Estimators for long-range dependence: an empirical study *

Murad S. Taqqu
Boston University

Vadim Teverovsky
Boston University

Walter Willinger
Bellcore

Abstract

Various methods for estimating the self-similarity parameter and/or the intensity of long-range dependence in a time series are available. Some are more reliable than others. To discover the ones that work best, we apply the different methods to simulated sequences of fractional Gaussian noise and fractional ARIMA$(0, d, 0)$. We also provide here a theoretical justification for the method of residuals of regression.

1 Introduction

Time series with long-range dependence appear in many contexts, for example, in the analysis of the traffic load in high speed networks (see Leland, Taqqu, Willinger and Wilson [6]). Many methods for estimating the self-similarity parameter H or the intensity of long-range dependence in a time series are available, some of which are described in detail in the recent monograph of Beran [1]. They are typically validated by appealing to self-similarity or to an asymptotic analysis where one supposes that the sample size of the time series converges to infinity. Leaving the question of robustness aside, we want to find out how well the methods function when applied to fractional Gaussian noise and fractional ARIMA. These are idealized time series for which the estimation methods should perform particularly well. We used a fairly large sample size $N = 10,000$, generated 50 different realizations for each of several values of H and compared the estimated values of H with the nominal ones, used in the simulation.

*This research was partially supported by the NSF grants NCR-9404931 and DMS-9404093 at Boston University.

The paper is organized as follows. In Section 2 we define both fractional Gaussian noise and fractional ARIMA, and in Section 3, we describe briefly the estimation methods. The results of the simulations are presented in Section 4.

2 The Time Series

We describe here fractional Gaussian noise (FGN) and fractional ARIMA (FARIMA), the two types of time series that we use in the simulations. These are the simplest models that display long-range dependence. The best way to introduce fractional Gaussian noise is through its "parent" *fractional Brownian motion* $\{B_H(t),\ t \geq 0\}$. Fractional Brownian motion is a Gaussian process with mean 0, stationary increments, variance $EB_H^2(t) = t^{2H}$ and covariance

$$EB_H(s)B_H(t) = \frac{1}{2}\{s^{2H} + t^{2H} - |s-t|^{2H}\} \tag{2.1}$$

It is statistically self-similar in the sense that $\{B_H(at),\ t \geq 0\}$ has the same finite-dimensional distributions as $\{a^H B_H(t),\ t \geq 0\}$ for all $a > 0$. The crucial index is H, a parameter which takes values between 0 and 1, called the *self-similarity* parameter.

Fractional Gaussian noise $\{X_i,\ i \geq 1\}$ is the increment of fractional Brownian motion, namely

$$X_i = B_H(i+1) - B_H(i),\ i \geq 1.$$

It is a mean zero, stationary Gaussian time series whose autocovariance function $\gamma(h) = EX_iX_{i+h}$ is given by $\gamma(h) = 2^{-1}\{(h+1)^{2H} - 2h^{2H} + |h-1|^{2H}\}$, $h \geq 0$. An important point about γ is that it satisfies

$$\gamma(h) \sim H(2H-1)h^{2H-2} \quad \text{as } h \to \infty \tag{2.2}$$

when $H \neq 1/2$ ($\sim$ means "asymptotic to"). Since $\gamma(h) = 0$ for $h \geq 1$ when $H = 1/2$, the X_i's are white noise in this case. The X_i's, however, are positively correlated when $\frac{1}{2} < H < 1$, and we say that they display *long-range dependence* or *long-memory.* The index H, in this context, measures the *intensity* of long-range dependence. The spectral density (Fourier transform of γ) is

$$f(\lambda) = C_H\Big(2\sin\frac{\lambda}{2}\Big)^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\lambda + 2\pi k|^{2H+1}} \sim C_H|\lambda|^{1-2H} \text{ as } \lambda \to 0, \tag{2.3}$$

where C_H is a constant.

Besides fractional Gaussian noise, we also consider *fractional ARIMA*$(0, d, 0)$. It is defined formally as

$$X_i = \Delta^{-d} \epsilon_i, \quad i \geq 1,$$

where the ϵ_i are independent, identically distributed normal random variables with mean 0 and variance 1, and where Δ is the differencing operator $\Delta \epsilon_i = \epsilon_i - \epsilon_{i-1}$. The way to interpret $X_i = \Delta^{-d} \epsilon_i$ with a fractional value of d is as a moving average:

$$X_i = \sum_{j=0}^{\infty} c_j \epsilon_{i-j}$$

where $c_j = \Gamma(j+d)\{\Gamma(d)\Gamma(j+1)\}^{-1}$, $j \geq 1$. Observe that $c_j \sim \Gamma(d)^{-1} j^{d-1}$ as $j \to \infty$, and that the autocovariance function $\gamma(h) = EX_iX_{i+h}$ of $X_i = \Delta^{-d}\epsilon_i$ satisfies for $0 < d < 1/2$,

$$\gamma(h) \sim C_d h^{2d-1} \quad \text{as} \quad h \to \infty \tag{2.4}$$

where $C_d = \pi^{-1}\Gamma(1-2d) \sin \pi d$. Thus, for large lags d, the autocovariance (2.4) has the same power decay as the autocovariance (2.2) of fractional Gaussian noise. Relating the exponents in (2.4) and (2.2) gives

$$d = H - \frac{1}{2}. \tag{2.5}$$

The advantage of fractional Gaussian noise over fractional ARIMA is that many of the asymptotic relations stated in the next section hold for finite sample sizes. This is because fractional Gaussian noise is the increment of the self-similar process fractional Brownian motion. The advantage of fractional ARIMA over fractional Gaussian noise is that it posesses a particularly simple spectral density

$$f(\lambda) = \frac{1}{2\pi}\Big(2 \sin \frac{\lambda}{2}\Big)^{-2d} \sim \frac{1}{2\pi}|\lambda|^{-2d} \text{ as } \lambda \to 0. \tag{2.6}$$

and that it is a particular case of fractional ARIMA(p, d, q), a versatile parametric family of models. Fractional ARIMA(p, d, q) is defined through the equations

$$\Phi(B)X_i = \Theta(B)\Delta^{-d}\epsilon_i \tag{2.7}$$

where $\Phi(B)$ and $\Theta(B)$ involve autoregressive and moving average coefficients respectively. Since we are not going to use (2.7) here, we refer the reader to Samorodnitsky and Taqqu [12] and to their monograph [13] for more details.

The fractional Gaussian noise series were simulated using the Durbin-Levinson algorithm, which was implemented in S-Plus using C routines. This algorithm, which is described for example in Brockwell and Davis [2], Chapter 8.2, provides an autoregressive representation of the Gaussian time series. The fractional ARIMA series were produced by the *arima.fracdiff.sim* command built into S-Plus Version 3.2 (S-Plus is a statistical package marketed by StatSci). The generation method is also based on the Durbin-Levinson algorithm and is described in Haslett and Raftery [4], p. 12-13.

To generate a Gaussian time series $X_i,\ i = 1, \cdots n$ with mean zero and autocovariance function γ, using the Durbin-Levinson algorithm, generate a sequence $\epsilon_i,\ i \geq 1$ of independent and identically distributed $N(0,1)$ random variables. Then set $X_1 = \gamma(0)^{1/2}\epsilon_1$ and suppose that $X_1, X_2, \ldots, X_n$ have been obtained. Then

$$X_{n+1} = \phi_{n,1}X_n + \cdots + \phi_{n,n}X_1 + v_n^{1/2}\epsilon_{n+1}. \tag{2.8}$$

The variances $v_i,\ i = 1, \cdots n$ and the coefficients $\{\phi_{n,i},\ i = 1, \cdots n\}$ are computed recursively. Set $v_0 = \gamma(0)$ and $v_n = v_{n-1}\left(1 - \phi_{n,n}\right)^2$. For $n \geq 1$,

$$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right]v_{n-1}^{-1}, \quad \phi_{n,i} = \phi_{n-1,i} - \phi_{n,n}\phi_{n-1,n-i} \text{ for } i < n.$$

3 Methods for Estimating H and d

3.1 Aggregated Variance Method

Divide the original time series $X = \{X_i, i \geq 1\}$ into blocks of size m and average within each block, that is consider the aggregated series

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i) \qquad k = 1, 2, \ldots, \tag{3.1}$$

for successive values of m. The index k, labels the block. Then take the sample variance of $X^{(m)}(k),\ k = 1, 2, \cdots$ within each block. This sample variance is an estimator of $\mathrm{Var}X^{(m)}$. Since, for fractional Gaussian noise and fractional ARIMA, $\mathrm{Var}X^{(m)} \sim {\sigma_0}^2 m^{\beta}$ as $m \to \infty$ where $\beta = 2H - 2 < 0$, we can obtain an estimate for β, or H, by proceeding as follows.

For a given m, divide the data, $X_1, \ldots, X_N$, into N/m blocks of size m, calculate

$X^{(m)}(k)$, for $k = 1, 2, \ldots, N/m$, and its sample variance

$$\widehat{\text{Var}}\; X^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k))^2 - \left(\frac{1}{N/m} \sum_{k=1}^{N/m} X^{(m)}(k) \right)^2 . \tag{3.2}$$

Repeat this procedure for different values of m and plot the logarithm of the sample variance versus $\log m$. Choose values of m, $\{m_i,\ i \geq 1\}$, that are equidistant on a log scale, so that $m_{i+1}/m_i = C$, where C is a constant which depends on the length of the series and the desired number of points.

Since $\widehat{\text{Var}}\; X^{(m)}$ is an estimate of $\text{Var} X^{(m)}$, the resulting points should form a straight line with slope $\beta = 2H - 2$, $-1 \leq \beta < 0$. In practice, the slope is estimated by fitting a line to the points obtained from the plot. It is assumed here that both m and N are large, and that $m \ll N$, so that both the length of each block, and the number of blocks is large. If X has (short-range or) no dependence, the slope obtained should equal -1 (this is the slope of the reference line in Figure 1, below).

3.2 Differencing the Variance

Common types of non-stationarity include jumps in the mean and slowly decaying trends. To distinguish them from long-range dependence, one can difference the variance (see Teverovsky and Taqqu [14] for details), that is, study

$$\widehat{\text{Var}} X^{(m_{i+1})} - \widehat{\text{Var}} X^{(m_i)}$$

where the m_i's are defined in the preceding subsection. Although differencing introduces additional fluctuations, it often provides a way of detecting the types of non-stationarity mentioned above. It is best used in conjunction with the basic aggregated variance method.

3.3 Absolute Values of the Aggregated Series

This method is very similar to the method of aggregated variance. The data is split in the same way, and the aggregated series (3.1) calculated. Instead of computing the sample variance, one finds the sum of the absolute values of the aggregated series, namely,

$$\frac{1}{N/m} \sum_{k=1}^{N/m} \left| X^{(m)}(k) \right| . \tag{3.3}$$

Then the logarithm of this statistic is plotted versus the logarithm of m. If the original series has long-range dependence with parameter H, the result should be a line with slope $H - 1$.

3.4 Higuchi's Method

This method was suggested by Higuchi [5]. It involves calculating the length of a path and, in principle, finding its fractal dimension D. The method is in fact very similar to the method of absolute values of the aggregated series discussed above. It involves taking the partial sums $Y(n) = \sum_{i=1}^{n} X_i$ of the original time series $\{X_i,\ i = 1, \ldots, N\}$, (e.g., producing fractional Brownian motion from fractional Gaussian noise) and then finding the normalized length of the curve, namely

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^{m} \left[\frac{N-i}{m}\right]^{-1} \sum_{k=1}^{[(N-i)/m]} \Big|Y(i+km) - Y(i+(k-1)m)\Big|,$$

where N is the length of the time series, m is essentially a block size and $[\,]$ denotes the greatest integer function. Then $EL(m) \sim C_H m^{-D}$, where $D = 2 - H$. Thus a log-log plot of $L(m)$ versus m should produce a straight line with a slope of $D = 2 - H$.

3.5 Residuals of Regression

This method which was used by Peng et al. [11] involves several steps. First, the series is broken up into blocks of size m. Then, within each of the blocks, the partial sums of the series are calculated. Call the partial sums within a block, $Y(i), i = 1, 2, ..., m$. Fit a least-squares line to the $Y(i)$ and compute the sample variance of the residuals. This procedure is repeated for each of the blocks, and the resulting sample variances are averaged. (Since the blocks are all of the same size, this is equivalent to calculating the sample variance of the entire series.) We prove in the Appendix that for large m, the resulting number is proportional to m^{2H} for fractional Gaussian noise; a similar result holds for fractional ARIMA. Thus, if the result is plotted on a log-log plot versus m, we should get a straight line with a slope of $2H$.

3.6 The R/S method

This is one of the better known methods. It is discussed in detail in Mandelbrot and Wallis [9], Mandelbrot [7] and Mandelbrot and Taqqu [8]. For a time series,

$X = \{X_i, i \geq 1\}$, with partial sum $Y(n) = \sum_{i=1}^{n} X_i$, and sample variance $S^2(n) := (1/n)\sum_{i=1}^{n} {X_i}^2 - (1/n)^2 Y(n)^2$, the R/S statistic, or the *rescaled adjusted range*, is given by :

$$\frac{R}{S}(n) := \frac{1}{S(n)} \left[\max_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0 \leq t \leq n} \left(Y(t) - \frac{t}{n} Y(n) \right) \right]. \tag{3.4}$$

For fractional Gaussian noise or fractional ARIMA,

$$E[R/S(n)] \sim C_H n^H, \tag{3.5}$$

as $n \to \infty$, where C_H is another positive, finite constant not dependent on n.

To determine H using the R/S statistic, proceed as follows. For a time series of length N, subdivide the series into K blocks, each of size N/K. Then, for each lag n, compute $R(k_i, n)/S(k_i, n)$, starting at points $k_i = iN/K + 1$, $i = 1, 2, \ldots$, such that $k_i + n \leq N$. For values of n smaller than N/K, one gets K different estimates of $R(n)/S(n)$. For values of n approaching N, one gets fewer values, as few as 1 when $n \geq N - N/K$.

Choosing logarithmically spaced values of n, plot $\log[R(k_i, n)/S(k_i, n)]$ versus $\log n$ and get, for each n, several points on the plot. This plot is sometimes called the pox plot for the R/S statistic. The parameter H can be estimated by fitting a line to the points in the pox plot. Since any short-range dependence in the series typically results in a transient zone at the low end of the plot, set a cut-off point, and do not use the low end of the plot for the purposes of estimating H. Usually, the very high end of the plot is not used as well, because there are too few points on the plot at the high end to make reliable estimates. The values of n that lie between the lower and higher cut-off points are used to estimate H.

3.7 Periodogram Method

One first calculates

$$I(\lambda) = \frac{1}{2\pi N} \Big| \sum_{j=1}^{N} X_j e^{ij\lambda} \Big|^2, \tag{3.6}$$

where λ is a frequency, N is the number of terms in the series, and X_j is the data. Because $I(\lambda)$ is an estimator of the spectral density, a series with long-range dependence should have a periodogram which is proportional to $|\lambda|^{1-2H}$ close to the origin. Therefore, a regression of the logarithm of the periodogram on the logarithm of the frequency λ should give a coefficient of $1-2H$. This provides an approximation to the parameter H. In practice, we use only the lowest 10% of the roughly $N/2 = 5000$ frequencies for the regression since the proportionality above only holds for λ close to the origin.

3.8 Modified Periodogram Method

This modification of the periodogram method compensates for the fact that on a log-log plot, most of the frequencies fall on the far right, thus exerting a very strong influence on the least-squared line fitted to the periodogram. The frequency axis is divided into logarithmically equally spaced boxes, and the periodogram values corresponding to the frequencies inside the box are averaged. Several of the values at very low frequencies are left untouched, since there are so few of them to begin with. In this study 0.1% of the points were left alone, and the rest divided into 60 boxes. Then a line was fit to the first 80% of the resulting points. Because the periodogram is very scattered we use a robustified least-squares (the *ltsreg* in Splus), formally a "least-trimmed squares regression" which minimizes the sum of the $q \simeq n/2$ smallest squared residuals.

3.9 Whittle Estimator

The Whittle estimator is also based on the periodogram. It involves the function

$$Q(\eta) := \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda;\eta)} d\lambda, \tag{3.7}$$

where $I(\lambda)$ is the periodogram (see (3.6)) and $f(\lambda;\eta)$ is the spectral density at frequency λ, and where η denotes the vector of unknown parameters. The Whittle estimator is the value of η which minimizes the function Q. When dealing with fractional Gaussian noise or fractional ARIMA, η is simply the parameter H or d. If the series is assumed to be FARIMA(p,d,q) (see (2.7)), then η includes also the unknown coefficients in the autoregressive and moving average parts. This estimator takes more time to obtain, but one also obtains confidence intervals. For details see Fox and Taqqu [3] and Beran [1]. Unlike the other estimators discussed here, the Whittle estimator is obtained through a non-graphical method. It also assumes that the parametric form of the spectral density is known.

4 Description of the Results

We generated a realization of fractional Gaussian noise with $H = 0.7$ and 10,000 points. Figures 1 and 2 illustrate the log-log plot corresponding to each of the estimation methods (except naturally Whittle).

However, no conclusions can be drawn on the basis of a single realization. For each value of $H = 0.6, 0.7, 0.8, 0.9$, we generated K=50 realizations, each 10,000 long, both for fractional Gaussian noise and fractional ARIMA$(0, d, 0)$. We also included $H = 0.5$ for fractional Gaussian noise, that is, a white noise sequence. The estimated values of d are expressed in terms of H by using (2.5).

For a given estimation method, we obtained K=50 estimated values of H, called $\{H_k,\ k = 1, \cdots, 50\}$. We computed their mean, standard deviation

$$\sigma^2 = \frac{1}{K-1}\left(\sum_{k=1}^{K} H_k^2 - \frac{1}{K}\left(\sum_{k=1}^{n} H_k\right)^2\right) \quad \text{and} \quad \text{MSE} = \frac{1}{K}\sum_{k=1}^{K}\left(H_k - \text{Nominal } H\right)^2.$$

The square root of MSE, the mean squared error, provides some information on the bias. The nominal H is the value of H we used to generate the fractional Gaussian noise (FGN) or fractional ARIMA (FARIMA). We did this for all the estimation methods discussed above. The results are presented in Table 1.

The box plots are a graphical representation of the results of the Table. The vertical axis in Figures 3 and 4 indicates the deviation from the nominal value of H. For each method we have – (1) a box representing the middle 50% of the data. That box contains a) a shaded region representing an approximate 95% confidence interval about the median (see McGill, Tukey and Larsen [10]); b) the median represented by the unshaded line in the middle of the shaded region; (2) "Whiskers" encompassing approximately 95% of the data, and designated by dashed lines; (3) Outliers that fell beyond the whiskers.

5 Appendix

We provide here a theoretical justification to the "Residuals of Regression" method described in Subsection 3.5. For simplicity, we suppose that the time series is fractional Gaussian noise with unit variance (the proof for fractional ARIMA is essentially the same).

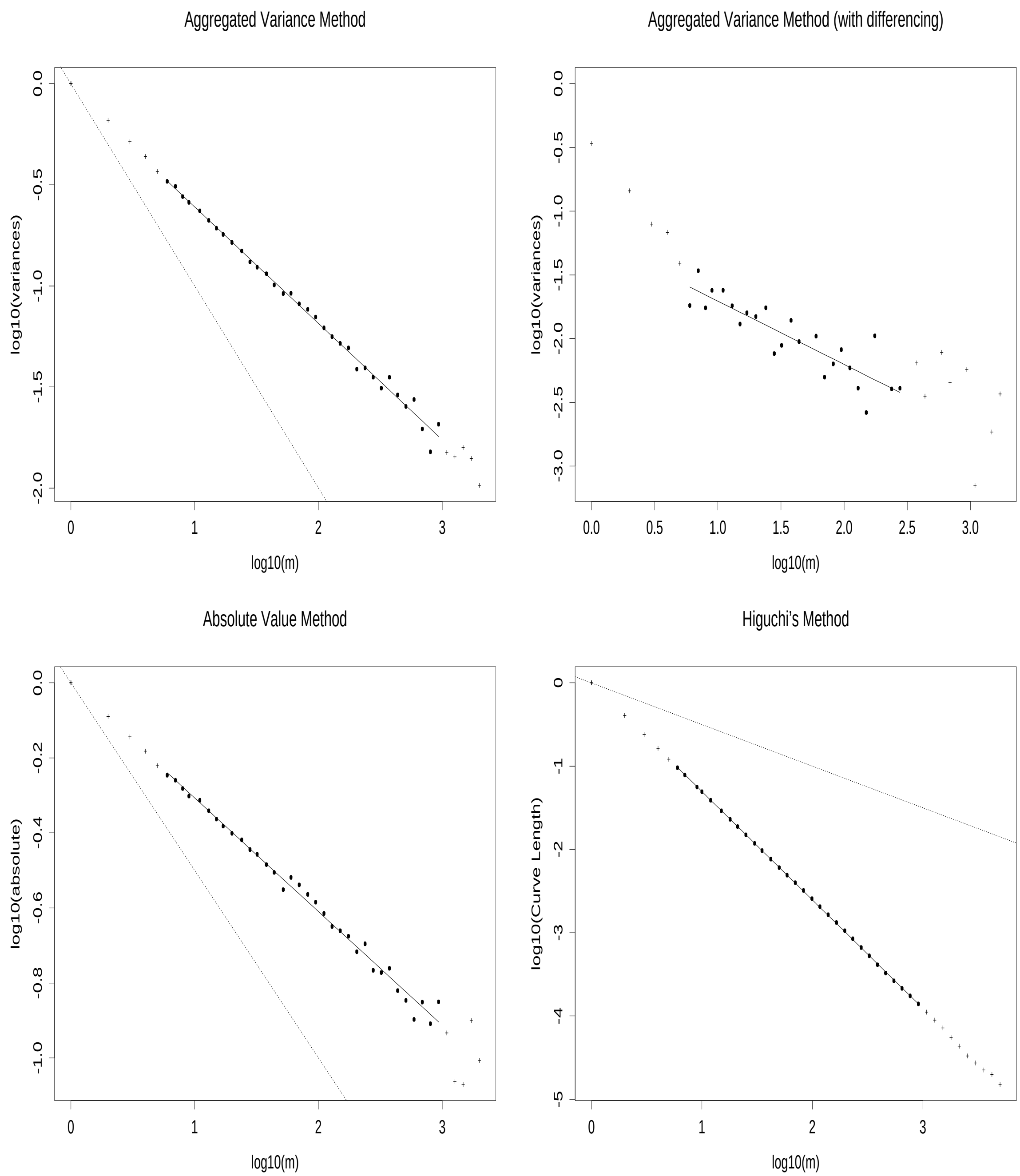

Figure 1: Estimating a simulated FGN ($H = 0.7$).

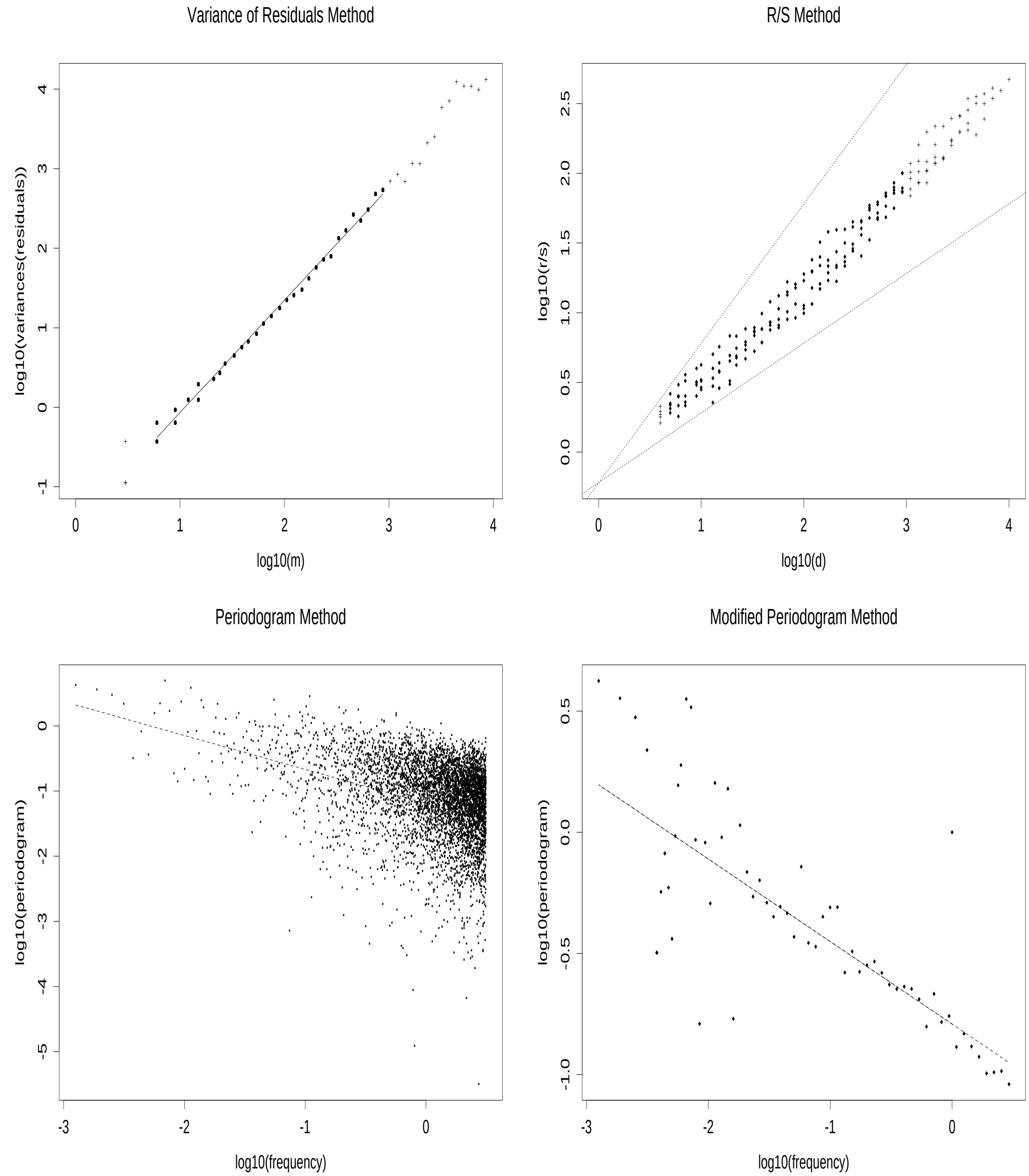

Figure 2: Estimating a simulated FGN ($H = 0.7$).

Estimation method		Nominal H								
		FGN					S-Plus FARIMA(0,d,0)			
		0.5	0.6	0.7	0.8	0.9	.6	.7	.8	.9
Variance	$\hat{H}$	.495	.588	.687	.772	.844	.591	.686	.773	.840
	$\hat{\sigma}$	.026	.027	.024	.022	.031	.031	.036	.028	.032
	$\sqrt{\text{MSE}}$	.026	.029	.027	.036	.063	.031	.038	.039	.068
DiffVar	$\hat{H}$	.483	.601	.694	.779	.878	.593	.686	.780	.872
	$\hat{\sigma}$	.057	.060	.077	.059	.076	.061	.071	.071	.070
	$\sqrt{\text{MSE}}$	.059	.059	.076	.062	.079	.061	.072	.073	.075
Absolute	$\hat{H}$	.497	.595	.700	.795	.896	.594	.698	.797	.888
	$\hat{\sigma}$	.028	.028	.024	.028	.049	.033	.040	.037	.040
	$\sqrt{\text{MSE}}$	.028	.028	.023	.028	.049	.033	.039	.037	.041
Higuchi	$\hat{H}$	.499	.595	.702	.795	.896	.596	.698	.797	.887
	$\hat{\sigma}$	.027	.027	.024	.028	.049	.032	.040	.037	.041
	$\sqrt{\text{MSE}}$	.027	.027	.024	.028	.049	.032	.040	.037	.042
Var. of Residuals	$\hat{H}$	.491	.589	.686	.782	.884	.583	.677	.772	.865
	$\hat{\sigma}$	.012	.015	.014	.018	.016	.016	.016	.017	.017
	$\sqrt{\text{MSE}}$	.015	.019	.020	.026	.022	.023	.028	.033	.039
R/S	$\hat{H}$	.535	.609	.687	.766	.821	.614	.688	.760	.823
	$\hat{\sigma}$	.023	.024	.020	.024	.027	.023	.022	.022	.026
	$\sqrt{\text{MSE}}$	.042	.025	.024	.042	.083	.027	.025	.046	.081
Periodogram	$\hat{H}$	.501	.601	.709	.812	.911	.609	.715	.816	.905
	$\hat{\sigma}$	.032	.029	.033	.025	.028	.039	.033	.031	.028
	$\sqrt{\text{MSE}}$	.032	.029	.034	.028	.030	.039	.034	.035	.028
Modified Periodogram	$\hat{H}$	.482	.595	.690	.796	.896	.598	.699	.808	.896
	$\hat{\sigma}$	.062	.048	.044	.058	.048	.051	.053	.049	.052
	$\sqrt{\text{MSE}}$	.064	.048	.044	.057	.048	.051	.052	.049	.051
Whittle	$\hat{H}$	.501	.601	.699	.800	.900	.601	.700	.800	.897
	$\hat{\sigma}$	.006	.006	.005	.007	.007	.008	.007	.009	.007
	$\sqrt{\text{MSE}}$	.006	.006	.005	.007	.007	.008	.007	.009	.008

Table 1: Estimation results for H using 50 independent realizations 10,000 long.

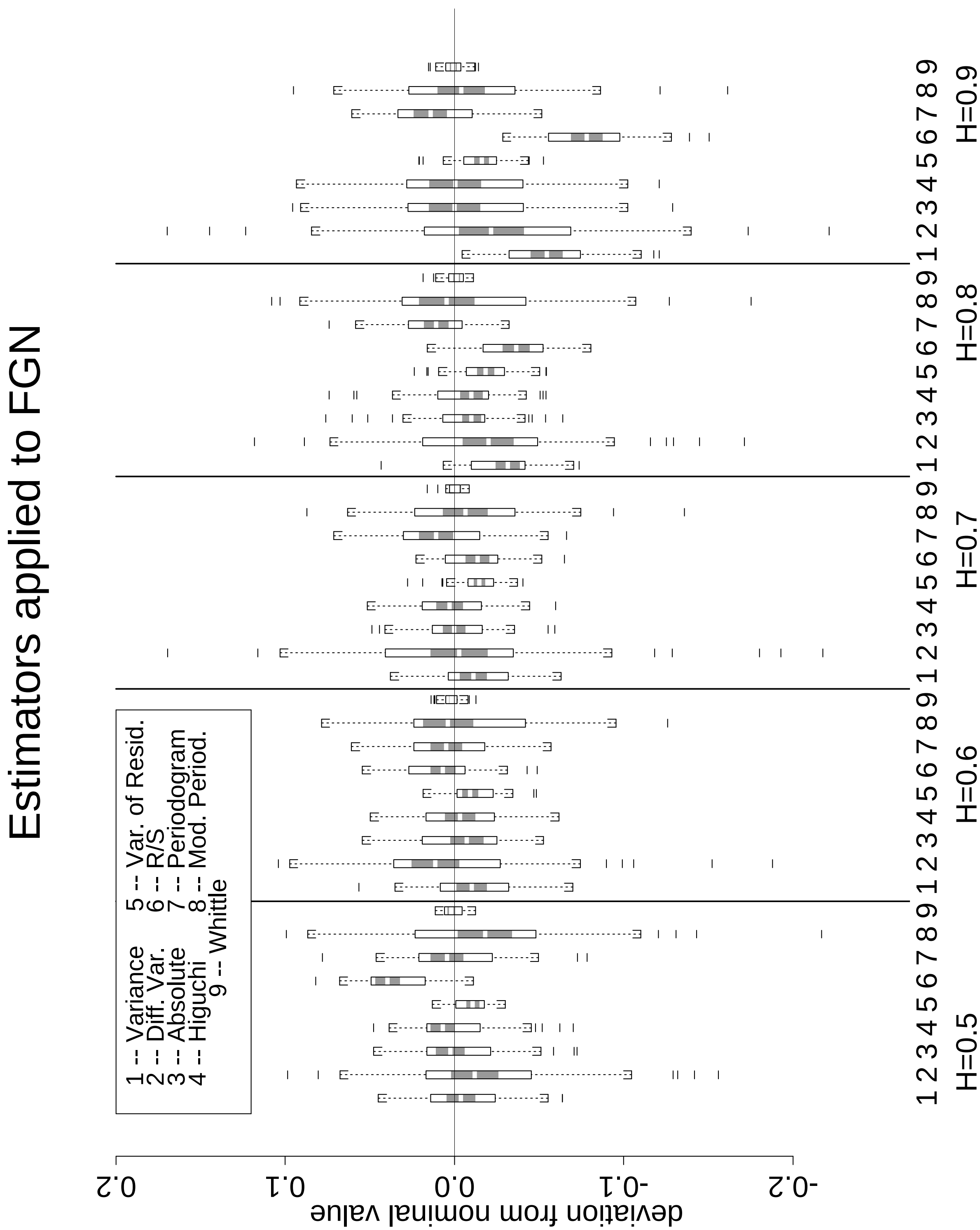

Figure 3: Boxplots of results based on 50 independent realizations for each H.

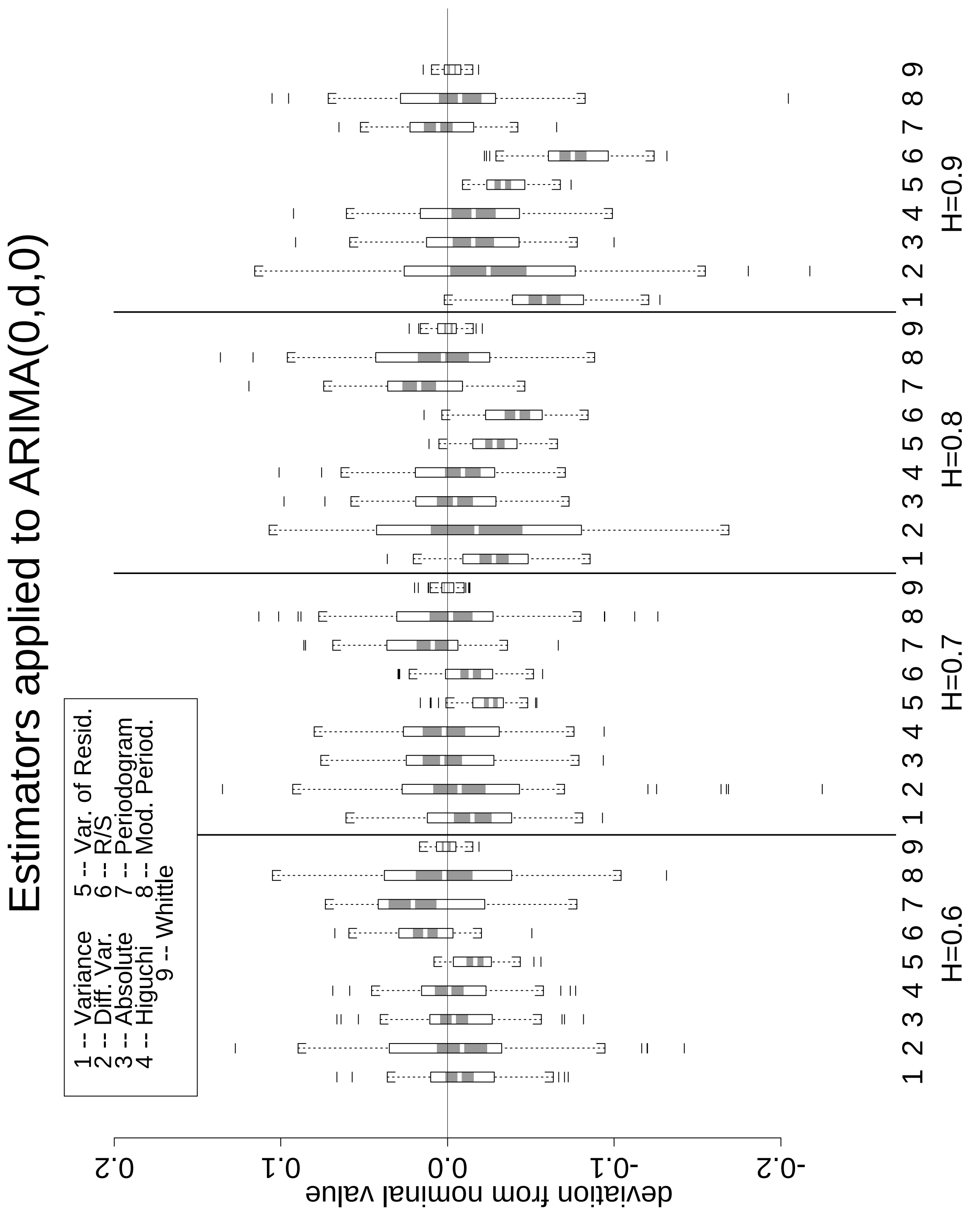

Figure 4: Boxplots of results based on 50 independent realizations for each H.

Recall that, to apply this method, one divides the time series into blocks of size m. Within each block one computes the partial sums $\{Y_t,\ t = 1, \cdots, m\}$, fits a regression $a + bt$ to these partial sums, and then computes the sample variance. The claim is that the expectation of the sample variance is asymptotically proportional to m^{2H}.

Theorem.

$$E\frac{1}{m}\sum_{t=1}^{m}(Y_t - a - bt)^2 \sim C_H m^{2H} \quad \text{as } m \to \infty,$$

where

$$C_H = \left(\frac{2}{2H+1} + \frac{1}{H+2} - \frac{2}{H+1}\right). \tag{5.1}$$

Proof: Since the data is supposed to be fractional Gaussian noise, the partial sums Y_i's are fractional Brownian motion, and thus their covariance is given by (2.1). Moreover, the slope b and intercept a of a least-square line on Y_t from 0 to m are given by:

$$b = \frac{\sum_{t=1}^{m} Y_t t - \frac{1}{m}\sum_{t=1}^{m} Y_t \sum_{t=1}^{m} t}{\sum_{t=1}^{m} t^2 - \frac{1}{m}\left(\sum_{t=1}^{m} t\right)^2} \simeq \frac{\sum_{t=1}^{m} Y_t t - \frac{m}{2}\sum_{t=1}^{m} Y_t}{m^3/12} \tag{5.2}$$

$$a = \frac{1}{m}\sum_{t=1}^{m} Y_t - \frac{1}{m}\sum_{t=1}^{m} bt \simeq \frac{1}{m}\sum_{t=1}^{m} Y_t - \frac{mb}{2}. \tag{5.3}$$

In the equation below, all of the sums are from $t = 1$ to $t = m$.

$$\begin{aligned}
&E\sum(Y_t - a - bt)^2 \\
&= E\left(\sum Y_t^2 + \sum a^2 + \sum (bt)^2 - 2\sum aY_t - 2\sum btY_t + 2\sum abt\right) \\
&\simeq E\left(\sum Y_t^2\right) + mE(a^2) + \frac{m^3}{3}E(b^2) - 2E(a\sum Y_t) - 2E(b\sum tY_t) + m^2E(ab) \\
&= E\left(\sum Y_t^2\right) + mE\left(\frac{1}{m}\sum Y_t - \frac{mb}{2}\right)^2 + \frac{m^3}{3}E(b^2) - 2E(b\sum tY_t) \\
&\quad -2E\left(\frac{1}{m}\left(\sum Y_t\right)^2 - \frac{m}{2}b\sum Y_t\right) + m^2E\left(\frac{1}{m}b\sum Y_t - \frac{mb^2}{2}\right) \\
&= E\left(\sum Y_t^2\right) + \frac{1}{m}E\left(\sum Y_t\right)^2 - mE\left(b\sum Y_t\right) + \frac{m^3}{4}E(b^2) + \frac{m^3}{3}E(b^2) \\
&\quad -\frac{2}{m}E\left(\sum Y_t\right)^2 + mE\left(b\sum Y_t\right) - 2E(b\sum tY_t) + mE\left(b\sum Y_t\right) - \frac{m^3}{2}E(b^2)
\end{aligned}$$

$$
\begin{aligned}
&= E\left(\sum Y_t^2\right) + \frac{m^3}{12}E(b^2) + mE\left(b\sum Y_t\right) - \frac{1}{m}E\left(\sum Y_t\right)^2 - 2E(b\sum tY_t) \\
&= E\left(\sum Y_t^2\right) + \frac{m^3}{12}\left[\left(\frac{12}{m^3}\right)^2 E\left(\sum tY_t - \frac{m}{2}\sum Y_t\right)^2\right] \\
&\quad + m\frac{12}{m^3}E\left(\sum tY_t \sum Y_t - \frac{m}{2}\left(\sum Y_t\right)^2\right) \\
&\quad - \frac{1}{m}E\left(\sum Y_t\right)^2 - 2\frac{12}{m^3}E\left(\left(\sum tY_t\right)^2 - \frac{m}{2}\sum tY_t \sum Y_t\right) \\
&= E\left(\sum Y_t^2\right) + \frac{12}{m^3}E\left(\sum tY_t\right)^2 - \frac{12}{m^2}E\left(\sum tY_t \sum Y_t\right) + \frac{3}{m}E\left(\sum Y_t\right)^2 \\
&\quad + \frac{12}{m^2}E\left(\sum tY_t \sum Y_t\right) - \frac{6}{m}E\left(\sum Y_t\right)^2 - \frac{1}{m}E\left(\sum Y_t\right)^2 - \frac{24}{m^3}E\left(\sum tY_t\right)^2 \\
&\quad + \frac{12}{m^2}E\left(\sum tY_t \sum Y_t\right) \\
&= E\left(\sum Y_t^2\right) - \frac{4}{m}E\left(\sum Y_t\right)^2 - \frac{12}{m^3}E\left(\sum tY_t\right)^2 + \frac{12}{m^2}E\left(\sum tY_t \sum Y_t\right) \\
&= A - \frac{4}{m}B - \frac{12}{m^3}D + \frac{12}{m^2}C. \qquad (5.4)
\end{aligned}
$$

In this equation we have represented several terms by A, B, C, and D respectively, and we will calculate them below.

$$
A := E\left(\sum_{t=1}^{m} Y_t^2\right) = \sum_{t=1}^{m} t^{2H} \simeq \frac{m^{2H+1}}{2H+1}.
$$

$$
\begin{aligned}
B &:= E(\sum_{t=1}^{m} Y_t)^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} E(Y_jY_k) = \sum_{j=1}^{m}\sum_{k=1}^{m}\frac{1}{2}\left(j^{2H} + k^{2H} - |j-k|^{2H}\right) \\
&= \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m}(j^{2H} + k^{2H}) - \sum_{j=1}^{m}\sum_{k<j}(j-k)^{2H} \\
&\simeq \frac{m^{2H+2}}{2H+1} - \sum_{j=1}^{m} j^{2H+1}\int_0^1 (1-x)^{2H}dx \\
&\simeq m^{2H+2}\left(\frac{1}{2H+1} - \frac{1}{(2H+2)(2H+1)}\right) = \frac{m^{2H+2}}{2H+2}.
\end{aligned}
$$

$$
\begin{aligned}
C &:= E\left(\sum_{t=1}^{m} Y_t \sum_{t=1}^{m} tY_t\right) = \sum_{j=1}^{m}\sum_{k=1}^{m}\frac{1}{2}\left(j^{2H+1} + k^{2H}j - j|j-k|^{2H}\right) \\
&= \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m}(j^{2H+1} + jk^{2H}) - \frac{1}{2}\sum_{j=1}^{m}\sum_{k<j} j(j-k)^{2H} - \frac{1}{2}\sum_{j=1}^{m}\sum_{k>j} j(j-k)^{2H}
\end{aligned}
$$

$$\simeq \frac{1}{2}\left(\frac{m^{2H+3}}{2H+2}+\frac{m^{2H+3}}{2(2H+1)}\right)-\frac{1}{2}\sum_{j=1}^{m} j^{2H+2}\left(\int_0^1 (1-x)^{2H}dx-\int_0^1 x(1-x)^{2H}dx\right)$$

$$\simeq \frac{m^{2H+3}}{4}\left(\frac{1}{H+1}+\frac{1}{2H+1}\right)-\frac{m^{2H+3}}{2(2H+3)}\left(\frac{1}{2H+1}-\frac{1}{(2H+2)(2H+1)}\right)$$

$$= \frac{m^{2H+3}}{4}\left(\frac{2}{H+1}-\frac{1}{2H+1}\right).$$

Next

$$D := E\left(\sum_{t=1}^{m} tY_t\right)^2 = \sum_{j=1}^{m}\sum_{k=1}^{m}\frac{1}{2}\left(j^{2H+1}k+jk^{2H+1}-jk|j-k|^{2H}\right)$$

$$= \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{m}(j^{2H+1}k+jk^{2H+1})-\sum_{j=1}^{m}\sum_{k<j} jk(j-k)^{2H}$$

$$\simeq \frac{1}{2}\left(\frac{m^{2H+4}}{2H+2}\right)-\sum_{j=1}^{m} j^{2H+3}\int_0^1 x(1-x)^{2H}dx$$

$$\simeq \frac{m^{2H+4}}{4(H+1)}-\frac{m^{2H+4}}{2H+4}\left(\frac{1}{2H+1}-\frac{1}{2H+2}\right)$$

$$= \frac{m^{2H+4}}{4(H+1)}\left(1-\frac{1}{(H+2)(2H+1)}\right).$$

Therefore, going back to (5.4):

$$E\sum_{t=1}^{m}(Y-a-bt)^2 = A-\frac{4}{m}B+\frac{12}{m^2}C-\frac{12}{m^3}D = C_H m^{2H+1},$$

where C_H is given in (5.1). Dividing the terms in this last expression by m yields the result.

References

[1] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, New York, 1994.

[2] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition, 1991.

[3] R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 14:517–532, 1986.

[4] J. Haslett and A. E. Raftery. Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *Applied Statistics*, 38:1–50, 1989. Includes discussion.

[5] T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 31:277–283, 1988.

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

[7] B. B. Mandelbrot. Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31:271–285, 1975.

[8] B. B. Mandelbrot and M. S. Taqqu. Robust R/S analysis of long-run serial correlation. In *Proceedings of the 42nd Session of the International Statistical Institute*, Manila, 1979. Bulletin of the International Statistical Institute. Vol.48, Book 2, pp. 69-104.

[9] B .B. Mandelbrot and J. R. Wallis. Computer experiments with fractional Gaussian noises, Parts 1,2,3˙ *Water Resources Research*, 5:228–267, 1969.

[10] R. McGill, J. W. Tukey, and W. A. Larsen. Variation of box plots. *The American Statistician*, 32:12–16, 1978.

[11] C. K. Peng, S. V. Buldyrev, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49:1685–1689, 1994.

[12] G. Samorodnitsky and M. S. Taqqu. Linear models with long-range dependence and finite or infinite variance. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. S. Taqqu, editors, *New Directions in Time Series Analysis, Part II*, pages 325–340, New York, 1992. IMA Volumes in Mathematics and its Applications, Volume 46, Springer-Verlag.

[13] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York, London, 1994.

[14] V. Teverovsky and M. S. Taqqu. Testing for long-range dependence in the presence of shifting means or a slowly declining trend using a variance-type estimator. Preprint.

Murad S. Taqqu and Vadim Teverovsky
Boston University
Department of Mathematics
111 Cummington Street
Boston, MA 02215-2411, USA
Email: murad@math.bu.edu, vt@math.bu.edu

Walter Willinger
Bellcore
Morristown, NJ 07960-6438, USA
Email: walter@bellcore.com