Preprint 1996. Appears on pages 177-217 in the book:
"A Practical Guide To Heavy Tails: Statistical Techniques and Applications"
Robert Adler, Raisa Feldman and Murad S. Taqqu, editors. Birkhauser, Boston, 1998

On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series

Murad S. Taqqu and Vadim Teverovsky [1] [2] [3]

Abstract

The goal of this paper is to provide benchmarks to the practitioner for measuring the intensity of long-range dependence in time series. It provides a detailed comparison of eight estimators for long-range dependence, using simulated FARIMA(p, d, q) time series with different finite and infinite variance innovations. FARIMA time series model both long-range dependence (through the parameter d) and short-range dependence (through the parameters p and q). We evaluate the biases and standard deviations of several estimators of d and compare them for each type of series used. We consider Gaussian, exponential, lognormal, Pareto, symmetric and skewed stable innovations. Detailed tables and graphs have been included. We find that the estimators tend to perform less well when p and q are not zero, that is, when there is additional short-range dependence structure. For most of the estimators, however, the use of infinite variance instead of finite variance innovations does not cause a great decline in performance.

1. Introduction

Long-range dependence has been observed in many time series, e.g. network traffic and finance (see [LTWW94], [Bai96], [GD] and also the special issue [JoE96]). Such dependence is also known as "long memory" or "$1/f$ noise". It is characterized, in the finite variance case, by slowly decaying covariances, a spectral density that tends to infinity as the frequencies tend to zero and by the self-similarity of aggregated summands. The intensity of these phenomena is measured by a parameter d. Various estimators of d have been

[1] Research partially supported by NSF grants NCR-9404931 and DMS-9404093 at Boston University.

[2] *AMS Subject classification:* 60G18, 62-09, 60E07.

[3] *Keywords:* Long-range dependence, long memory, robustness, infinite variance, time series, periodogram, Whittle.

proposed. Even though some have known large sample properties, it is important to test their accuracy by using simulated series. Such a study was started in Taqqu, Teverovsky and Willinger [TTW95], using "ideal" models that display long-range dependence, namely Fractional Gaussian Noise (FGN) or Gaussian Fractional ARIMA$(0, d, 0)$. Both of these models are Gaussian and have a well-controlled short-range dependence structure. Since real-world data possess a more complex structure, the *robustness* of the estimators must also be examined. In this study we will analyze the performance of various estimators of long-range dependence when there is additional short-range dependence structure and/or infinite variance.

The FARIMA(p, d, q) family of models is widely used in the modeling of time series with long-range dependence (see, for example, Brockwell and Davis [BD91]). These are moving averages $X_n = \sum_{i=-\infty}^{n} c_{n-i}\epsilon_i$, where c_k behaves like k^{d-1} for large k and the ϵ_i's are independent, identically distributed random variables (innovations). Here we will focus specifically on FARIMA$(1, d, 1)$ series, with both finite and infinite variance innovations.

The estimation methods we consider in this paper are : 1) The Absolute Value Method 2) Regression on the Periodogram 3) R/S 4) Aggregated Variance 5) Variance of Residuals 6) Whittle Estimator 7) Aggregated Whittle 8) Local Whittle. (In addition we will briefly discuss three other estimators without presenting detailed results. These are: Higuchi's Fractal Dimension Method, Differenced Variance, Modified Periodogram.) These estimators are also described in Taqqu, Teverovsky, Willinger [TTW95]. All are graphical except the Whittle-based ones. The latter are related to Whittle's approximate MLE and discussed at greater length in Taqqu and Teverovsky [TT95a] and Taqqu and Teverovsky [TT96]. We do not use exact (Gaussian) MLE methods because these strongly depend on the series being Gaussian, which is often not the case in practice. The exact MLE is, moreover, much more computationally intensive than the approximate MLE (Whittle), which is, itself, much more computationally intensive than the graphical methods.

The innovations that appear in the FARIMA$(1, d, 1)$ series have either finite or infinite variance. We will use Gaussian, exponential and lognormal innovations in the finite variance case and Pareto and stable innovations in the infinite variance case. Both Pareto and stable distributions are characterized by a parameter α which describes the heaviness of the distribution tails. If ϵ is either Pareto or stable with parameter α, then $P(\epsilon > x) \sim Cx^{-\alpha}$ as $x \to \infty$, that is, the probability tails decrease slowly, like a power function. Moreover, $\mathrm{Var}(\epsilon) = \infty$ if $\alpha < 2$ and $E|\epsilon| = \infty$ if $0 < \alpha \leq 1$.

The inclusion of infinite variance innovations forces us to carefully differentiate between the two parameters d and H that are used to characterize long-range dependence. The parameter d, which plays the role of a differencing parameter in the FARIMA model, is described in Section 2. The parameter H is a scaling parameter, that can be characterized as follows: given a time series

X_i, of length N, define the corresponding aggregated series,

$$X^{(m)}(k) := \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k = 1, 2, ..., [N/m]. \tag{1.1}$$

Then, for large enough N/m and m,

$$X^{(m)} \overset{d}{\sim} m^{H-1} S \tag{1.2}$$

(where $\overset{d}{\sim}$ means "asymptotically equal to" in the sense of distributions). Here S is a process which depends on the distribution of X but does not depend on m (the term *asymptotically self-similar* is used to describe property (1.2)). In the finite variance case S is Fractional Gaussian Noise (FGN) and in the infinite variance case it is Linear Fractional Stable Noise (LFSN). These processes are exactly self-similar in the sense that Relation (1.2) holds for all $m \geq 1$ and not just asymptotically. FGN and LFSN are introduced in the next section.

The scaling relation (1.2) defines the parameter H. For finite variance processes,

$$H = d + 1/2,$$

and for infinite variance processes,

$$H = d + 1/\alpha.$$

It is important therefore to know whether an estimator is estimating H or d.

The paper is structured as follows. The time series are introduced in Section 2 and the experimental set-up is described in Section 3. The estimators are defined in Section 4 and the results presented in Section 5. Several tables and graphs are included for reference. Our conclusions are stated in Section 6.

2. The Time Series

In this section we present the various types of series which are used in this paper. These include fractional Gaussian noise and FARIMA with innovations that are either Gaussian, non-Gaussian with finite variance or that have infinite variance, including stable innovations.

2.1 Fractional Gaussian Noise (FGN)

We start with the FGN series, which serves as a reference point for the FARIMA series introduced below. The importance of FGN stems from its exact self-similarity. If $X = \{X_i, i \geq 1\}$ is FGN and if $X^{(m)}$ is defined as in (1.1), then $X^{(m)} \overset{d}{=} m^{H-1} X$ ($\overset{d}{=}$ denotes equality in distribution). This relation is much stronger than the asymptotic relation (1.2).

The FGN series $\{X_i,\ i \geq 1\}$ is a zero mean, stationary, Gaussian time series whose autocovariance function at lag h is:

$$\gamma(h) = EX_iX_{i+h} = \frac{1}{2}\{(h+1)^{2d+1} - 2h^{2d+1} + |h-1|^{2d+1}\},\ h \geq 0,\ -1/2 < d < 1/2. \tag{2.3}$$

For $d \neq 0$, the autocovariance satisfies

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

($\sim$ means here "asymptotic to"), that is, for large lags, it decreases to zero like a power function. d measures the intensity of the long-range dependence of the series X and it is related to the scaling exponent H by the relation $H = d + 1/2$. In addition, the spectral density (Fourier transform of γ) is given by:

$$f(\nu) = C_d\Big(2\sin\frac{\nu}{2}\Big)^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\nu + 2\pi k|^{2d+2}} \sim C_d|\nu|^{-2d} \text{ as } \nu \to 0, \tag{2.5}$$

where C_d is a constant. The spectral density behaves like a power function at the origin. Long-range dependence corresponds to $d > 0$ ($H > 1/2$), because in this case, the series spectral density diverges to infinity at the origin. For graphs of the autocorrelation function and the spectral density for FGN with $d = 0.2$ ($H = 0.7$) see Figure 1.

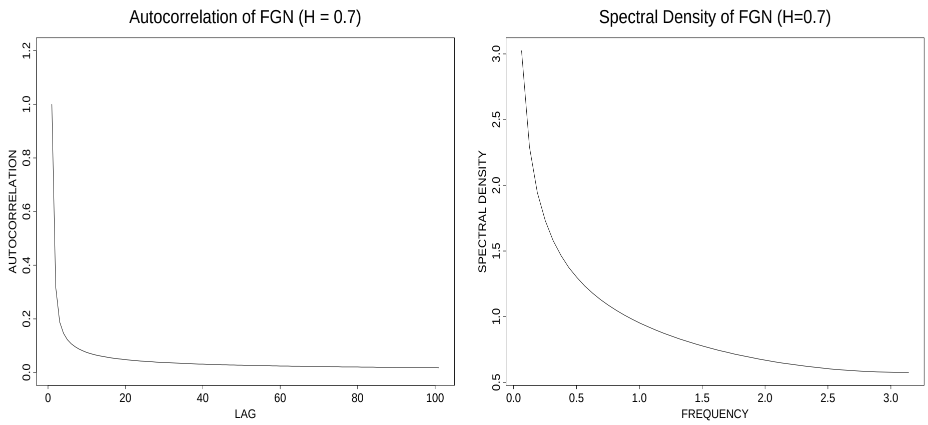


Figure 1: Autocorrelation function and spectral density of FGN with $d = 0.2$ ($H = 0.7$).

2.2 Gaussian FARIMA

Let us now turn to the Gaussian FARIMA(p, d, q) processes. A Gaussian FARIMA$(0, d, 0)$ process is defined by $X_i = \Delta^{-d}\epsilon_i, \quad i \geq 1$, where the ϵ_i are independent, identically distributed Gaussian random variables with mean 0, and where Δ is the differencing operator $\Delta\epsilon_i = \epsilon_i - \epsilon_{i-1}$. For fractional d we interpret Δ^{-d} by using formal power series expansion, as follows:

$$\Delta^{-d} = \sum_{i=0}^{\infty} b_i(-d)B^i, \tag{2.6}$$

where B is the backward operator, $B\epsilon_i = \epsilon_{i-1}$, and

$$b_i(-d) = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}, \quad i = 1, 2, \ldots .$$

Γ denotes the gamma function. The autocovariance function of this process satisfies, for $-1/2 < d < 1/2$,

$$\gamma(h) \sim C_d h^{2d-1} \quad \text{as} \quad h \to \infty \tag{2.7}$$

where $C_d = \pi^{-1}\Gamma(1-2d)\sin \pi d$. Thus, for large lags d, the autocovariance (2.7) of a Gaussian FARIMA$(0, d, 0)$ process has the same power decay as the autocovariance (2.4) of FGN. Correspondingly, the spectral density of a Gaussian FARIMA$(0, d, 0)$ process behaves in the same way as that of a FGN process, i.e. it is like a power function at the origin.

A Gaussian FARIMA$(0, d, 0)$ is a particular case of Gaussian FARIMA(p, d, q) processes. These are defined through the equation

$$X_i - \phi_1 X_{i-1} - \cdots - \phi_p X_{i-p} = \Delta^{-d}\epsilon_i - \theta_1 \Delta^{-d}\epsilon_{i-1} - \cdots - \theta_q \Delta^{-d}\epsilon_{i-q}, \tag{2.8}$$

where the $\Delta^{-d}\epsilon_i$ are as above and the ϕ_i and the θ_i are the autoregressive and the moving average coefficients respectively. The Gaussian FARIMA(p, d, q) series follow the same asymptotic relations for their autocovariance and spectral density as do the Gaussian FARIMA$(0, d, 0)$ processes, although the actual functions are more complicated, and have additional short-term components. These short-term components can influence significantly the estimates of d. While some estimators are relatively insensitive to deviations from "ideal" long-range dependence, others are extremely sensitive to them, and thus lack robustness. The goal of this study is to determine which estimators are robust and which ones are not.

2.3 FARIMA with other finite variance innovations

In addition to Gaussian innovations ϵ_i, we also consider innovations ϵ_i which have *exponential* and *lognormal* distributions. An exponential random variable ϵ has the cumulative distribution function,

$$F(x) = P(\epsilon \le x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\lambda x} & x \ge 0. \end{cases} \tag{2.9}$$

We use $\lambda = 1$ in our series. A lognormal random variable ϵ has the cumulative distribution function (c.d.f.),

$$F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right), \tag{2.10}$$

where Φ is the c.d.f. of a standard Normal random variable. Here we use $\mu = 0$, $\sigma = 1$. We only analyze FARIMA$(0, d, 0)$ series for these types of innovations.

2.4 FARIMA with infinite variance innovations

For infinite variance innovations ϵ, we consider symmetric and skewed stable distributions and Pareto distributions. They are characterized by the parameter α and satisfy

$$P(\epsilon > x) \sim Cx^{-\alpha}, \quad \text{as } x \to \infty, \tag{2.11}$$

implying that the variance is infinite if $\alpha < 2$. The *symmetric α-stable* distribution, denoted $S_\alpha(\sigma)$ has the characteristic function,

$$E[\exp(i\theta\epsilon)] = \exp(-\sigma^\alpha|\theta|^\alpha).$$

where σ is the scale parameter. We will concentrate on the case $1 < \alpha \leq 2$, with the understanding that for $\alpha = 2$ the process is Gaussian. A stable random variable which is totally skewed to the right has the characteristic function,

$$E[\exp(i\theta\epsilon)] = \exp\left[-\sigma^\alpha|\theta|^\alpha(1 - i(\text{sign } \theta)\tan\frac{\pi\alpha}{2})\right],$$

where $i = \sqrt{-1}$, and sign $\theta = -1, 1, 0$ if $\theta < 0, > 0, = 0$ respectively. We use $\sigma = 1$ in all of the simulated series. Two of the important properties of stable distributions are: (1) If ϵ is α-stable with scale parameter σ then $a\epsilon$ is also α-stable with scale parameter $|a|\sigma$. (2) If ϵ_1 and ϵ_2 are independent α-stable with σ_1 and σ_2 respectively, then $\epsilon_1 + \epsilon_2$ is α-stable with parameter $(\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}$.

Finally, a Pareto random variable ϵ has the c.d.f.,

$$F(x) = 1 - \frac{1}{x^\alpha}, \quad x \geq 1. \tag{2.12}$$

Pareto random variables also have infinite variance, and in fact satisfy (2.11) exactly for all $x \geq 1$.

When generating series with infinite variance, we will use independent symmetric α-stable variables as innovations in FARIMA(p, d, q) series (see (2.8)), and skewed stable and *Pareto* distributions as the innovations in a FARIMA$(0, d, 0)$ series. The parameter d is restricted to the interval $[0, 1-1/\alpha)$. The resulting process X will be long-range dependent and α-stable if the innovations are α-stable, and asymptotically α-stable if the innovations are Pareto.

2.5 Linear Fractional Stable Noise (LFSN)

FGN is the only exactly self-similar Gaussian process, i.e. such that

$$X^{(m)} \stackrel{d}{=} m^{H-1}X. \tag{2.13}$$

If X is α-stable with $\alpha < 2$, then there are many different processes satisfying (2.13), and LFSN is one of them. It is usually defined through an integral representation, but, since this representation is not used here, we refer the reader to Chapter 7 of [ST94]. The only property of LFSN used in this paper is Relation (2.13) (exact self-similarity).

For more information on Fractional Gaussian Noise or Gaussian FARIMA series, see Brockwell and Davis [BD91] or Beran [Ber94] and, for stable processes, Samorodnitsky and Taqqu [ST94]. The FARIMA is a useful family of models because any short-range dependence can be modeled by choosing the appropriate autoregressive and moving average coefficients in (2.8). Unfortunately these coefficients can also affect the estimates of long-range dependence. These effects may be considerable, even for the series of length $10,000$ used in this study.

3. Set-Up

In [TTW95] we analyzed

a) Fractional Gaussian Noise,
b) Gaussian FARIMA$(0, d, 0)$,

and the results are summarized in Table 1 and the top row of Figure 5. These series provide a baseline for the present research. Here we focus on situations where there is additional short-range dependence and/or finite or infinite variance innovations. The ten types of series examined are:

1) Gaussian FARIMA$(1, d, 0)$, with $\phi_1 = 0.5$.
2) Gaussian FARIMA$(0, d, 1)$ with $\theta_1 = 0.5$.
3) Gaussian FARIMA$(1, d, 1)$ with $\phi_1 = -0.3,\ \theta_1 = -0.7$.
4) Gaussian FARIMA$(1, d, 1)$ with $\phi_1 = 0.3,\ \theta_1 = 0.7$.
5) FARIMA$(0, d, 0)$ with exponential innovations.
6) FARIMA$(0, d, 0)$ with lognormal innovations.
7) Symmetric Stable FARIMA$(0, d, 0)$ with $\alpha = 1.5,\ 1.2$.
8) Symmetric Stable FARIMA$(1, d, 1)$ with $\alpha = 1.5,\ 1.2,\ \phi_1 = 0.3,\ \theta_1 = 0.7$.
9) FARIMA$(0, d, 0)$ with Pareto innovations, $\alpha = 1.5$.
10) FARIMA$(0, d, 0)$ with skewed stable innovations, $\alpha = 1.5$.

The parameters ϕ_1 and θ_1 generate short-range dependence. Extreme values of these parameters (close to 1) can create very large deviations from pure long-range dependence. Our choices of values for ϕ_1 and θ_1 are a compromise between showing what can go wrong and totally obscuring the long-range dependence (for more information on the role of ϕ_1 and θ_1 see [TT96]).

For each type of series, we take d in the range $[0, 1 - 1/\alpha)$. This results in an $H < 1$, which is necessary for X to be a stationary series. Specifically,

a) for Gaussian series we chose $d = 0.0(0.1)0.4$,
b) for $\alpha = 1.5$ stable and Pareto, we chose $d = 0.0(0.1)0.3$,
c) for $\alpha = 1.2$ stable, we chose $d = 0.0, 0.1$.

The Gaussian FARIMA 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), except for the case $d = 0$ ($H = 0.5$) when there is no long-range dependence. The method of generation is based on the Durbin-Levinson algorithm (see for example Brockwell and Davis [BD91], Chapter 8.2) and is described in

Haslett and Raftery [HR89], p. 12-13. For $d = 0$, we wrote our own routine to generate the processes. The $d = 0$ case is very simple because no fractional differencing is necessary.

The stable innovation series were generated using the *rstab* command in S-Plus (in the skewed case this was modified to produce non-negative random variables). The stable FARIMA(p, d, q) processes were then produced by using Eqs. (2.6) and (2.8). The exponential and lognormal distributions were produced by the S-Plus commands *rexp* and *rlnorm* respectively. The Pareto innovations were produced by using the formula $U^{-1/\alpha}$ where U is a Uniform$(0, 1)$ random variable. The FARIMA$(0, d, 0)$ series were then produced by using Eq. (2.6). The Durbin-Levinson algorithm can not be used for infinite variance processes since it depends heavily on the existence of the covariance function. We used instead the infinite sums in (2.6), truncated at $i = 10,000$.

For each type of finite variance series and each value of d, 50 series of length $10,000$ were simulated. Because of the potential for greater variability, 1000 independent series of length $10,000$ were produced for the infinite variance series. We used the estimation methods described in Section 4 below to estimate d or H, tabulated the average biases of the estimates as well as the in-sample standard deviations (standard errors) and the square roots of the sample MSEs. The following notation is used in the tables. If d_i is the estimate for sample i out of I total samples, then,

$$\overline{d} = \frac{1}{I}\sum_{i=1}^{I} d_i, \quad BIAS = d_0 - \hat{d}, \quad \hat{\sigma}^2 = \frac{1}{I-1}\sum_{i=1}^{I}(d_i - \overline{d})^2, \quad \mathrm{MSE} = \frac{1}{I}\sum_{i=1}^{I}(d_i - d_0)^2,$$

where d_0 is the nominal value of d. When H is estimated, we computed all of the statistics on the basis of $H_0 = d_0 + 1/\alpha$, and proceeded accordingly.

The results are presented in Tables 1 - 4 for the finite variance series and in Tables 6 - 8 for the infinite variance series. Detailed boxplots, for the finite variance series, are given in Figures 7, 8. In these boxplots, the vertical axis shows the deviations from the nominal values of d or H. Each boxplot has the following components: (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 [MTL78]); 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 fall beyond the whiskers.

The greater the number of realizations, the more outliers are present. In the infinite variance case, we generated 1000 realizations for each type of series. The boxplots, in this case, are displayed **without** the outliers, because otherwise they would be too crowded. To give an idea of the variability, we present histograms (Figure 9) and detailed boxplots (Figure 10) for the Local Whittle Estimator applied to FARIMA$(0, d, 0)$ series with the following innovations: Gaussian, stable with $\alpha = 1.5$ and stable with $\alpha = 1.2$.

In Table 5 we show how this estimator behaves for these time series as the number of independent realizations grows from 50 to 1000. The table shows that the mean and standard deviation of the estimates do not change significantly as the number of realizations grows. However, with infinite variance innovations, there are instances of very extreme outliers in the distribution of the estimates. Thus, when $\alpha = 1.2$, $d = 0.1$, there was one extreme instance out of 1000 when the estimate was biased by 0.33. (We note that this particular realization caused problems for other periodogram based estimators as well. Both the Whittle method and the periodogram method had biases of at least 0.3, much larger than for any other realization.) To see the effect such an outlier can have on the mean and standard deviation, we divided the 1000 estimates obtained for each type of series into 20 blocks of 50 estimates each. When $\alpha = 1.2$, we found that $\hat{\sigma}$ was approximately 0.02 for most of the 20 blocks, but it was .059 for the block where the outlier was included. The mean was only slightly affected. In the Gaussian case, $\hat{\sigma}$ did not vary by more than 0.005 from its average value of 0.030 over the 20 blocks. Therefore, whereas 50 realizations is enough in the finite variance case, more are needed in the infinite variance case. Our studies suggest using I of at least 300 or 400 if infinite variance innovations are used.

4. Estimators

The various estimators used here are described briefly below (see also [TT95a], [TTW95]). A number of them are obtained through graphical methods. They are illustrated in Figures 2 - 6. In most of the figures, a line indicating the best fit to the data, whose slope provides the estimates of d or H has been included. The choice of cut-offs is somewhat arbitrary and reflects our experience. Slight changes in the values of the cut-offs tend not to change significantly the estimates. The figures include also a second line whose slope is the correct "nominal" slope (the intercept of this second line was chosen so as not to interfere with the data points themselves).

4.1 Absolute Value Method

Consider the aggregated series (1.1), obtained by dividing a given series of length N into blocks of length m, and averaging the series over each block. We take the first absolute moment of this series,

$$AM^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} \left| X^{(m)}(k) - \overline{X} \right|, \tag{4.14}$$

where $X^{(m)}$ is defined in (1.1) and $\overline{X}$ is the overall series mean. The centered aggregated series $X^{(m)} - EX^{(m)}$ is asymptotically equal to $m^{H-1}S_{\alpha,d}$ where $S_{\alpha,d}$ is an α-stable random sequence (in fact, it is LFSN). In particular, if X has finite variance, we get $S_{2,d}$ which is FGN.

Thus, $AM^{(m)}$ behaves like m^{H-1} for large m. Since $H = d + 1/2$ for finite variance series, $m^{H-1} = m^{d-1/2}$. For infinite variance series, $H = d + 1/\alpha$ (see Samorodnitsky and Taqqu [ST94]) and hence $m^{H-1} = m^{d+1/\alpha-1}$. This estimation method gives H and not d.

For successive values of m, the sample absolute first moment of the aggregated series is plotted versus m on a log-log plot. The result should be a straight line with a slope of $H-1$. In practice, the slope is estimated by fitting a least-squares line to the points of the plot. We assume that both N and N/m are large. This ensures that both the length of each block and the number of blocks is large. If the series has no long-range dependence and finite variance, then $H = 0.5$ and the slope of the fitted line should be $-1/2$. In practice, the points at the very low and high ends of the plots are not used to fit the least-squares line. Indeed, short-range effects can distort the estimates of H if the low end of the plot is used, and at the very high end of the plot there are too few blocks to get reliable estimates of AM. In this study, the cut-offs are $10^{0.7}$ and $10^{2.5}$.

A technique suggested by Higuchi [Hig88] is very similar to the Absolute Value method. It uses,

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^{m} \left[\frac{N-1}{m}\right]^{-1} \sum_{k=1}^{[(N-i)/m]} \left| \sum_{j=i+(k-1)m+1}^{i+km} X(j) \right|. \tag{4.15}$$

Here, N is the length of the time series, m is the equivalent of a block size, and [ ] denotes the greatest integer function. For long-range dependent series, $EL(m) \sim Cm^{H-2}$. Essentially, the difference with the absolute value method lies in using a sliding window to compute the aggregated series, instead of using non-intersecting blocks. Because of this, the method is much more computationally intensive. While this modification may result in increased accuracy in shorter time series, there seems to be no advantage in using it for the series length of $10,000$. Hence, the results of this method are not presented here.

4.2 Variance Method

We again consider the aggregated series $X^{(m)}$ defined in (1.1), but instead of the first absolute moment, we compute its sample variance,

$$\widehat{\mathrm{Var}} X^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k) - \overline{X})^2. \tag{4.16}$$

As stated earlier, the series $X^{(m)} - EX^{(m)}$ scales like m^{H-1}, thus if the series is Gaussian, or at least finite variance, the sample variance will be asymptotically proportional to $m^{2H-2} = m^{2d-1}$ for large N/m and m.

Deriving the asymptotic behavior of $\widehat{\mathrm{Var}} X^{(m)}$ is more complicated when the series has infinite variance. Here is a sketch. In view of (1.2),

$$\begin{aligned} \widehat{\mathrm{Var}} X^{(m)} &\overset{d}{\sim} \widehat{\mathrm{Var}}(m^{H-1} S_{\alpha,d}) \qquad (4.17)\\ &= m^{2H-2} \left[\frac{1}{N/m} \sum_{k=1}^{N/m} S_{\alpha,d}^2(k) - \left(\frac{1}{N/m} \sum_{k=1}^{N/m} S_{\alpha,d}(k) \right)^2 \right] \end{aligned}$$

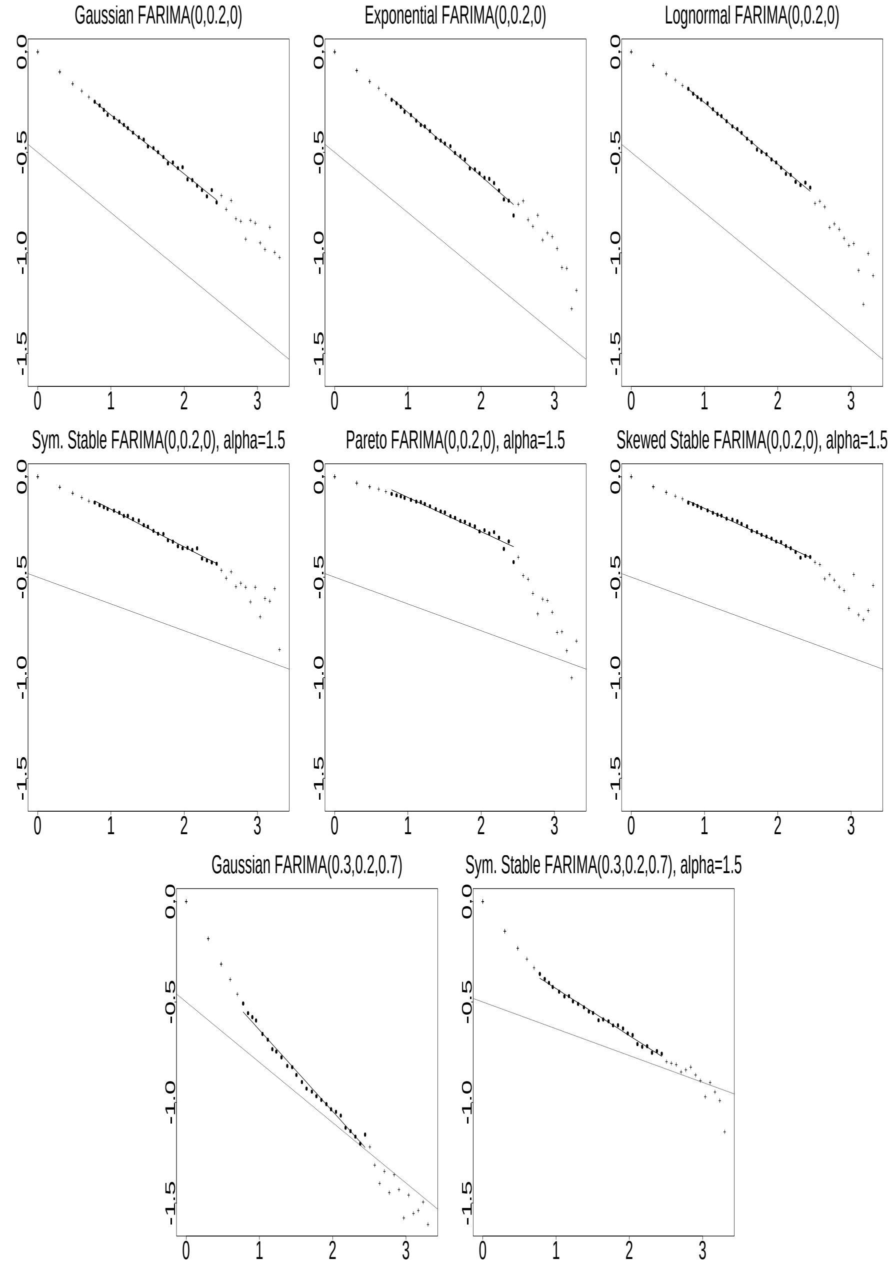


Figure 2: Absolute Value Method. $d = 0.2$. Vertical scale, $\log_{10}(AM^{(m)})$. Horizontal scale, $\log_{10}(m)$. *Top row*: Finite variance, $\mathrm{FARIMA}(0, d, 0)$ series. *Second row*: Infinite variance, $\mathrm{FARIMA}(0, d, 0)$ series. *Third row*: $\mathrm{FARIMA}(1, d, 1)$ series. Reference lines with the theoretical slope $d+1/\alpha-1 = H-1$ are included. The innovations in the infinite variance series (stable, Pareto) have $\alpha = 1.5$.

where $S_{\alpha,d}$ is an α-stable sequence (as stated earlier, it is LFSN). Therefore,

$$\widehat{\mathrm{Var}} X^{(m)} \overset{d}{\sim} m^{2H-2}\left[\left(\frac{N}{m}\right)^{2/\alpha-1} U_{\alpha/2} - C^2\right] \tag{4.18}$$

where $U_{\alpha/2}$ is a positive $\alpha/2$-stable random variable and C is a constant (see Kokoszka and Taqqu [KT96]).

Thus, for large N/m and large m,

$$\widehat{\mathrm{Var}} X^{(m)} \overset{d}{\sim} C(N) m^{2H-2+1-2/\alpha} U_{\alpha/2} = C(N) m^{2d-1} U_{\alpha/2}, \tag{4.19}$$

where $C(N)$ is a constant dependent on N. For fixed sample size $N >> m$, the Variance method thus produces a random variable proportional to m^{2d-1}, in the finite variance as well as in the infinite variance case, providing an estimate for d.

As for the Absolute Value method, a log-log plot versus m should give a straight line, this time with a slope of $2d-1$. The cut-offs of $10^{0.7}$ and $10^{2.5}$ were used in this study.

Jumps in the mean and slowly decaying trends are two types of non-stationarity which can occur. To distinguish these from long-range dependence, we can difference the variance (see Teverovsky and Taqqu [TT95b] for details), and study

$$\widehat{\mathrm{Var}} X^{(m_i+1)} - \widehat{\mathrm{Var}} X^{(m_i)} \tag{4.20}$$

where the m_i's are the successive values of m as defined above. Differencing introduces more fluctuation, but it can provide a way of detecting the presence of the two types of non-stationarity mentioned above. The method should be used together with the original aggregated variance method. For stable FARIMA$(1, d, 1)$ series, the scattering of the estimates produced by this method is so great that the estimator does not produce very useful results. Since we do not focus here on non-stationarity, results for the differenced variance method are not included.

4.3 Variance of Residuals Method

The Variance of Residuals method was introduced by Peng et al. [PBS$^+$94]. First the series is divided into blocks of size m. Then, within each block, the partial sums of the series are calculated, $Y(t) = \sum_{i=1}^{t} X_i$. A least-squares line, $a + bt$, is fitted to the partial sums within each block, and the sample variance of the residuals is computed,

$$\frac{1}{m}\sum_{t=1}^{m}\left(Y(t) - a - bt\right)^2 .$$

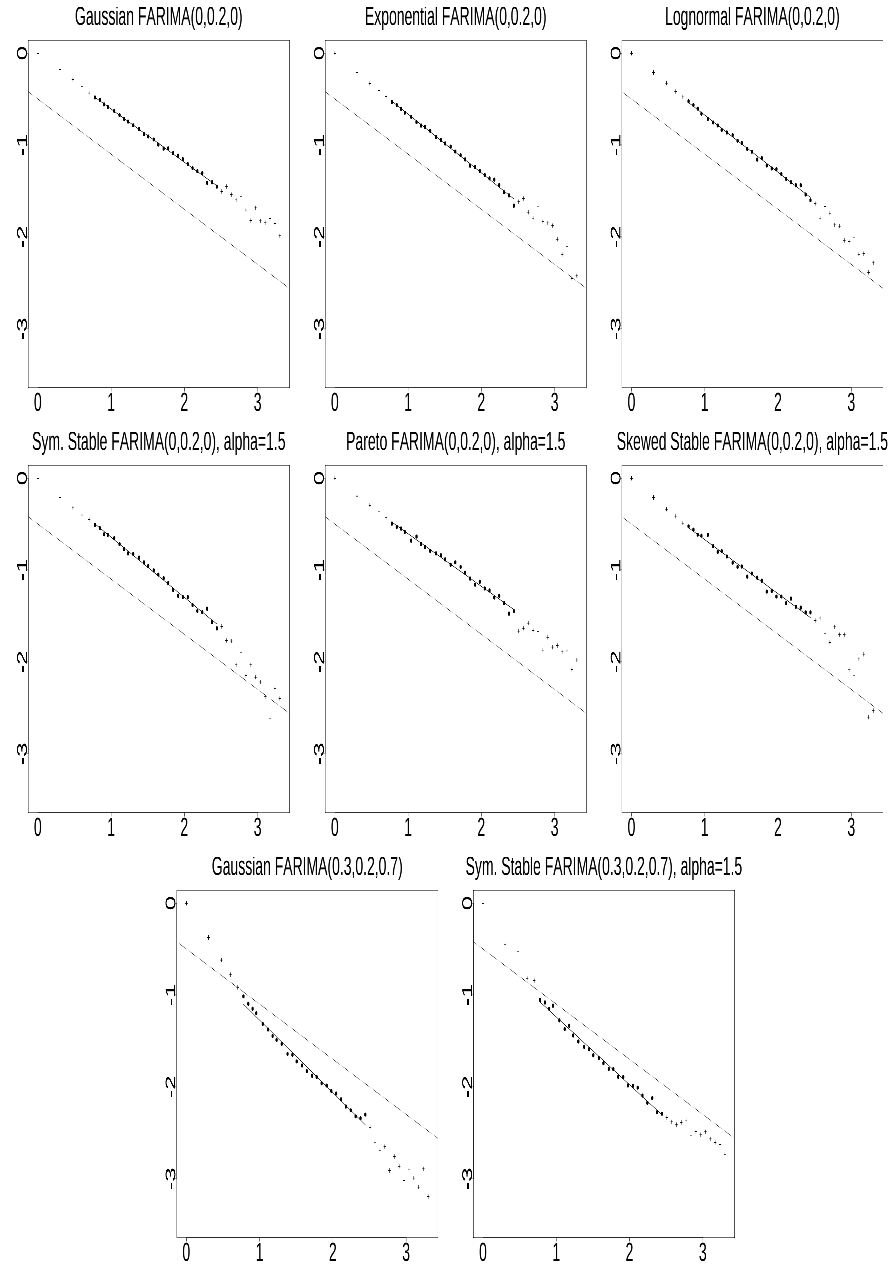


Figure 3: Variance Method. $d = 0.2$. Vertical scale, $\log_{10} \widehat{\mathrm{Var}} X^{(m)})$. Horizontal scale, $\log_{10}(m)$. *Top row*: Finite variance, FARIMA$(0, d, 0)$ series. *Second row*: Infinite variance, FARIMA$(0, d, 0)$ series. *Third row*: FARIMA$(1, d, 1)$ series. Reference lines with the theoretical slope $2d - 1$ are included. The innovations in the infinite variance series (stable, Pareto) have $\alpha = 1.5$.

The following is a sketch of how this estimator behaves in the presence of long-range dependence. Define two random variables,

$$A_1 = \int_0^1 L_{\alpha,H}(t)dt, \qquad A_2 = \int_0^1 tL_{\alpha,H}(t)dt.$$

with $L_{\alpha,H}(t)$, $H = d + 1/\alpha$, denoting α-stable Fractional Stable Motion (see [ST94] for a definition). Since, for large m,

$$Y(mt) \overset{d}{\sim} m^H L_{\alpha,H}(t),$$

we can use the following approximations, where all of the sums are from 1 to m:

$$\sum Y(t) \sim \int_0^m Y(t)dt \overset{d}{\sim} m^{H+1}A_1,$$
$$\sum tY(t) \sim \int_0^m tY(t)dt \overset{d}{\sim} m^{H+2}A_2.$$

Then, the least-squares coefficients can be calculated as:

$$b = \frac{\sum Y(t)t - \frac{1}{m}\sum Y(t)\sum t}{\sum t^2 - \frac{1}{m}(\sum t)^2} \overset{d}{\sim} \frac{m^{H+2}A_2 - m^{H+2}A_1/2}{m^3/3 - m^3/4} = m^{H-1}[12A_2 - 6A_1],$$

$$a = \frac{1}{m}\sum Y(t) - \frac{1}{m}\sum bt \overset{d}{\sim} m^H A_1 - m^H[12A_2 - 6A_1]/2 = m^H[4A_1 - 6A_2],$$

where $\overset{d}{\sim}$ means approximately equal in distribution. Therefore,

$$\frac{1}{m}\sum_{t=1}^{m}(Y(t) - a - bt)^2 \overset{d}{\sim} m^{2H}\int_0^1 (L_{\alpha,H}(t) - (4A_1 - 6A_2) - 6t(2A_2 - A_1))^2 \, dt. \tag{4.21}$$

Since the integral in (4.21) does not depend on m, the sample variance of residuals within each block scales like m^{2H}. The above calculations apply for both the infinite and finite variance case. For a related more rigorous proof in the Gaussian case, see [TTW95].

The sample variance of residuals is computed for each block, and its median is then obtained over the N/m blocks. If we assume, as a first approximation, that the median also behaves like m^{2H}, then a log-log plot versus m should follow a straight line with a slope of $2H$. The cut-offs used here are $(10^{0.7}, 10^{2.5})$.

If, instead of the median, the average over the blocks is computed, some complications arise. The problem stems from the fact that the integral in (4.21) behaves like an $\alpha/2$-stable random variable (hence has infinite mean) if the original series has infinite variance with parameter α. Thus, if the average of an $\alpha/2$-stable random variable over N/m blocks is computed, the result is proportional to $m^{1-2/\alpha}$, and the average of the corresponding variance of residuals is proportional to $m^{2H+1-2/\alpha} = m^{2d+1}$. Therefore, taking the average produces an estimator of d, not H. If d is the parameter of interest, taking the average over the blocks would be a possible procedure. In practice, this is not recommended because the scatter is too large for the infinite variance series.

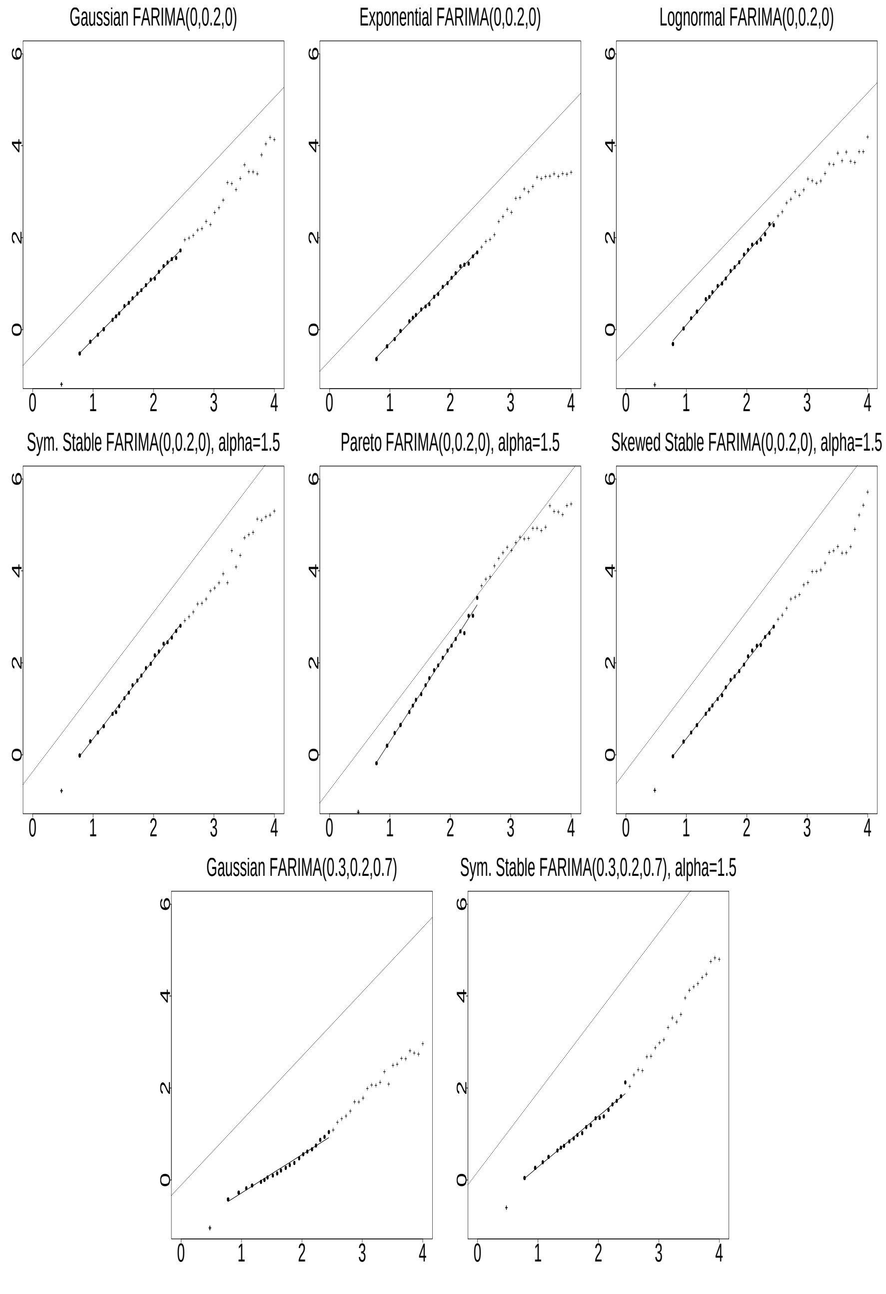


Figure 4: Variance of Residuals Method. $d = 0.2$. Vertical scale, $\log_{10}(statistic)$. Horizontal scale, $\log_{10}(m)$. *Top row*: Finite variance, $\mathrm{FARIMA}(0, d, 0)$ series. *Second row*: Infinite variance, $\mathrm{FARIMA}(0, d, 0)$ series. *Third row*: $\mathrm{FARIMA}(1, d, 1)$ series. Reference lines with the theoretical slope $2(d + 1/\alpha) = 2H$ are included. The innovations in the infinite variance series (stable, Pareto) have $\alpha = 1.5$.

4.4 R/S Method

The R/S method is one of the oldest methods for estimating H and is discussed in detail in Mandelbrot and Wallis [MW69], Mandelbrot [Man75] and Mandelbrot and Taqqu [MT79]. Given the partial sums of a time series $Y(t) = \sum_{i=1}^{t} X(i)$, and its sample variance $S^2(n)$, the R/S statistic is defined:

$$\frac{R}{S}(n) := \frac{1}{S(n)} \left[\max_{0\le t\le n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0\le t\le n} \left(Y(t) - \frac{t}{n} Y(n) \right) \right]. \quad (4.22)$$

Because of the asymptotic scaling property of $Y(t)$, $R(n)$behaves like $n^H = n^{d+1/\alpha}$. On the other hand, $S(n)$ is just the square root of the sample variance, which, as in (4.18), is proportional to $n^{2/\alpha-1}$. Thus, $S(n)$ behaves like $n^{1/\alpha-1/2}$. Since, in fact, one has joint convergence of $(R(n), S(n))$, $R/S(n)$ behaves like $n^{d+1/2}$, as $n \to \infty$. This provides a way of estimating d, whatever the value of α. A rigorous proof, based on the results of Mandelbrot [Man75], can be found in Avram [Avr86].

The actual method divides the original time series into K blocks, each of size N/K. Then, for each lag n, we compute $R(k_i, n)/S(k_i, n)$, starting at points $k_i = iN/K + 1, i = 1, 2, \ldots$, such that $k_i + n \le N$.

We plot R/S versus the lag n on a log-log plot. A line is fitted to the plot, and its slope should equal $d + 1/2$. The cut-offs are $10^{0.7}$ and $10^{3.5}$.

4.5 Periodogram Method

The periodogram is defined as

$$I(\nu) = \frac{1}{2\pi N} \left| \sum_{j=1}^{N} X(j) e^{ij\nu} \right|^2, \quad (4.23)$$

where ν is the frequency, N is the length of the series, and X is the time series. In the finite variance case, $I(\nu)$ is an estimator of the spectral density of X, and a series with long-range dependence will have a spectral density proportional to $|\nu|^{-2d}$ close to the origin. A log-log regression thus provides an estimate of d. In the infinite variance case the problem is significantly more complicated. Although extensive work has been done in this general area (see Klüppelberg and Mikosch [KM96a], [KM96b] and Kokoszka and Mikosch [KM]), there are no theoretical results for the periodogram regression method. The proportionality to $|\nu|^{-2d}$ as $\nu \to 0$, however, seems to hold empirically in the infinite variance case as well, suggesting the need for theoretical research in this direction.

We thus expect a log-log plot of the periodogram versus the frequency to display a straight line with a slope of $-2d$. In practice we use only the lowest 10% of the frequencies for the calculation, since this behavior holds only for frequencies close to zero.

We do not include here the "smoothed" periodogram method discussed in [TTW95], because in our studies the original periodogram has consistently performed better.

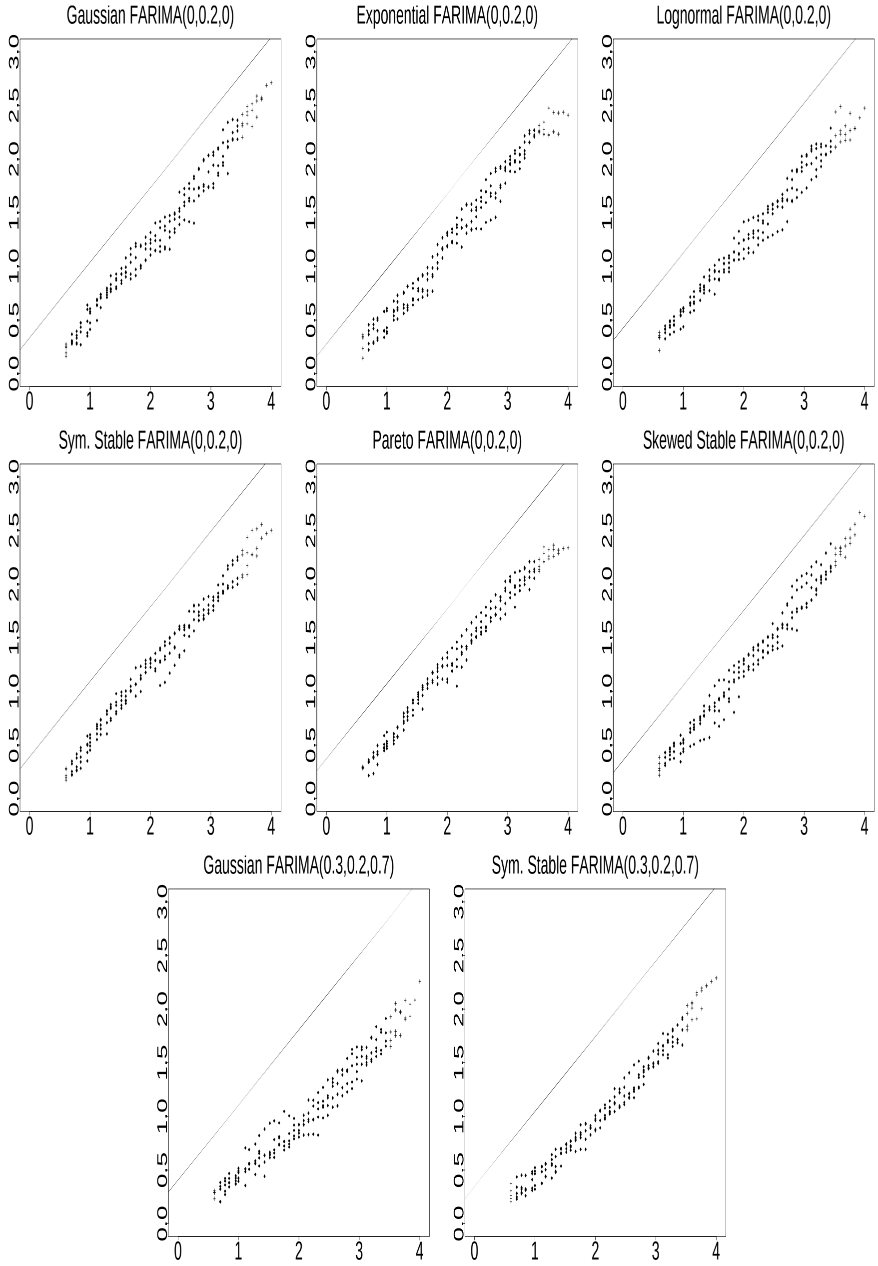


Figure 5: R/S Method. $d = 0.2$. Vertical scale, $\log_{10}(statistic)$. Horizontal scale, $\log_{10}(m)$. *Top row*: Finite variance, FARIMA$(0, d, 0)$ series. *Second row*: Infinite variance, FARIMA$(0, d, 0)$ series. *Third row*: FARIMA$(1, d, 1)$ series. Reference lines with the theoretical slope $d + 1/2$ are included. The innovations in the infinite variance series (stable, Pareto) have $\alpha = 1.5$.

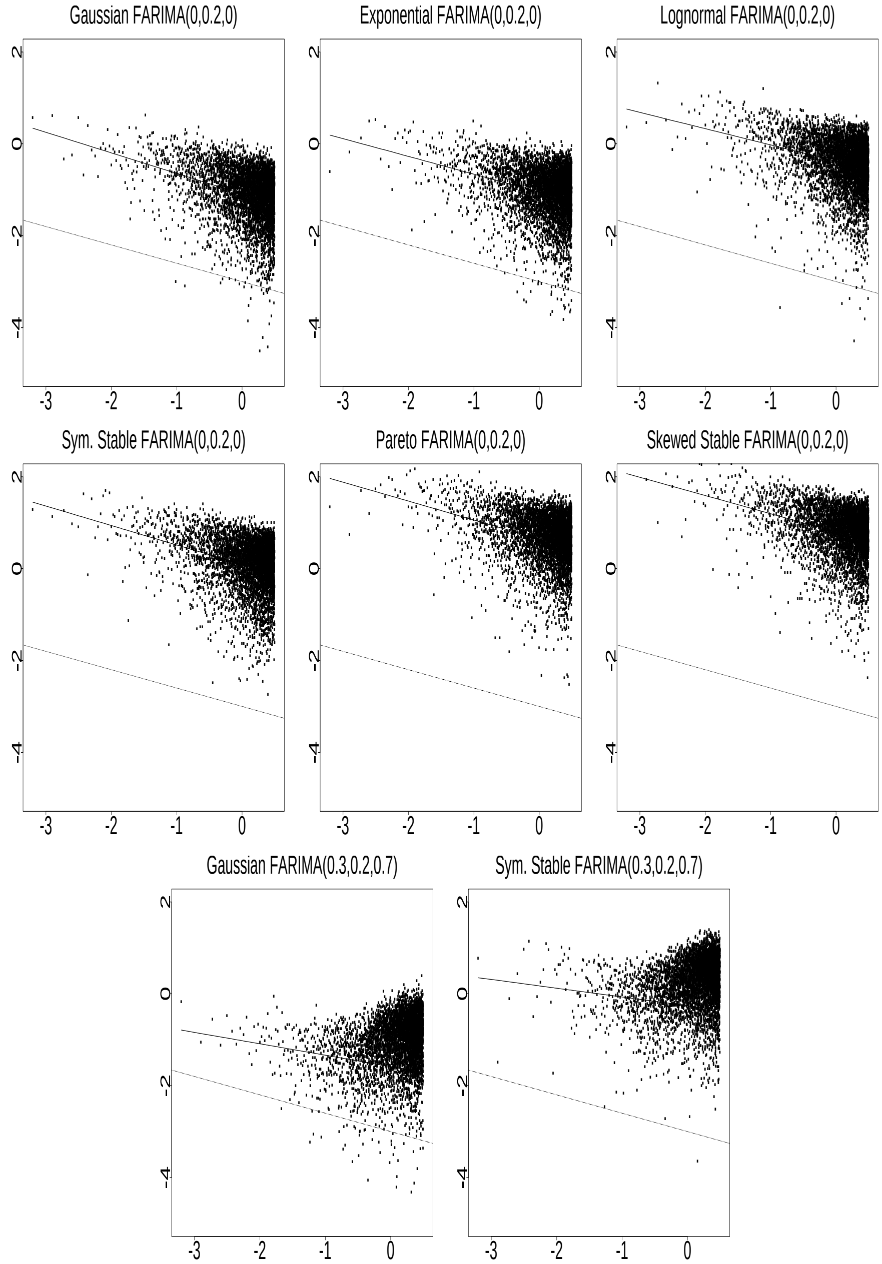


Figure 6: Periodogram Method. $d = 0.2$. Vertical scale, $\log_{10}(periodogram)$. Horizontal scale, $\log_{10}(frequency)$. *Top row*: Finite variance, FARIMA$(0, d, 0)$ series. *Second row*: Infinite variance, FARIMA$(0, d, 0)$ series. *Third row*: FARIMA$(1, d, 1)$ series. Reference lines with the theoretical slope $-2d$ are included. The innovations in the infinite variance series (stable, Pareto) have $\alpha = 1.5$.

4.6 Whittle Method

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

$$Q(\eta) := \int_{-\pi}^{\pi} \frac{I(\nu)}{f(\nu;\eta)} d\nu + \int_{-\pi}^{\pi} \log f(\nu;\eta) d\nu. \tag{4.24}$$

where η is the vector of unknown parameters, $I(\nu)$ is the periodogram (see (4.23)) and $f(\nu;\eta)$ is the spectral density at frequency ν. The term $\int_{-\pi}^{\pi} \log f(\nu;\eta)d\nu$ can be set equal to 0 by renormalizing $f(\nu;\eta)$. The normalization only depends on a scale parameter, and not on the rest of the components of η. Thus, we replace f by f^*, such that $f^* = cf$, and $\int_{-\pi}^{\pi} \log f^*(\nu;\eta)d\nu = 0$.

The Whittle estimator is defined as the value of η which minimizes the function Q. In practice, (4.24) is replaced by a sum over Fourier frequencies. η is the parameter d if the series is FARIMA$(0, d, 0)$, and it also includes the autoregressive and moving average parts if the series is FARIMA(p, d, q) (see (2.8)). It is d and not H which is estimated even in the infinite variance case.

The main disadvantages of the Whittle estimator are: (1) the need to know the parametric form of the spectral density and (2) a greater computational effort than for the graphical methods listed above. If the exact form of the spectral density is not known, the estimator can become very biased. Moreover, one may not notice that bias exists because the output is non-graphical. Thus, this estimator is not very robust. For details see Fox and Taqqu [FT86] and Beran [Ber94] in the Gaussian case and Klüppelberg and Mikosch [KM96b] and Kokoszka and Taqqu [KT96] in the infinite variance case. In this study we assume for all of the series that the true model is FARIMA(p, d, q) with $p = q = 1$. This assumption is reasonable because we do not generate series with p or q higher than 1. The estimator would become biased if the generated series had a p or q larger than 1 because of underspecification.

4.7 Aggregated Whittle Method

The Aggregated Whittle estimator provides a way of robustifying the Whittle estimator in the absence of exact parametric information about the spectral density of the given data series. It can be used if the time series is long enough. The idea is to aggregate the data, which creates a new, shorter series

$$X_k^{(m)} := \frac{1}{m} \sum_{i=m(k-1)+1}^{mk} X_i. \tag{4.25}$$

Of course the shortened length of the aggregated series increases the standard deviation of the estimator, but, on the other hand, for high enough aggregation level m, the centered aggregated series $X^{(m)} - EX^{(m)}$ will be close to Fractional Gaussian Noise and the bias can be reduced by using the Whittle estimator with an underlying FGN model, in the finite variance case. The limit is LFSN in the infinite variance case, but the results of [KT96] suggest that it is reasonable to still use the Whittle estimator with an underlying FGN model.

4.8 Local Whittle Method

Unlike the Whittle estimator, the Local Whittle estimator is a semi-parametric estimator, in that it only specifies the parametric form of the spectral density when ν is close to zero, namely, it assumes that

$$f(\nu) \sim G(d)|\nu|^{-2d} \qquad \text{as } \nu \to 0. \tag{4.26}$$

where $G(d)$ is some function of d. This estimator is essentially equivalent to the Whittle, except that it only assumes the behaviour of the spectral density up to some frequency $2\pi M/N$. One estimates d by minimizing

$$R(d) = \log\left(\frac{1}{M}\sum_{j=1}^{M}\frac{I(\nu_j)}{\nu_j^{-2d}}\right) - 2d\frac{1}{M}\sum_{j=1}^{M}\log\nu_j. \tag{4.27}$$

The Local Whittle estimator was developed by Robinson [Rob95] and was further studied, along with the Whittle and Aggregated Whittle estimators, by Taqqu and Teverovsky [TT95a] in the Gaussian case. As we will show here, it works well in the infinite variance case, although there are as yet no corresponding theoretical results.

5. Results

We will first present some general features of our results, and then comment on the behavior of each estimator separately.

5.1 General Observations

As can be seen in Table 1 and Figure 7, most estimators perform well when FGN or Gaussian FARIMA$(0,d,0)$ series are used. This provides a comparative benchmark for what happens when one adds deviations from "pure" Gaussian long-range dependence.

When Gaussian FARIMA(p,d,q) series are used the results depend very strongly on the values of the AR and MA parameters [4]. If the parameters are negative, as in the left part of Table 3, most estimators are essentially unbiased, and the results are comparable to those for FARIMA$(0,d,0)$ series. On the other hand, if the parameters are positive, as in Table 2 and on the right-hand side of Table 3, as well as the bottom two-thirds of Figure 7, then most estimators are quite biased. The exception is the Whittle estimator when assuming a FARIMA$(1,d,1)$ model because the series we generate have p and q no higher than 1. The next best estimator is the Local Whittle.

Using exponential and lognormal innovations instead of Gaussian ones does not affect most of the estimators (see Table 4 and the boxplots in Figure 8).

[4] When p and/or q are 1, the notation FARIMA(ϕ_1, d, θ_1) is used in the tables in order to display the values of ϕ_1 and/or θ_1.

The exceptions are the Absolute Value Method which is biased downward for high d, and the Variance of Residuals method which is very biased for the lognormal-based series.

Most of the estimators performed well for symmetric stable FARIMA$(0, d, 0)$ series (see the left sides of Tables 6 and 8 and the top halves of Figures 11 and 13). The exception again was the Absolute Value estimator which appeared both more scattered and more biased. In addition, while the standard deviations and confidence intervals for most of the methods are similar to the Gaussian cases, more outliers appear. This is because the series have infinite variance. When stable FARIMA$(1, d, 1)$ are used, the results are very similar to the Gaussian $(1, d, 1)$ case with the same values of the parameters ($\phi = 0.3$, $\theta = 0.7$) (see the right sides of Tables 6 and 8 and the bottom halves of Figures 11 and 13). The exception is the Absolute Value method which is affected by the infinite variance of the sample series. We note again that while most of the estimators estimate the parameter d, the Variance of Residuals and the Absolute Value methods estimate H. Also, the Whittle method tends to have infrequent but very large outliers in the FARIMA$(1, 0, 1)$ cases. Whereas the estimate of d should be 0, approximately 1% of the time the boundary value of 0.5 is returned. (The Whittle estimator software restricts the values of d to the interval $(-0.5, 0.5)$ corresponding to the stationary range for finite variance series.) Because of these outliers, the standard errors for the Whittle estimator are larger in the FARIMA$(1, 0, 1)$ cases, both for $\alpha = 1.5$ and $\alpha = 1.2$.

Using Pareto and skewed stable innovations for the FARIMA$(0, d, 0)$ series gives results very similar to the symmetric stable FARIMA$(0, d, 0)$, with $\alpha = 1.5$ (see Table 7 and Figure 12). The exception is the Variance of Residuals method, which is very biased for Pareto innovations.

5.2 Graphical Methods

The *Absolute Value method* is one of the better graphical methods for Gaussian series. Unfortunately its performance declines for infinite variance series, particularly as α gets closer to 1 (and also as d grows for non-Gaussian finite variance series) (see Figure 2). An intuitive explanation is that as α gets closer to 1, the series become more and more heavy-tailed, and even the mean diverges when $\alpha = 1$. Thus, outliers in the original series can have an inordinate effect on the estimate of H, and therefore the estimates are very scattered, as seen in the boxplots. Therefore, while this method can be used if the data series is known to be Gaussian or has finite variance, it is not recommended if the series is likely to be infinite variance.

The *Variance method* is biased downward for high values of d even if Gaussian FARIMA$(0, d, 0)$ series are used. This bias is not especially large unless $d > 0.3$, so this should not be an over-riding consideration. For Gaussian FARIMA$(1, d, 1)$ series, the performance declines, as with the other methods, but the bias is never greater than 0.1. The biases and standard deviations remain approximately the same for Gaussian and non-Gaussian series. Although the method is never extremely accurate like the Whittle-type methods, it appears quite robust (see Figure 3).

The *Periodogram method* is perhaps the best of the graphical methods examined here. While it is not extremely accurate for the "ideal" series, it suffers the smallest losses in efficiency as deviations from the "ideal" are introduced. For the FARIMA$(1,d,1)$ series it has some of the smallest MSEs along with the Absolute Value method, but unlike the latter, its efficiency does not suffer when non-Gaussian series are introduced. Even for stable FARIMA$(1,d,1)$ its $\sqrt{\text{MSE}}$ is never larger than .095 (see Figure 6).

The *R/S method* has long been known to be biased towards $d = 0.2$. It is biased upward for small values of d and downward for large values of d. The method is also very biased when FARIMA$(1,d,1)$ series are used, and while not suffering a loss of efficiency for non-Gaussian series, it is already so inaccurate that it should be used only to get a very rough idea of the intensity of long-range dependence (see Figure 5).

The *Variance of Residuals method* is one of only two methods which give H, the scaling parameter, instead of d. It works very well for both Gaussian and stable FARIMA$(0,d,0)$ series, giving the smallest MSEs of any of the graphical methods, and is better than the Whittle estimator for several parameter values. Unfortunately, while its standard error remains small, its bias becomes quite large for lognormal, Pareto and also FARIMA$(1,d,1)$ series. Apparently the effects of the short-range dependence persist to a greater degree here than for the other graphical estimators (see Figure 4). Its accuracy for "ideal" series makes it desirable to attempt to make this method more robust. A possible solution is to only use the larger values of m for our estimates by moving the lower cut-off to the right. This does result in decreased bias, but since the interval where H is estimated becomes shorter, the estimates are more scattered and the standard error increases. Another way in which this method can be robustified is by again aggregating the original series, just as in the Aggregated Whittle, and for the same reasons. The aggregated series should become closer and closer to FGN (LFSN) as the aggregation level increases, and the estimator is quite good for those kinds of series. Clearly the level of aggregation is very important. If it is not large enough, the estimates will be very biased. If it is too large, the estimates will be very scattered, and the standard error will be large. To use the Variance of Residuals effectively, one would need to run the estimator at several cut-off values or aggregation levels. The techniques proposed for the Whittle-type estimators in [TT96] would also be reasonable here.

5.3 Whittle-type estimators

The *Whittle estimator* with the assumed model of FARIMA$(1,d,1)$ gave the best performance for the series used in this study. Recall that all of the series used here fall within the FARIMA$(1,d,1)$ family of models. Certainly, if the parametric form of a time series is known, then the Whittle estimator is to be recommended. Even if the exact form is not known, but the maximum order (p, q) is known, this estimator can give good results. There is no problem with symmetric stable series, as the estimator's performance is essentially unaffected. When other infinite variance series are introduced, the estimator gives very incorrect estimates for a few realizations. For the majority of realizations, however, it gives good results.

We note here that in addition to d, the Whittle estimator also provides values for ϕ and θ, the AR and MA coefficients. In general, the estimates of these coefficients are approximately as accurate as the estimates of d. However, a problem can arise if the actual series is FARIMA$(0, d, 0)$ or FGN and if a FARIMA$(1, d, 1)$ model is assumed. The estimator will often give non-zero values for ϕ and θ. These non-zero values are usually almost equal, and thus essentially cancel each other. They can be quite large and this can affect, somewhat, the estimation of d. It is, however, better to overspecify the order of the model than to underspecify it. In the latter case, the biases will be very large.

Aggregated Whittle is in general fairly unbiased and although its standard error is larger than for the Whittle estimator, it is smaller than for the graphical methods for the FARIMA$(1, d, 1)$ series. Just as for other methods, the level of aggregation can be an important factor in the performance of the estimator. We used an aggregation level of $m = 50$, which seemed to provide reasonable results for the time series of length $10,000$ considered here. An aggregation level which is much smaller would not eliminate the bias, and a much larger aggregation level would introduce too much uncertainty.

The *Local Whittle estimator* seems to perform quite well. Overall, it is better than the other methods, except the original Whittle (under the correct underlying model). Unlike the Whittle it does not estimate the AR and MA coefficients. The Local Whittle may be the preferred estimator to use when the parametric form of a series is completely unknown, except for its long memory. Just as for the Whittle method, some realizations yield very incorrect estimates, but, in the vast majority of realizations this problem does not arise. As with the other methods, the choice of M in (4.27) is an important one. $M = N/32$ seems to be a good choice. For more details about the choice of M (Local Whittle) or m (Aggregated Whittle), see Taqqu and Teverovsky [TT96].

As mentioned at the end of Section 3, we analyzed the Local Whittle in particular detail. The results for series with Gaussian, stable $\alpha = 1.5$ and stable $\alpha = 1.2$ innovations are illustrated in Figures 9 and 10. These figures clearly show the difference between finite and infinite variance series. Although the estimates for the stable series have more outliers, they are in fact more concentrated in the center of the distribution. This is best seen in the boxplots (Figure 10) for the stable series, as the box representing the middle 50% of the estimates is narrower, but more outliers are present. While the standard deviation of the estimates is smaller for the stable series (see Table 5), one needs a larger number of realizations, I, to account for the greater variability.

	Gaussian Innovations								
	Nominal d								
	FGN					FARIMA(0,d,0)			
	0.0	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
	ABSOLUTE								
BIAS	-.003	-.005	.000	-.005	-.004	-.006	-.002	-.003	-.012
$\hat{\sigma}$	.028	.028	.024	.028	.049	.033	.040	.037	.040
$\sqrt{\text{MSE}}$	.028	.028	.023	.028	.049	.033	.039	.037	.041
	PERIODOGRAM								
BIAS	.001	.001	.009	.012	.011	.009	.015	.016	.005
$\hat{\sigma}$	.032	.029	.033	.025	.028	.039	.033	.031	.028
$\sqrt{\text{MSE}}$	.032	.029	.034	.028	.030	.039	.034	.035	.028
	R/S								
BIAS	.035	.009	-.013	-.034	-.079	.014	-.012	-.040	-.077
$\hat{\sigma}$	.023	.024	.020	.024	.027	.023	.022	.022	.026
$\sqrt{\text{MSE}}$	.042	.025	.024	.042	.083	.027	.025	.046	.081
	VARIANCE								
BIAS	-.005	-.012	-.013	-.028	-.056	-.009	-.014	-.027	-.060
$\hat{\sigma}$	.026	.027	.024	.022	.031	.031	.036	.028	.032
$\sqrt{\text{MSE}}$	.026	.029	.027	.036	.063	.031	.038	.039	.068
	VARIANCE OF RESIDUALS								
BIAS	-.009	-.011	-.014	-.018	-.016	-.017	-.023	-.028	-.035
$\hat{\sigma}$	.012	.015	.014	.018	.016	.016	.016	.017	.017
$\sqrt{\text{MSE}}$	.015	.019	.020	.026	.022	.023	.028	.033	.039
	WHITTLE (FARIMA$(1, d, 1)$)								
BIAS	-.004	-.006	-.002	-.005	-.002	-.002	.000	-.001	-.004
$\hat{\sigma}$	.022	.015	.017	.019	.015	.022	.025	.020	.018
$\sqrt{\text{MSE}}$	.022	.016	.017	.019	.015	.022	.025	.020	.018
	AGGREGATED WHITTLE ($m = 50$)								
BIAS	-.007	-.017	-.005	-.002	-.017	-.008	-.005	-.011	-.024
$\hat{\sigma}$	.054	.048	.041	.042	.047	.052	.056	.046	.054
$\sqrt{\text{MSE}}$	.054	.050	.041	.041	.050	.052	.056	.047	.058
	LOCAL WHITTLE ($N/m = 32$)								
BIAS	-.003	-.005	.006	.002	.006	.003	.000	.004	-.004
$\hat{\sigma}$	.034	.028	.030	.026	.031	.031	.033	.022	.025
$\sqrt{\text{MSE}}$	.034	.028	.031	.026	.031	.031	.033	.022	.025

Table 1: Estimation results for d using 50 independent copies of time series of length $10,000$. These are the "ideal" results for FGN and for Gaussian FARIMA(0,d,0).

	Gaussian Innovations									
	Nominal d									
	FARIMA$(.5, d, 0)$					FARIMA$(0, d, .5)$				
	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
	ABSOLUTE									
BIAS	.026	.017	.013	.012	-.014	-.061	-.040	-.034	-.022	-.029
$\hat{\sigma}$	.028	.029	.031	.044	.045	.017	.031	.033	.041	.052
$\sqrt{\text{MSE}}$	.038	.033	.033	.045	.047	.064	.051	.047	.046	.060
	PERIODOGRAM									
BIAS	.022	.029	.027	.031	.040	-.029	-.015	-.028	-.014	-.016
$\hat{\sigma}$	.027	.026	.025	.032	.033	.028	.029	.035	.033	.029
$\sqrt{\text{MSE}}$	.035	.039	.037	.044	.052	.040	.033	.044	.036	.032
	R/S									
BIAS	.103	.074	.027	-.016	-.064	-.076	-.092	-.113	-.122	-.141
$\hat{\sigma}$	.020	.022	.021	.024	.021	.017	.022	.021	.026	.021
$\sqrt{\text{MSE}}$	.105	.078	.034	.029	.067	.077	.095	.115	.124	.143
	VARIANCE									
BIAS	.025	.013	.001	-.022	-.052	-.063	-.044	-.051	-.050	-.073
$\hat{\sigma}$	.027	.024	.026	.032	.019	.031	.029	.029	.034	.033
$\sqrt{\text{MSE}}$	.036	.028	.026	.039	.059	.065	.053	.058	.060	.080
	VARIANCE OF RESIDUALS									
BIAS	.121	.101	.087	.077	.070	-.192	-.195	-.198	-.187	-.177
$\hat{\sigma}$	.011	.013	.014	.017	.019	.011	.013	.013	.016	.016
$\sqrt{\text{MSE}}$	.121	.102	.088	.079	.073	.192	.195	.198	.188	.177
	WHITTLE (FARIMA$(1, d, 1)$)									
BIAS	-.002	-.002	-.005	-.004	.000	-.006	.002	-.012	-.001	-.004
$\hat{\sigma}$	.023	.022	.021	.026	.026	.020	.021	.023	.026	.022
$\sqrt{\text{MSE}}$	.023	.022	.021	.026	.026	.021	.021	.026	.026	.022
	AGGREGATED WHITTLE ($m = 50$)									
BIAS	.001	.007	.005	-.009	-.021	-.032	-.007	-.024	-.024	-.027
$\hat{\sigma}$	.045	.057	.044	.050	.046	.034	.054	.044	.048	.047
$\sqrt{\text{MSE}}$	.044	.057	.044	.051	.050	.047	.054	.049	.053	.053
	LOCAL WHITTLE ($N/m = 32$)									
BIAS	.010	.013	.007	.010	.016	-.010	-.008	-.015	-.012	-.014
$\hat{\sigma}$	.036	.027	.025	.034	.031	.025	.030	.031	.032	.032
$\sqrt{\text{MSE}}$	.037	.030	.026	.035	.035	.027	.030	.034	.033	.035

Table 2: Estimation results for d using 50 independent copies of FARIMA time series of length $10,000$. The series are Gaussian FARIMA$(0.5, d, 0)$ and $(0, d, 0.5)$.

	Gaussian Innovations									
	Nominal d									
	FARIMA$(-.3, d, -.7)$					FARIMA$(.3, d, .7)$				
	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
	ABSOLUTE									
BIAS	.001	.006	.011	-.006	-.010	-.098	-.077	-.055	-.043	-.029
$\hat{\sigma}$	.031	.029	.029	.044	.039	.028	.029	.040	.046	.052
$\sqrt{\text{MSE}}$	.031	.029	.031	.044	.040	.102	.082	.068	.062	.059
	PERIODOGRAM									
BIAS	-.003	.004	.017	.013	.017	-.068	-.067	-.063	-.063	-.051
$\hat{\sigma}$	.032	.033	.032	.032	.031	.033	.030	.028	.033	.028
$\sqrt{\text{MSE}}$	.032	.033	.036	.034	.035	.076	.073	.069	.071	.058
	R/S									
BIAS	.062	.028	-.010	-.040	-.091	-.106	-.123	-.145	-.157	-.173
$\hat{\sigma}$	.023	.024	.024	.028	.034	.018	.024	.024	.022	.024
$\sqrt{\text{MSE}}$	.066	.036	.025	.049	.097	.108	.126	.147	.158	.175
	VARIANCE									
BIAS	.000	-.002	-.008	-.034	-.054	-.099	-.082	-.069	-.074	-.086
$\hat{\sigma}$	.028	.029	.025	.032	.028	.027	.029	.031	.030	.031
$\sqrt{\text{MSE}}$	.028	.029	.026	.027	.060	.102	.087	.076	.079	.091
	VARIANCE OF RESIDUALS									
BIAS	.014	.002	-.002	-.008	-.010	-.235	-.246	-.253	-.249	-.237
$\hat{\sigma}$	.014	.014	.015	.018	.017	.011	.012	.015	.017	.016
$\sqrt{\text{MSE}}$	.020	.014	.015	.020	.019	.235	.246	.254	.249	.238
	WHITTLE (FARIMA$(1, d, 1)$)									
BIAS	-.003	-.005	-.002	-.002	.001	-.004	-.004	-.007	-.009	-.006
$\hat{\sigma}$	.012	.013	.012	.012	.010	.033	.032	.029	.031	.033
$\sqrt{\text{MSE}}$	.013	.013	.012	.012	.010	.033	.032	.029	.030	.033
	AGGREGATED WHITTLE ($m = 50$)									
BIAS	.002	.003	.000	-.016	-.020	-.040	-.024	-.015	-.026	-.031
$\hat{\sigma}$	.051	.051	.043	.055	.043	.046	.047	.048	.050	.046
$\sqrt{\text{MSE}}$	.050	.050	.042	.057	.047	.061	.053	.050	.056	.055
	LOCAL WHITTLE ($N/M = 32$)									
BIAS	-.003	-.001	.008	-.002	.007	-.032	-.024	-.033	-.031	-.030
$\hat{\sigma}$	.029	.030	.029	.31	.031	.031	.032	.030	.030	.032
$\sqrt{\text{MSE}}$	.029	.029	.030	.031	.031	.044	.039	.044	.043	.044

Table 3: Estimation results for d using 50 independent copies of Gaussian FARIMA$(1, d, 1)$ time series of length $10,000$.

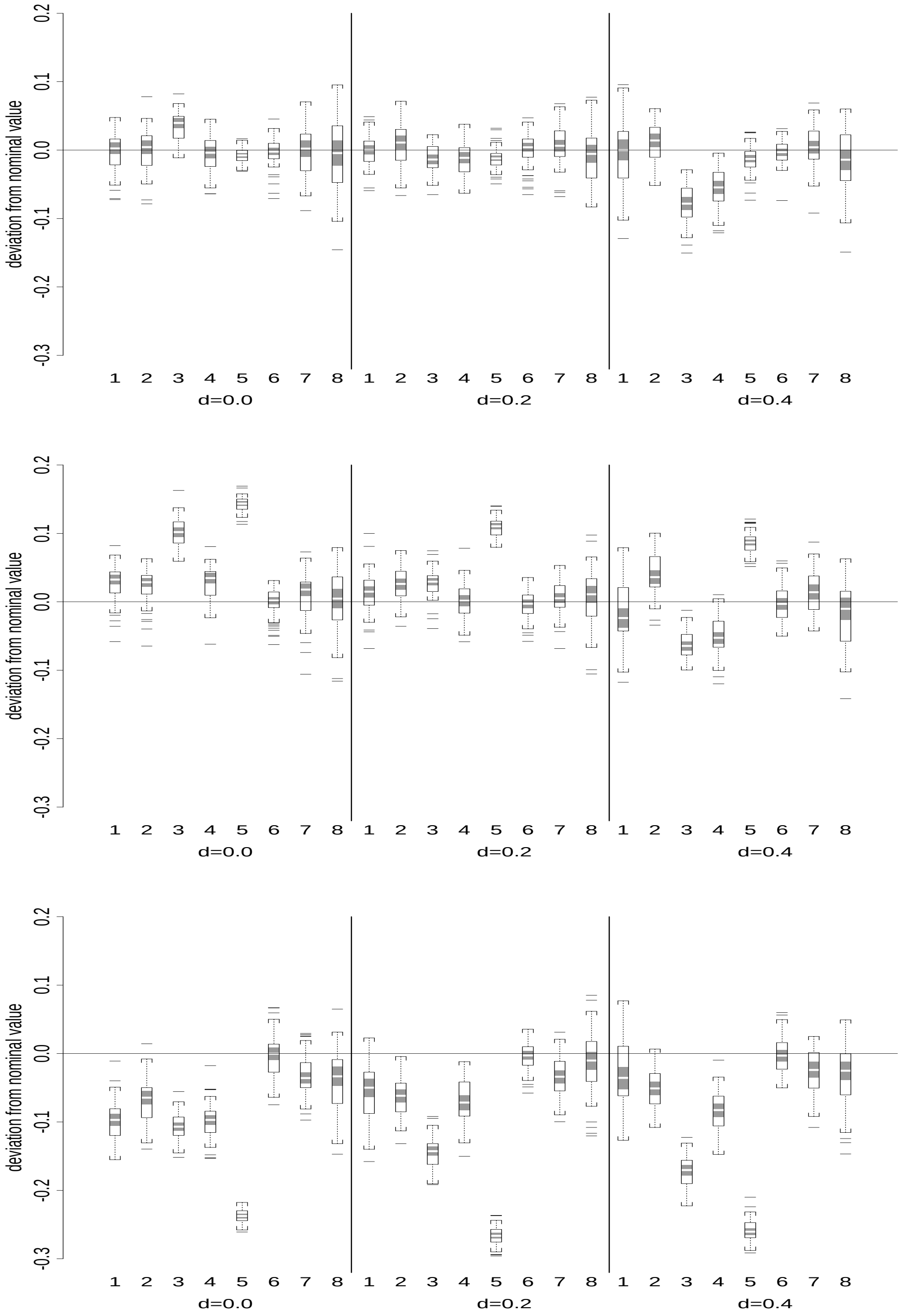
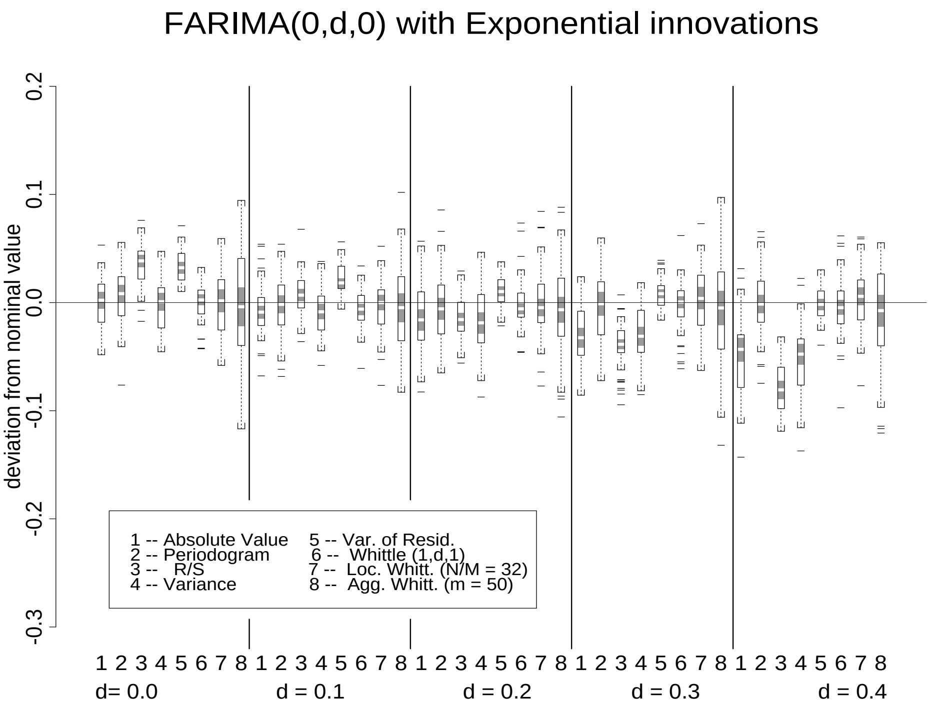


Figure 7: *Top row*: FARIMA$(0, d, 0)$, *Middle row*: FARIMA$(0, d, 0.5)$, *Bottom row*: FARIMA$(0.3, d, 0.7)$. Gaussian innovations. *Boxplots:* 1 – Absolute Value. 2 – Periodogram. 3 – R/S. 4 – Variance. 5 – Variance of Residuals. 6 – Whittle. 7 – Local Whittle ($N/M = 32$). 8 – Aggregated Whittle ($m = 50$).

	Exponential Innovations FARIMA$(0,d,0)$ 0.0	0.1	0.2	0.3	0.4	Lognormal Innovations FARIMA$(0,d,0)$ 0.0	0.1	0.2	0.3	0.4
	Nominal d									
	ABSOLUTE									
BIAS	-.001	-.008	-.013	-.029	-.050	.027	.012	.005	-.010	-.040
$\hat{\sigma}$	.024	.024	.033	.027	.036	.026	.028	.027	.029	.031
$\sqrt{\text{MSE}}$	.023	.025	.035	.039	.061	.037	.030	.027	.030	.050
	PERIODOGRAM									
BIAS	.006	-.003	-.003	-.005	-.001	.004	-.004	-.002	.003	.001
$\hat{\sigma}$	.026	.027	.031	.031	.032	.031	.032	.030	.029	.027
$\sqrt{\text{MSE}}$	.026	.027	.031	.031	.032	.031	.032	.030	.029	.027
	R/S									
BIAS	.036	.007	-.014	-.040	-.079	.036	.005	-.018	-.047	-.085
$\hat{\sigma}$	.020	.019	.021	.022	.024	.019	.019	.020	.022	.019
$\sqrt{\text{MSE}}$	.041	.020	.025	.045	.083	.040	.019	.027	.051	.087
	VARIANCE									
BIAS	-.004	-.008	-.014	-.030	-.051	.003	-.013	-.013	-.027	-.052
$\hat{\sigma}$	.025	.022	.031	.025	.033	.025	.026	.029	.026	.029
$\sqrt{\text{MSE}}$	.025	.024	.034	.039	.060	.025	.029	.031	.037	.059
	VARIANCE OF RESIDUALS									
BIAS	.035	.021	.010	.007	-.002	.125	.105	.097	.086	.081
$\hat{\sigma}$	.015	.014	.015	.014	.015	.015	.016	.015	.019	.022
$\sqrt{\text{MSE}}$	.038	.026	.018	.016	.015	.126	.106	.098	.088	.084
	WHITTLE (FARIMA$(1,d,1)$)									
BIAS	.000	-.004	-.001	-.003	-.003	.005	-.005	-.005	.001	-.005
$\hat{\sigma}$	.017	.017	.023	.024	.028	.021	.020	.023	.021	.025
$\sqrt{\text{MSE}}$	.017	.018	.023	.024	.028	.021	.021	.023	.021	.026
	AGGREGATED WHITTLE ($m=50$)									
BIAS	-.003	-.005	-.004	-.009	-.015	.000	-.012	-.001	-.011	-.018
$\hat{\sigma}$	.051	.040	.047	.047	.048	.043	.051	.049	.039	.044
$\sqrt{\text{MSE}}$	.051	.040	.046	.047	.050	.043	.052	.049	.041	.047
	LOCAL WHITTLE ($N/M=32$)									
BIAS	.000	-.003	-.001	.003	.002	.004	-.010	.003	.003	-.002
$\hat{\sigma}$	.031	.024	.033	.031	.029	.028	.027	.030	.031	.028
$\sqrt{\text{MSE}}$	.031	.024	.032	.031	.029	.028	.029	.030	.031	.028

Table 4: Estimation results for d using 50 independent copies of FARIMA$(0,d,0)$ time series of length $10,000$. The innovations are exponential and lognormal.



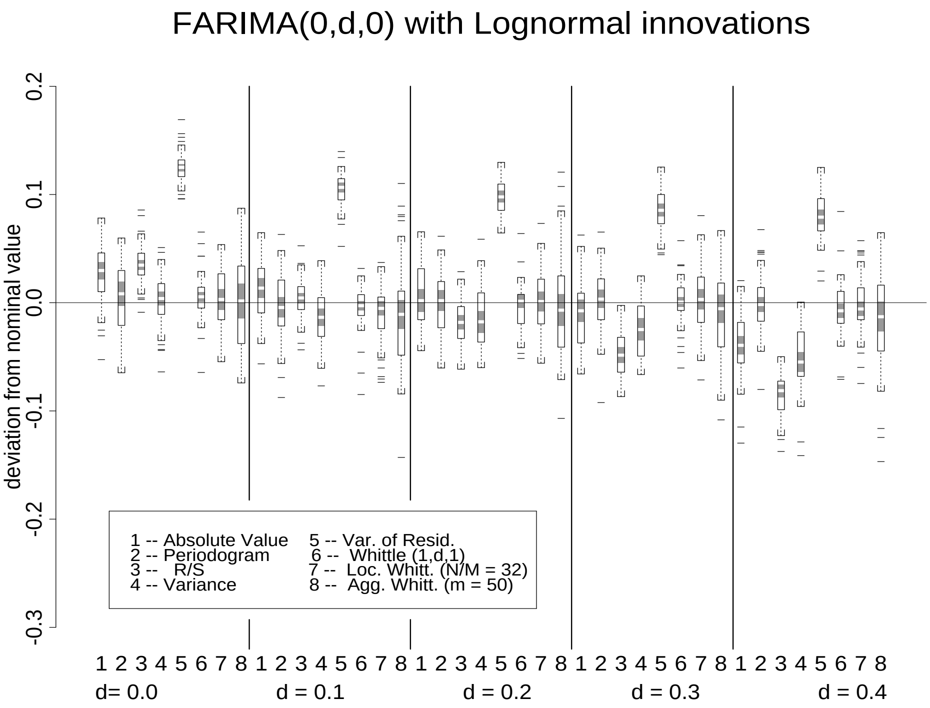


Figure 8: Boxplots for FARIMA$(0, d, 0)$ series with finite variance innovations.

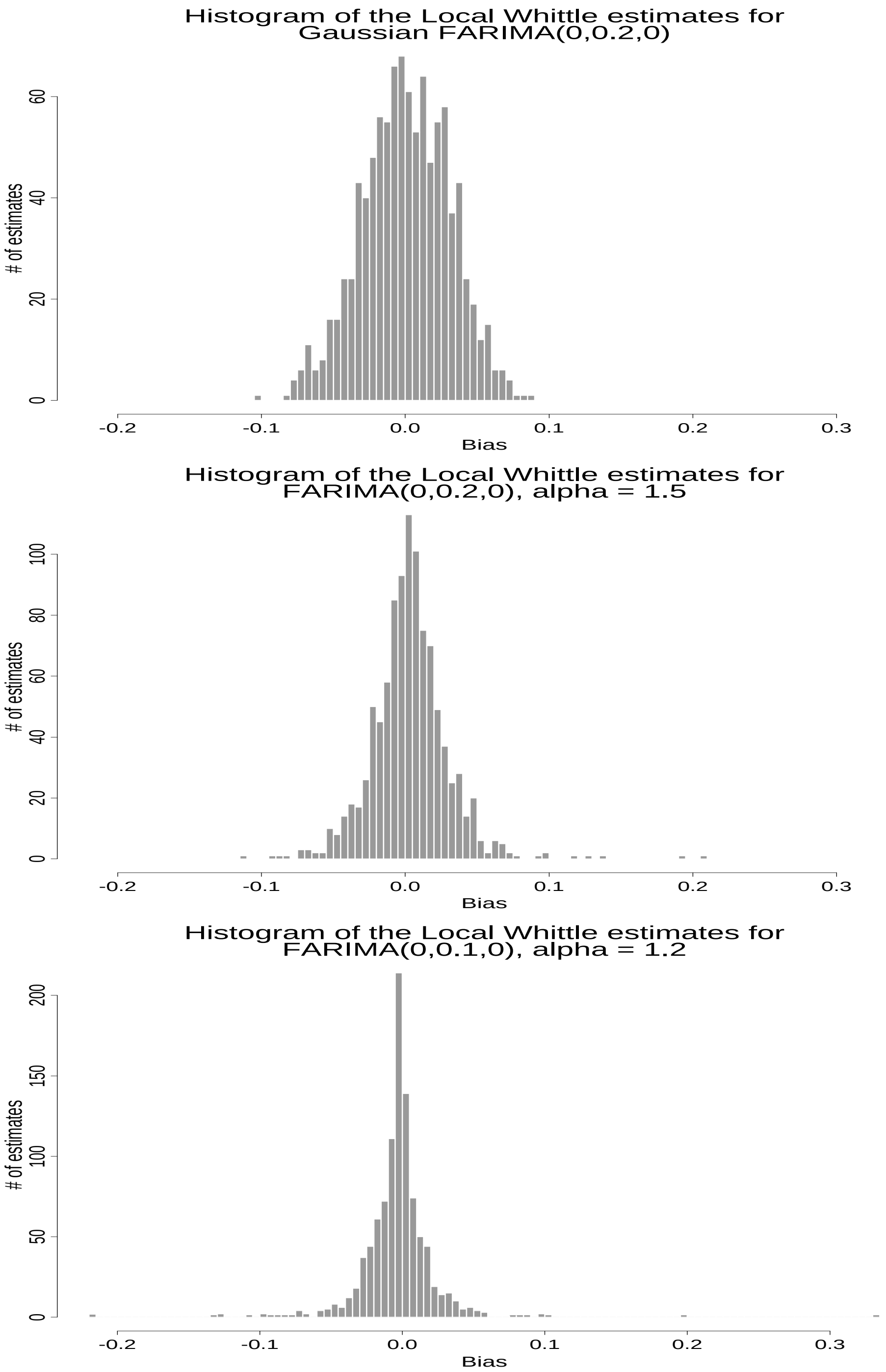


Figure 9: Histograms of the Local Whittle estimates for series with different α's, using $I = 1000$ realizations. The width of each bar is 0.005.

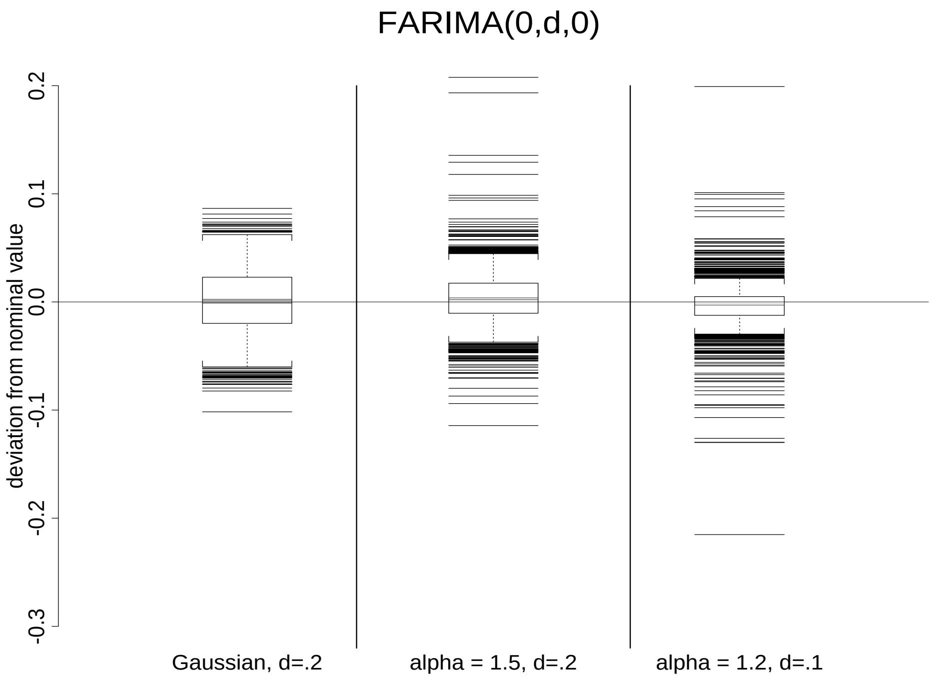


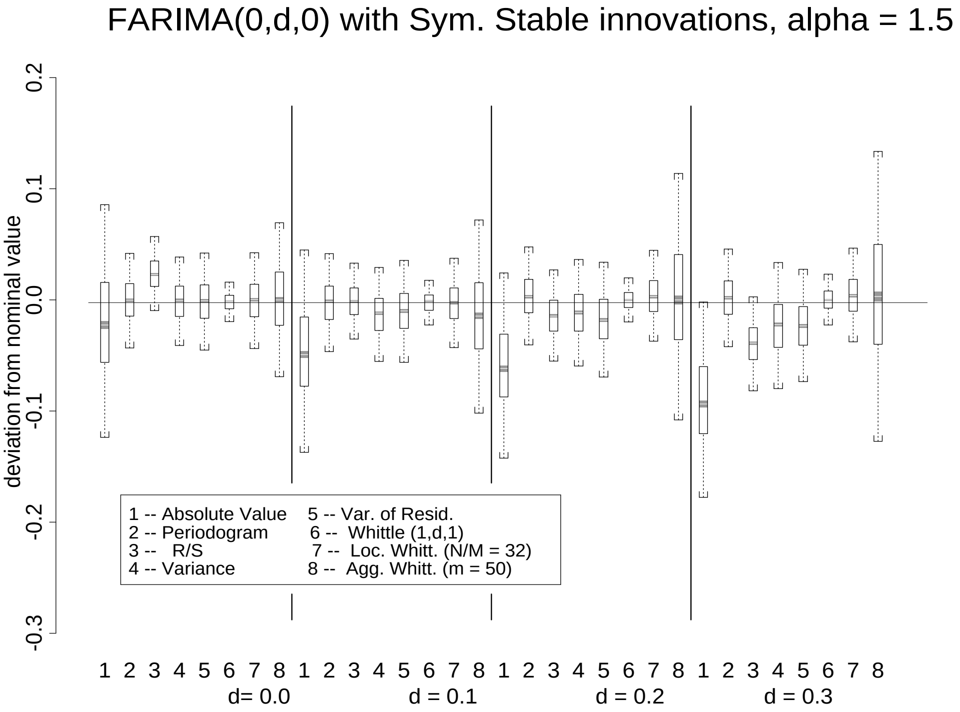
Figure 10: Detailed boxplots for the Local Whittle estimates corresponding to the histogram in Figure 9. Outliers are included.

	Gaussian		$\alpha = 1.5$		$\alpha = 1.2$	
I	BIAS	$\hat{\sigma}$	BIAS	$\hat{\sigma}$	BIAS	$\hat{\sigma}$
50	-.007	.030	-.001	.027	-.005	.020
100	-.001	.032	.002	.027	-.008	.021
200	.000	.031	.002	.025	-.007	.020
300	.002	.031	.003	.024	-.006	.022
400	.002	.030	.004	.024	-.005	.021
500	.002	.031	.004	.026	-.005	.021
600	.002	.030	.004	.026	-.005	.022
700	.002	.030	.004	.027	-.005	.023
800	.001	.030	.004	.027	-.004	.023
900	.001	.030	.003	.027	-.004	.023
1000	.001	.030	.003	.027	-.003	.026

Table 5: Mean and standard deviation of the Local Whittle estimator applied to FARIMA$(0, d, 0)$ series with various α's, using various numbers of independent realizations, I. The d's are as in Figure 10.

	Symmetric Stable Innovations with $\alpha = 1.5$							
	Nominal d							
	FARIMA$(0,d,0)$				FARIMA$(.3,d,.7)$			
	0.0	0.1	0.2	0.3	0.0	0.1	0.2	0.3
	ABSOLUTE							
BIAS	-.015	-.041	-.056	-.089	-.111	-.110	-.100	-.110
$\hat{\sigma}$	.064	.056	.048	.045	.071	.066	.059	.049
$\sqrt{\text{MSE}}$	.066	.070	.073	.100	.132	.129	.116	.120
	PERIODOGRAM							
BIAS	.000	-.003	.003	.003	-.063	-.065	-.062	-.061
$\hat{\sigma}$	.026	.029	.027	.033	.027	.027	.026	.028
$\sqrt{\text{MSE}}$	.026	.029	.027	.033	.069	.071	.067	.067
	R/S							
BIAS	.023	-.001	-.014	-.039	-.041	-.117	-.149	-.156
$\hat{\sigma}$	.017	.018	.021	.021	.017	.015	.020	.023
$\sqrt{\text{MSE}}$	.029	.018	.025	.044	.044	.118	.150	.157
	VARIANCE							
BIAS	-.002	-.014	-.012	-.024	-.087	-.085	-.070	-.063
$\hat{\sigma}$	.024	.025	.028	.038	.037	.030	.029	.032
$\sqrt{\text{MSE}}$	.024	.029	.031	.045	.095	.090	.075	.070
	VARIANCE OF RESIDUALS							
BIAS	-.001	-.010	-.017	-.023	-.241	-.257	-.263	-263
$\hat{\sigma}$	.023	.024	.025	.026	.021	.022	.024	.025
$\sqrt{\text{MSE}}$	.023	.026	.030	.035	.242	.258	.264	.264
	WHITTLE (FARIMA$(1,d,1)$)							
BIAS	-.003	-.004	-.001	.000	.001	-.010	-.004	.000
$\hat{\sigma}$	.021	.022	.023	.026	.063	.026	.030	.029
$\sqrt{\text{MSE}}$	.021	.023	.023	.026	.063	.027	.030	.029
	AGGREGATED WHITTLE $(m = 50)$							
BIAS	.001	-.014	.005	.005	-.030	-.041	-.017	-.008
$\hat{\sigma}$	.045	.049	.060	.067	.053	.053	.052	.062
$\sqrt{\text{MSE}}$	.045	.051	.060	.068	.061	.067	.055	.062
	LOCAL WHITTLE $(N/M = 32)$							
BIAS	-.001	-.003	.003	.004	-.031	-.034	-.027	-.025
$\hat{\sigma}$	.027	.027	.027	.032	.025	.026	.026	.029
$\sqrt{\text{MSE}}$	.027	.027	.027	.032	.040	.042	.037	.038

Table 6: Estimation results for d using 1000 independent copies of FARIMA time series of length $10,000$. The innovations are symmetric Stable with $\alpha = 1.5$.



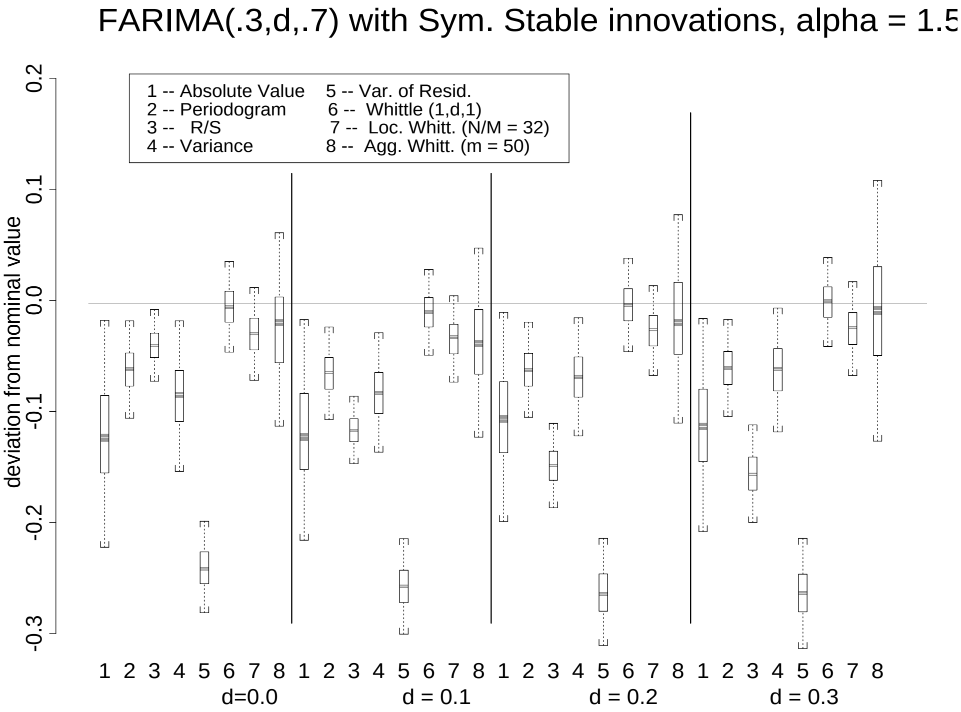
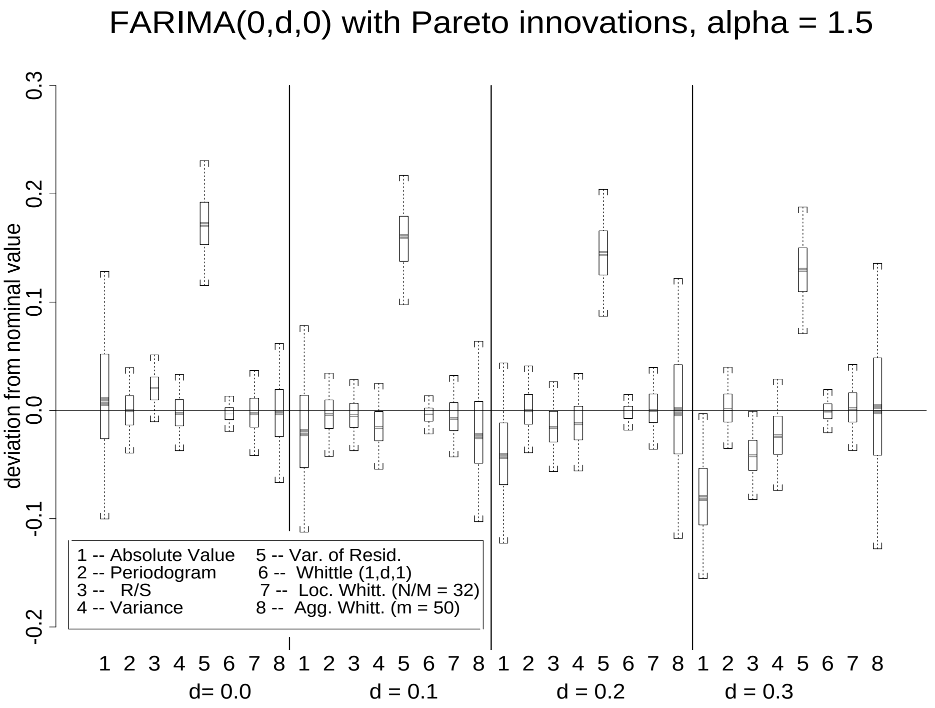


Figure 11: Boxplots for FARIMA$(0, d, 0)$ and FARIMA$(1, d, 1)$ ($\phi_1 = 0.3$, $\theta_1 = 0.7$) series with symmetric Stable innovations, $\alpha = 1.5$.

	Heavy tailed Innovations with $\alpha = 1.5$							
	Nominal d							
	Pareto				Skewed Stable			
	0.0	0.1	0.2	0.3	0.0	0.1	0.2	0.3
	ABSOLUTE							
BIAS	.019	-.014	-.037	-.079	-.026	-.051	-.062	-.094
$\hat{\sigma}$	.067	.055	.046	.040	.065	.059	.051	.045
$\sqrt{\text{MSE}}$	.070	.057	.059	.088	.070	.078	.080	.104
	PERIODOGRAM							
BIAS	.001	-.003	.002	.004	-.001	-.004	.002	.004
$\hat{\sigma}$	.026	.024	.031	.027	.028	.027	.028	.027
$\sqrt{\text{MSE}}$	.026	.024	.031	.028	.028	.027	.028	.027
	R/S							
BIAS	.020	-.005	-.015	-.041	.025	-.001	-.014	-.037
$\hat{\sigma}$	.016	.018	.021	.021	.018	.019	.021	.023
$\sqrt{\text{MSE}}$	.026	.018	.025	.046	.031	.019	.025	.044
	VARIANCE							
BIAS	-.001	-.014	-.011	-.022	-.002	-.015	-.012	-.022
$\hat{\sigma}$	.021	.022	.028	.029	.022	.022	.026	.030
$\sqrt{\text{MSE}}$	.021	.026	.031	.037	.022	.027	.028	.037
	VARIANCE OF RESIDUALS							
BIAS	.172	.160	.145	.131	-.003	-.009	-.019	-.023
$\hat{\sigma}$	.030	.030	.030	.030	.022	.023	.024	.026
$\sqrt{\text{MSE}}$	.175	.163	.148	.134	.022	.026	.030	.034
	WHITTLE (FARIMA$(1,d,1)$)							
BIAS	-.001	-.003	.000	.002	-.002	-.003	.000	.001
$\hat{\sigma}$	.020	.018	.023	.024	.019	.019	.022	.023
$\sqrt{\text{MSE}}$	.020	.018	.023	.024	.020	.020	.022	.023
	AGGREGATED WHITTLE ($m = 50$)							
BIAS	.000	-.017	.004	.006	-.001	-.018	.004	.006
$\hat{\sigma}$	.039	.048	.061	.065	.042	.048	.061	.064
$\sqrt{\text{MSE}}$	.039	.048	.061	.065	.042	.052	.061	.064
	LOCAL WHITTLE ($N/M = 32$)							
BIAS	-.001	-.004	.004	.005	-.001	-.004	.003	.006
$\hat{\sigma}$	.025	.025	.029	.028	.026	.027	.027	.030
$\sqrt{\text{MSE}}$	.025	.026	.030	.029	.026	.027	.028	.031

Table 7: Estimation results for d using 1000 independent copies of FARIMA$(0,d,0)$ time series of length $10,000$. The innovations are Pareto and skewed Stable, with $\alpha = 1.5$.



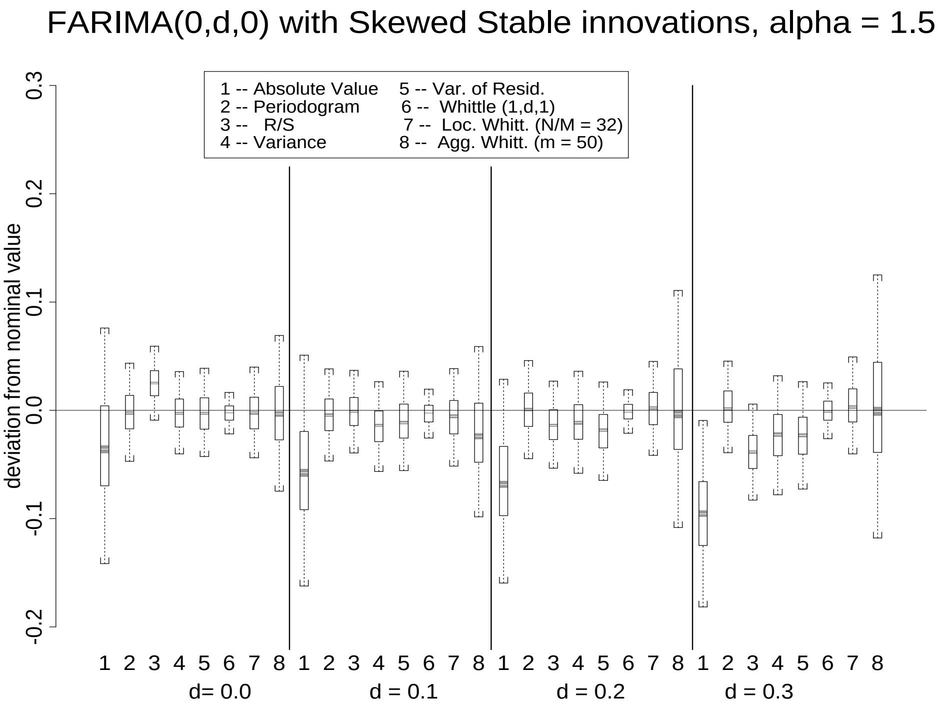


Figure 12: Boxplots for FARIMA$(0, d, 0)$ series with Pareto and skewed Stable innovations, $\alpha = 1.5$.

	Symmetric Stable Innovations with $\alpha = 1.2$ Nominal d			
	FARIMA$(0, d, 0)$		FARIMA$(.3, d, .7)$	
	0.0	0.1	0.0	0.1
	ABSOLUTE			
BIAS	-.058	-.101	-.156	-.163
$\hat{\sigma}$	.074	.059	.079	.071
$\sqrt{\text{MSE}}$	.093	.117	.175	.178
	PERIODOGRAM			
BIAS	.000	-.002	-.063	-.067
$\hat{\sigma}$	.024	.027	.027	.026
$\sqrt{\text{MSE}}$	.024	.027	.069	.072
	R/S			
BIAS	.014	-.008	-.025	-.113
$\hat{\sigma}$	.015	.016	.014	.013
$\sqrt{\text{MSE}}$	.020	.018	.029	.114
	VARIANCE			
BIAS	-.002	-.013	-.080	-.084
$\hat{\sigma}$	.022	.026	.037	.031
$\sqrt{\text{MSE}}$	.022	.029	.088	.089
	VARIANCE OF RESIDUALS			
BIAS	.007	-.003	-.239	-.251
$\hat{\sigma}$	.034	.036	.033	.032
$\sqrt{\text{MSE}}$	.034	.036	.241	.253
	WHITTLE (FARIMA$(1, d, 1)$)			
BIAS	-.003	-.003	.001	-.010
$\hat{\sigma}$	.020	.025	.058	.031
$\sqrt{\text{MSE}}$	.021	.025	.058	.032
	AGGREGATED WHITTLE ($m = 50$)			
BIAS	-.002	-.011	-.022	-.041
$\hat{\sigma}$	.039	.054	.056	.053
$\sqrt{\text{MSE}}$	.039	.055	.061	.067
	LOCAL WHITTLE ($N/M = 32$)			
BIAS	-.001	-.003	-.030	-.034
$\hat{\sigma}$	.024	.026	.025	.024
$\sqrt{\text{MSE}}$	.024	.026	.039	.042

Table 8: Estimation results for d using 1000 independent copies of FARIMA time series of length $10,000$. The innovations are symmetric Stable with $\alpha = 1.2$.

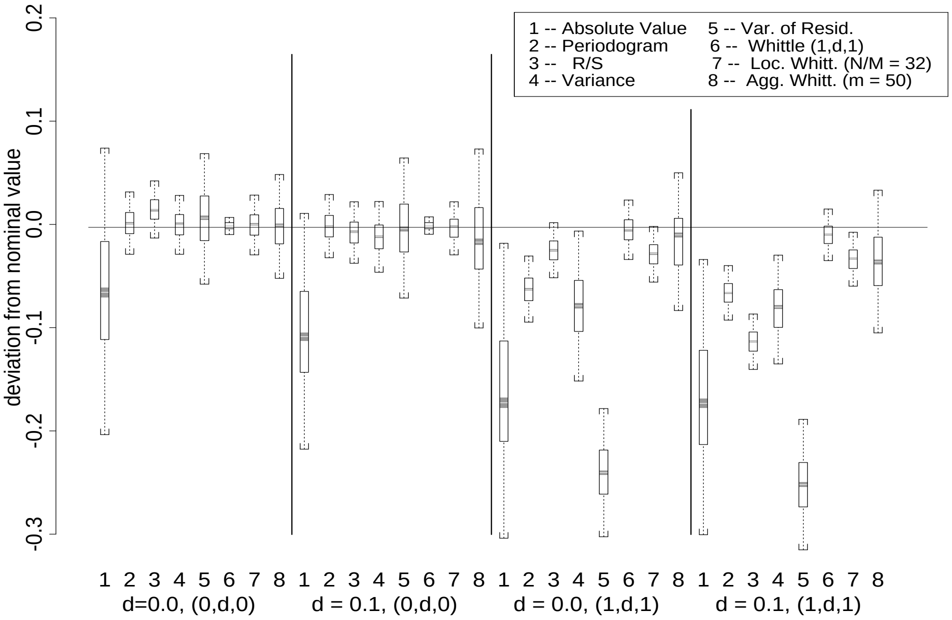


Figure 13: Boxplots for FARIMA$(0, d, 0)$ and FARIMA$(1, d, 1)$ ($\phi_1 = 0.3$, $\theta_1 = 0.7$) series with symmetric Stable innovations, $\alpha = 1.2$.

6. Conclusions

We have studied in this paper the robustness of several estimators. Both finite and infinite variance FARIMA series were considered, using Gaussian, stable, exponential, lognormal and Pareto distributions, as well as different combinations of $p, q = 0, 1$. Since most of the estimators had been developed for Gaussian (or at most finite variance) series, testing whether they still work in the infinite variance case is important. We have sketched the theoretical behavior of some estimators in the infinite variance case, and used simulations to test their robustness. The inclusion of infinite variance processes requires making a sharp distinction between the two parameters H and d which are used almost indistinguishably in the finite variance case. We have found that most of the estimators estimate d (or $d + 1/2$) and not $H = d + 1/\alpha$. It is best, in practice, to use several different methods to estimate long-range dependence, since this can provide a better perspective on the structure of the time series.

There are two main conclusions to be gathered from our simulations. First, many of the estimators are relatively robust with respect to deviations from Gaussian series. There tends to be somewhat more scattering of the estimates in the infinite variance case, but the standard errors and MSEs are not very different. In fact, in many cases the standard errors are smaller in the infinite variance case, as discussed for the Local Whittle estimator.

The exceptions are the Absolute Value and Variance of Residuals methods, which tend to be quite sensitive to the distribution of the innovations. However, we have found that for FARIMA$(0, d, 0)$ series the Variance of Residuals method still works well (for estimating H), while the Absolute Value method loses efficiency, especially for values of α close to 1. Note that these methods estimate $H = d + 1/\alpha$ and not d.

Our second general conclusion is that the estimators are strongly affected by non-zero AR and MA components. Even for the non-extreme values used in this study, the biases of most of the estimators are considerable. The best of the methods, Whittle, only worked because its assigned model was correct (or overspecified). If the model is underspecified, the Whittle estimator becomes more biased than any of the others used here. The Local Whittle and the Periodogram are the two methods which seem the most robust and accurate. One can improve them further by using several different cut-off values, as described in Taqqu and Teverovsky [TT96].

References

[Avr86] F. Avram. *Some limit theorems for stationary sequences with infinite or finite variance.* PhD thesis, Cornell University, Ithaca, New York, 1986.

[Bai96] R. T. Baillie. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5–59, 1996.

[BD91] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, 2nd edition, 1991.

[Ber94] J. Beran. *Statistics for Long-Memory Processes.* Chapman & Hall, New York, 1994.

[FT86] 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.

[GD] C. W. J. Granger and Z. Ding. Stylized facts on the temporal and distributional properties of daily data from speculative markets. Preprint.

[Hig88] T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 31:277–283, 1988.

[HR89] 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.

[JoE96] Fractional differencing and long memory processes, R.T. Baillie and M.L.King, editors. Special issue of the *Journal of Econometrics*, Vol. 73, No. 1, July 1996.

[KM] P. Kokoszka and T. Mikosch. The integrated periodogram for long-memory processes with finite or infinite variance. Preprint 1995. To appear in *Stochastic Processes and their Applications.*

[KM96a] C. Klüppelberg and T. Mikosch. The integrated periodogram for stable processes. *The Annals of Statistics*, 1996. To appear.

[KM96b] C. Klüppelberg and T. Mikosch. Self-normalized and randomly centered spectral estimates. In P. M. Robinson and M. Rosenblatt, editors, *Athens Conference on Applied Probability and Time Series Analysis. Volume II: Time Series Analysis in Memory of E. J. Hannan*, pages 259–271, New York, 1996. Springer-Verlag. Lecture Notes in Statistics, **115**.

[KT96] P. S. Kokoszka and M. S. Taqqu. Parameter estimation for infinite variance fractional ARIMA. To appear in *The Annals of Statistics*, 1996.

[LTWW94] 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.

[Man75] 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.

[MT79] 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.

[MTL78] R. McGill, J. W. Tukey, and W. A. Larsen. Variation of box plots. *The American Statistician*, 32:12–16, 1978.

[MW69] B .B. Mandelbrot and J. R. Wallis. Computer experiments with fractional Gaussian noises, Parts 1,2,3' *Water Resources Research*, 5:228–267, 1969.

[PBS$^+$94] 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.

[Rob95] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23:1630–1661, 1995.

[ST94] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, London, 1994.

[TT95a] M. S. Taqqu and V. Teverovsky. Robustness of Whittle-type estimates for time series with long-range dependence. Preprint, 1995.

[TT95b] 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, 1995.

[TT96] M. S. Taqqu and V. Teverovsky. Semi-parametric graphical estimation techniques for long-memory data. In P. M. Robinson and M. Rosenblatt, editors, *Athens Conference on Applied Probability and Time Series Analysis. Volume II: Time Series Analysis in Memory of E. J. Hannan*, pages 420–432, New York, 1996. Springer-Verlag. Lecture Notes in Statistics, **115**.

[TTW95] 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.

Mathematics Department, Boston University, Boston, MA 02215, USA
Email: murad@math.bu.edu, vt@math.bu.edu