple enough so that closed-form expressions can be found for the variance-covariance function of the process. For our model, we show that the variance of the cumulative input is a *non-linear* function of time in log-log scale, which explains the curved appearance of VT plots for simulated and empirical data (see Figure 3.1). This suggests that nonlinear regression might yield more accurate estimates of the model parameters. These theoretical results are confirmed by simulations, real data, and the use of residual plots, the details of which were first reported in a separate paper [15]. Following the referee’s suggestion we have incorporated some of the numerical and graphical results presented there into Section 4 and summarized in Tables ((4.1), (4.2)). Our results, which apply to all SRD and LRD $M/G/\infty$ processes, generalize those obtained by Krunkz and Matta ([8]) for an important special case. A more detailed analysis of this model, given in Section 3, explains, simply and rigorously, why fitting a straight line to a VT plot is *always* biased towards LRD, even when the process is known to be SRD.

A separate, but closely related question, is obtaining a limit theorem for the cumulative input traffic under “heavy traffic” that is consistent with empirical internet traffic measurements, a topic that has attracted much attention in the current literature ([5],[10],[14],[16],[17]). In some of these papers the limit process is not unique; it depends on how two parameters, the arrival rate (also called the connection rate) and the time scale, go to infinity. In this paper, we derive a central limit type theorem for the cumulative input process that differs significantly from those obtained by others. We show that as the arrival rate of network users tends to infinity the cumulative input process, when suitably normalized, tends (in the sense of weak convergence of processes) to a Gaussian limit *which is not fBm* and with a covariance function that depends on the service time distribution (Section 5). As a consequence of this result,

*Received July 3, 2002; accepted for publication February 6, 2003.

[†]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305, USA (rkrantz@math.umass.edu).

[‡]Department of Mathematics and Statistics, Lederle Graduate Research Center, University of Massachusetts, 710 N. Pleasant Street Amherst, MA 01003, USA (joeh@math.umass.edu).

when the intensity of user arrivals is large, the cumulative input process is approximately a deterministic increasing function of time plus a zero-mean Gaussian noise. Although this is similar in spirit to some results in [10] and in [16], our results are not comparable because their limit theorems concern the asymptotic behaviour of the input process over large time scales, whereas we study its asymptotic behavior over a fixed time interval.

2. The $M/G/\infty$ input process. One of the simplest plausible models of network traffic possibly exhibiting LRD is undoubtedly the $M/G/\infty$ input process, where customers arrive according to a homogeneous Poisson process with arrival rate λ (also called the connection rate), service time distribution $G(y) = P(S \leq y)$, finite mean $\mu = E(S)$, but infinite variance (Parulekar and Makowski ([11])). Instead of the discrete time model considered by them, we consider a continuous time, $M/G/\infty$ queueing system of the sort described in section 5.6(iii) of Cox and Isham ([4]), but interpreted as follows: Network users enter the system according to a homogeneous Poisson process of rate λ and remain in it for a random time S , having distribution G , during which they generate packets at a constant unit rate. We denote by $Z_\lambda(t)$ the number of users in the system at time t . In queueing parlance, $Z_\lambda(t)$ corresponds to the number of busy servers, and the time that each user is in the system corresponds to the service time. We then use a *fluid approximation* to the total number $X_\lambda(t)$ of packet arrivals during the time interval $[0, t]$, which is given by

$$X_\lambda(t) = \int_0^t Z_\lambda(s) ds \tag{1}$$

In Sections (1 – 4) we consider the situation in which $Z_\lambda(t)$ is stationary, which is the case relevant for the analysis of VT plots; in Section (6) we indicate how the main limit theorem can be extended to the nonstationary case.

It is shown below that $X_\lambda(t)$ is LRD when the distribution of S is “heavy tailed” in the sense of Pareto:

$$\bar{G}(y) = P(S > y) \sim Ky^{-\theta}, \quad 1 < \theta < 2, \quad K > 0 \tag{2}$$

We note that $E(S) < \infty$ if $\theta > 1$ and $E(S^2) < \infty$, if $\theta > 2$.

Notation: Here, and elsewhere, we use the letters K, K_1 etc., to denote constants whose exact values are of no particular interest, and we use the usual notation for asymptotic equivalence:

$$f(t) \sim g(t) \text{ means } \lim_{t \rightarrow \infty} \frac{f(t)}{g(t)} = 1$$

The asymptotic behaviour of the variance of the sample mean $v(t)$, where

$$v(t) := \text{var}(X_\lambda(t)/t), \tag{3}$$

determines whether the process $X_\lambda(t)$ is LRD or SRD. A process is LRD when $v(t)$ decreases to zero at a rate $t^{-\alpha}$, where $0 < \alpha < 1$; that is, the variance of the sample mean decreases to zero at a rate slower than t^{-1} ([1], p.42). In particular, when the distribution of S is heavy tailed in the sense of Pareto then Lemma 2.1 implies that $\alpha = \theta - 1$, that is,

$$v(t) \sim K_1 t^{1-\theta}, \quad 1 < \theta < 2.$$

These results follow from the following well known formulas for the mean, auto-covariance and variance functions of the processes $Z_\lambda(t)$ and $X_\lambda(t)$ (Cox and Isham [4], p.138-139 and Parzen [12], Equation 3.13, p.80, modified here for the stationary case).

$$E(Z_\lambda(t)) = \lambda\mu; E(X_\lambda(t)) = \lambda\mu t \tag{4}$$

$$\text{cov}(Z_\lambda(t), Z_\lambda(t + u)) = C_\lambda(u) = \lambda \int_{|u|}^\infty P(S > y)dy \tag{5}$$

$$\text{var}(X_\lambda(t)) = 2 \int_0^t du \int_0^u C_\lambda(u - v)dv \tag{6}$$

Formulas (5) and (6) for the auto-covariance and variance functions are most conveniently expressed in terms of the function $g(u)$, defined as follows:

$$g(u) = \int_0^u ds \int_0^s C(r)dr, C(u) = \int_{|u|}^\infty P(S > y)dy \tag{7}$$

Notice that $C(u)$ is the covariance function of $Z_\lambda(t)$ for $\lambda = 1$, and that $C(0) = E(S)$. Moreover,

$$\text{cov}(Z_\lambda(t), Z_\lambda(t + u)) = \lambda C(u) \tag{8}$$

$$v(t) = \text{var}\left(\frac{X_\lambda(t)}{t}\right) = \frac{2\lambda g(t)}{t^2} \tag{9}$$

$$\text{var}(X_\lambda(t) - X_\lambda(s)) = 2\lambda g(t - s) \tag{10}$$

$$\text{cov}(X_\lambda(s), X_\lambda(t)) = \lambda(g(s) + g(t) - g(t - s)) \tag{11}$$

It follows at once from Equation (5) that $Z_\lambda(t)$ is covariance stationary and $C_\lambda(u) < \infty$ if and only if $E(S) < \infty$; so, throughout this paper, we assume the service time distribution has a finite mean.

The properties of the function $g(u)$, listed below, follow immediately from its definition in Equation (7); they play a useful role in our statistical analysis of VT plots.

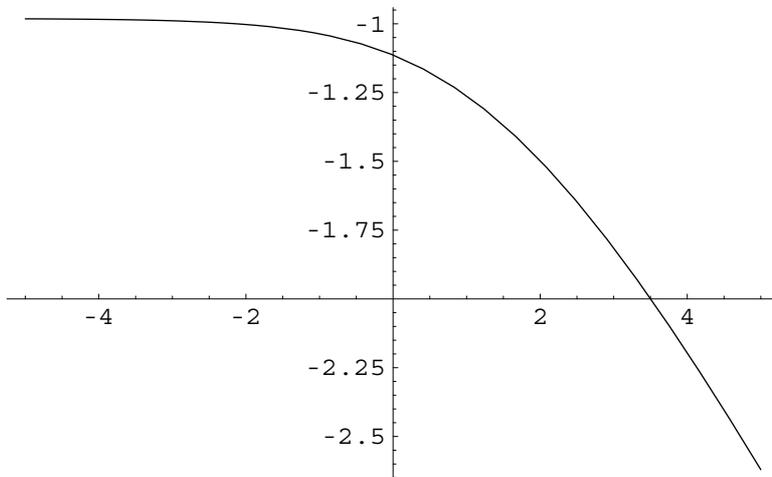
$$g(0) = g'(0) = 0, g''(u) = C(u) > 0, g'''(u) = C'(u) < 0 \tag{12}$$

$$\frac{d}{du} \left(\frac{g(u)}{u^2}\right) < 0 \tag{13}$$

$$1 < \frac{ug'(u)}{g(u)} < 2, 0 < u < \infty \tag{14}$$

$$\lim_{u \rightarrow 0} g(u)/u^2 = C(0)/2, \lim_{u \rightarrow \infty} g(u)/u^2 = 0 \tag{15}$$

It follows from Equation (7) that $C(u)$ is absolutely continuous. Consequently, the functions g, g', g'' , always exist, are each continuous, and g'' is absolutely continuous with $g'''(y) = P(S > y)$, (*a.e.*). Equation (15) follows at once from a suitable application of l'Hopital's rule. Equation (14) follows from Equation (12) which implies that $g(u)$ is positive, convex, increasing and $g(0) = 0$; consequently, $g(u)/u < g'(u)$, from which the first inequality follows. The derivations of the other inequalities listed in (13) and (14) are given in Section 6. We shall use the inequality given in Equation (14)

FIG. 3.1. Graph of $\ln[v(e^u)]$.

to prove that the widely used technique of estimating model parameters via simple linear regression is always biased towards detecting LRD, even when the underlying model is SRD (see Section 3).

LEMMA 2.1. (i) Assume $\overline{G}(y) \sim Ky^{-\theta}$, $1 < \theta < 2$; then

$$g(t) \sim K_1 t^{3-\theta}, \quad g''(t) \sim K_2 t^{1-\theta}, \quad 1 < \theta < 2 \quad (16)$$

(ii) Assume $\mu_2 = E(S^2) < \infty$; then

$$g(t) \sim \mu_2 t \quad (17)$$

Proof. To prove (16) and (17) apply l'Hopital's rule to compute $\lim_{t \rightarrow \infty} g(t)/t^{3-\theta}$, $\lim_{t \rightarrow \infty} g''(t)/t^{1-\theta}$, and $\lim_{t \rightarrow \infty} g(t)/(\mu_2 t)$, respectively.

3. Variance-Time Plots. A Variance-Time (VT) plot is a graph, in log-log scale, of $v(t) = \text{var}(X_\lambda(t)/t)$ against t , i.e., it is a plot of the points

$$(u, \ln[v(e^u)]), \quad -\infty < u < \infty \quad (18)$$

where $u = \ln t$. We call the graph of Eq. (18) a theoretical VT plot. Figure (3.1) is a theoretical VT plot for the case $\overline{G}(y) = (1+y)^{-\theta}$, $\theta = 1.5$, and λ chosen so that $\lambda = (\theta - 1)(2 - \theta)(3 - \theta)/2$.

Figure (3.1) suggests that $\ln[v(e^u)]$ is asymptotically linear in u . This is a consequence of the fact that $\ln v(t) \sim b_0 + b_1 \ln t$, which we prove below [cf. Lemma(3.4)]. More precisely, the graph of $v(t)$ on a log-log scale is asymptotically linear in $\ln t$, with asymptotic slope $b_1 = -1$ when the input process is SRD, and $b_1 > -1$ otherwise. Thus, detecting LRD is equivalent to showing that $b_1 > -1$. It is natural to estimate b_1 via linear regression, that is by fitting a straight line to that portion of the curve

which is nearly linear. Unfortunately, as we now show, this method is always positively biased, even when the input process is SRD. The proofs of Lemmas (3.1) and (3.2) below are given in Section 6.

LEMMA 3.1. *Suppose we use the method of least squares to fit a straight line to the data set (u_i, y_i) , $i = 0, 1, \dots, n$, obtained by sampling a continuous function $f(u)$, $a \leq u \leq b$ at the equally spaced points $u_i = a + i(b - a)/n$, $y_i = f(u_i)$. Denote the slope of the fitted line by $\beta_{1,n}(f)$ and $\bar{x} = (a + b)/2$. Then,*

$$\lim_{n \rightarrow \infty} \beta_{1,n}(f) = \beta_1(f) = \frac{12 \int_a^b (x - \bar{x})f(x)dx}{(b - a)^3} \tag{19}$$

LEMMA 3.2. *If f is continuous, monotone increasing on $[a, b]$, and $f(b) - f(a) > 0$, then $\beta_1(f) > 0$. In particular, if $f'(x) > c$ on $[a, b]$, then $\beta_1(f) > c$.*

LEMMA 3.3. $\beta_1(\ln[v(e^u)]) > -1$

Proof. We first note that

$$\frac{d \ln[v(e^u)]}{du} = \frac{e^u g'(e^u)}{g(e^u)} - 2 > 1 - 2 = -1; \tag{20}$$

which is a consequence of the fact that $tg'(t)/g(t) > 1$ (see Equation (14)). The result follows by applying Lemma 3.2 to the function $f(u) = \ln[v(e^u)] = \ln(2\lambda) + \ln[g(e^u)] - 2u$.

From Equation (20), Lemmas (3.1) and (3.2), we see that the slope of $\ln[v(e^u)]$ on any finite time interval is always strictly greater than -1 , and that estimating b_1 via simple linear regression is always biased towards LRD, even when the underlying model is SRD. In other words, the estimate of b_1 is biased even when we are sampling from the exact model.

We now give a more explicit description of the asymptotic behavior of $v(t)$.

LEMMA 3.4. *(i) Assume $E(S^2) = \infty$ with heavy tailed Pareto service time distribution $\bar{G}(y) \sim Ky^{-\theta}$, $1 < \theta < 2$; then*

$$\lim_{t \rightarrow \infty} \ln(v(t)) - (b_0 + b_1 \ln t) = 0 \tag{21}$$

$$b_0 = \ln \left(\frac{2\lambda K}{(3 - \theta)(2 - \theta)(\theta - 1)} \right), b_1 = 1 - \theta \tag{22}$$

(ii) Assume $\mu_2 = E(S^2) < \infty$; then

$$\lim_{t \rightarrow \infty} \ln(v(t)) - (b_0 + b_1 \ln t) = 0 \tag{23}$$

$$b_0 = \ln(2\lambda\mu_2), b_1 = -1$$

(iii) When $\theta = 2$ in Equation (2) then Equation (21) is modified as follows:

$$\lim_{t \rightarrow \infty} \ln(v(t)) - b_0 - \ln t + \ln(\ln(t) - 1) = 0 \tag{24}$$

$$b_0 = \ln(2\lambda K)$$

Proof. Equations (21) and (23) follow immediately from Eq.(16) and the fact that $v(t) = 2\lambda g(t)/t^2$.

We next derive a closed form expression for $\ln(v(t))$ when the tail of the service time distribution is given by $\overline{G}(y) = (1 + y)^{-\theta}$.

LEMMA 3.5. *Assume*

$$G(y) = 1 - (1 + y)^{-\theta}, \quad 0 < y < \infty,$$

then

$$\text{var}(X_\lambda(t)) = 2\lambda a(\theta)((1 + |t|)^{3-\theta} - (3 - \theta)|t| - 1), \quad (1 < \theta < 2) \quad (25)$$

$$\text{var}(X_\lambda(t)) = 2\lambda((1 + t) \log(1 + t) - t), \quad \theta = 2 \quad (26)$$

$$\text{var}(X_\lambda(t)) = \lambda(t - \log(1 + t)), \quad \theta = 3 \quad (27)$$

Proof. In this case the functions $g(u)$ and $C(u)$ can be explicitly computed by substituting $\overline{G}(y)$ into Equation (7); thus,

$$C(u) = \frac{1}{\theta - 1} (1 + |u|)^{-\theta+1} \quad (28)$$

$$\begin{aligned} g(u) &= \int_0^u ds \int_0^s \frac{1}{\theta - 1} (1 + |r|)^{-\theta+1} dr \\ &= a(\theta)((1 + |u|)^{3-\theta} - (3 - \theta)|u| - 1) \end{aligned} \quad (29)$$

$$a(\theta) = \frac{1}{(\theta - 1)(2 - \theta)(3 - \theta)}$$

Inserting these expressions for $g(u)$ into Equation (9) yields the exact formulas for $\text{var}(X_\lambda(t))$.

A nonlinear asymptotic expansion of $\ln(v(t))$: Equation (25) suggests that a nonlinear asymptotic expansion might be more appropriate. The following more refined asymptotic expansion of Equation (25),

$$\ln v(t) \approx b_0 + b_1 \ln(t) + b_2 t^{-(b_1+1)}, \quad t \rightarrow \infty \quad (30)$$

is obtained in the usual way by noting that

$$\ln((1 + t)^{3-\theta} - (3 - \theta)t - 1) = (3 - \theta) \ln(t) - (3 - \theta)t^{\theta-2} + o(t^{\theta-2}), \quad t \rightarrow \infty,$$

where we used the approximations

$$\begin{aligned} (1 + t^{-1})^{3-\theta} &= 1 + (3 - \theta)t^{-1} + o(t^{-1}), \quad t \rightarrow \infty \\ \ln(1 - (3 - \theta)t^{\theta-2}) &= -(3 - \theta)t^{\theta-2} + o(t^{\theta-2}), \quad t \rightarrow \infty \end{aligned}$$

A graphical and numerical comparison of estimates of the model parameters based on empirical VT plots using linear and non linear regression is presented in Section (4).

4. Empirical Variance-Time Plots . The $X_\lambda(t)$ process, as previously noted, is LRD when $1 < \theta < 2$, and SRD when $\theta > 2$. It is, therefore, a question of no small importance to be able to determine whether or not the process is LRD by estimating from the empirical data the unknown parameter θ . This does not appear to be an easy problem. Resnick ([13]), for example, has studied the tail behaviour of generalized Pareto distributions using a variety of methods; unfortunately, in 6 out of the 11 cases studied he obtained negative values for θ , which suggests the problem is still open. The differences between LRD and SRD are revealed by studying the asymptotic behavior of the variances of the discrete time series $Z^T(l)$, ($l = 1, 2, \dots$) obtained by averaging the original time series over non overlapping blocks of size T , that is

$$Z^T(l) = \frac{1}{T} \int_{(l-1)T}^{lT} Z_\lambda(u)du, \quad (l = 1, 2, \dots) \tag{31}$$

It is easy to verify that the sequence $Z^T(l)$ ($l = 1, 2, \dots$) is also covariance stationary, and *asymptotically second order self similar*— we omit the details.

Variance-time (VT) plots are a widely used method for estimating the Hurst parameter $H = (3-\theta)/2$ that characterizes the tail behaviour of the service time distribution ([1], [7], [17]). A VT plot is an empirical sample of the graph $(\log(T), \log(v(T)))$, where T is the block size defined above (cf. Equation (31)) and $v(T)$ has been defined in Equation (3). We estimate $\log v(T)$ from the empirical data by computing the sample variances obtained from the time series $Z^T(l)$, ($l = 1, 2, \dots$) (Equation(31)).

The variance-time plot is the scatterplot of the data set (x_i, y_i) ($i = 1, 2, \dots, n$), where

$$x_i = \log T_i, \quad i = 1, \dots, n \tag{32}$$

$$y_i = \log s^2(T_i), \tag{33}$$

where $s^2(T_i)$ is the sample variance of $Z^{T_i}(l)$, $l = 1, \dots, m_{T_i}$.

The traditional implementation of the VT plot is to fit a straight line $b_0 + b_1x$ to the scatterplot, using simple linear regression, and then estimate H via the equation $\hat{H} = (2+b_1)/2$, ([1], p.92). This is equivalent to replacing the nonlinear function $v(T)$ with its asymptotic linear approximation defined by Equation (21). However, Krnunz and Makowski have shown that this implementation of the VT plot can incorrectly suggest LRD even in cases when the simulated data, generated using a discrete time $M/G/\infty$ model, is short range dependent ([7]). Their work suggests that the usual VT plot is strongly biased towards LRD. This should not come as a surprise because, as noted earlier, the function $v(T)$ is a nonlinear function of the block size T , which suggests that one can obtain better estimates of the model parameters using nonlinear regression analysis. The following example illustrates why using linear regression yields a positively biased estimate for the Hurst parameter H .

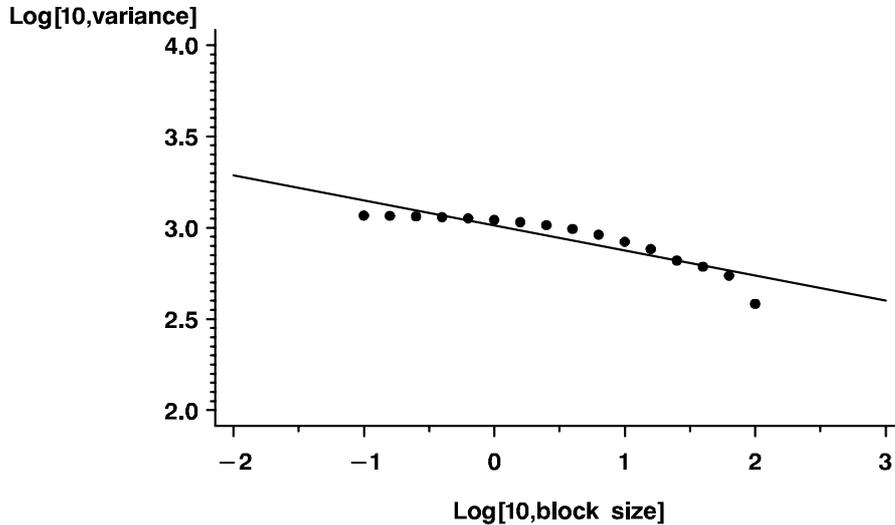
Note: For the example considered in this paper we chose the time blocks T_i according to the following scheme:

$$\log T_i = \log_{10}(i \times 0.2) - 2, \quad i = k, \dots, 20, \quad \text{where}$$

k is chosen large enough to avoid numerical singularities in the nonlinear curve fitting algorithm.

Example: Figure 4.1 is the variance-time plot for a simulated trace of the $M/G/\infty$ input process with $\lambda = 300$, $\theta = 1.5$, and $H = 0.75$. Here $b_0 =$

V – T Plot, lambda = 300 theta = 1.5



Regression Equation:
 $V = 3.01142 - 0.137348 * M$

Estimated Regression Line and Scatter plot

FIG. 4.1.

3.01142, $b_1 = -0.1483$, so the estimate for the Hurst parameter in this case is $\hat{H} = (2 - 0.1483)/2 = 0.9259 > 0.75$, a very poor estimate, indeed. The reason for the positive bias for the estimated value of H is clear from Figure (4.1): The points of the scatter plot lie on a concave curve, which tends to make the slope of the fitted line greater than the asymptotic slope, which equals $2H - 2$. Consequently,

$$\hat{H} = \frac{b_1 + 2}{2} > \frac{(2H - 2) + 2}{2} = H$$

In Figure 4.1, $V = \log v(T)$ and $M = \log T$.

Looking at Figure 4.1 we see a small but non-negligible curvature in the scatter plot suggesting that the linear model may not be appropriate. This is confirmed by the *studentized residual plot*(cf. [9]) displayed in Figure 4.2. In our situation the usual assumptions for simple linear regression definitely *do not* hold, but we use the (studentized) residuals as an informal tool to assess the adequacy of the linear fit. The residual plot in Figure 4.2 provides additional evidence for using nonlinear regression to estimate the model parameters instead of the linear (in $\log T$) model of Equation (21). In particular, we used Mathematica’s nonlinear regression program to fit the exact, closed form, expression for $\log v(T)$ obtained from Equation (18). Similar results were obtained for the Bellcore pAugTL data set (source:[6]). These examples confirm our theoretical conclusions that the linear regression estimate of the Hurst parameter H is biased towards LRD and is greater than the non-linear estimate (see Tables (4.1) and (4.2)). For example, the linear regression estimate for

Residual Plot, $\lambda = 300, \theta = 1.5$

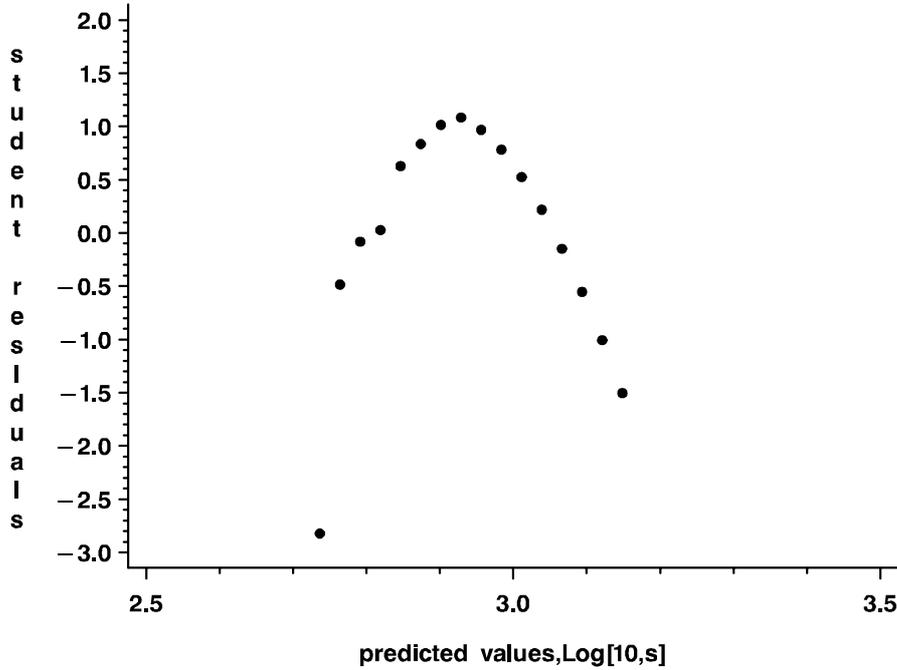


FIG. 4.2.

H from the simulated $M/G/\infty$ data is 0.9259, which is about 23% larger than the correct value $H = 0.75$; it is also 11% larger than 0.8357, the estimate obtained via nonlinear regression. The discrepancy between the two estimates for the Bellcore data is even more striking: The linear regression estimate is 0.7679 and the nonlinear one is 0.6030. In this case the linear regression estimate is approximately 27% greater than the nonlinear one. This may explain why the variance-time plot sometimes incorrectly indicates the presence of LRD.

Model	slope	$\hat{\theta}$	\hat{H}	$\hat{\lambda}$
linear	-0.1483	1.1483	0.9259	120.057
non-linear	na	1.3286	0.8357	399.59

TABLE 4.1. *Parameter estimates for simulated data.*

Model	slope	$\hat{\theta}$	\hat{H}	$\hat{\lambda}$
linear	-0.4642	1.4642	0.7679	1.09×10^6
non-linear	na	1.7940	0.6030	6.94×10^7

TABLE 4.2. *Parameter estimates for Bellcore data.*

5. A limit theorem for the rescaled $X_\lambda(t)$ process . The marginal distributions of the $Z_\lambda(t)$ process (in the stationary case) are Poisson with mean $\lambda\mu$; this is a standard result for the $M/G/\infty$ queue (Parzen [12], Ex. 5E, p.147). Consequently, as $\lambda \rightarrow \infty$ the distribution of the normalized variable $Z_\lambda^*(t)$ defined by

$$Z_\lambda^*(t) = \frac{Z_\lambda(t) - \lambda\mu}{\sqrt{\lambda}} \tag{34}$$

converges weakly, for each t , to a normal variable with mean 0 and variance μ . Indeed, it can be shown that the finite dimensional distributions of the process $Z_\lambda^*(t)$ converge weakly to those of a Gaussian process denoted by $Z^*(t)$. More precisely, we can prove the following lemma.

LEMMA 5.1. For $0 \leq t_1 < t_2 < \dots < t_k$,

$$(Z_\lambda^*(t_1), \dots, Z_\lambda^*(t_k)) \Rightarrow (Z^*(t_1), \dots, Z^*(t_k)), \lambda \rightarrow \infty,$$

where $(Z^*(t_1), \dots, Z^*(t_k))$ is a Gaussian random vector, with mean zero, and covariance matrix

$$\Gamma(t_i, t_j) = E(Z^*(t_i)Z^*(t_j)) = \int_{|t_i-t_j|}^\infty (1 - G(y)) dy \tag{35}$$

and \Rightarrow denotes weak convergence.

Proof. See Section (6).

This suggests, and we shall prove, that the finite dimensional distributions of the normalized processes $X_\lambda^*(t)$ defined by

$$X_\lambda^*(t) = \frac{X_\lambda(t) - \lambda\mu t}{\sqrt{\lambda}} \tag{36}$$

converge weakly, as $\lambda \rightarrow \infty$, to a Gaussian process $X^*(t)$.

LEMMA 5.2. For $0 \leq t_1 < t_2 < \dots < t_k$,

$$(X_\lambda^*(t_1), \dots, X_\lambda^*(t_k)) \Rightarrow (X^*(t_1), \dots, X^*(t_k)), \lambda \rightarrow \infty,$$

where $(X^*(t_1), \dots, X^*(t_k))$ is a Gaussian random vector, with mean zero, and covariance matrix K given by

$$K(s, t) = g(s) + g(t) - g(t - s) \tag{37}$$

Proof. See Section (6).

It is to be observed that the covariance function of the limit Gaussian process $X^*(t)$ is not independent of the service time distribution G . This is a consequence of the fact that $g(t)$ is also a function of G . In the special case when $\overline{G}(y) = (1 + y)^{-\theta}$, discussed in Lemma 3.5, it follows from Equation (37) that

$$K(s, t) = a(\theta) ((1 + |s|)^{2H} + (1 + |t|)^{2H} - (1 + |t - s|)^{2H} - 2H(|s| + |t| - |t - s|) - 1),$$

where $H = (3 - \theta)/2$ is the so-called Hurst parameter.

Proof. See Section (6).

It follows from Equation (37) that all weak limits have the same covariance function $K(s, t)$. This, together with tightness of the family of processes $\{X_\lambda^*(t) : \lambda > 0\}$, which we shall prove momentarily, shows that the finite-dimensional convergence in Lemma 4.2 can be strengthened to weak convergence of $X_\lambda^*(t)$ to the Gaussian process $X^*(t)$ having mean zero and covariance $K(s, t)$.

THEOREM 5.1. *The family $\{X_\lambda^*(t) : \lambda > 0\}$ is relatively compact in the space $D_R[0, \infty)$.*

Proof. The proof is a straightforward application of Theorem 15.6, p.128 of Billingsley ([2]). It suffices to show, for $t_1 < t < t_2$, that

$$E(|X_\lambda^*(t) - X_\lambda^*(t_1)||X_\lambda^*(t_2) - X_\lambda^*(t)|) \leq \frac{\mu}{4}(t_2 - t_1)^2 \tag{38}$$

The inequality (38) implies that the hypotheses of Billingsley’s Theorem 15.6 are satisfied with $\gamma = 1$, $\alpha = 1$, and $F(t) = (\sqrt{\mu}/2)t$. In detail:

$$\begin{aligned} E(|X_\lambda^*(t) - X_\lambda^*(t_1)||X_\lambda^*(t_2) - X_\lambda^*(t)|) &= \frac{1}{\lambda} \int_t^{t_2} \int_{t_1}^t E((|Z_\lambda(s) - \lambda\mu|)(|Z_\lambda(v) - \lambda\mu|)) ds dv \\ &\leq \frac{1}{\lambda} \int_t^{t_2} \int_{t_1}^t \lambda\mu ds dv \\ &= \mu(t - t_1)(t_2 - t) \leq \mu \frac{(t_2 - t_1)^2}{4} \end{aligned}$$

The inequality is a consequence of Schwarz’s inequality and the fact that the univariate distributions of the $Z_\lambda(t)$ process are Poisson with parameter $\lambda\mu$.

We have thus proved the following theorem:

THEOREM 5.2. *The normalized processes $X_\lambda^*(t)$ converge weakly to the Gaussian process $X^*(t)$ with covariance function given by Equation (37).*

We observe that this result is valid for an arbitrary service-time distribution G . As a consequence of Theorem 5.2, when the intensity λ of customer arrivals is large, the equation for the cumulative input process

$$X_\lambda(t) = \lambda\mu t + \sqrt{\lambda}X_\lambda^*(t),$$

which follows from Equation (36), shows that $X_\lambda(t)$ is an increasing linear deterministic function of t plus a term that is approximately Gaussian noise. In the more general nonstationary case, considered in Section 6, the deterministic term above is replaced by $\lambda \int_0^t \mu(s) ds$, where $\mu(t)$ is defined in section 6, Equation (54). It is noteworthy that $K(Ts, Tt)/T^{2H}$, the covariance functions of the normalized processes $X_T(t) = X^*(Tt)/T^H$, has an asymptotic limit independent of the service time distribution, that is,

$$\lim_{T \rightarrow \infty} \frac{K(Ts, Tt)}{T^{2H}} = K_1 (|s|^{2H} + |t|^{2H} - |t - s|^{2H}) \tag{39}$$

This expression is the covariance function of fractional Brownian motion (fBm). In fact, it can be shown that X_T converges weakly to fBm as $T \rightarrow \infty$.

There are results in the same spirit due to Mikosch et al. [10] dealing with the cumulative input during the time interval $[0, Tt]$, but with the customer arrival intensity $\lambda = \lambda(T)$ being an increasing function of T . Depending on the tail behavior of the service-time distribution, the limit process as $T \rightarrow \infty$ can be either fBm or a stable Lévy motion. These results differ from the case considered here, where the time interval is fixed, and only the arrival rate goes to infinity. Finally, we note that, contrary to a statement in [5], section 2.5, the result of our Theorem 5.2 was not obtained by Taquq et al, whose paper [16] deals only with the so-called ON/OFF model.

6. Mathematical details and proofs. Derivation of inequalities (13) and (14): We derive inequality (13) by noting that $g'(u)$ is a concave, increasing function for $0 < u < \infty$; this, together with the fact that $g'(0) = 0$, implies that $g''(u) < g'(u)/u$; equivalently,

$$ug''(u) - g'(u) < 0, \quad 0 < u < \infty$$

It is easy to see that

$$\frac{d}{du} \left(\frac{g(u)}{u^2} \right) = \frac{ug'(u) - 2g(u)}{u^2}$$

Consequently, it suffices to show that $h(u) = ug'(u) - 2g(u) < 0$, $0 < u < \infty$. This follows at once from the fact that

$h(0) = 0$ and $h'(u) = ug''(u) - g'(u) < 0$. The inequality $g'(u) < 2g(u)/u$ is equivalent to the assertion $2g(u) - ug'(u) = -h'(u) > 0$ that we just derived. This completes the derivation of the second inequality of (14).

Proof of Lemma 3.1. The following equation for $\beta_{1,n}(f)$, the slope of the fitted line, is well known.

$$\beta_{1,n}(f) = \frac{\sum (x_i - \bar{x})f(x_i)}{\sum (x_i - \bar{x})^2} \quad (40)$$

$$x_i = a + i(b-a)/n, \quad \bar{x} = \left(\sum_{0 \leq i \leq n} x_i \right) / (n+1) = (a+b)/2$$

Dividing the numerator and denominator of the expression on the right hand side by n and letting $n \rightarrow \infty$ we see that the numerator and denominator are Riemann sums whose limits are $\int_a^b (x - \bar{x})f(x)dx$ and $(b-a)^3/12$, respectively. This completes the proof.

Proof of Lemma 3.2. It suffices to prove this for the case $a = 0$ and $b = 1$, in which case we have

$$\beta_1(f) = 12 \int_0^1 (t - 1/2)f(t) dt$$

The change of variables $s = t - 1/2$ yields

$$\beta_1(f) = 12 \int_0^1 (t - 1/2)f(t) dt = 12 \int_0^{1/2} t[f(1/2 + t) - f(1/2 - t)] dt$$

It is easy to see that f continuous, monotone increasing, with $f(1) - f(0) > 0$ on $[0, 1]$ implies that $f(1/2 + t) - f(1/2 - t) > 0$ on an open subinterval of $[0, 1/2]$; therefore $\beta_1(f) > 0$. Finally, we note that $f'(x) > c$ implies $f(x) - cx$ is monotone increasing; therefore $\beta_1(f(x) - cx) = \beta_1(f) - c > 0$.

Proof of Theorem 5.2. We consider the stationary case first, and then indicate the minor modifications required in order to extend the proof to the nonstationary case. Our proof requires the computation of the joint moments for both the $X_\lambda^*(t)$ and $X^*(t)$ processes, which are given by the following expressions.

$$E(X_\lambda^*(t_1) \cdots X_\lambda^*(t_k)) = \int_{[0,t_1]} \cdots \int_{[0,t_k]} E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right) du_1 \dots du_k \quad (41)$$

Replacing $X_\lambda^*(t)$ and $Z_\lambda^*(t)$ by $X^*(t)$ and $Z^*(t)$, respectively, we obtain a similar expression for the joint moments of the $X^*(t)$ process:

$$E(X^*(t_1) \cdots X^*(t_k)) = \int_{[0,t_1]} \cdots \int_{[0,t_k]} E \left(\prod_{i=1}^k Z^*(u_i) \right) du_1 \dots du_k \quad (42)$$

The proof is a consequence of lemmas (5.1), (5.2), and Theorem 5.1, the latter having been already proved.

Proof of Lemma 5.1. Using the reproductive properties of the Poisson process we can represent the Poisson arrival process $N_\lambda(t)$ feeding the $M/G/\infty$ queue as the sum

$$\sum_{i=1}^{[\lambda]} M_i(t) + M_\delta(t),$$

where $M_i(t)$, $i = 1, \dots, [\lambda]$ are independent Poisson processes, with rate 1 and $M_\delta(t)$ is a Poisson process with rate $\delta = \lambda - [\lambda]$, independent of $M_i(t)$, $i = 1, \dots, [\lambda]$. Thus the input stream of users is the sum of $[\lambda] + 1$ independent Poisson input streams. Let $\{T_{im} : m \geq 1\}$ denote the sequence of arrival times for the process $M_i(t)$, $i = 1, \dots, [\lambda]$, and similarly for $\{T_{\delta m} : m \geq 1\}$. Also let S_{im} be the time in the system for the m th user from stream i , $i = 1, \dots, [\lambda]$, and similarly for $\{S_{\delta m} : m \geq 1\}$. The families of random variables $\{T_{im} : m \geq 1\}$, $\{S_{im} : m \geq 1\}$, $\{T_{\delta m} : m \geq 1\}$, $\{S_{\delta m} : m \geq 1\}$ are all mutually independent, and, moreover, the random variables $\{S_{im}, S_{\delta m} : i = 1, \dots, [\lambda], m \geq 1\}$ are iid with service time distribution G .

Denote the number of users in the system at time t corresponding to the input processes $M_i(t)$ and $M_\delta(t)$ by $Y_i(t)$ and $Y_\delta(t)$, respectively. Then

$$Y_i(t) = \sum_{m=1}^{\infty} I_{\{T_{im} \leq t < T_{im} + S_{im}\}},$$

with a similar expression for $Y_\delta(t)$. The processes $Y_i(t)$, $i = 1, \dots, [\lambda]$, $Y_\delta(t)$ are therefore independent, and $Y_i(t)$, $i = 1, \dots, [\lambda]$ are iid.

Observe that the normalized process $Y_i^*(t)$ is identical in law to $Z_1^*(t)$. Consequently, we have the following representation of the $Z_\lambda^*(t)$ process:

$$Z_\lambda^*(t) = \sum_{i=1}^{[\lambda]} Y_i^*(t) + Y_\delta^*(t), \tag{43}$$

Thus, the $Y_i^*(t)$, $i = 1, \dots, [\lambda]$ are $[\lambda]$ iid copies of the process $Z_1^*(t)$, $Y_\delta^*(t)$ is independent of the $Y_i^*(t)$ processes and has the same distribution as $Z_{\lambda-[\lambda]}^*(t)$. Applying Chebychev's inequality to $Y_\delta(t_i)$ yields the result that

$$\lim_{\lambda \rightarrow \infty} P(|Y_\delta(t) - \delta\mu| > \lambda^{1/2}\epsilon) \leq \lim_{\lambda \rightarrow \infty} \frac{\delta\mu}{\lambda\epsilon^2} = 0$$

Consequently,

$$\frac{(Y_\delta^*(t_1), \dots, Y_\delta^*(t_k))}{\sqrt{\lambda}} \rightarrow_p \mathbf{0} \text{ as } \lambda \rightarrow \infty,$$

where \rightarrow_p denotes convergence in probability.

The proof is completed by applying the multidimensional central limit theorem to the iid sequence of random vectors $(Y_i(t_1), \dots, Y_i(t_k))$, $i = 1, \dots, [\lambda]$ with covariance function $\Gamma(t_i, t_j)$ given in Equation (35).

LEMMA 6.1. *There exists λ_0 , $0 < \lambda_0 < \infty$, depending only on k , such that*

$$|E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right)| \leq E \left(\prod_{i=1}^k |Z_\lambda^*(u_i)| \right) \leq 2k! \exp(\mu/2)$$

for all $\mathbf{u} = (u_1, \dots, u_k)$, $u_i \geq 0$, and $\lambda \geq \lambda_0$

By letting some of the factors in the preceding expression coincide we can obtain a similar bound for all joint moments of higher order, stated in the following corollary:

LEMMA 6.2. *For any set of non negative integers $\mathbf{r} = (r_1, \dots, r_k)$, the following inequality holds:*

$$E(|Z_\lambda^*(u_1)|^{r_1} \dots |Z_\lambda^*(u_k)|^{r_k}) \leq 2r! \exp(\mu/2),$$

for all $\mathbf{u} = (u_1, \dots, u_k)$, $u_i \geq 0$, and $\lambda \geq \lambda_0$, where $r = \sum_{1 \leq i \leq k} r_i$.

Proof of Lemma 6.1. Using a generalized version of Hölder's inequality we see that

$$|E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right)| \leq E \left(\prod_{i=1}^k |Z_\lambda^*(u_i)| \right) \leq \prod_{i=1}^k (E(|Z_\lambda^*(u_i)|^k))^{1/k} \tag{44}$$

Since $|z|^k \leq k! \exp(|z|)$, it follows that

$$\begin{aligned} |E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right)| &\leq \prod_{i=1}^k (k!)^{1/k} (E \exp(|Z_\lambda^*(u_i)|))^{1/k} \\ &\leq k! \prod_{i=1}^k (E \exp(Z_\lambda^*(u_i)) + E \exp(-Z_\lambda^*(u_i)))^{1/k} \end{aligned} \tag{45}$$

Using the specific form of the moment generating function of the Poisson distribution, we have

$$\lim_{\lambda \rightarrow \infty} E \exp(\pm Z_\lambda^*(u_i)) = \exp(\mu/2)$$

Notice that in the stationary case, the distribution of $Z_\lambda^*(u)$ does not depend on $u \geq 0$. Thus, for sufficiently large $\lambda_0 > 0$,

$$|E(\exp(Z_\lambda^*(u_i)) + E(\exp(-Z_\lambda^*(u_i)))| \leq 2 \exp(\mu/2) \tag{46}$$

for $\lambda \geq \lambda_0$. Inserting this into Equation (45) we get

$$|E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right)| \leq k! \prod_{i=1}^k 2^{1/k} \exp(\mu/2k) = 2k! \exp(\mu/2) \tag{47}$$

LEMMA 6.3. For any \mathbf{u} and any set of non negative integers $\mathbf{r} = (r_1, \dots, r_k)$

$$\lim_{\lambda \rightarrow \infty} E(Z_\lambda^*(u_1)^{r_1} \dots Z_\lambda^*(u_k)^{r_k}) = E(Z^*(u_1)^{r_1} \dots Z^*(u_k)^{r_k})$$

Proof. This follows from Lemmas (5.1) and (6.2).

Proof of Lemma 5.2. We begin our proof by applying the Lebesgue dominated convergence theorem to the integrand on the right hand side of Equation (41), which is justified by lemmas (6.1) and (6.3). Consequently,

$$\lim_{\lambda \rightarrow \infty} E(X_\lambda^*(u_1)^{r_1} \dots X_\lambda^*(u_k)^{r_k}) = E(X^*(u_1)^{r_1} \dots X^*(u_k)^{r_k})$$

The proof is completed by recalling the following standard results on weak convergence of probability measures in \mathcal{R}^k .

LEMMA 6.4. Let ν_n, ν be probability measures on \mathcal{R}^k such that ν is uniquely determined by its joint moments $m_\nu(\mathbf{r})$, where

$$m_\nu(\mathbf{r}) = \int_{\mathcal{R}^k} x_1^{r_1} \dots x_k^{r_k} \nu(d\mathbf{x}),$$

$\mathbf{x} = (x_1, \dots, x_k)$ and $\mathbf{r} = (r_1, \dots, r_k)$. Suppose for every \mathbf{r}

$$\lim_{n \rightarrow \infty} m_{\nu_n}(\mathbf{r}) = m_\nu(\mathbf{r}),$$

then $\nu_n \Rightarrow \nu$.

Proof. See Billingsley ([3], p.352, problem 30.5).

LEMMA 6.5. A k dimensional normal distribution is determined by its joint moments.

Proof. See Billingsley ([3], p.351, problem 30.3).

It follows from lemmas 5.1, 6.2, 6.3 that the joint moments of $X_\lambda^*(t)$ in Equation (41) converge to those of a k dimensional normal distribution, hence that

$$(X_\lambda^*(t_1), \dots, X_\lambda^*(t_k)) \Rightarrow (X^*(t_1), \dots, X^*(t_k)), \lambda \rightarrow \infty$$

Weak convergence in the nonstationary case: We begin with the observation that $Z_\lambda(t)$ is again Poisson, but now

$$E(Z_\lambda(t)) = \lambda \int_0^t (1 - G(y)) dy = \lambda\mu(t) \tag{48}$$

The normalized process $Z_\lambda^*(t)$ of Equation (34) is defined as follows:

$$Z_\lambda^*(t) = \frac{Z_\lambda(t) - \lambda\mu(t)}{\sqrt{\lambda}} \tag{49}$$

The multidimensional central limit theorem given in Lemma 5.1, with minor modifications, remains valid. More precisely, the same method of proof yields the following extension of Lemma 5.1.

LEMMA 6.6. For $0 \leq t_1 < t_2 < \dots < t_k$

$$(Z_\lambda^*(t_1), \dots, Z_\lambda^*(t_k)) \Rightarrow (Z^*(t_1), \dots, Z^*(t_k)), \lambda \rightarrow \infty$$

where $(Z^*(t_1), \dots, Z^*(t_k))$ is a Gaussian random vector, with mean zero, and covariance matrix

$$\Gamma(t_i, t_j) = E(Z^*(t_i)Z^*(t_j)) = \int_0^{\min(t_i, t_j)} (1 - G(|t_i - t_j| + y)) dy \tag{50}$$

To extend Lemma (6.3) to the nonstationary case, it is only necessary to modify inequality (46) as follows.

$$|E(\exp(Z_\lambda^*(u_i)) + E(\exp(-Z_\lambda^*(u_i)))| \leq 2 \exp(\mu(u_i)/2) \tag{51}$$

And this yields the following extension of Equation (47).

For $\max_{1 \leq i \leq p} \mu(u_i) \leq B$

$$|E \left(\prod_{i=1}^k Z_\lambda^*(u_i) \right) | \leq p! \prod_{i=1}^k 2^{1/p} \exp(B/2p) = 2p! \exp(B/2) \tag{52}$$

The rest of the proof is unchanged.

Acknowledgement: The authors thank the referees for their helpful suggestions for improving the paper and T. Hosmer, J. Nataf, and C. Babyaa for help with computing. It is a pleasure and honor for us to contribute a paper to this special issue of Methods and Applications of Analysis in honor of Professors Stroock and Varadhan.

REFERENCES

- [1] JAN BERAN, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.
- [2] P. BILLINGSLEY, *Convergence of Probability Measures*, J. Wiley & Sons, Inc. New York (1968).
- [3] P. BILLINGSLEY, *Probability and Measure*, J. Wiley & Sons, Inc. New York (1979).
- [4] D.R. COX AND V. ISHAM, *Point Processes*, Chapman and Hall, London and New York (1980).
- [5] C.A. GUERIN, H. NYBERG, O. PERRIN, S. RESNICK, H. ROOTZÉN, AND C. STĂRICĂ, *Empirical Testing of the Infinite Source Poisson Data Traffic Model*, Preprint, available as TR 1257 at www.orie.cornell.edu/trlist.
- [6] DAVID L. JAGERMAN, BENJAMIN MELAMED, AND WALTER WILLINGER, *Frontiers in Queueing, Stochastic Modeling of Traffic Processes*, J. H. Dshalalow, ed., CRC Press, 1997, pp. 271–320.
- [7] M. KRUNZ, A.M. MAKOWSKI, Modeling Video Traffic Using $M/G/\infty$ Input Processes: A Compromise between Markovian and LRD Models, *IEEE Journal on Selected Areas in Communications (JSAC)*, 16:5 (1998), pp.733–748.
- [8] M. KRUNZ AND I.MATTA, *Analytical Investigation of the Bias Effect in Variance Type Estimators for Inference of Long-Range Dependence*, *Computer Network Journal*, Special issue on Advances in Modeling and Engineering of Long-Range Dependent Traffic, 40:3(2002), pp.445–458.
- [9] J. NETER, M.H. KUTNER, C.J. NACHTSHEIM, W. WASSERMAN, *Applied Linear Statistical Models*, 4e, R.D. Irwin, 1996.
- [10] T. MIKOSCH, S. RESNICK, H. ROOTZÉN, AND A. STEGEMAN, *Is Network Traffic approximated by Stable Lévy Motion or Fractional Brownian Motion?*, *Annals of Applied Probability*, 12:1(2002), pp.23–68.
- [11] M. PARULEKAR, A.M. MAKOWSKI, *$M/G/\infty$ input Processes: A versatile class of models for network traffic*, preprint.
- [12] E. PARZEN, *Stochastic Processes*, Holden-Day, 1962.
- [13] S. RESNICK, *Applied Probability Modeling of Data Networks*, MAA, Gehman lecture, 15 April 2000, Slides in PDF format. Available at: <http://www.orie.cornell.edu/sid>.
- [14] R.H. RIEDI, M.S. CROUSE, V. RIBEIRO, AND R.G. BARANIUK, *A Multifractal Wavelet Model with Applications to Network Traffic*, *IEEE Trans. Info. Theory*, 45(1999), pp. 992–1018.
- [15] W. A. ROSENKRANTZ & J. HOROWITZ, *Statistical Analysis of Variance Time Plots Used to Estimate Parameters of a Long Range Dependent Process*, Invited Paper, 10th INFORMS Applied Probability Conference, University of Ulm, July 1999 (revised July 2002).
- [16] M. TAQQU, W. WILLINGER, AND R. SHERMAN, *Proof of a Fundamental Result in Self-Similar Traffic Modeling*, *Computer Communication Review*, 27(1997), pp. 5–23.
- [17] W. WILLINGER, M.S. TAQQU, W.E. LELAND AND D.V. WILSON, *Self Similarity in High Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements*, *Statistical Science*, 10(1995), pp. 67–85.

