Journal of Statistical Computation and Simulation
Vol. 78, No. 12, December 2008, 1179–1198

Taylor & Francis
Taylor & Francis Group

Wavelet-based confidence intervals for the self-similarity parameter

AGNIESZKA JACH and PIOTR KOKOSZKA*

Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, USA

(Received 25 June 2007)

We propose and compare several methods of constructing wavelet-based confidence intervals for the self-similarity parameter in heavy-tailed observations. We use empirical coverage probabilities to assess the procedures by applying them to Linear Fractional Stable Motion with many choices of parameters. We find that the asymptotic confidence intervals provide empirical coverage often much lower than nominal. We recommend the use of resampling confidence intervals. We also propose a procedure for monitoring the constancy of the self-similarity parameter and apply it to Ethernet data sets.

Keywords: Bootstrap; Subsampling; Confidence intervals; Self-similarity; Wavelets

1. Introduction

Empirical studies have shown [1–3] that many network traffic traces exhibit self-similarity or are long-range dependent (LRD). It is of importance in modeling the flow of information through a network to estimate the self-similarity parameter. Significant contribution in this direction have been made by several authors, see [4, 5] for a review and references. The existing theoretical and numerical studies are chiefly concerned with point estimation of the self-similarity parameter. As discussed later in this paper, it is known that the appropriate estimators are asymptotically unbiased and normal, but the asymptotic variance depends in a very complex way on unknown parameters. Therefore, useful approximations to the asymptotic variance have been derived, but these are based on the assumption of the Gaussianity of observations. As will be seen in real data examples discussed in section 5, the latter assumption is not always justified and often the data exhibit heavy tails characterized by the tail index α. Suitable approximations to the asymptotic variance have also been derived in the case of heavy-tailed observations. One should hope that these approximations would yield useful

*Corresponding author. Email: piotr@stat.usu.edu

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00949650701587855

confidence intervals for the self-similarity parameter, but our research has shown that this is typically not the case, even if the tail index α is correctly specified. This index is very difficult to estimate in practice, and a mis-specification of α worsens the empirical coverage probability of the asymptotic confidence intervals. Boundary effects are an important source of bias in finite samples. However, even after removing the boundary wavelet coefficients, it is not unusual for an asymptotic confidence interval with nominal coverage probability of 95% to have an empirical coverage probability lower than 75%. A first-order bias corrected estimator has been derived in the case of Gaussian observations by Veitch and Abry [6], but this modification relies on some very specific properties of the normal distribution and is not directly applicable to heavy-tailed time series.

In this paper, we propose a number of alternative methods of constructing confidence intervals for the self-similarity parameter. We compare five different methods and provide practical recommendations. We also propose a procedure for monitoring the constancy of the self-similarity parameter and apply it to Ethernet traffic measurements.

In our simulation study, we focus on self-similar (motion-type) processes and use the linear fractional stable motion (LFSM) as an archetype. Recall that a stochastic process $\boldsymbol{X} = \{X_t\}_{t\in\mathcal{R}}$ is self-similar with self-similarity parameter $H > 0$ if for any positive constant c

$$\{X_{ct}\}_{t\in\mathcal{R}} =_d \{c^H X_t\}_{t\in\mathcal{R}},$$

where $=_d$ denotes the equality of the finite-dimensional distributions. We say that the process $\boldsymbol{X} = \{X_t\}_{t\in\mathcal{R}}$ is a Linear Fractional Stable Motion (LFSM) if X_t is defined by the integral

$$X_t = \int_{\mathcal{R}} \left[((t-s)_+)^d - ((-s)_+)^d\right] M_\alpha(\mathrm{d}s), \tag{1}$$

where $\alpha \in (0, 2]$, and $d = H - 1/\alpha$, for some $H \in (0, 1)$, $H \neq 1/\alpha$, Mα is an α-stable symmetric random measure on $\mathcal{R}$ with Lebesgue control measure, see section 3.3 of ref. [7], $x_+ \equiv \max\{x, 0\}$. This process has stationary increments and is self-similar with parameter H. When $\alpha = 2$, *i.e.*, when the process is Gaussian, and when $H > 1/2$ the increments of $\boldsymbol{X}$ are LRD. In the case of $\alpha < 2$, the process defined by equation (1) has heavy tails and even though there is no universally agreed upon definition for LRD when the variance is infinite, the increment process of $\boldsymbol{X}$, called a Linear Fractional Stable Noise, is said to be LRD, if $H > 1/\alpha$. For further details about LFSM and heavy-tailed self-similar and LRD processes, see Chapter 7 of [7]. The LFSM has been used for modeling large network traffic fluctuations [8, 9].

The objective of this paper is to develop and compare several wavelet-based methods of constructing confidence intervals for the parameter H in the LFSM. Wavelets are known to be an efficient tool in the context of LRD and self-similarity, see [4, 10–12] among others. When a discrete realization of a process is available, Mallat's algorithm can be used to produce the set of discrete wavelet transform (DWT) coefficients, making these procedures applicable to very long time series obtained, for example, from network traffic measurements.

Two features of the DWT coefficients of an LRD process, decorrelation and scaling, are utilized in the estimation of H and the construction of confidence intervals through the asymptotic and resampling approaches. We assess the procedures under consideration using empirical coverage probabilities. Our overall assessment of the relative performance of the methods enables us to provide home useful guidance for their practical application.

The paper is organized as follows. We first describe in section 2 a wavelet-based estimator of the self-similarity parameter of the LFSM. This estimator forms the basis for the construction of the confidence intervals. In section 3, we introduce several methods of constructing confidence intervals. We then give in section 4 a detailed description of the simulation study and discuss the results. In section 5, an application of our techniques to Ethernet traces is presented.

2. Estimation of the self-similarity parameter of LFSM

In this section, we present a wavelet-based method of estimating the self-similarity parameter H of a LFSM, which serves as a cornerstone for a confidence interval construction. This method was derived in [13], see also [4].

The wavelet transform coefficients of a continuous time process $\boldsymbol{X} = \{X_t\}_{t\in\mathcal{R}}$ is a collection of the quantities

$$d_{j,k} = \int_{\mathcal{R}} X_t \psi_{j,k}(t)\mathrm{d}t, \quad j \in \mathcal{Z}, \quad k \in \mathcal{Z}, \tag{2}$$

where $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, $t \in \mathcal{R}$. The function $\psi\colon \mathcal{R} \to \mathcal{R}$, called the mother wavelet, satisfies some regularity assumptions (see ref. [14] for the details). In this setting, 2^j and k are called scale and location, respectively. In our paper, we work with the DWT coefficients, which are obtained as the output of Mallat's algorithm applied to discrete time observations $X_0, X_1, \ldots, X_{N-1}$. This scheme produces a set of DWT coefficients $\bar{d}_{j,k}$; N_j of them are available at scale

$$n_j = 2^j, \quad j = 1, 2, \ldots, J,$$

where $N_j = 2^{J-j}$, $J = \lfloor \log_2(N) \rfloor$ and $\lfloor\cdot\rfloor$ denotes the integer part. The coefficients $\bar{d}_{j,k}$ can be viewed as approximations to the $d_{j,k}$, see Chapter 11 of [10].

The regression-based approach for estimating the self-similarity parameter H in LFSM involves the statistic

$$Y(N_j) = \frac{1}{N_j} \sum_{k=0}^{N_j-1} \log_2 |\bar{D}_{j,k}|, \tag{3}$$

where

$$\bar{D}_{j,k} \equiv 2^{-j/2}\bar{d}_{j,k}. \tag{4}$$

The generalized least squares regression estimator for H is defined by

$$\hat{H} = \sum_{j=j_{\min}}^{j_{\max}} w_j Y(Nj), \tag{5}$$

where $j_{\min}$ and $j_{\max}$ dictate a range of scales upon which the estimator is constructed. Stoev *et al.* [12] refer to equation (5) as the "log" estimator. In most applications, due to bias-variance trade-off, the number of scales used in equation (5) is smaller than the total number of available scales (see section 4). The weights w_j in equation (5) satisfy the following conditions

$$\sum_{j=j_{\min}}^{j_{\max}} w_j = 0, \quad \sum_{j=j_{\min}}^{j_{\max}} j w_j = 1.$$

The specific choice of the w_j is discussed in section 4.1.

Detailed explanation of the estimation technique along with the asymptotic results are given in [12].

3. Confidence intervals for the parameter H

In this section, we present several methods of constructing confidence intervals for the self-similarity parameter H in LFSM. These methods can be grouped into two broad categories: methods based on an asymptotic approximation, and resampling methods. We describe them in sections 3.1 and 3.2, respectively.

3.1 *Asymptotic approach*

The decorrelation property of wavelet coefficients allows us to approximate the variance of the self-similarity estimator $\hat{H}$. Abry *et al.* [11] derived the following approximate formula

$$\mathrm{Var}\{\hat{H}\} \equiv \sigma^2(\alpha) \approx \frac{(\log_2(\mathrm{e}))^2\pi^2}{12}\left(1+\frac{2}{\alpha^2}\right)\left(\sum_{j=j_{\min}}^{j_{\max}}\frac{w_j^2}{N_j}\right), \tag{6}$$

where the summation extends over the octaves used to construct the estimator $\hat{H}$. Consequently, the $100(1-\beta)\%$ *asymptotic confidence interval* for H is defined as

$$(\hat{H} + q_Z(\beta/2)\sigma(\alpha),\ \hat{H} + q_Z(1-\beta/2)\sigma(\alpha)), \tag{7}$$

where $q_Z(\beta)$ denotes the βth quantile of the standard normal distribution. Note that the definition of the asymptotic confidence intervals for H requires the knowledge of the stability index α and thus we assume that this parameter is known. We comment on this issue when we discuss our conclusions and recommendations in section 4.4.

Our study of the estimator $\hat{H}$ showed a significant effect of the boundary wavelet coefficients on the bias of this estimator (see section 4.2). To study this bias, we define the $100(1-\beta)\%$ confidence interval for the bias as

$$\left((\overline{\hat{H}} - H) + \frac{q_Z(\beta/2)s}{\sqrt{R}},\ (\overline{\hat{H}} - H) + \frac{q_Z(1-\beta/2)s}{\sqrt{R}}\right), \tag{8}$$

where $\overline{\hat{H}}$ and s are, respectively, the sample mean and the sample standard deviation of R independent replications of $\hat{H}$.

Recall that if a wavelet filter of length L is used, the first $K_j = \min(\lceil (L-2)(1-2^{-j})\rceil, N_j)$ wavelet coefficients at octave j are affected by circularly filtering the data and the remaining $M_j = N_j - K_j$ coefficients,

$$\tilde{d}_{j,k} \equiv \bar{d}_{j,K_j+k}, \quad j = 1, 2, \ldots, J, \quad k = 0, 1, \ldots, M_j - 1, \tag{9}$$

called the non-boundary wavelet coefficients are not influenced (for more details see Comments and Extensions to section 4.11 of ref. [10]). The corresponding estimator of H based on the non-boundary wavelet coefficients is, therefore, defined as

$$\hat{H}^{nb} = \sum_{j=j_{\min}}^{j_{\max}} w_j Y(Mj), \tag{10}$$

where

$$Y(M_j) = \frac{1}{M_j}\sum_{k=0}^{M_j-1}\log_2|\tilde{D}_{j,k}| \tag{11}$$

and

$$\tilde{D}_{j,k} \equiv 2^{-j/2}\tilde{d}_{j,k}. \tag{12}$$

Consequently, the $100(1-\beta)\%$ *asymptotic confidence interval* using the non-boundary wavelet coefficients is defined as

$$\left(\hat{H}^{nb} + q_Z\left(\frac{\beta}{2}\right)\sigma(\alpha),\ \hat{H}^{nb} + q_Z\left(1-\frac{\beta}{2}\right)\sigma(\alpha)\right). \tag{13}$$

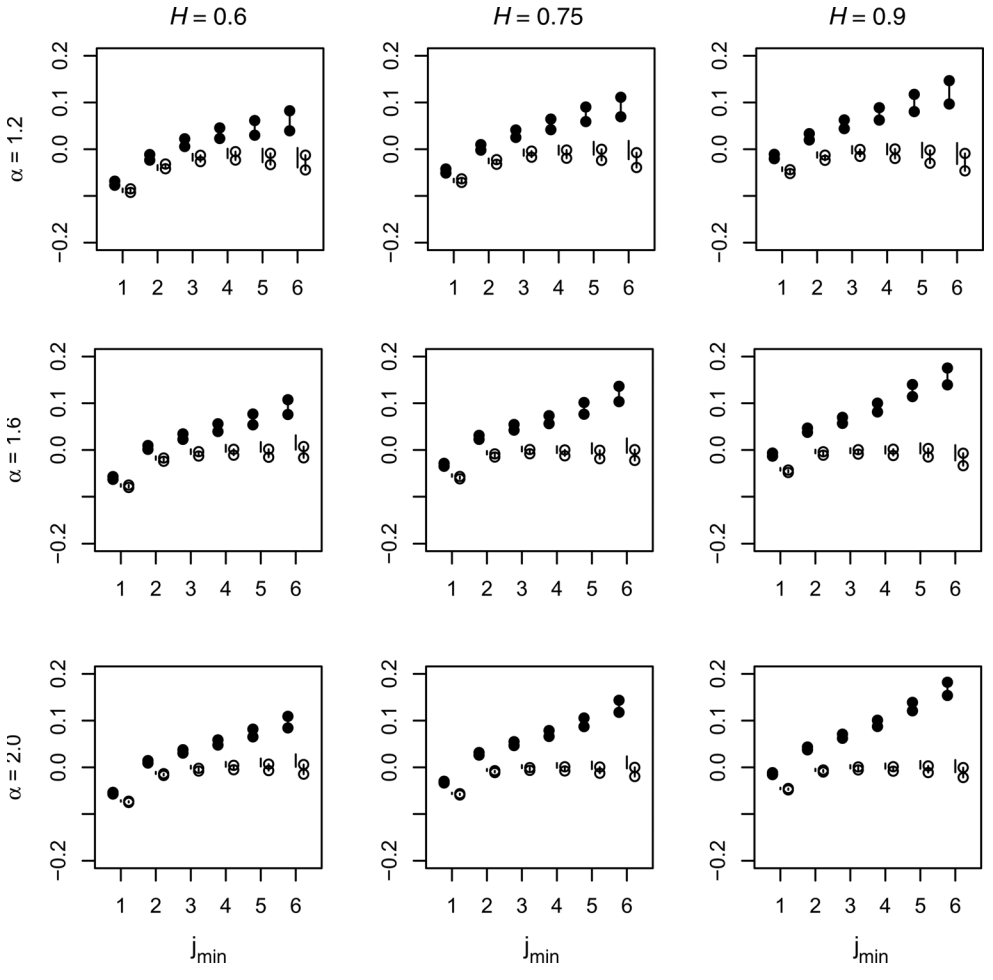

Figure 1. The 95% confidence intervals (8) for the bias of the asymptotic estimator as a function of $j_{\min}$ based on realizations of length $N = 2^{13}$. Estimators $\hat{H}$ were calculated using periodic (•) and reflection (∘) boundary rules, and excluding boundary coefficients (plain bars). All estimators were calculated using scales $2^{j_{\min}}, \ldots, 2^{10}$.

The use of the non-boundary wavelet coefficients $\tilde{d}_{j,k}$ combined with the appropriate selection of $j_{\min}$ and $j_{\max}$ provide confidence intervals for bias, which contain 0. Such confidence intervals can be calculated using formula (8) with $\hat{H}$ replaced by $\hat{H}^{nb}$ (see figure 1).

3.2 *Resampling approach*

In this section, we focus on the resampling methods of constructing confidence intervals. The resampling procedures are applied to DWT coefficients within a given scale. These methods are heuristically justified by the approximate decorrelation property of these coefficients.

3.2.1 Subsampling confidence intervals with non-overlapping blocks. Consider the coefficients $\tilde{D}_{j,k}$ defined by equation (12) at some fixed scales $n_1 = 2^{j_{\min}}$, $n_2 = 2^{j_{\min}+1}, \ldots, n_m = 2^{j_{\max}}$, $m = j_{\max} - j_{\min} + 1$, and $k = 0, 1 \ldots, M_j - 1$ at scale n_j. For each

octave j, $j = j_{\min}, j_{\min}+1, \ldots, j_{\max}$, we choose a number of blocks B_j of lengths $l_0 = l_1 = \cdots = l_{B_j-2} = \lfloor M_j/B_j \rfloor$, $l_{B_j-1} = \lfloor M_j/B_j \rfloor + (M_j \bmod B_j)$ at scale n_j. In other words, if M_j is not divisible by B_j, the first $B_j - 1$ blocks are of the same length, l_0, and the last one is longer by $(M_j \bmod B_j)$. At each octave j, we consider the following blocks of the rescaled non-boundary wavelet coefficients

$$\{\tilde{D}_{j,0}, \ldots, \tilde{D}_{j,l_0-1}\}, \{\tilde{D}_{j,l_0}, \ldots, \tilde{D}_{j,l_0+l_1-1}\}, \ldots, \{\tilde{D}_{j,\sum_{i=0}^{B_j-2} l_i}, \ldots, \tilde{D}_{j,M_j-1}\}$$

Based on each of these blocks, we compute

$$Y(M_j^m) = \frac{1}{l_m} \sum_{k=\sum_{r=0}^{m-1} lr}^{\sum_{r=0}^{m} l_r - 1} \log_2 |\tilde{D}_{j,k}|, \quad m = 0, 1, \ldots, B_j - 1.$$

The index m indicates the position of the block, its range depends on the octave j. Next we select $j_{\max} - j_{\min} + 1$ blocks, one on each scale, and compute the estimator

$$\hat{H}_{SN} = \hat{H}_{SN}(m(j_{\min}), \ldots, m(j_{\max})) = \sum_{j=j_{\min}}^{j_{\max}} w_j Y(M_j^{m(j)}), \tag{14}$$

where $m(j)$ is the index of the block selected on scale j. There are $S = \Pi_{j=j_{\min}}^{j_{\max}} B_j$ estimators (14) based on one realization of LFSM. Consequently, the $100(1-\beta)\%$ *subsampling confidence interval* for H is

$$\big(q_s(\beta/2), q_s(1-\beta/2)\big), \tag{15}$$

where $q_s(\beta)$ denotes the βth empirical quantile of the empirical distribution of the S estimators $\hat{H}_{SN}$. The subscript SN in equation (14) stands for 'Subsampling with Non-overlapping blocks'.

3.2.2 Naive block bootstrap and bootstrap confidence intervals. We first describe the naive block bootstrap method, subsequently called the brevity block bootstrap. We use the notation introduced in section 3.2.1. For each octave j we choose a block length l_j. It follows that there are $M_j - l_j + 1$ (overlapping) blocks of the form

$$\{\tilde{D}_{j,0}, \ldots, \tilde{D}_{j,l_j-1}\}, \{\tilde{D}_{j,1}, \ldots, \tilde{D}_{j,l_j}\}, \ldots, \{\tilde{D}_{j,M_j-l_j}, \ldots, \tilde{D}_{j,M_j-1}\}.$$

Next, by drawing with replacement from the set $\{0, 1, \ldots, M_j - l_j\}$, we select $B_j = \lfloor M_j/l_j \rfloor$ integers and denote them by $\{s_0, s_1, \ldots, s_{B_j-1}\}$. Based on this selection, we construct the bootstrap series of coefficients

$$\{\tilde{D}_{j,s_0}, \ldots, \tilde{D}_{j,s_0+l_j-1}\}, \{\tilde{D}_{j,s_1}, \ldots, \tilde{D}_{j,s_1+l_j-1}\}, \ldots, \{\tilde{D}_{j,s_{B_j-1}}, \ldots, \tilde{D}_{j,s_{B_j-1}+l_{j-1}}\}.$$

If $(M_j \bmod l_j) \neq 0$ the length of this series exceeds the number of the non-boundary coefficients at scale j, therefore, we truncate it to obtain a series of length M_j, *i.e.*, of the

Table 1. Methods of constructing wavelet-based confidence intervals for the self-similarity parameter H.

METHOD	ABBR.	CHARACTER
Asymptotic (correct α)	AP	$\times$
Asymptotic ($\alpha = 1.5$)	A	$\diamond$
Subsampling with Non-overlapping blocks	SN	$\circ$, ∇
Bootstrap	B	$\triangle$
Block Bootstrap	BB	$+$, $\bullet$, $*$

same length as the original series of the non-boundary coefficients at scale 2^j. Denoting these coefficients by $\tilde{D}^*_{j,k}$ we can compute the estimator

$$\hat{H}^*_{BB} = \sum_{j=j_{\min}}^{j_{\max}} w_j Y^*(M_j), \tag{16}$$

where

$$Y^*(M_j) = \frac{1}{M_j} \sum_{k=0}^{M_j-1} \log_2 |\tilde{D}^*_{j,k}|. \tag{17}$$

If we draw one resample at each octave j, we obtain one estimator (16). Resampling B times (one resample on each scale) yields B block bootstrap estimators $\hat{H}^*_{BB}$. Consequently, the $100(1-\beta)\%$ *block bootstrap confidence interval* for H is

$$\left(q_b\left(\frac{\beta}{2}\right), q_b\left(1-\frac{\beta}{2}\right)\right), \tag{18}$$

where $q_b(\beta)$ denotes the β-th empirical quantile of the empirical distribution of the B estimators $\hat{H}^*_{BB}$.

Note that when the blocks of length 1 are used, the block bootstrap method is equivalent to the DWT-based bootstrap method described in [15]. We refer to confidence intervals obtained in this way as *bootstrap confidence intervals* even though it is a special case of the naive block bootstrap procedure.

For ease of reference, we list in table 1 the methods investigated in this paper and their abbreviations as well as plotting characters used for marking the empirical coverage probabilities in section 6. In case of subsampling method and block bootstrap we tried several choices of the blocks and thus we have more than one plotting character (see sections 4.3.1 and 4.3.2 for more details).

4. Simulation study

This section contains the main results of the paper. In section 4.1, we describe in detail the procedures for generating realizations of LFSM and for estimating H. Sections 4.2 and 4.3 focus on the performance of the asymptotic and resampling methods, respectively. Conclusions are discussed in section 4.4. We present only the results that best illustrate our findings. Tables with numerical values of the empirical coverage probabilities and a large number of additional graphs are presented in an extended version of this paper available from the authors.

4.1 *Generation of LFSM and estimation of H*

Consider a discrete realization of LFSM $X_0, X_1, \ldots, X_N$, $(X_0 = 0)$. The random variables X_t, $t = 1, 2, \ldots, N$ can be expressed as $X_t = \sum_{k=1}^{t} U_k$, where

$$U_k = \int_{\mathcal{R}} \left[((k-s)_+)^d - ((k-1-s)_+)^d\right] M_\alpha(\mathrm{ds}). \tag{19}$$

As in [7], section 7.11, the random variable U_k is approximated as

$$U_k \approx \sum_{i=1, i\neq m}^{Mm} \left[((i/m)_+)^d - ((i/m-1)_+)^d\right] m^{-1/\alpha} Z_\alpha(i-mk),$$

for some choice of discretization parameters m and M, which correspond to the mesh size and the truncation level in integral (19), respectively. It is possible to implement an algorithm for calculating the last summation directly (see section 7.11 [7]) or indirectly via the Fast Fourier Transform (see [16]). However, since our goal is to compare methods of constructing confidence intervals, the computational issues concerning generating paths of LFSM are beyond the scope of this paper. We refer to [16] for more details. In our study, we used the indirect implementation with discretization parameters $m = 64$ and $M = 2^{14}$. According to Stoev and Taqqu [16], the choice of $m = 64$ when $\alpha = 1.5$ provides good results for several values of $M = 60, 600, 6000$, and for various choices of H. In our study M equals the length of the generated series.

We used nine pairs (α, H) of the parameters obtained as combination of the following sets

$$\alpha \in \{1.2, 1.6, 2.0\} \quad \text{and} \quad H \in \{0.6, 0.75, 0.9\}.$$

For each pair, we generated $R = 300$ independent realizations of LFSM of length 2^{14} via the indirect method, and truncated them to our target length $N = 10,000$ for further computations. The generated series were originally of length 2^{14} and not 10,000, in order to utilize the Fast Fourier transform. A same choice for M, *i.e.*, $M = 2^{14}$ combined with $m = 64$ yielded m and $M +$ length of generated series as powers of 2.

To compute wavelet-based estimators of the self-similarity parameter H and to construct confidence intervals, we computed discrete wavelet transform coefficients using Mallat's algorithm with a wavelet filter of width $L = 6$, corresponding to Daubechies wavelet with 3 vanishing moments. The definition and the role of vanishing moments are explained in section 11.9 of [10].

Practical selection of scales upon which the estimators are built depends on the length of the series, as well as the bias-variance trade-off [12]. When the series length is 10,000, there are at most $J = 13$ dyadic scales available; however, not all of them should be included in the estimation. Based on the analysis of the bias of the asymptotic estimator (8) presented in section 4.2 and rationale behind the resampling methods, we made the following selection. We used two sets of scales $n_1 = 2^3, n_2 = 2^4, \ldots, n_6 = 2^8$ and $n_1 = 2^3, n_2 = 2^4, \ldots, n_7 = 2^9$ to see which was the better option. In addition, we considered series of length $N = 2^{12}$ to compare methods of constructing confidence intervals for H with respect to the length of the series (*longer series*, $N = 2^{13}$ and *shorter series*, $N = 2^{12}$). For shorter series, we computed the estimators using two collections of scales $n_1 = 2^3, n_2 = 2^4, \ldots, n_5 = 2^7$ and $n_1 = 2^3, n_2 = 2^4, \ldots, n_6 = 2^8$. The weights $w_1, w_2, \ldots, w_m$ (m is the number of octaves involved in the estimation) can be determined by an $m \times m$ strictly positive definite matrix $\boldsymbol{G}$. The connection between the weights and the matrix $\boldsymbol{G}$ is described in [12]. We used $\boldsymbol{G} = \mathrm{diag}\{n_1, \ldots, n_m\}$, where $n_1 = 2^{j_{\min}}, n_2 = 2^{j_{\min}+1}, \ldots, n_m = 2^{j_{\max}}$, which is the only practically available choice of $\boldsymbol{G}$ when the autocovariances of the observations are unknown.

4.2 *Asymptotic confidence intervals for H*

In this section, we discuss the results of simulations for the asymptotic methods.

Based on confidence intervals for the bias of the asymptotic estimator (8), calculated for two boundary rules, periodic and reflection (see [15] for more details), and as a function of $j_{\min} = 1, 2, \ldots, 6$, we conclude that; first, it is necessary to drop the first few scales to avoid bias ($j_{\min} = 3$ is a reasonable choice), but not too many to keep the variance low; secondly, the bias caused by the periodic rule is greater than that for the reflection rule; thirdly, the exclusion of the boundary wavelet coefficients yields the desirable bias containing 0, located between the biases introduced by the two boundary rules (see figure 1 for the longer series, similar results were obtained for the shorter series).

The choice of the upper cut-off is not as critical and the behaviour of the confidence intervals, when all wavelet coefficients are included, is very similar for three choices of $j_{\max} = 10, 11, 12$ we considered. We present the comparison of the confidence intervals for bias (8) of the asymptotic estimator computed with and without the boundary coefficients, for $j_{\max} = 10$, since for $J = 13$ and wavelet filter of length $L = 6$ this is the largest scale containing the non-boundary coefficients (for $J = 12$ the largest scale is 9). On the other hand, since the number of the non-boundary coefficients at this scale is low, it is reasonable to use $j_{\max} < 10$, especially for the resampling methods.

However, even after excluding the boundary coefficients and carefully selecting $j_{\min}$ and $j_{\max}$, the asymptotic method does not provide satisfactory results. The empirical coverage probabilities for the two choices of J, 13 and 12, are very similar with values in the 40–75% range, and increasing with H, for all values of α considered (see figures 2 and 3, $\times$). The variability among the confidence intervals based on different replications but for the same (α, H) is large when $\alpha = 1.2$, and small when $\alpha = 2.0$, and overall smaller for the longer series.

A practical implementation of the asymptotic method requires the knowledge of the tail index α. Our simulations have shown, however, that α need not be known precisely, as long as it is not too close to zero. The variance (6) as a function of α changes very slowly for $\alpha > 0.5$. The empirical coverage probabilities with the variance (6) computed with a possibly mis-specified $\alpha = 1.5$ behave very similarly to those based on the correct α (see figures 2 and 3, $\diamond$).

We summarize our conclusions for the asymptotic approach in the following points.

- Asymptotic intervals (7) are strongly influenced by the boundary coefficients.
- Asymptotic intervals (13) based on the non-boundary wavelet coefficients have coverage in the range 40–75%.
- Asymptotic intervals (13) based on the variance (6) computed for incorrect $\alpha = 1.5$ provide similar coverage to those based on the correct α.
- Asymptotic intervals (13) computed for different replications but fixed (α, H) vary a lot among each other when the series length is 2^{12} and α small.

4.3 *Resampling confidence intervals*

4.3.1 Subsampling confidence intervals with non-overlapping blocks. We continue the discussion of the simulation results with the resampling methods, focusing first on the subsampling procedure with non-overlapping blocks. Choosing the lengths of the blocks in procedures of this type is always a difficult task. We tried two strategies. First, we partitioned different scales into blocks with lengths roughly the same across all scales. At the same time we aimed for the total number of estimators, S, which is not too high. Then, we used blocks

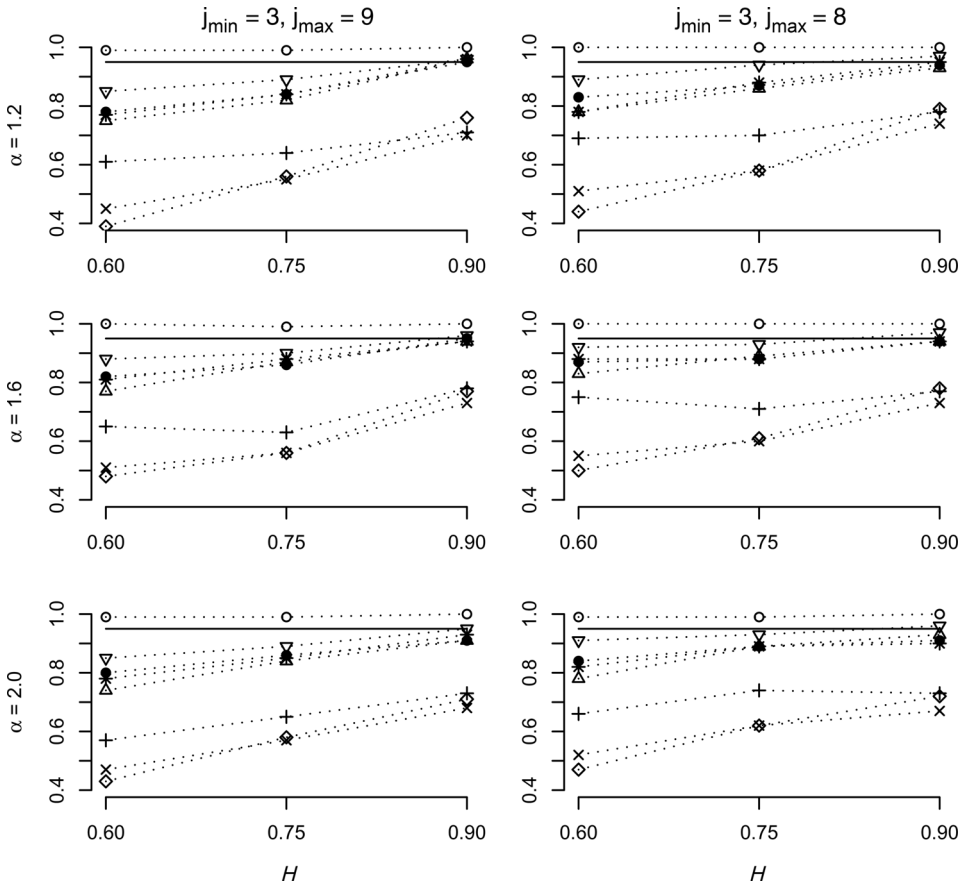

Figure 2. Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{13}$. Estimators were calculated using indicated range of scales.

of length equal to approximately 25% of the coefficients at the lowest scales, and to all the coefficients at the remaining scales, this time making sure that S was sufficiently large to construct the percentile-type confidence intervals.

Using the first strategy, for the shorter series ($N = 2^{12}$) the number of blocks were as follows:

$$\begin{aligned} &B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^7, \\ &B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8. \end{aligned} \tag{20}$$

For the longer series ($N = 2^{13}$), we used

$$\begin{aligned} &B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8, \\ &B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = B_9 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^9. \end{aligned} \tag{21}$$

In each case the number of estimators (14) was $S = 1024$.

For both lengths, and for both choices of scales, the empirical coverage probabilities are very similar and very close to 100% (see figures 2 and 3, ∘). The intervals are much longer than those based on the asymptotic method, especially for the shorter series, and thus not very informative. This strategy is thus not recommended.

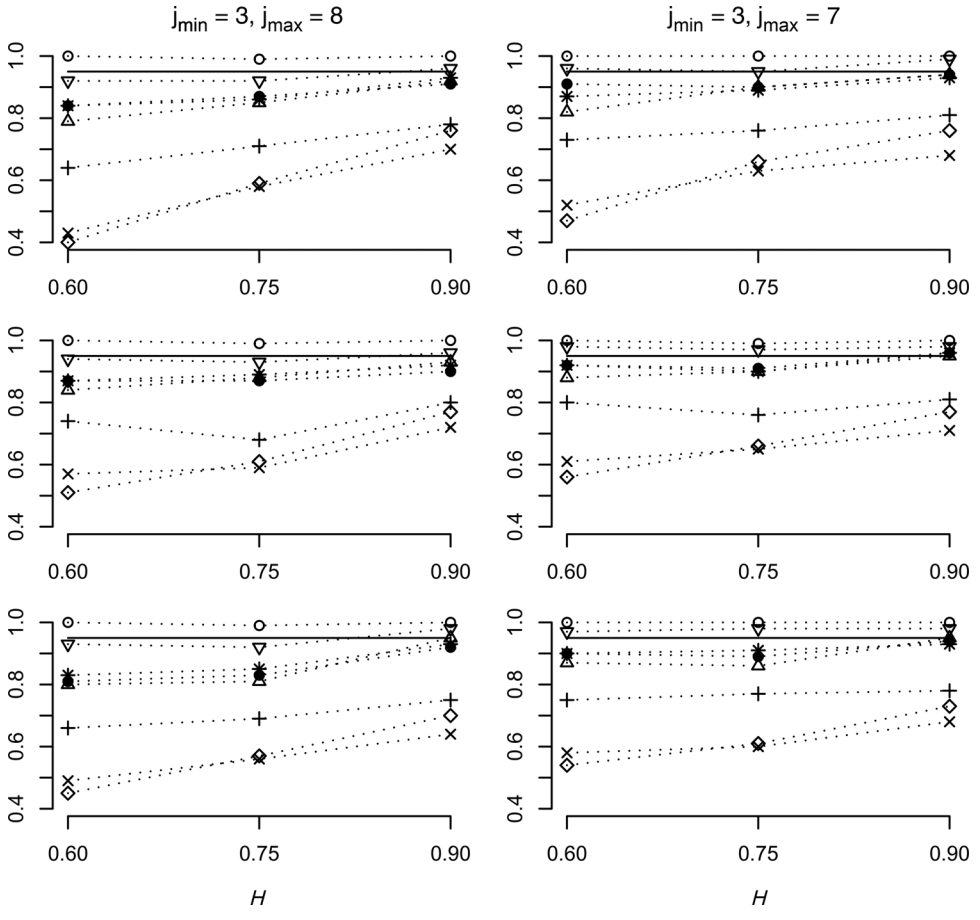

Figure 3. Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{12}$. Estimators were calculated using indicated range of scales.

Using the second strategy, for the shorter series ($N = 2^{12}$) the number of blocks were as follows:

$$\begin{aligned} &B_3 = B_4 = \cdots = B_6 = 4, B_7 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^7, \\ &B_3 = B_4 = \cdots = B_6 = 4, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8. \end{aligned} \tag{22}$$

For the longer series ($N = 2^{13}$), we used

$$\begin{aligned} &B_3 = B_4 = \cdots = B_6 = 4, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8, \\ &B_3 = B_4 = \cdots = B_6 = 4, B_7 = B_8 = B_9 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^9. \end{aligned} \tag{23}$$

In each case the number of estimators (14) was $S = 256$. Such a choice of the number of blocks, not only yields shorter confidence intervals, but also provides coverage probabilities very close to nominal (see figures 2 and 3, ∇).

We observe an increase in the coverage probability for the longer series with $j_{\min} = 3$ and $j_{\max} = 9$ as a function of H for all α, with the values approaching 95% as H increases. This slight undercoverage can be improved by choosing $j_{\min} = 3$ and $j_{\max} = 8$. The empirical coverage is then very close to the nominal value. The empirical coverage probability for the

shorter series (both sets of scales) is more stable and is close to 95%; however, intervals are longer and so less informative.

The results for the subsampling method can be summarized as follows.

- The subsampling method yields longer confidence intervals and higher coverage than the asymptotic method.
- When the subsamples are too short, coverage is close to 100%, the intervals are too long.
- Sampling about 25% of the coefficients at the lowest scales provides coverage of 90–95% (with a target value of 95%).
- Confidence intervals for the longer series are shorter. The choice of $j_{\min} = 3$ and $j_{\max} = 8$ for both lengths is optimal.

A referee of this paper suggested to study hybrid 'asymptotic-subsampling' $100(1-\alpha)\%$ confidence intervals defined as

$$\left(\hat{H}^{nb} + q_Z\left(\frac{\beta}{2}\right) s_{SN},\, \hat{H}^{nb} + q_Z\left(1-\frac{\beta}{2}\right) s_{SN}\right), \tag{24}$$

where s_{SN} is the sample standard deviation based on the 256 estimators $\hat{H}_{SN}$ with the number of blocks determined by equations (22) and (23). This method gives results very similar, the difference in empirical coverage probabilities is around 1–2% in either direction, to the percentile based subsampling method discussed in this section. All conclusions listed above remain valid for this modification.

4.3.2 Block bootstrap and bootstrap confidence intervals. In this section, we focus on confidence intervals based on block bootstrap. In a special case of block size equal to 1, following Percival *et al.* [15], we refer to them as bootstrap intervals.

For comparison with the subsampling method, we used the lengths of blocks implied by equation (20)–(23). We also used much shorter blocks: $l_j = 4$ and $l_j = 2$, $j = j_{\min}, \ldots, j_{\max}$, for both series lengths and both sets of scales.

Using l_j dictated by equations (20)–(23) leads to undercoverage. For $J = 13$ and blocks of lengths $l_3 = l_4 = l_5 = 63$, $l_6 = 62$, $l_7 = 60$, $l_8 = 28$, $l_9 = 12$ and $l_3 = l_4 = l_5 = 63$, $l_6 = 62$, $l_7 = 60$, $l_8 = 28$, the coverage is in the range 60–75%, while for the shorter series with $l_3 = l_4 = l_5 = 31$, $l_6 = 30$, $l_7 = 28$, $l_8 = 12$ and $l_3 = l_4 = l_5 = 31$, $l_6 = 30$, $l_7 = 28$, the coverage falls in the range 65–80% interval (see figures 2 and 3, $+$). These results can be improved dramatically if shorter blocks are used (see figures 2 and 3, $\bullet$). By taking $l_j = 4$ at all scales, the coverage for $N = 2^{13}$ and $N = 2^{12}$, for both sets of scales, increases to 80–95%. Very similar results are obtained for $l_j = 2$ (see figures 2 and 3, $*$). The bootstrap method yields similar coverage of about 75–95%, (see figures 2 and 3, $\triangle$). It has the advantage of not relying on a single block-size selection.

The results of simulations for the block bootstrap and bootstrap methods can be summarized as follows:

- Block bootstrap with short blocks and bootstrap methods generally provide good coverage, which is slightly below the nominal value of 95% – often by 5% and no more than 20%.
- When longer blocks are used, the undercoverage is more serious.

4.4 *Summary and conclusions*

Based on all our experiments the following overall conclusions can be drawn. Recall that the target coverage is 95%.

(1) The asymptotic method provides quite low empirical coverage probabilities, in the range 40–75%, even if only non-boundary wavelet coefficients are used. The use of all wavelet coefficients is not recommended as they introduce large biasas, whose direction depends on the boundary rule applied. An approximate value of α must be known to determine the length of the interval; however, a reasonable choice of this value, for example $\alpha = 1.5$, leads to confidence intervals yielding similar coverage probabilities to those based on correct α. The asymptotic method, even with correctly specified α, is outperformed by all resampling methods considered.
(2) Bootstrap method gives coverage between 75% and 95%.
(3) Block bootstrap and subsampling with non-overlapping blocks yield best results (coverage of 80–95%) for appropriate choice of the block lengths. If the blocks in the block bootstrap method are too long, the coverage is too low (60–80%). If the blocks in the subsampling method are too short, the coverage is too high (98–100%).
(4) Resampling confidence intervals are longer than the asymptotic, especially for the shorter series.

Subsampling with appropriate choice of blocks yields coverage probabilities typically some 1–5% below the nominal coverage. Block bootstrap gives shorter, more informative, confidence intervals, but at the expense of lower empirical coverage, 1–15% below the nominal coverage. The bootstrap method is an attractive alternative, leading to only slightly greater undercoverage than the block bootstrap, but not requiring block-size selection. These three methods provide coverage reasonably close to nominal and are much better than the asymptotic method, which has been used so far. To achieve better coverage it is necessary to use longer confidence intervals than the asymptotic ones.

We conclude this section with a heuristic explanation of the conclusions stated in point (3) above. The block bootstrap method reconstructs the series of DWT coefficients at octave j from blocks of these coefficients. If these blocks are too long, the reconstructed series look much like the original series, not enough variability is introduced, the estimators of H are too close to the estimator computed from the original sample, consequently, the confidence intervals based on the empirical distribution of these estimators are too short. By contrast, the subsampling method treats blocks in the same way as the original series of coefficients. These blocks are not put together, so in order for them to 'imitate' the original coefficients, they must be long enough. If they are too short, too much variability is introduced, resulting in confidence intervals, which are too long.

5. Application to Ethernet traffic

In this section, we present an application of our techniques to four Ethernet data sets. Each set contains a million packet arrival times together with the packet sizes in bytes recorded at an Ethernet link at the Bellcore Morristown Research and Engineering Facility. In these two-column data sets, the first column gives arrival time in seconds since the start of the trace, and the second gives the corresponding Ethernet packet size in bytes (for more information about these traces see [1] or http://ita.ee.lbl.gov/html/contrib/BC.html). To construct the discrete time series we used in our application, we computed the total number of bytes transmitted during consecutive time intervals of constant lengths 12, 10, 1000, and 1000 ms, thus obtaining four traces 'pAug', 'pOct', 'OctExt', and 'OctExt4', respectively. More specifically, let δ denote the length of the time interval and Z_t the size of a packet arriving at time t; the index t can take any value in the interval $[t_0, T]$, where t_0 and T are the first and the last values in the time

column, respectively. The discrete time process $\{Z_n^{(\delta)}\}$ based on a given trace is obtained by placing

$$Z_n^{(\delta)} = \sum_{\{t: t-t_0 \in [(n-1)\delta, n\delta]\}} Z_t, \quad n = 1, 2, \ldots, N, \quad N = \lfloor (T - t_0)/\delta \rfloor.$$

The four time series are plotted in figures 5–8.

Veitch and Abry [17, 18] developed a test for the time constancy of scaling exponents in self-similar or LRD Gaussian time series and applied it to Ethernet sequences similar to the sequences $\{Z_n^{(\delta)}\}$ defined above. If the observations can be modeled as increments of a fractional Brownian motion with self-similarity parameter H, then the scaling exponent γ is defined as $\gamma = 2H - 1$.

Their procedure can be summarized as follows.

1. Choose an appropriate $m > 1$ and divide the time series into m adjacent blocks.
2. Use a common range of octaves $[j_1(m), j_2(m)]$ to compute the estimates $\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_m$, which can be considered as uncorrelated Gaussian variables with unknown means γ_i and known variances σ_i^2, $i = 1, 2, \ldots, m$.
3. For a given significance level β, reject

$$H_0 : \text{all scaling exponents are equal } (\gamma_i = \gamma^0, i = 1, 2, \ldots, m)$$

in favour of

$$H_1 : \text{some scaling exponents are different } (\gamma_i \neq \gamma_j, \text{ for some } i \text{and } j),$$

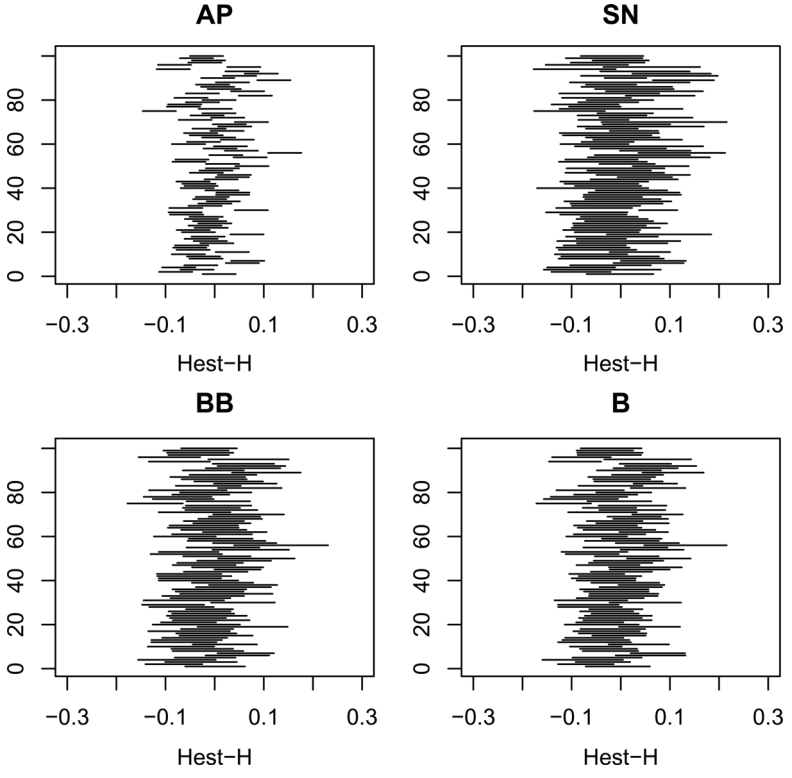

Figure 4. Examples of 95% confidence intervals (AP, SN, BB, B with subtracted H) based on $R = 100$ series of length $N = 2^{13}$ generated with $\alpha = 1.6$ and $H = 0.75$; scales $2^3, \ldots, 2^9$.

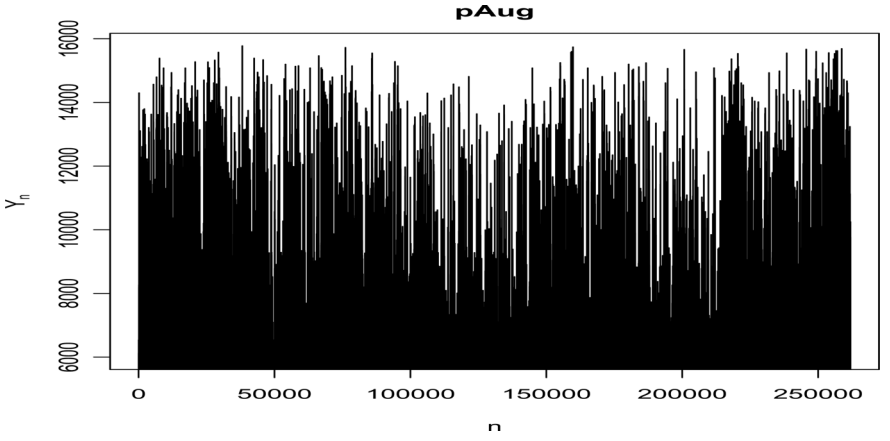

Figure 5. The time series of bytes per 12 ms based on 'pAug' trace.

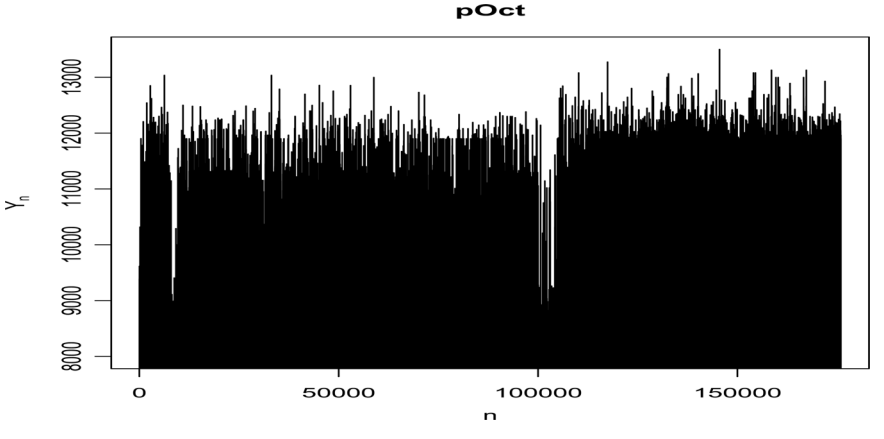

Figure 6. The time series of bytes per 10 ms based on 'pOct' trace.

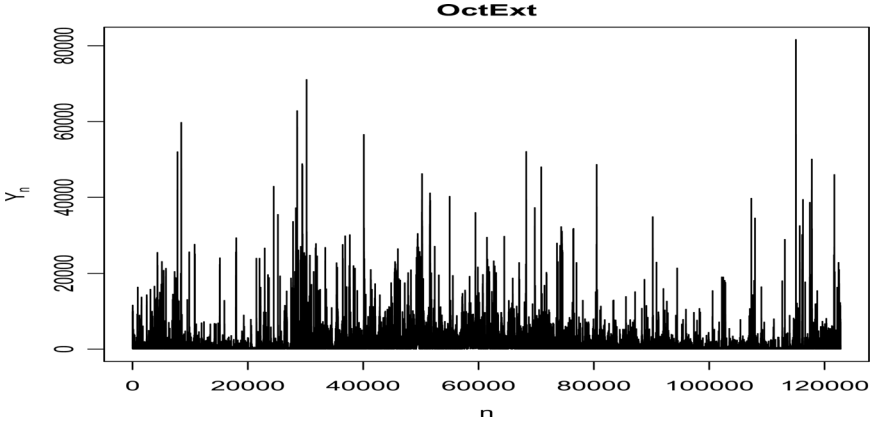

Figure 7. The time series of bytes per 1000 ms based on 'OctExt' trace.

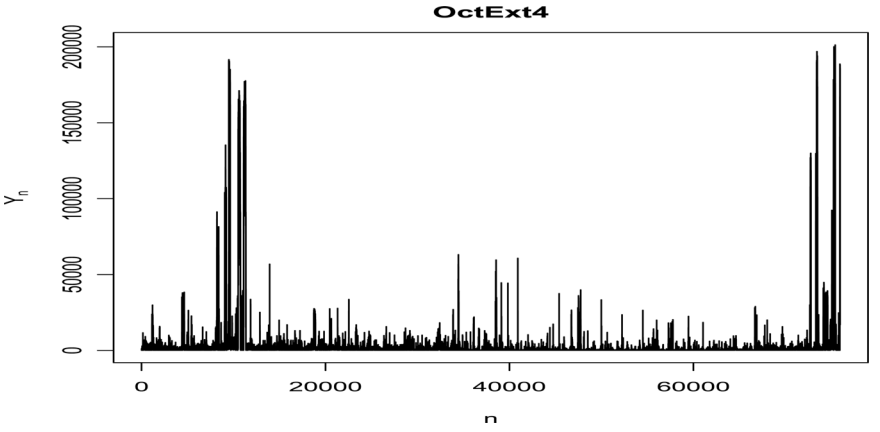

Figure 8. The time series of bytes per 1000 ms based on 'OctExt' trace.

if $V_m > C_m(\beta)$. Under the null hypothesis, the distribution of the test statistic

$$V_m = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left(\hat{\gamma}_i - \frac{\sum \hat{\gamma}_i/\sigma_i^2}{\sum 1/\sigma_i^2} \right)^2$$

is the chi–squared distribution with $m - 1$ degrees of freedom.

The procedure of Veitch and Abry [17] is essentially a one-way analysis of variance and relies on the fact that the estimators $\hat{\gamma}_i$, $i = 1, 2, \ldots, m$, are approximately independent normal random variables with identical known variance, which is approximated assuming that the observations themselves are normal.

The procedure we propose is applicable when the data can be modeled as realizations of a LFSM and does not require the assumptions that the observations are normal. Its practical implementation is based on one of the resampling methods of constructing confidence intervals for H. Similar procedures have been proposed in other contexts, see *e.g.*, [19].

Suppose then that $X_0, X_1, \ldots, X_N$ is a realization of a self-similar (motion-type) process with stationary increments sampled at equi-spaced time points. Our approach can be summarized as follows:

1. Divide a time series into m adjacent blocks of the same length.
2. For a given significance level β, use a common to all blocks range of octaves $[j_{\min}, j_{\max}]$ to construct $100(1-\beta)\%$ bootstrap (or block bootstrap, or subsampling) confidence intervals for the self-similarity parameter H_i, $i = 1, 2, \ldots, m$ in each block; denote this interval by (l_i, r_i).
3. Denote by k the largest number of intervals (l_i, r_i) with non-empty intersection.
4. At level of significance β reject

$$H_0 : H_i = H^0, \quad i = 1, 2, \ldots, m$$

in favour of

$$H_1 : H_i \neq H_j, \quad \text{for some } i \text{ and } j,$$

if $\lfloor 100((m-k)/m) \rfloor \% > \beta$ and conclude that H is not constant. Otherwise accept H_0 and conclude that there is no evidence that H is not constant.

For example, when $\beta = 0.05$, we fail to reject the null hypothesis if at least 95% of the confidence intervals (with a confidence level of 95%) overlap. Note that since the bootstrap and block bootstrap confidence intervals have a tendency to undercover the true H (section 4.4), the acceptance of the null hypothesis will yield a very strong evidence supporting the claim that there is no change in the self-similarity parameter. Rejection, on the other hand, has to be treated with caution.

We applied our test with B, BB, and SN methods to the cumulative sums of the four time series plotted in figures 5–8. After subtracting a linear trend, these cumulative sums can be modeled as realizations of a self-similar process with stationary increments. Recall that the wavelet coefficients, and hence our procedure, are not affected by a linear trend. We split the four series into 26, 17, 12, and 7 blocks, respectively, of lengths approximately equal to 10,000. We used scales $2^3, \ldots, 2^9$; however, the same conclusions for all series and all types of intervals (with slightly different intersections) are drawn when smaller set of scales $2^3, \ldots, 2^8$ is used.

Nominal B, BB, SN, 95% confidence intervals based on the first two series, 'pAug' and 'pOct', are presented in figures 9 and 10. For the series 'pAug', all 26 SN confidence intervals cover the same range (0.8330, 0.8690) and, therefore, we accept the null hypothesis supporting the claim about constancy in H. Less conservative tests based on B and BB methods reject this claim. Slightly different conclusions hold for the series 'pOct'. At least $100–\lfloor 100(1/17)\rfloor\%$ of BB and SN confidence intervals overlap, yielding intersections (0.7706, 0.7819) and

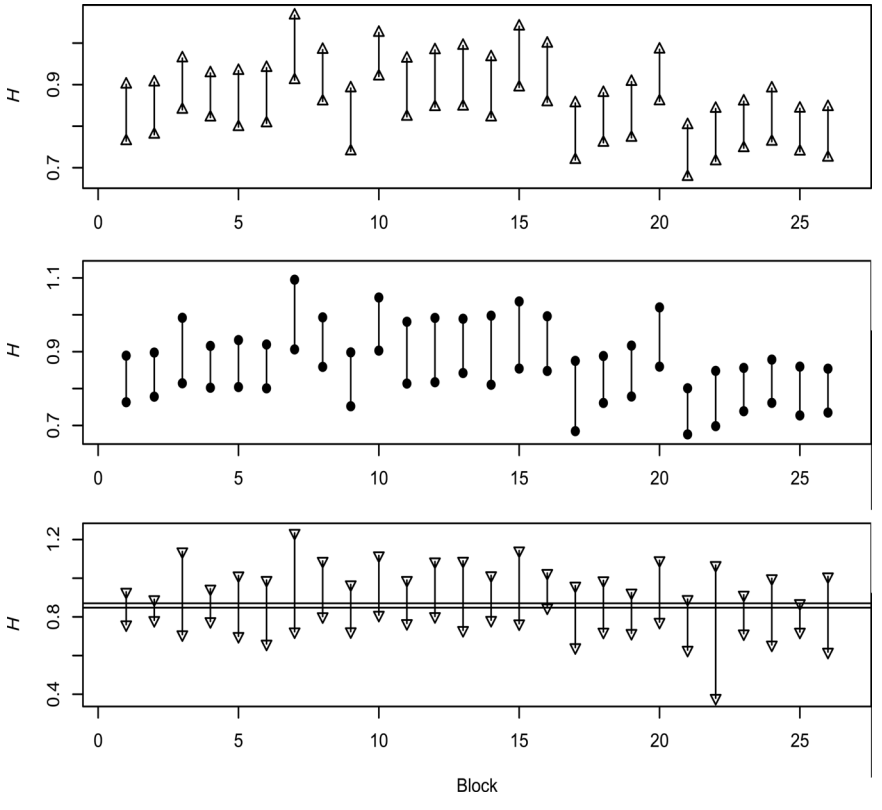

Figure 9. The 95% confidence intervals for H (B, BB, SN, respectively) from 26 adjacent blocks obtained from 'pAug' time series; scales: $2^3, \ldots, 2^9$.

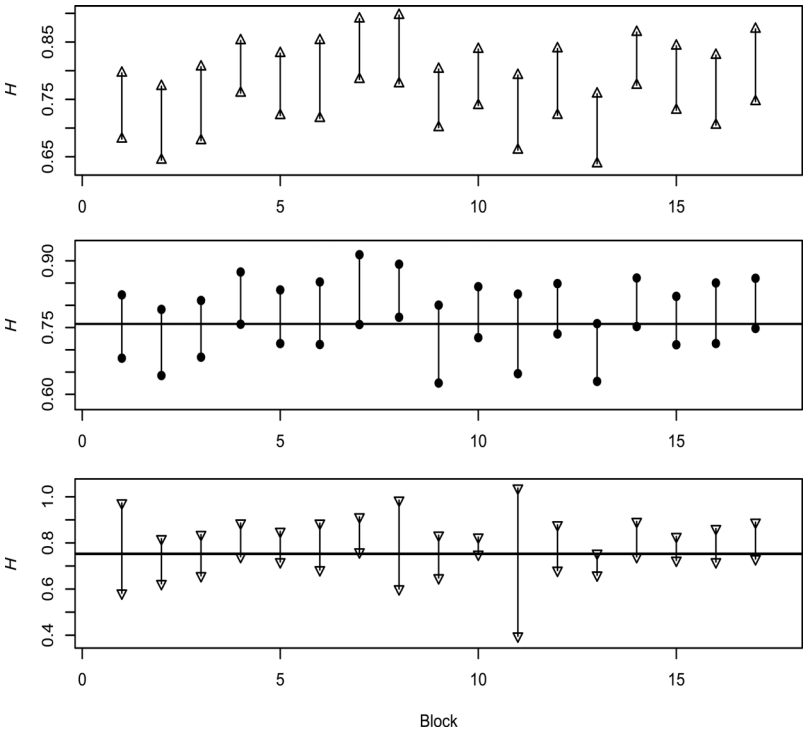

Figure 10. The 95% confidence intervals for H (B, BB, SN, respectively) from 17 adjacent blocks obtained from 'pOct' time series; scales: $2^3, \ldots, 2^9$.

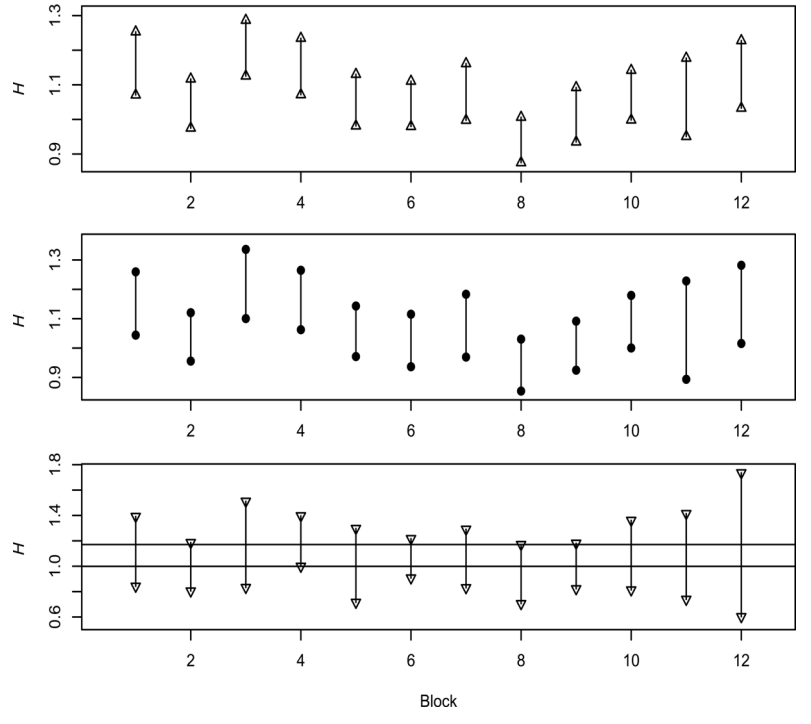

Figure 11. The 95% confidence intervals for H (B, BB, SN, respectively) from 12 adjacent blocks obtained from 'OctExt' time series; scales: $2^3, \ldots, 2^9$.

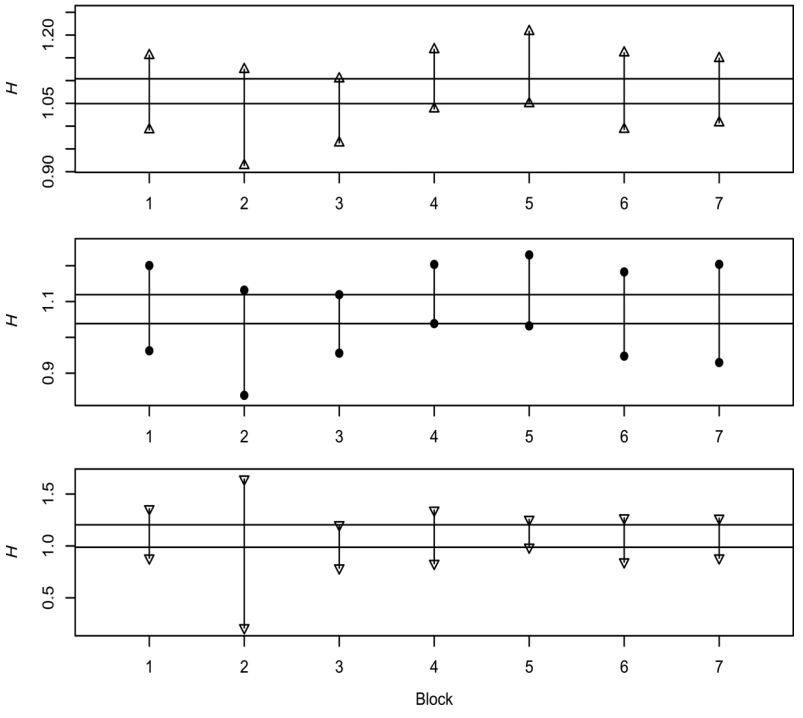

Figure 12. The 95% confidence intervals for H (B, BB, SN, respectively) from 7 adjacent blocks obtained from 'OctExt4' time series; scales: $2^3, \ldots, 2^9$.

(0.7422, 0.7468), respectively. Our results for the series 'pAug' and 'pOct', based on the SN method, thus accord with those arrived at by Veitch and Abry [17].

The 95% confidence intervals for the series 'OctExt' and 'OctExt4' are plotted in figures 11 and 12. The intersection of SN confidence intervals for the series 'OctExt' is (0.9829, 1.1855) and for the series 'OctExt4' is (0.9675, 1.2308). For the former series the constancy in H is not confirmed by the tests based on B and BB methods, however for the 'OctExt4' series, B and BB-based tests also indicate that H is constant, yielding intersections (1.0404, 1.1072) and (1.0163, 1.1365), respectively. If we were to accept the constancy in H for both series, the estimated value of H would fall into the range including values greater than 1. These series thus cannot be viewed as increments of a LFSM, for which $0 < H < 1$, but might well be assumed to be increments of a different self-similar process with $0 < \alpha < 1$. Both series, especially 'OctExt4', are seen to have very heavy tails. Recall, (see Corollary 7.1.11 of [7]); that if $0 < \alpha < 1$ the upper bound on H is $1/\alpha$ and not 1. Recall also that the test of Veitch and Abry [17] used an approximation to the variance of γ_i, which was based on the assumptions that the observations are approximately normal. Such an assumption is questionable for the series 'OctExt' and 'OctExt4', and the rejection reported by Veitch and Abry [17] might be spurious. We note, however, that our numerical experiments discussed in section 4 considered only LFSM with $1 < \alpha \leq 2$, so our conclusion of the constancy of H in the series 'OctExt' and 'OctExt4' must be treated with caution.

Acknowledgements

Partially supported by NSF grants DMS-0413653 and INT-0223262 and NATO grant PST.EAP.CLG 980599.

References

[1] Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V., 1994, On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, **2**, 1–15.

[2] Paxson, V. and Floyd, S., 1995, Wide area traffic: the failure of Poisson modelling. *IEEE/ACM Transactions on Networking*, **3**, 226–224.

[3] Park, K. and Willinger, W. (eds.), 2000, *Self-similar Network Traffic and Performance Evaluation* (New Jersey: John Willey & Sons).

[4] Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D., 2000, Wavelets for the analysis, estimation and synthesis of scaling data. In: K. Park and W. Willinger (Eds) *Self-Similar Network Traffic and Performance Evaluation* (New York: Wiley (Interscience Division)), pp. 39–88.

[5] Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D., 2002, Self-similarity and long-range dependence through the wavelet lens. In: P. Doukhan, G. Oppenheim and M. S. Taqqu (Eds.) *Theory and Applications of Long Range Dependence* (Boston: Birkhäuser).

[6] Veitch, D. and Abry, P., 1999, A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory special issue on* "*Multiscale Statistical Signal Analysis and its Applications*", **45**(3), 878–897.

[7] Samorodnitsky, G. and Taqqu, M.S., 1994, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman & Hall).

[8] Willinger, W., Paxson, V. and Taqqu, M.S., 1998, Self-similarity and heavy tails: structural modeling of network traffic. In: R. Feldman, R. Adler and M.S. Taqqu (Eds.) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (Boston: Birkhäuser).

[9] Karasaridis, A. and Hatzinakos, D., 2001, Network heavy traffic modeling using α-stable self-similar processes. *IEEE Transactions on Communications*, **49**(7), 1203–1214.

[10] Percival, D.B. and Walden, A.T., 2000, *Wavelet Methods for Time Series Analysis* (Cambridge: Cambridge University Press).

[11] Abry, P., Taqqu, M.S. and Pesquet-Popescu, B., 2000b, Wavelet based estimators for self similar α-stable processes. In: International Conference on Signal Processing, 16th World Computer Congress, Beijing, China. Available online at: http://www.ens-lyon.fr/∼pabry/.

[12] Stoev, S., Pipiras, V. and Taqqu, M.S., 2002, Estimation of the self-similarity parameter in linear fractional stable motion. *Signal Processing*, **80**, 1873–1901.

[13] Abry, P., Delbeke, L. and Flandrin, P., 1999, Wavelet-based estimator for the self–similarity parameter of α-stable processes. In: *Proceedings of the IEEE-International Conference on Acoustics, Speech and Signal Processing*, pp. 1729–1732.

[14] Daubechies, I., 1992, *Ten lectures on wavelets*, CBMS-NSF series, Vol. 61 (Philadelphia: SIAM).

[15] Percival, D.B., Sardy, S. and Davison, A.C., 2000, Wavestrapping time series: adaptive wavelet based bootstrapping. In: W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young (eds.) *Nonlinear and Nonstationary Signal Processing* (Cambridge University Press), pp. 442–471.

[16] Stoev, S. and Taqqu, M., 2004, Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform. *Fractals*, **12**, 95–121.

[17] Veitch, D. and Abry, P., 2001, A statistical test for the time constancy of scaling exponents. *IEEE Transactions on Signal Processings* **49**(10), 2325–2334.

[18] Veitch, D. and Abry, P., 2002, A statistical test for the time constancy of scaling exponents-extended version. Technical Report. University of Melbourne. Available online at http://www.emulab.ee.mu.oz.au/∼darryl/

[19] Katkovnik, V., 1999, A new method for varying adaptive bandwidth selection. *IEEE Transactions on Signal Processing*, **47**(9), 2567–2571.