A critical look at Lo's modified R/S statistic *†‡

Vadim Teverovsky, Boston University
Murad S. Taqqu, Boston University
Walter Willinger, AT&T Labs-Research

December 11, 1998

Abstract

We report on an empirical investigation of the modified *rescaled adjusted range* or R/S statistic that was proposed by Lo [15] as a test for long-range dependence with good robustness properties under "extra" short-range dependence. In contrast to the classical R/S statistic that uses the standard deviation S to normalize the rescaled range R, Lo's modified R/S-statistic V_q is normalized by a modified standard deviation S_q which takes into account the covariances of the first q lags, so as to discount the influence of the short-range dependence structure that might be present in the data. Depending on the value of the resulting test-statistic V_q, the null hypothesis of no long-range dependence is either rejected or accepted. By performing Monte Carlo simulations with "truly" long-range and short-range dependent time series, we study the behavior of V_q, as a function of q, and uncover a number of serious drawbacks to using Lo's method in practice. For example, we show that as the truncation lag q increases, the test statistic V_q has a strong bias toward accepting the null hypothesis (i.e., no long-range dependence), even in ideal situations of "purely" long-range dependent data.

1 Introduction

The phenomenon of long-range dependence has a long history and has remained a topic of active research in the study of economic and financial time series (e.g., see Lo [15], and Cutland, Kopp and Willinger [8], and references therein). It is widespread in other areas in the physical and natural sciences (e.g., see Mandelbrot [16] and Beran [5] for details) and has been extensively documented in hydrology, meteorology and geophysics (see for example Mandelbrot and Wallis [18, 19, 21]). More recently, long-range dependence has also started to play an important role in the engineering sciences, especially in the analysis and performance modeling of traffic measurements from modern high-speed communications networks (for a recent bibliographical survey of this area, see Willinger, Taqqu and Erramilli [25]).

Long-range dependent processes provide an elegant explanation and interpretation of an empirical law that is commonly referred to as *Hurst's law* or the *Hurst effect*. In short, for a given set of observations $(X_i, i \geq 1)$, with partial sum $Y(n) = \sum_{i=1}^{n} X_i, n \geq 1$, and sample variance $S^2(n) =$

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

†*AMS Subject classification:* 60G18, 62-09.

‡*Keywords and phrases:* Long-range dependence, Fractional Gaussian Noise, Fractional ARIMA, long memory, R/S.

$n^{-1}\sum_{i=1}^{n}(X_i - n^{-1}Y(n))^2, n \geq 1$, the *rescaled adjusted range statistic* or *R/S-statistic* is defined by

$$\frac{R}{S}(n) = \frac{1}{S(n)}\left[\max_{0\leq t\leq n}\left(Y(t) - \frac{t}{n}Y(n)\right) - \min_{0\leq t\leq n}\left(Y(t) - \frac{t}{n}Y(n)\right)\right], \quad n \geq 1. \tag{1.1}$$

Hurst [13] found that many naturally occurring empirical records appear to be well represented by the relation

$$E\left[\frac{R}{S}(n)\right] \sim c_1 n^H, \quad \text{as } n \to \infty, \tag{1.2}$$

with typical values of the *Hurst parameter H* in the interval $(0.5, 1.0)$, and c_1 a finite positive constant that does not depend on n. On the other hand, if the observations X_i come from a short-range dependent model, then it is known [9, 3] that

$$E\left[\frac{R}{S}(n)\right] \sim c_2 n^{0.5}, \quad \text{as } n \to \infty, \tag{1.3}$$

where c_2 is independent of n, and finite and positive. The discrepancy between relations (1.2) and (1.3) is generally referred to as the *Hurst effect* or the *Hurst phenomenon*.

Classical R/S-analysis aims at inferring from an empirical record the value of the Hurst parameter in relation (1.2) for the long-range dependent process that presumably generated the record at hand. In practice, classical R/S-analysis is based on a heuristic graphical approach, originally developed by Mandelbrot and Wallis [19] (see also Mandelbrot and Taqqu [17] and Bassingthwaighte and Raymond [4]), that computes the R/S-statistic in (1.1) at many different lags n and for a number of different points (for details, see Section 2), plots the resulting estimates versus the lags on log-log scale, and yields an estimate of the Hurst parameter via the slope of the resulting "pox plot". Classical R/S-analysis is clearly not reliable for small samples, but can be highly effective and useful as a graphical or "eye balling" method for reasonably large samples, where it often provides a rather accurate picture of the presence or absence of long-range dependence in a given empirical record and, in the former case, about the intensity of long-range dependence as measured by the Hurst parameter. For practical purposes, the most useful feature of the classical R/S-analysis is its relative robustness under changes in the marginal distribution of the data, particularly if the marginals exhibit heavy tails with infinite variance (see for example Mandelbrot and Wallis [20] and Mandelbrot and Taqqu [17]).

The less attractive features of the classical R/S-analysis are its sensitivity to the presence of explicit short-range dependence structures, its bias, and a lack of a distribution theory for the underlying statistic (1.1). These characteristics stand in the way of using classical R/S as a rigorous statistical inference method. To overcome some of these shortcomings, Lo [15] proposed a *modified R/S-statistic* that is obtained by replacing the denominator S in (1.1), i.e., the sample standard deviation, by a consistent estimator of the square root of the variance of the partial sum $Y(n)$. The motivation for this modification is that in the case of dependent random variables, the variance of the partial sum is not simply the sum of the variances of the individual X_i's but also includes their autocovariances up to some lag, for a judicious choice of the truncation lag. Lo derives the limiting distribution of his modified R/S-statistic under both short-range and long-range dependence, claims that it is robust to short-range dependence, and illustrates through Monte Carlo simulations that it has reasonable power against certain long-range dependence alternatives.

In this paper, we revisit Lo's modified R/S-method for testing whether or not there is long-range dependence in a given time series. Using synthetically generated time series (long-range as well as short range dependent), we show that although Lo's method is a significant improvement over the original R/S-method, in contrast to the prevailing view in the econometrics literature (Hauser, Kunst, Reschenhofer [11], Huang and Yang [12], Campbell, Lo and MacKinlay [7]), it does not appear to provide the "ultimate" test for long-range dependence. In fact, we show here that there are a number of problems associated with Lo's method and it's use in practice. The most important finding is that Lo's method has a strong preference for accepting the null hypothesis of no long-range dependence, irrespective of whether long-range dependence is present in the data or not. This means that when Lo's method indicates that there is no evidence of long-range dependence in a given data set, one still has to conduct further investigations. Based on this experience as well as on additional evidence, we do not recommend to use Lo's method in isolation, i.e., as the sole technique to test for long-range dependence, but to always rely on a diverse set of graphical and statistical methods for checking for long-range dependence, as described for example in Beran [5], Taqqu, Teverovsky and Willinger [24], Taqqu and Teverovsky [23] and Abry and Veitch [1].

In Willinger, Taqqu and Teverovsky [26], we illustrate the main points of our study with an application of Lo's modified R/S statistic to historical records of daily stock return indices from the Center for Research in Security Prices (CRSP). In the context of asset returns, the classical R/S-analysis was first considered by Mandelbrot, who suggested H-values of around 0.55 to be representative for stock returns; later on, Greene and Fielitz [10] used it in a large-scale empirical study of 200 daily stock return series of securities listed on the New York Stock Exchange and claimed to have found significant evidence for the presence of long-range dependence in many of these series or asset returns. In contrast, when using his modified R/S statistic, Lo [15] finds no evidence of long-range dependence in the CRSP data. He attributes the findings in [10] to the fact that the classical R/S method is sensitive to the presence of short-range dependence and concludes that the dynamic behavior of asset returns may be adequately described by traditional, short-range dependent models. In [26], we revisit the CRSP data sets that Lo used in his empirical study of long memory in stock returns and discuss his findings in view of the new evidence presented here. By performing an in-depth analysis of the data, we identify the causes that led Lo to the acceptance of the null hypothesis (no long-range dependence), and arrive at a much less conclusive picture. In fact, we find empirical evidence that the CRSP data exhibit long-range dependence, but because the corresponding Hurst parameters are typically very low (around 0.60), this evidence is not absolutely conclusive.

The paper is structured as follows. In Section 2, we review the classical R/S-method and give a detailed description of Lo's modified R/S-statistic. Section 3 describes the results of our Monte Carlo simulation studies that used synthetically generated short-range and long-range dependent time series and provide new insights into the behavior of Lo's modified R/S-statistic. We conclude with some constructive suggestions for dealing with the long-range dependence issue in a given data set in Section 4.

2 The Classical and Modified R/S-Methods

In this section, we give a brief description of the classical R/S-analysis methodology introduced by Mandelbrot and Wallis [19], and outline the basic ideas behind Lo's [15] modified R/S-statistic and its main properties.

2.1 Graphical R/S-Method

The graphical implementation of the classical R/S-statistic given by equation (1.1) attempts to exploit as fully as possible the information in a given historical record. To this end, a given sample of N observations is subdivided into K *blocks*, each of size N/K. Then, for each *lag* n, $n \leq N$, estimates $R(k_m, n)/S(k_m, n)$ of $R(n)/S(n)$ are computed by starting at the points, $k_m = (m-1)N/K + 1$, $m = 1, 2, \ldots, K$, and such that $k_m + n \leq N$. Thus, for any given m, all the data points before $k_m = mN/K + 1$ are ignored. For values of n smaller than N/K, there are K different estimates of $R(n)/S(n)$; for values of n approaching N, there are fewer values, as few as 1 when $n \geq N - N/K$. Also note that the $R(k_m, n)/S(k_m, n)$ values corresponding to neighboring values of k_m and n are strongly interdependent; for a given $n > N/K$, the various estimates of $R(n)/S(n)$ involve overlapping observations, and so do the estimates when evaluated at different lags but for a fixed starting point k_m.

The graphical R/S-approach consists of calculating the estimates $R(k_m, n)/S(k_m, n)$ for logarithmically spaced values of n, and plotting $\log(R(k_m, n)/S(k_m, n))$ versus $\log(n)$, for all starting points k_m. This results in the *rescaled adjusted range plot*, also known as the *pox plot of* R/S. When the Hurst parameter H in relation (1.2) is defined, a typical rescaled adjusted range plot starts with a transient regime indicating the presence of a particular short-range dependence structure in the data, but will eventually settle down and fluctuate around a straight "street" of slope H. The use of the graphical R/S-analysis in practice is to determine whether such an asymptotic behavior appears to be supported by the data and, in the affirmative, to estimate the asymptotic Hurst exponent H as the street's slope, usually by a simple least-squares fit. The freedom in choosing the various parameters associated with the graphical R/S-analysis (i.e., the number of blocks K, the number of lags n, and the degree of overlapping) can often be used very effectively to provide a coherent picture of the underlying short-range dependence structure as well as of the nature of the apparent asymptotic regime governed by the Hurst parameter H. Examples of synthetically generated traces, each of length $N = 10{,}000$, of a "purely" long-range dependent process (i.e., fractional Gaussian noise with $H = 0.7$ – see Section 3.1) and of a generic "hybrid" long-range/short-range dependent process (i.e., fractional ARIMA$(1, d, 1)$ with AR-parameter 0.3, MA-parameter 0.7, and $d = 0.4$ – see Section 3.1), with corresponding pox plots of R/S are shown in Figure 1. Notice the presence of a distinct transient regime for the hybrid long-range/short range series (bottom right), while for the purely long-range dependent time series (bottom left), the asymptotic regime is evident over the whole range of lag-values (bottom left).

2.2 Definition of Lo's Modified R/S-Statistic $V_q(N)$

Following Lo [15], define R as in (1.1), but instead of considering multiple lags, only focus on lag $n = N$, the length of the series. Furthermore, instead of simply using the sample standard deviation, S, to normalize R, use a weighted sum of autocovariances, namely,

$$S_q(N) = \left(\frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X}_N)^2 + \frac{2}{N} \sum_{j=1}^{q} \omega_j(q) \left[\sum_{i=j+1}^{N} (X_i - \bar{X}_N)(X_{i-j} - \bar{X}_N) \right] \right)^{1/2}, \qquad (2.1)$$

where $\bar{X}_N$ denotes the sample mean of the time series, and the weights $\omega_j(q)$ are given by

$$\omega_j(q) = 1 - \frac{j}{q+1}, \quad q < N.$$

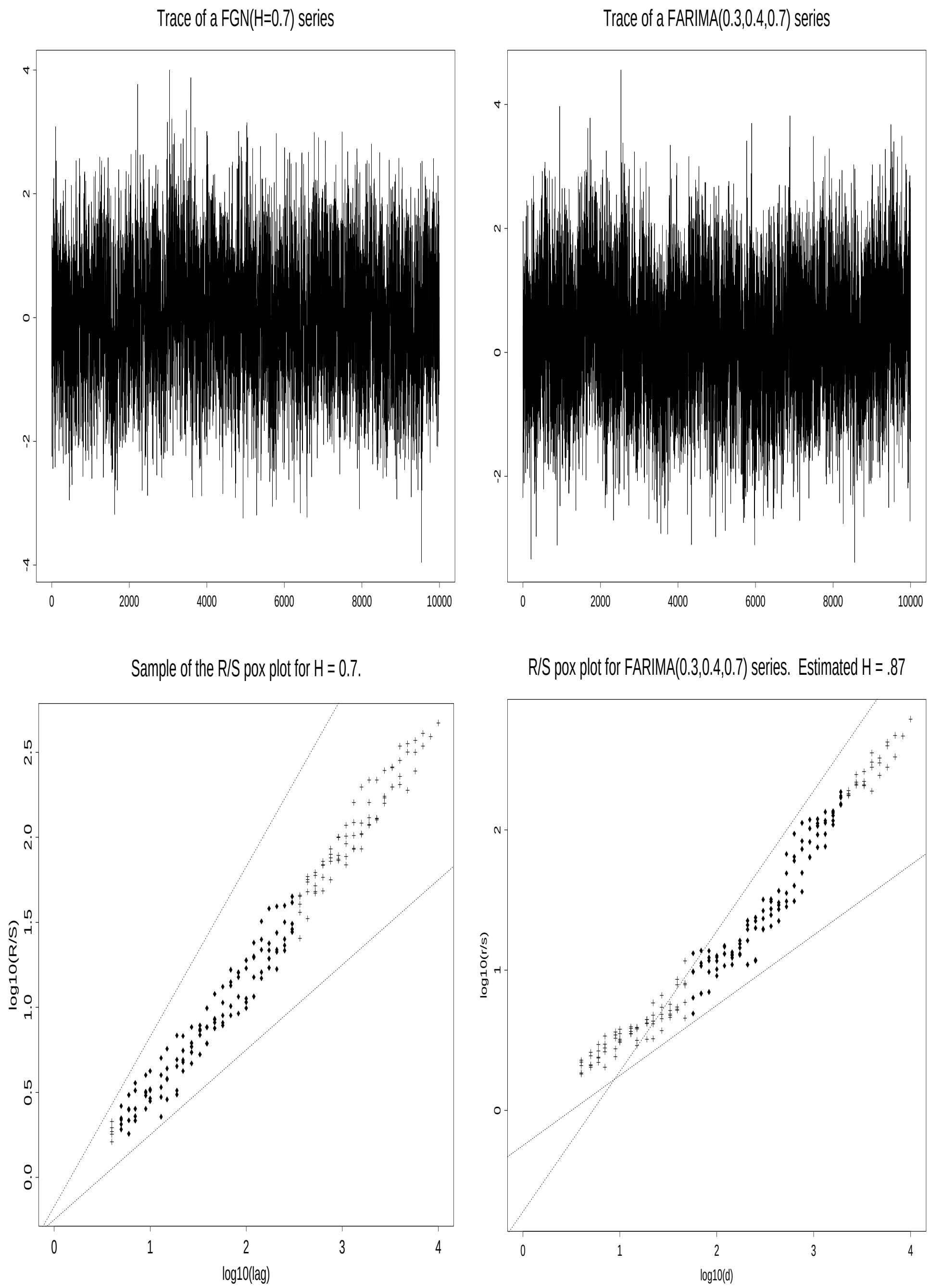

Figure 1: Sample traces and corresponding graphical R/S output for FGN with $H = 0.7$ (left) and FARIMA$(0.3, 0.4, 0.7)$ (right).The straight lines serve as reference: their slopes are 0.5 and 1.0 respectively.

Note that $S_0^2(N)$ is the sample variance $S^2(N)$ defined earlier in Section 1. Another way of representing $S_q(N)$ is:

$$S_q(N) = \sqrt{S^2 + 2\sum_{j=1}^{q} \omega_j(q)\hat{\gamma}_j}, \tag{2.2}$$

where $\hat{\gamma}_j$ are the sample autocovariances. As we will show in Section 3.2, the weights ω_j are chosen so that S_q^2 is the sample variance of an aggregated or averaged series.

We now compute Lo's modified R/S-statistic, $V_q(N)$, defined by

$$V_q(N) = N^{-1/2}R(N)/S_q(N). \tag{2.3}$$

If a series has no long-range dependence, Lo shows that given the right choice of q, the distribution of $V_q(N)$ is asymptotic to that of

$$W_1 = \max_{0\le t\le 1} W_0(t) - \min_{0\le t\le 1} W_0(t), \tag{2.4}$$

where W_0 is a standard Brownian bridge ($W_0(t) = B(t) - tB(1)$, where B denotes standard Brownian motion). Since the distribution of the random variable W_1 is known,

$$P[W_1 \le x] = 1 - 2\sum_{n=1}^{\infty}(4x^2n^2 - 1)e^{-2x^2n^2} \tag{2.5}$$

(Kennedy [14]), it follows that

$$P\{W_1 \in [.809, 1.862]\} = 0.95. \tag{2.6}$$

Lo uses the interval $[0.809, 1.862]$ as the 95% (asymptotic) acceptance region for testing the null hypothesis

$$H_0 = \{\text{no long-range dependence, i.e., } H = 0.5\}$$

against the composite alternative

$$H_1 = \{\text{there is long-range dependence, i.e., } 0.5 < H < 1\}.$$

The main innovation here is, of course, using $S_q(N)$ in the denominator instead of $S_0(N) = S$ as in the classical R/S-statistic. Lo's motivation for using the former statistic instead of S is as follows. If $q = 0$, the "range" of the time series, R, is normalized by the sample standard deviation, S. If the series is short-range dependent, e.g. if the process is actually an AR(1) process (i.e., no long-range dependence), then as $N \to \infty$, the statistic $V_0(N)$ will converge to a constant times W_1, where the constant depends on the particular short-range dependence structure imposed by the underlying AR(1) process. This can cause a problem if one uses $V_0(N)$ solely as a test statistic instead of relying on the full power of the graphical R/S-method. Because of the presence of this constant, the value of the statistic could easily fall outside the confidence interval (2.6), and then a false conclusion of long-range dependence would be drawn. On the other hand, given an appropriate choice of q, normalizing R by $S_q(N)$ will compensate for any "extra" short-range dependencies that may be present in the data, and $V_q(N)$ will then converge to W_1 for large N. If a time series has both short- and long-range dependence, Lo's statistic V_q should typically fall outside the confidence interval (2.6), implying correctly that long-range dependence is present. Notice that unlike the graphical R/S-method, which usually provides a rough estimate of the Hurst parameter H, Lo's method only indicates whether long-range dependence is present or not.

Lo's results are asymptotic, in that they assume $N \to \infty$ and $q = q(N) \to \infty$. In practice, the sample size N is finite, and the question naturally arises as to whether there is a "right" choice of $q(N)$. The choice of $q(N)$ influences both the actual size of the test, $P(\text{reject } H_0 \mid H_0)$, and its power, $P(\text{accept } H_0 \mid H_1)$. The right choice of q in Lo's method is essential, as will be seen below. The actual proof of Lo's result concerning the asymptotic distribution of his modified R/S-statistic $V_q(N)$ requires that $q(N) = o(N^{1/4})$. However, it is mentioned in a footnote that the results have actually been shown to hold if $q = o(N^{1/2})$ (see Andrews [2]). None of this, of course, provides a method for actually choosing q. Among the several suggestions offered in [15], a data-driven formula for choosing q given by

$$q_{opt} = \left[\left(\frac{3N}{2} \right)^{1/3} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2} \right)^{2/3} \right] \tag{2.7}$$

is especially appealing; here, N is the length of the data, $\hat{\rho}$ is the estimated first-order autocorrelation coefficient, and [ ] is the greatest integer function. Formula (2.7) is based on the assumption that the true model is an AR(1) process. Some simulation studies performed in [15] used series of length 100 to 1000 and q's ranging up to 50. For any of these series, the probability of accepting the null hypothesis varied significantly with q. In general, for the larger sample lengths, the larger the q, the less likely was the null hypothesis to be rejected. This is in accord with our results below. For example, for a long-range dependent process with $H = 0.833$ and $N = 1000$, Lo shows that the probability of (wrongly) accepting the null hypothesis is approximately 0.95 with $q = 50$, and even for q_{opt} which in this case is approximately 12, the probability of it is still 0.37, which is still much too high. In contrast, a careful estimation based on the graphical R/S-method consistently yields H-estimates that are within a reasonable range of the theoretical H-value and are clearly distinct from 0.5 for these types of series.

3 Experimental Results

In this section, we first introduce Fractional Gaussian noise (FGN) and Fractional ARIMA (FARIMA), the two types of time series that we use in our simulation studies to synthetically generate traces of "purely" long-range dependent observations and "hybrid" short-range/long-range dependent data. We then report on our results and experience with using Lo's modified R/S-method in the context of these "truly" long-range and/or short-range dependent data.

3.1 The Time Series

Fractional Gaussian noise $\{X_i,\ i \geq 1\}$ is a mean zero, stationary Gaussian time series whose autocovariance function $\gamma(h) = EX_iX_{i+h}$ is given by

$$\gamma(h) = \sigma^2 2^{-1}\{(h+1)^{2H} - 2h^{2H} + |h-1|^{2H}\}, \quad h \geq 0,$$

where $\sigma^2 = \text{Var } X_i$. An important point about γ is that it satisfies

$$\gamma(h) \sim \sigma^2 H(2H-1)h^{2H-2}, \quad \text{as } h \to \infty \tag{3.1}$$

when $H \neq 0.5$ ($\sim$ means "asymptotic to"). The X_i's are white noise when $H = 0.5$, since $\gamma(h) = 0$ for $h \geq 1$. 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.

Besides Fractional Gaussian noise, we also consider *Fractional ARIMA*(p, d, q). A series X_i is a FARIMA(p, d, q) series if it satisfies:

$$\Phi(B)X_i = \Theta(B)\Delta^{-d}\epsilon_i, \tag{3.2}$$

where the ϵ_i are independent, identically distributed normal random variables with mean 0 and variance 1, B is the backward operator, $BX_i = X_{i-1}$, and where Δ is the differencing operator $\Delta\epsilon_i = \epsilon_i - \epsilon_{i-1}$. Φ and Θ are polynomials in B of order p and q respectively. For example, $(1-\phi_1 B)X_i = (1-\theta_1 B)\Delta^{-d}\epsilon_i$ denotes a FARIMA$(1, d, 1)$ time series.

The autocovariance function of X satisfies for $0 < d < 0.5$,

$$\gamma(h) \sim \sigma^2 C h^{2d-1}, \quad \text{as} \quad h \to \infty, \tag{3.3}$$

where $\sigma^2 =$Var X_i and C depend on d, Φ, and Θ. Thus, asymptotically, the FARIMA(p, d, q) covariances behave in the same way as the FGN covariances, with $H = d + 0.5$.

We refer the reader to Samorodnitsky and Taqqu [22] and Brockwell and Davis [6] for more details on FGN and FARIMA, respectively. In the following, we will use synthetically generated FGN traces, as well as simulated FARIMA(p, d, q) series with $p, q \leq 1$.

3.2 Properties of the Modified R/S-statistic V_q

The theoretical results concerning Lo's modified R/S-statistic V_q suggest the following practical procedure for testing for long memory in a given data set: evaluate the test-statistic V_q for a wide range of different q-values and plot their values versus q, together with the 95% confidence interval given by equation (2.6). Lo's findings seem to indicate that there will be an initial region of small q-values where the values of the test-statistic will vary considerably, depending on the strength of the short-range dependence structure in the data; for q-values that are large enough to allow V_q to compensate for the "extra" short-range dependence that may be present in the data, the plot should show a period of stability, reflecting the fact that the only remaining effects are due to the presence (or absence) of long-range dependence in the data.

To see how well this procedure works in practice, we first applied it to synthetically generated FGN traces. Since FGN is the canonical example of a "purely" long-range dependent process with no "extra" short-range dependence structure, Lo's findings suggest that the values of V_q would stabilize for even small q's and that they should fall well outside the 95% confidence interval for $H > 0.5$ and safely within the 95% confidence region for H-values around 0.5. This, it turns out, is not the case; instead, for most of the FGN traces (each of length $N = 10,000$) that we used in our study with $H > 0.5$, V_q is strictly decreasing with q (see Figure 2). Moreover, when $\log V_q$ is plotted versus $\log q$, the result is an approximately straight line whose estimated slope is linear in H and approximately equal to $0.5 - H$ (see Figure 2). In the case of i.i.d. data (i.e., $H = 0.5$), V_q is approximately constant.

This empirically observed power decrease associated with Lo's modified R/S-statistic, V_q, is easily explained in terms of the behavior of the variance of the aggregated process $X^{(q)} = (X_i^{(q)} : i \geq 1)$,

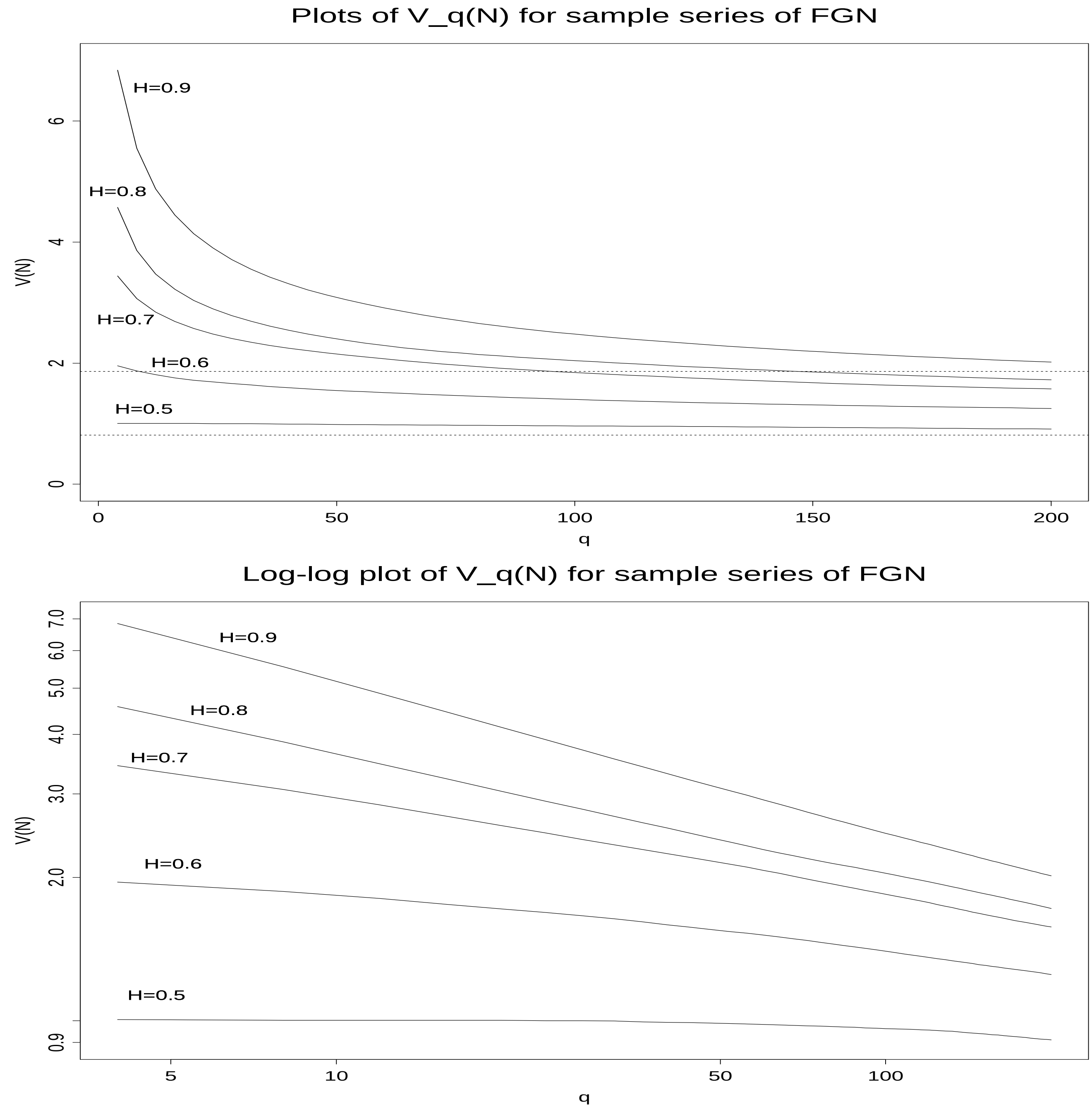

Figure 2: $V_q(N)$ for sample series of FGN.

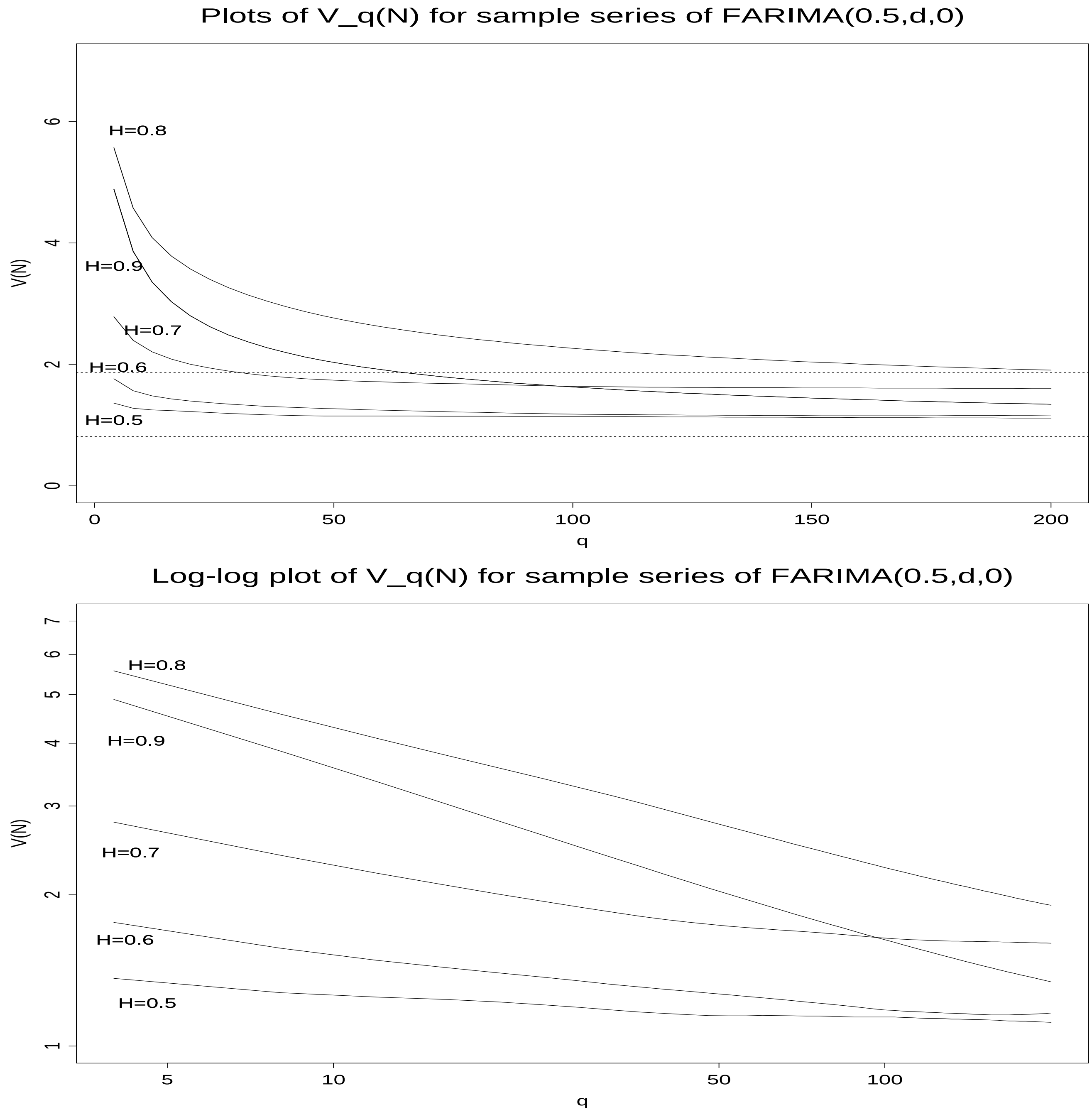

Figure 3: $V_q(N)$ for sample series of FARIMA$(0.5, d, 0)$.

where $X^{(q)}$ is defined by

$$X_k^{(q)} := \frac{1}{q} \sum_{i=(k-1)q+1}^{kq} X_i, \quad k \geq 1. \tag{3.4}$$

Now, if $\sigma^2 = \mathrm{Var} X_i$ and $\gamma_j = \mathrm{Cov}(X_i, X_{i+j})$, we have

$$\mathrm{Var} X^{(q+1)} = \frac{1}{q+1} \left[\sigma^2 + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{q+1} \right) \gamma_j \right].$$

In view of (2.1), $\widehat{\mathrm{Var}} X^{(q+1)} = (q+1)^{-1} S_{q+1}^2$ can be regarded a consistent estimator of Var $X^{(q+1)}$, and thus, for large q

$$S_q^2 \simeq q \widehat{\mathrm{Var}} X^{(q)}. \tag{3.5}$$

Next recall (e.g., see Taqqu, Teverovsky and Willinger [24]) that for a process with long-range dependence the variance of the aggregated process $X^{(q)}$ behaves asymptotically as q^{2H-2}. Thus $S_q^2(N) \sim q^{2H-1}$ for large enough q, and since $R(N)$ does not depend on q, we have

$$V_q(N) \simeq q^{1/2-H}. \tag{3.6}$$

Thus, as q increases, V_q decreases (for $H > 0.5$) and, for large enough q, it will be well within the confidence interval for the null hypothesis, i.e., $V_q \in [.809, 1.862]$ (see Equation (2.6)).

This finding points to a serious problem when relying on V_q as the sole indicator for whether or not a given data set is consistent with long-range dependence. Even though for small enough q, Lo's modified R/S-statistic can correctly distinguish between long- and short-range dependence, it stops doing so for larger q's. As Table 1 illustrates, if $N = 10,000$ and $q = 10$, i.e., $q/N = 0.001$, and the nominal H-value of the underlying FGN traces is 0.6, then the modified R/S-statistic V_q wrongly accepts the null hypothesis of no long-range dependence over 60% of the time. In the left part of Table 1 we show the results of using V_q at various values of q for simulated FGN series with different H-values ($H = 0.5(0.1)0.9$); each series is of length $N = 10,000$. For $H = 0.5$, we generated i.i.d. $N(0,1)$ random variables. For each H-value, we used 50 independent FGN realizations. The table gives the fraction of the time that the null hypothesis was accepted. Note that if larger q's are used then the null hypothesis is almost always wrongly accepted. In Table 2, we present the information from Table 1 in a more graphical form. A "+" sign indicates that the correct hypothesis (no long-range dependence) was accepted more than 90% of the time.

Next we apply the same procedure to synthetically generated traces of different FARIMA$(1, d, 0)$ processes; we choose two different values for the AR(1)-coefficient, $\phi = 0.5, 0.9$, and 5 different values for the long-range dependence parameter $d = 0(0.1)0.4$; i.e., $H = 0.5(0.1)0.9$). Again, each trace consists of $N = 10,000$ observations. Note that the presence of the AR(1) term in these models allows for the possibility to account for a wide range of "extra" short-range dependence structure; the case $\phi = 0.9$ is somewhat extreme and was included to illustrate what happens when the "extra" short-range dependence is very strong and when one cannot easily distinguish between short- and long-range dependence. In the tables, these traces are referred to as FARIMA$(0.5, d, 0)$ and FARIMA$(0.9, d, 0)$. As can be seen from Figure 3 and Tables 1 and 2, with the exception of the cases $d = 0$ (i.e., $H = 0.5$) and $q = 0$, the results obtained for the FARIMA traces are very similar to those obtained for FGN; for small q's, the null hypothesis of no long-range dependence is rejected for most of the sample series, even if $H = 0.5$. On the other hand, for large enough q, the hypothesis is accepted for almost all of

q	q/N	Nominal H														
		FGN					FARIMA$(0.5, d, 0)$					FARIMA$(0.9, d, 0)$				
		.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
0	.000	.98	.32	.00	.00	.00	.32	.02	.00	.00	.00	.00	.00	.00	.00	.00
10	.001	.98	.60	.10	.00	.00	.94	.46	.12	.00	.00	.58	.08	.01	.00	.00
20	.002	.98	.72	.24	.00	.00	1.00	.56	.26	.04	.02	.86	.38	.12	.00	.00
50	.005	.98	.90	.42	.14	.04	1.00	.80	.54	.14	.08	.98	.68	.32	.12	.02
100	.010	.96	.94	.68	.34	.18	1.00	.92	.74	.38	.24	.96	.78	.62	.42	.20
200	.020	.98	.94	.82	.76	.42	1.00	.92	.88	.74	.48	.96	.92	.80	.68	.50
500	.050	1.00	.98	.98	.98	.90	1.00	1.00	.96	.98	.92	.98	1.00	.96	.96	.94

Table 1: Fraction of $V_q(N)$ statistics which is inside the 95% confidence interval for the null hypothesis H_0 of no long-range dependence. 50 series of size $N = 10,000$ were used for each value of H and each type of series.

q	q/N	Nominal H														
		FGN					FARIMA$(0.5, d, 0)$					FARIMA$(0.9, d, 0)$				
		.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
0	.000	+	-	+	+	+	-	+	+	+	+	-	+	+	+	+
10	.001	+	-	+	+	+	+	-	-	+	+	-	+	+	+	+
20	.002	+	-	-	+	+	+	-	-	+	+	-	-	-	+	+
50	.005	+	-	-	-	+	+	-	-	-	+	+	-	-	-	+
100	.010	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-
200	.020	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-
500	.050	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-

Table 2: This table duplicates the previous one, in a slightly more graphical form. The + signs indicate that the correct hypothesis was accepted at least 90% of the time.

AR(0.5)	MA(0.5)	ARMA(0.3, 0.7)	ARMA(0.7, 0.3)	ARMA(−0.3, −0.7)
1.00	.98	.90	.96	.92
+	+	+	+	+

Table 3: Fraction of $V_q(N)$ statistics which is inside the 95% confidence interval for the null hypothesis of no long-range dependence using $q = q_{opt}$. 50 series of size $N = 10,000$ were used for each value of H and each type of short-range dependent ARMA series. A "+" sign indicates that the correct hypothesis was accepted more than 90% of the time.

the series, irrespective of the "true" nature of the dependence structure. In fact, in Table 2, there are no rows with all "+" 's. Moreover, the only time that the V_q-versus-q-plots level out is when V_q has decreased far enough to be within the 95% confidence interval corresponding to the null hypothesis. Thus, if one assumes that a "leveling out" or a region of "stability" of these plots is a sign that the statistics can be trusted, then the conclusion will almost always be to accept the null hypothesis of no long-range dependence, even when the series is "purely" long-range dependent, e.g., a FGN trace with a fairly high H. The only effect that the different d's (H's) have on this finding is in determining how large q has to be for the cross-over from the alternative to the null hypothesis to occur. Also, in almost all of the cases, the only time that V_q was fairly constant as q changed was when V_q was well within the acceptance region of the null hypothesis.

All indications so far strongly suggest that picking a single value of q to determine, on the basis of V_q, whether or not to reject or accept the null hypothesis of no long-range dependence in a given data set is highly problematic and often leads to erroneous conclusions. Let us also examine what happens if one chooses q on the basis of the data-driven formula (2.7). First note that this q_{opt} increases as the lag-1 autocorrelation coefficient increases. Thus, as d increases for a FARIMA$(1, d, 0)$ series, q_{opt} will in general also increase. Therefore, for large d (or H), where we should reject the null hypothesis almost all of the time, we will instead accept it fairly often, since the q_{opt} we use will be quite large. In Table 3, we show the results of using q_{opt} on several types of short-range dependent series: an AR(1) process with $\phi = 0.5$ (referred to as AR(0.5) in the table), a MA(1) process with $\theta = 0.5$, and a number of different ARMA(1, 1) with $\phi = 0.3, 0.7, -0.3$, $\theta = 0.7, 0.3, -0.7$, respectively. For each series we generated 50 independent realizations, each of length $N = 10,000$. The observed values of q_{opt} ranged from 17 to 30, which is fairly small compared to the values obtained for long-range dependent series (see below). Table 3 shows the fraction of the time that the null hypothesis is accepted, where the last line uses the same convention as in Table 2, that is, a "+" sign indicates that the correct hypothesis was accepted more than 90% of the time. As can be seen, the methodology works fairly well for these series that lack any long-range dependence and only exhibit "genuine" short-range dependence – the correct (null) hypothesis is almost always accepted.

In contrast, Table 4 shows the results of our empirical study when q is chosen according to the data-driven formula (2.7), and when the data sets are generated from some genuinely long-range dependent series. In particular, we use the FARIMA versions of the AR(0.5) and ARMA(0.7, 0.3) models considered in Table 3, that is, we consider FARIMA$(0.5, d, 0)$ and FARIMA$(0.7, d, 0.3)$ processes with $d = 0.1(.1)0.4$ (i.e., $H = 0.6(.1)0.9$). On top of the "extra" short-range dependence structure provided by the AR(0.5) and ARMA(0.7, 0.3) models, their corresponding FARIMA versions account for dif-

	FARIMA$(0.5, d, 0)$				FARIMA$(0.7, d, 0.3)$			
d	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
H	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
$\simeq q_{opt}$	40	50	65	100	40	50	70	120
Accept H_0	.74	.54	.20	.24	.54	.30	.20	.24
Power	-	-	-	-	-	-	-	-

Table 4: Fraction of $V_q(N)$ statistics which is inside the 95% confidence interval for the null hypothesis of no long-range dependence using $q = q_{opt}$. The order of magnitude of q_{opt} is indicated for each type of series. 50 series of size $N = 10,000$ were used for each value of H and each type of long-range dependent FARIMA series. A "-" sign indicates that the estimated power of the test is less than 90%.

ferent degrees of long memory via the d parameter. In addition to the fraction of time where the null hypothesis is accepted, Table 4 also gives the order of magnitude of q_{opt} for each type of series. Since q_{opt} is data-driven, its value, for a given model, varies from realization to realization. However, the observed variations are never very large and typically vary between ± 5 from the q_{opt}-values shown in Table 4. The "+" and "-" signs are as in Table 2. Note that in contrast to Table 3, in the FARIMA case, the modified R/S-statistic $V_{q_{opt}}$ has generally very low power; even for $d = 0.4$, which is equivalent to very strong long-range dependence, the null hypothesis is incorrectly accepted almost a quarter of the time! Things are even worse in the case of a FARIMA$(0.9, d, 0)$ process (not shown in Table 3), where q_{opt} is typically very large, about 600 for $d = 0.4$, and where the null hypothesis is accepted over 90% of the time, even for $d = 0.4$.

In summary, Lo's test based on the modified R/S-statistic V_q tends to be very conservative in rejecting the null hypothesis of no long-range dependence: When a series exhibits only short-range dependence, we have seen that the test performs very well and gives the correct results; however, when the series is, in fact, long-range dependent, the test still tends to accept the null hypothesis, this time incorrectly. Clearly, the choice of the truncation lag q is crucial and involves a delicate trade-off. On the one hand, to compensate for the "extra" short-range dependence that may be present in a given data set, q has to be relatively large and has to increase as the lag-1 autocorrelation coefficient increases. On the other hand, a higher degree of long-range dependence will also lead to larger values of q, resulting in an "over-compensation" as d increases which, in turn, tends to completely obscure even genuine long-range dependence. For example, d-values of about 0.4 give rise to q_{opt}-values of 100 and larger, large enough to often obscure any dependence whatsoever, not just the "extra" short-range dependence, but also the existing long memory. By contrast, a careful analysis of the types of time series considered above using the traditional graphical R/S-method consistently leads to a correct assessment of the presence or absence of long-range dependence in the data; moreover, if long-range dependence is present, it provides reasonably accurate estimates of the true Hurst parameter (usually within approximately 0.05 of the nominal H-value), even when the long-range dependence structure is "contaminated" by "extra" short-range dependence, i.e., for FARIMA models with non-zero AR and MA components. As an illustration, the right portion of Figure 1 shows the trace ($N = 10,000$) and corresponding pox plot of R/S of a FARIMA$(0.3, 0.4, 0.7)$ process, where the shaded region of the pox plot is used to obtain an H-estimate of about 0.87 for the true Hurst parameter $H = 0.9$.

From our empirical study, we conclude that while Lo's modified R/S-statistic represents a theoretical improvement over the classical rescaled adjusted range statistic, its practical application requires care and has a number of problematic features. The most important is the strong dependence between the outcome of the test (based on the test-statistic V_q) and the choice of the truncation lag q, with a definite bias toward accepting the null hypothesis of no long-range dependence for large q's. This happens even in ideal scenarios of "purely" long-range dependent data with high values of the Hurst parameter. Our recommendations are therefore to always rely on a wide range of different q-values and associated V_q-values and to use plots such as the ones shown in Figures 2 and 3 when using Lo's modified R/S-statistic to test for long-range dependence in a given data set. Moreover, one ought never to use Lo's method in isolation, but always in conjunction with other graphical and statistical techniques for checking for long memory, especially when Lo's method results in accepting the null hypothesis of no long-range dependence.

4 Conclusion

The idea of modifying the classical R/S-statistic by using $S_q(N)$ instead of S to compensate for the presence of any "extra" short-range dependence that may be present in a given data set is attractive. Unfortunately, in practice, it seems difficult to determine what value of q or range of q-values to choose as truncation lag(s): too small a q is not likely to account for all the extra short-range effects, and too large a q is bound to over compensate and destroy any long-range effect that might be in the data. Waiting for the 'stability" of the statistic with respect to q does not seem to work, as it only becomes stable within the acceptance region for the null hypothesis.

Any choice of "q_{opt}", the optimal value of q, whether data-driven (as in formula (2.7)) or not, will have to depend on the underlying model for the time series, which is not known *a priori* for actual data. In addition, any reasonable choice of q_{opt} will require a delicate compromise: On the one hand, q_{opt} ought to be large when the small-lag autocorrelations are large – so as to compensate for the strong short-range dependence; on the other hand, for "truly" long-range dependent series, a large Hurst parameter will tend to yield a large q_{opt}, thus causing automatically the acceptance of the null hypothesis.

Based on our findings described in Section 3 concerning the properties of Lo's modified R/S-statistic, we strongly advise against its use as the sole technique for testing for long memory in a given data set and advocate instead the use of a diverse portfolio of time domain-based and frequency domain-based graphical and statistical methods described in Taqqu, Teverovsky and Willinger [24] and Taqqu and Teverovsky [23]. These include the graphical R/S-method, the modified R/S-statistic with corresponding V_q-versus-q-plots, and Whittle's approach. Although Lo's modified R/S-statistic is a conceptual improvement over the classical R/S-statistic, it should not be used blindly nor in isolation. In particular, an acceptance of the null hypothesis of no long-range dependence based on the modified R/S-statistic should never be viewed as the "final word", mainly because of the serious difficulties that the test-statistic V_q has in identifying "genuine" long-range dependence (e.g., FGN with high H-values). Instead, an acceptance of the null based on the test-statistic V_q should always be accompanied and supported by further analysis of the data. For an illustration of some practical implications of our findings, and for examples of additional data analytic techniques that might be appropriate when investigating the long-range dependence phenomenon in historical records, we refer to Willinger, Taqqu and Teverovsky [26], who report on an in-depth analysis of the Center for Research in Security Prices (CRSP) daily stock return data.

References

[1] P. Abry and D. Veitch. Wavelet analysis of long range dependent traffic. To appear in *IEEE Transactions on Information Theory*, 1997.

[2] D. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

[3] A. A. Annis and E. H. Lloyd. The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika*, 63:111–116, 1976.

[4] J. B. Bassingthwaighte and G. M. Raymond. Evaluating rescaled rangle analysis for time series. *Annals of Biomedical Engineering*, 22:432–444, 1994.

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

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

[7] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.

[8] N. J. Cutland, P. E. Kopp, and W. Willinger. Stock price returns and the Joseph effect: a fractional version of the Black-Scholes model. In E. Bolthausen, M. Dozzi, and F. Russo, editors, *Seminar on Stochastic Analysis, Random Fields and Applications*, pages 327–351. Birkhäuser, Boston, 1995.

[9] W. Feller. The asymptotic distributions of the range of sums of independent random variables. *Annals of Mathematical Statistics*, 22:427–432, 1951.

[10] M. T. Greene and B. D. Fielitz. Long-term dependence in common stock returns. *Journal of Financial Economics*, 4:339–349, 1977.

[11] M. A. Hauser, R. M. Kunst, and E. Reschenhofer. Modelling exchange rates: Long-run dependence versus conditional heterscedasticity. *Applied Financial Economics*, 4:233–239, 1994.

[12] B. N. Huang and C. W. Yang. The fractal structure in multinational stock returns. *Applied Economics Letters*, 2:67–71, 1995.

[13] H. E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineershh*, 116:770–808, 1951.

[14] D. Kennedy. The distribution of the maximum Brownian excursion. *Journal of Applied Probability*, 13:371–376, 1976.

[15] A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59:1279–1313, 1991.

[16] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman and Co., New York, 1983.

[17] 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.

[18] B. B. Mandelbrot and J. R. Wallis. Noah, Joseph and operational hydrology. *Water Resources Research*, 4:909–918, 1968.

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

[20] B. B. Mandelbrot and J. R. Wallis. Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence. *Water Resources Research*, 5:967–988, 1969.

[21] B. B. Mandelbrot and J. R. Wallis. Some long-run properties of geophysical records. *Water Resources Research*, 5:321–340, 1969.

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

[23] M. S. Taqqu and V. Teverovsky. On estimating the intensity of long-range dependence in finite and infinite variance series. In R. Adler, R. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 177–217, Boston, 1998. Birkhäuser.

[24] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(4):785–798, 1995. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

[25] W. Willinger, M. S. Taqqu, and A. Erramilli. A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks. In F. P. Kelly, S. Zachary, and I. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 339–366. Clarendon Press, Oxford, 1996.

[26] W. Willinger, M. S. Taqqu, and V. Teverovsky. Stock market prices and long-range dependence. Preprint. To appear in 1999 in *Finance and Stochastics*, 1997.

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

Walter Willinger, AT&T Labs-Research, Florham Park, NJ 07932-0971, USA. *Email:* walter@research.at.com