"To appear in "Stochastic Models" in 1997

Robustness of Whittle-type Estimators for Time Series with Long-Range Dependence. *†‡

Murad S. Taqqu
Boston University

Vadim Teverovsky
Boston University

Abstract

We study the robustness of the "standard Whittle", "local Whittle" and "aggregated Whittle" estimators by using a large number of simulated Gaussian time series with long-range dependence. We also consider what happens when the Gaussian innovations are replaced by infinite variance symmetric stable ones. The standard Whittle estimator is a parametric estimator, the local Whittle estimator is a semi-parametric one recently developed by Robinson (1995) and the aggregated Whittle estimator smoothes out the high frequencies. The goal is to estimate H, the intensity of long-range dependence. We investigate the standard deviation and bias of these estimators in order to determine when they are reliable. These estimators are then applied to real-life Ethernet data.

1 Introduction

The recent discovery that long-range dependence is present in high-speed networks has given a new urgency to the development of reliable methods for estimating the intensity of long-range dependence in extremely long time-series (a sample size of 10^5 is not uncommon in Ethernet networks). While many of the available estimators are graphical in nature (Beran (1994), Taqqu, Teverovsky

*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, long memory, semi-parametric, maximum likelihood.

and Willinger (1995)), some involve the minimization of a likelihood-type function. These include the "standard Whittle", "local Whittle" and "aggregated Whittle" estimators. We perform here an extensive Monte Carlo investigation in order to find out how reliable these estimators are. We use fractional Gaussian noise (FGN) and fractional ARIMA (FARIMA) with parameters (p, d, q) where $(p, q) = (0,0), (0,1), (1,0), (1,1)$, $d = 0(.1).4$ and sample sizes up to $10,000$. Our results complement those of Robinson (1995) who applied the local Whittle estimator to FGN series with lengths $N = 64, 128$, and 256. In other work (Taqqu and Teverovsky(1996)), we have used a graphical approach to these estimators by plotting them as a function of a parameter which balances bias versus variance. In that paper, only a FARIMA $(1, d, 1)$ model was considered, with a fixed value of ϕ, θ and d. Here, we use different series with various degrees of long-range dependence in order to investigate how robust the different estimators are. We consider, in addition, infinite variance symmetric stable innovations. The results are presented numerically, in tables and through boxplots. The goal of these studies is to provide a benchmark that will be useful to practitioners.

The standard Whittle method for estimating the parameter H of long-range dependence has been widely used (see Fox and Taqqu (1986), Beran (1994)). It is a parametric method in that it assumes that the spectral density of the series is known with the exception of a few parameters, which are to be estimated. This assumption allows for very precise estimation when the series being examined fits the assumed model exactly. If, on the other hand, the actual series is not of the exact form specified in the model, the parametric estimators may give incorrect results.

To increase robustness, one can use "aggregated Whittle" by summing the series within blocks before applying Whittle. Alternatively, one can use *semi-parametric* estimators. For long-range dependence estimation, the semi-parametric estimators usually assume that the autocorrelations of the series behave as h^{2H-2}, where h is the lag, and H is the long-range dependence parameter. Equivalently, they can assume that the spectral density is proportional to $|\nu|^{1-2H}$ for small frequencies ν. Such an estimator has recently been developed by Robinson (1995). It is obtained through a non-graphical method by minimizing a function similar to that in standard Whittle.

The purpose of this paper is to examine the relative performances of the standard Whittle, henceforth called "Whittle", local Whittle and aggregated Whittle estimators. The Whittle estimator requires that one make an assumption on the underlying model. We will first suppose that it is the correct one and then, in order to investigate robustness, we will suppose that it is misspecified. Sometimes, the misspecification will be relatively mild, for example assuming a FARIMA $(1, d, 1)$ when the underlying model is FARIMA $(0, d, 0)$. In other cases the misspecification will be more serious, for example, assuming FARIMA $(0, d, 0)$ when the underlying model is FARIMA $(1, d, 0)$, FARIMA $(0, d, 1)$ or FARIMA $(1, d, 1)$. The aggregated Whittle estimator, on the other hand, is always used under the assumption that the underlying model is FGN. The local Whittle estimator, being semi-parametric, makes only an assumption on the behaviour of the spectrum at the origin.

H (or d) is related to the exponent of the spectral density at the origin. Its estimation presents problems analogous to the estimation of the parameter α in heavy-tailed distributions. One can also consider long-range dependent series where the innovations are themselves heavy-tailed. We provide some results in this direction in Section 7.

The paper is organized as follows. We introduce the time series in Section 2 and the estimators in Section 3. In Section 4, we describe the experimental set-up and study how the local Whittle estimator performs under ideal circumstances, when the periodogram is replaced by the theoretical spectral density. We conclude the section with a "rule of thumb". The results of our Monte Carlo experiments are stated and discussed in Section 5. The methods are then applied to Ethernet data in Section 6. Section 7 treats an infinite variance case, where the innovations are symmetric stable with $\alpha = 1.5$. Section 8 provides a conclusion.

2 The Time Series

We first describe the FGN and FARIMA series, the two types of time series that we use in the simulations. These are the simplest models that display long-range dependence. The best way to introduce FGN is through its "parent" *fractional Brownian motion* $\{B_H(t),\ t \geq 0\}$. Fractional Brownian motion is a Gaussian process with mean 0, stationary increments, variance $EB_H^2(t) = t^{2H}$ and covariance $EB_H(s)B_H(t) = 2^{-1}\{s^{2H} + t^{2H} - |s-t|^{2H}\}$. The FGN series $\{X_i,\ i \geq 1\}$ is the increment of fractional Brownian motion, namely $X_i = B_H(i+1) - B_H(i),\ i \geq 1$. It is a zero mean, stationary, Gaussian time series whose autocovariance function $\gamma(h) = EX_iX_{i+h}$ is given by $\gamma(h) = 2^{-1}\{(h+1)^{2H} - 2h^{2H} + |h-1|^{2H}\}$, $h \geq 0$. An important point about γ is that it satisfies

$$\gamma(h) \sim H(2H-1)h^{2H-2} \quad \text{as } h \to \infty \tag{2.1}$$

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

$$f(\nu) = C_H\Big(2\sin\frac{\nu}{2}\Big)^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\nu + 2\pi k|^{2H+1}} \sim C_H|\nu|^{1-2H} \quad \text{as } \nu \to 0, \tag{2.2}$$

where C_H is a constant. It blows up at the origin like a power function.

Besides FGN, we also consider FARIMA $(0, d, 0)$. It is defined formally as $X_i = \Delta^{-d}\epsilon_i,\quad i \geq 1$, where the ϵ_i are independent, identically distributed Gaussian random variables with mean 0 and variance 1, and where Δ is the differencing operator $\Delta\epsilon_i = \epsilon_i - \epsilon_{i-1}$. A way to interpret $X_i = \Delta^{-d}\epsilon_i$ with a fractional value of d is as a moving average:

$$X_i = \sum_{j=0}^{\infty} c_j\epsilon_{i-j} \tag{2.3}$$

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

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

where $C_d = \pi^{-1}\Gamma(1-2d)\sin\pi d$. Thus, for large lags h, the autocovariance (2.4) of a Gaussian FARIMA $(0, d, 0)$ process has the same power decay as the autocovariance (2.1) of FGN. Relating the exponents in (2.4) and (2.1) gives

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

The spectral density $f(\nu)$ of FARIMA $(0, d, 0)$ also blows up at the origin like a power function:

$$f(\nu) = (2\sin(\nu/2))^{-2d}/(2\pi) \sim \nu^{-2d}/(2\pi) \text{ as } \nu \to 0.$$

A FARIMA $(0, d, 0)$ is a particular case of FARIMA (p, d, q) processes, a versatile parametric family of models. FARIMA (p, d, q) is defined through the equations

$$X_i - \phi_1 X_{i-1} - \cdots - \phi_p X_{i-p} = \Delta^{-d}\epsilon_i - \theta_1 \Delta^{-d}\epsilon_{i-1} - \cdots - \theta_q \Delta^{-d}\epsilon_{i-q}, \quad (2.6)$$

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 FARIMA (p, d, q) series follow the same type of asymptotic relations for their autocovariance and spectral density as do the FARIMA $(0, d, 0)$ processes, although the actual functions are more complicated, and have additional short-term components. For more information on FARIMA or FGN series, see Brockwell and Davis (1991), Beran (1994) or Samorodnitsky and Taqqu (1994).

The FGN series were simulated using the Durbin-Levinson algorithm, which was implemented in S-Plus using C routines. This algorithm, which is described in Brockwell and Davis (1991), Chapter 8.2, provides an autoregressive representation of the Gaussian time series.

3 The Estimators

We shall introduce here the Whittle, aggregated Whittle and local Whittle estimators. Let X_j, $j = 1, \ldots, N$ be a mean zero time series with spectral density $f(\nu; \eta)$, $-\pi < \nu < \pi$. For *parametric* estimators, such as Whittle's, the function f is assumed known, except for the values of the parameter vector η. The goal is then to estimate η. For *semi-parametric* estimators, such as local Whittle, only the long-range dependence of the time series is assumed, with the corresponding assumptions on the spectral density. In that case, the goal is to estimate the long-range dependence parameter, usually denoted by H or d.

All of the estimators discussed here are based on the periodogram of the time series,

$$I(\nu) = \frac{1}{2\pi N}\Big|\sum_{j=1}^{N} X_j e^{ij\nu}\Big|^2. \tag{3.1}$$

3.1 The Whittle Estimator

The Whittle estimator is the value of the vector η which minimizes the function

$$Q(\eta) := \int_{-\pi}^{\pi} \frac{I(\nu)}{f(\nu;\eta)} d\nu + \int_{-\pi}^{\pi} \log f(\nu;\eta) d\nu. \tag{3.2}$$

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 $f^* = \beta f$, and $\int_{-\pi}^{\pi} \log f^*(\nu;\eta)d\nu = 0$. The first term in (3.2) stays unchanged except for division by β. The parameter β is regarded as an extra parameter, which we do not care to estimate. Since β does not depend on the other parameters, it will not affect the minimization of Q over these parameters. In actual applications, instead of an integral, the corresponding sum over the Fourier frequencies, $\nu_j = 2\pi j/N$ is computed, where $j = 1, 2, \ldots [(N-1)/2]$ and N is the length of the series. Thus, the actual function which the algorithm minimizes is:

$$Q^*(\eta) = \sum_{j=1}^{[(N-1)/2]} \frac{I(\nu_j)}{f^*(\nu_j;\eta)}. \tag{3.3}$$

When dealing with fractional Gaussian noise or FARIMA $(0, d, 0)$, η is the parameter H or d respectively. If the series is assumed to be FARIMA (p, d, q), where p and q are known, then η also includes the unknown coefficients in the autoregressive and moving average parts. For such series the Whittle estimate $\hat{H}$ of H converges to its true value at the rate of $N^{1/2}$ and the asymptotic distribution of $\sqrt{N}(\hat{H} - H)$ is Gaussian. For details see Fox and Taqqu (1986).

3.2 The Aggregated Whittle Estimator

The aggregated Whittle estimator provides an additional technique for obtaining a robust estimate of H. It can be used if the time series is long enough. The idea is to aggregate the data, which creates a new, shorter series

$$X_i^{(m)} := \frac{1}{m} \sum_{j=m(i-1)+1}^{mi} X_j. \tag{3.4}$$

The shorter series increases the standard deviation of the estimator. On the other hand, $m^{1-H}X_i^{(m)}$ converges weakly to fractional Gaussian noise. If the aggregation level m is high enough and long-range dependence is present, then the new series will be close to fractional Gaussian noise. The Whittle estimator with an underlying FGN model can thus be used on the series.

3.3 The Local Whittle Estimator

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,

$$f(\nu) \sim G(H)|\nu|^{1-2H} \qquad \text{as } \nu \to 0. \tag{3.5}$$

The local Whittle estimator is also based on the periodogram (3.1). Its computation involves the introduction of an additional parameter m, which is an integer less than $N/2$, and satisfying, as $N \to \infty$, $1/m + m/N \to 0$. For a spectral density satisfying (3.5), the discrete analogue of the function Q in (3.2) is

$$Q(G,H) := \frac{1}{m}\sum_{j=1}^{m}\left(\frac{I(\nu_j)}{G\nu_j^{1-2H}} + \log G\nu_j^{1-2H}\right) \tag{3.6}$$

where $\nu_j = 2\pi j/N$. Note the two differences with the Whittle estimator: (a) the frequencies are integrated (summed) not up to $2\pi[(N-1)/2]/N$ but only up to $2\pi m/N$; (b) the spectral density f is replaced by its asymptotic form (3.5).

Replacing in (3.6) the constant G by its estimate $\hat{G} = \frac{1}{m}\sum_{j=1}^{m}\frac{I(\nu_j)}{\nu_j^{1-2H}}$, one obtains

$$R(H) := Q(\hat{G},H) - 1 = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{I(\nu_j)}{\nu_j^{1-2H}}\right) - (2H-1)\frac{1}{m}\sum_{j=1}^{m}\log\nu_j. \tag{3.7}$$

The value $\hat{H}$ which minimizes $R(H)$ converges in probability to the actual value H_0, as $N \to \infty$. Robinson (1995) shows that

$$m^{1/2}(\hat{H} - H_0) \to_d \text{Normal } (0, 1/4), \qquad \text{as } N \to \infty. \tag{3.8}$$

This last relation yields asymptotic ("theoretical") values for the standard deviations, confidence intervals, etc.

As is typical in the usual bias-versus-variance trade-off, the choice of m can be quite important. The larger the value of m, the faster $\hat{H}$ converges to H_0 (see (3.8)). On the other hand, if the series is not "ideal", e.g. if it is FARIMA $(1, d, 1)$, then we would want m to be small. In such a case we only want to use the frequencies which are close to zero, since at larger frequencies the short-range behaviour of the series will affect the form of the spectral density.

4 Experimental Set-Up

In this section we present the types of data and the specific estimators (including model specifications and parameter choices) used in our simulation experiments. We also describe the "theoretical" estimates we would expect to see when using the local Whittle estimator on FGN and FARIMA series.

4.1 Data

We use the following types of series for our Monte-Carlo simulation experiments.

FGN		$H = 0.5(0.1)0.9$
FARIMA$(0, d, 0)$		$d = 0.1(0.1)0.4$
FARIMA$(1, d, 0)$	$\phi_1 = 0.5$	$d = 0.0(0.1)0.4$
FARIMA$(0, d, 1)$	$\theta_1 = 0.5$	$d = 0.0(0.1)0.4$
FARIMA$(1, d, 1)$	$\phi_1 = 0.3, \theta_1 = 0.7$	$d = 0.0(0.1)0.4$

The case $H = 0.5$ for FGN or $d = 0$ for FARIMA $(0, d, 0)$ corresponds to i.i.d. $N(0, 1)$ data (white noise). Since the parameter d corresponds to $H - 1/2$, H is $0.5(0.1)0.9$ for all of the FARIMA series except FARIMA $(0, d, 0)$. For consistency, we will use H throughout, even for FARIMA . For each type of series and for each value of H, 50 series, each $10,000$ long were generated. We also used 50 series of length 100, to provide a contrast with the longer time series. Since the theoretical value of the variance of $\hat{H}$ is equal to $1/(4m)$ by (3.8), the theoretical variance of the *average* estimator over $n = 50$ series will be $1/(4mn)$. Therefore, if we take $m = N/32$ (see Section 5) and $N = 10,000$, the theoretical standard deviation of the average will be approximately 0.004. Since this value is quite small, $n = 50$ should be an adequate number of repetitions for detecting the presence of bias, at least for the longer time series.

4.2 Preliminary Results: "Theoretical" Estimates for the Local Whittle Estimator

Even for "ideal" series such as FGN or FARIMA $(0, d, 0)$, the theoretical spectral density is only proportional to ν^{1-2H} as ν tends to 0. Therefore, for any finite length series, where the value of m/N is not negligible, the actual value of $\hat{H}$ may deviate from H_0, even if the periodogram is replaced by the spectral density. We want to examine the magnitude of this "theoretical" bias.

In Figure 1 we display the graphs of $f(\nu;\eta)/G\nu^{1-2H}$ for FGN, FARIMA $(0, d, 0)$, FARIMA $(1, d, 0)$, FARIMA $(0, d, 1)$, and FARIMA $(1, d, 1)$. We chose a value of H equal to 0.9 ($d = 0.4$), and we used for G the theoretical proportionality constant for each type of series. The graph is in log-log coordinates. In Figure 2 we provide an enlarged view for the two "ideal" series, FGN and FARIMA $(0, d, 0)$. The scale of the two figures is quite different. It can be seen that while for small frequencies all of the series have spectral densities fairly close to their theoretical form, the differences increase significantly as the frequency increases. The situation worsens as the series go from FGN and FARIMA $(0, d, 0)$ series, to FARIMA $(0, d, 1)$ and FARIMA $(1, d, 0)$ series, to FARIMA $(1, d, 1)$. As can be seen in Figure 2, even the "ideal" series have spectral densities that are quite different from the reference line at higher frequencies.

In order to see how these differences affect the local Whittle estimator, we replace the periodogram in (3.6) by the theoretical spectral densities of FGN(H) and FARIMA (p, d, q). In this case, no random series needs to be generated in order to estimate the corresponding H. The biases are given in Tables 1 and 2.

As seen in Table 1, for $m = N/4$ and high H there can be a fairly significant bias for both FGN (positive bias) and FARIMA $(0, d, 0)$ (negative bias). If $m = N/16$ or $m = N/32$, there is virtually no theoretical bias in estimating FGN or FARIMA $(0, d, 0)$ with the local Whittle estimator. There is also virtually no difference between the values for $N = 10,000$, $N = 1000$ and $N = 100$.

Table 2 shows the results for FARIMA $(1, d, 0)$, $(0, d, 1)$, and $(1, d, 1)$ series. The biases are now significantly larger. For $N = 10,000$ and $N = 1000$, only for $m = N/32$ are the biases reasonable. For $N = 100$, the theoretical biases

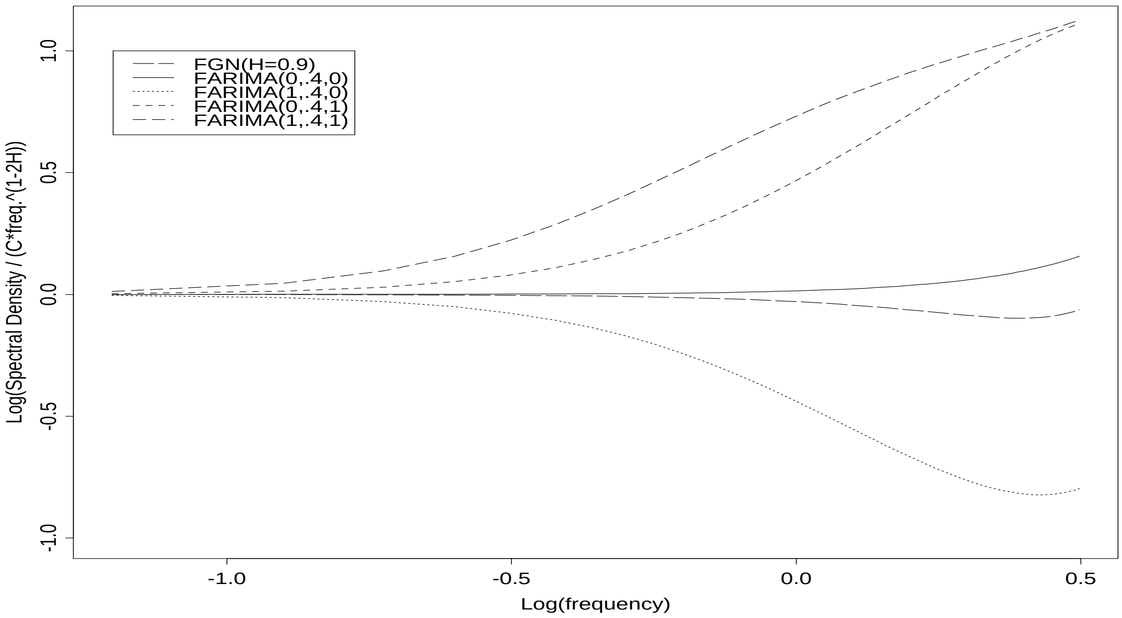


Figure 1: Log-Log plot of the ratios of the spectral densities of the various series to $G\nu^{1-2H}$, for $H = 0.9$, and $N = 10,000$. G is chosen for each series type to give a ratio of 1 at $\nu = 0$.

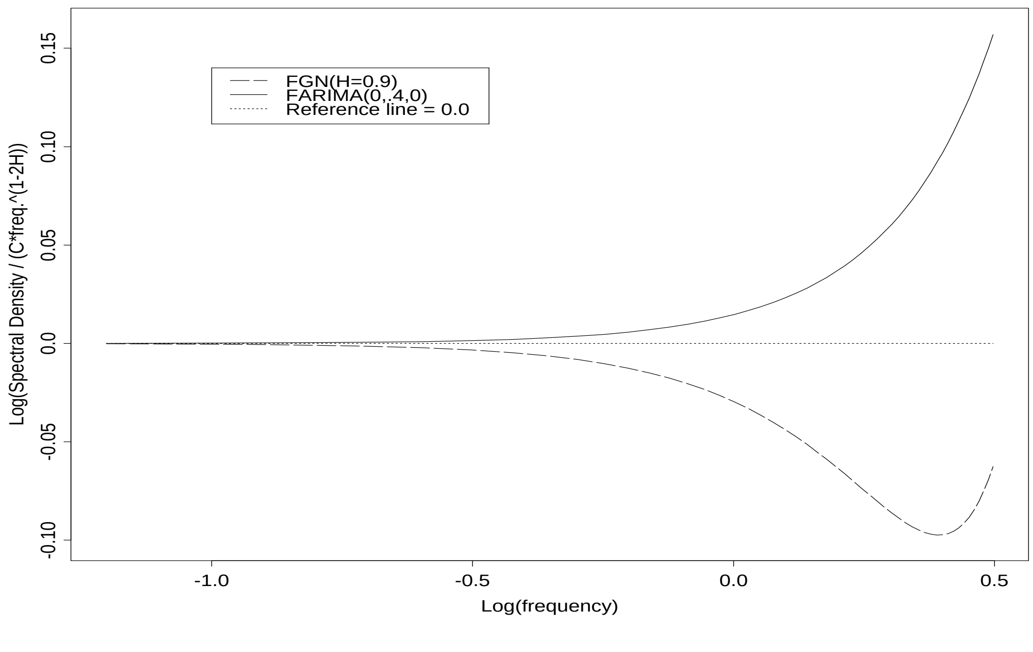


Figure 2: Vertical enlargement of Fig. 1 for FGN and FARIMA $(0, d, 0)$ and inclusion of the reference line, 0.0.

are even greater, because of the small number of frequencies being used in the estimation of H. It seems that for such a short series, there is very little that can be done if the series is a FARIMA (p, d, q), since the biases are quite large, and the standard errors will be quite large as well (larger for smaller m's). For the longer series, with $N \geq 1000$, it seems that $m \leq N/32$ needs to be used to produce a reasonably small bias.

These theoretical biases indicate that, in practice, in order to achieve any degree of accuracy, a relatively small m has to be used. The penalty for a smaller m is a larger standard deviation, so the effects on the bias and on the standard deviation have to be balanced against each other.

The preceding analysis suggests the following **rule of thumb** for simulated series: use Tables 1 and 2 to get a rough idea of the bias and use $\sqrt{1/(4m)}$ (see Relation (3.8)) to get a rough idea of the standard deviation. We shall investigate when this "rule of thumb" is applicable.

5 Experimental Results

We present the results for a given type of time series in the following order: Whittle, local Whittle and aggregated Whittle. The Whittle and local Whittle results are then included in one table while the results for the aggregated Whittle are presented in a separate table. Several types of estimates are produced.

The Whittle estimator was applied to all of the series listed in Section 4.1 using the correctly specified model. To study the effect of misspecification, we also applied it to all the FARIMA series assuming a FARIMA $(0, d, 0)$ model (*underspecified Whittle*) and then to all the series assuming a FARIMA $(1, d, 1)$ model (*overparametrized Whittle*).

Going on to the local Whittle estimator, we used $m = N/4, N/8, N/32$ to compute our estimates for each of the series above.

To investigate the effectiveness of the Whittle aggregation technique, we used all the series of length $N = 10,000$ at levels of aggregation $m = 1, 10, 50,$ and 100, ($m = 1$ is just the original series). For each value of m, we estimated H assuming the underlying model is FGN.

For each kind of data series, we generated 50 independent samples and

Estimation method	Nominal H								
	FGN					FARIMA$(0, d, 0)$			
	.5	.6	.7	.8	.9	.6	.7	.8	.9
$N = 10,000$									
$m = N/4$	.000	.007	.012	.015	.018	-.002	-.005	-.007	-.009
$m = N/8$	.000	.002	.004	.005	.005	-.001	-.001	-.002	-.002
$m = N/16$	.000	.001	.001	.001	.001	.00	.000	.000	-.001
$m = N/32$	.000	.000	.000	.000	.000	.000	.000	.000	.000
$N = 1000$									
$m = N/4$	.000	.008	.013	.016	.019	-.002	-.005	-.007	-.010
$m = N/8$	.000	.003	.004	.005	.006	-.001	-.001	-.002	-.003
$m = N/16$	.000	.001	.001	.001	.002	.000	.000	-.001	-.001
$m = N/32$	.000	.000	.000	.000	.000	.000	.000	.000	.000
$N = 100$									
$m = N/4$	.000	.010	.016	.021	.024	-.003	-.006	-.010	-.013
$m = N/8$	.000	.004	.006	.007	.008	-.001	-.002	-.003	-.004
$m = N/16$	.000	.001	.002	.003	.003	.000	-.001	-.001	-.001
$m = N/32$	.000	.001	.001	.001	.001	.000	.000	.000	.000

Table 1: The local Whittle estimator using the non-random FGN and FARIMA$(0, d, 0)$ spectral densities instead of the periodogram.

reported the average values of the biases of estimates of H, as well as the sample standard deviations (standard errors) and the square roots of the sample MSEs. The following notation is used. If H_i is the estimate for sample i then,

$$\bar{H} = \frac{1}{50}\sum_{i=1}^{50} H_i, \quad \hat{\sigma}^2 = \frac{1}{49}\sum_{i=1}^{50}(H_i - \bar{H})^2, \quad \text{MSE} = \frac{1}{50}\sum_{i=1}^{50}(H_i - H_0)^2, \quad \text{Bias} = \bar{H} - H_0,$$

where H_0 is the nominal value of H.

5.1 FGN and FARIMA $(0, d, 0)$

The top part of Table 3 displays the Whittle and local Whittle results for the FGN and FARIMA $(0, d, 0)$ series of length $N = 10,000$ and Table 4 displays the corresponding aggregated Whittle results.

For FGN, it is clear that a correctly specified Whittle estimator produces

Table 2: The local Whittle estimator using the non-random FARIMA spectral density instead of the periodogram.

Estimation method	Nominal H: FARIMA$(1,d,0)$ $(\phi_1 = 0.5)$					FARIMA$(0,d,1)$ $(\theta_1 = -0.5)$					FARIMA$(1,d,1)$ $(\phi_1 = 0.3, \theta_1 = -0.7)$				
	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
$N = 10,000$															
$m = N/4$	.746	.843	.940	1.037	1.134	.301	.400	.498	.596	.694	.241	.339	.438	.536	.634
$m = N/8$	.602	.701	.801	.900	1.000	.406	.506	.605	.704	.804	.320	.420	.519	.619	.718
$m = N/16$	.532	.632	.732	.832	.932	.468	.568	.668	.768	.868	.414	.514	.614	.714	.813
$m = N/32$	.509	.609	.709	.809	.909	.491	.591	.691	.791	.891	.471	.571	.671	.771	.871
$N = 1000$															
$m = N/4$	.754	.851	.948	1.045	1.142	.284	.382	.479	.577	.675	.216	.314	.412	.510	.608
$m = N/8$	.610	.709	.809	.908	1.007	.396	.495	.595	.694	.793	.298	.397	.497	.596	.696
$m = N/16$	.536	.636	.736	.836	.936	.464	.564	.664	.763	.863	.401	.501	.600	.700	.800
$m = N/32$	.511	.611	.711	.811	.911	.489	.589	.689	.789	.889	.464	.564	.664	.764	.864
$N = 100$															
$m = N/4$	.803	.900	.996	1.093	1.189	.217	.314	.411	.508	.605	.135	.232	.329	.426	.523
$m = N/8$	.649	.748	.847	.946	1.045	.353	.452	.551	.650	.749	.214	.313	.412	.512	.611
$m = N/16$	.563	.663	.762	.862	.962	.437	.537	.636	.736	.836	.330	.430	.529	.629	.729
$m = N/32$	.526	.626	.726	.826	.926	.474	.573	.673	.773	.873	.415	.515	.615	.715	.815

Estimation method and assumed Model		Nominal H								
		True Model FGN					True Model FARIMA$(0,d,0)$			
		.5	.6	.7	.8	.9	.6	.7	.8	.9
$N = 10,000$										
Whittle Correct	Bias	.001	.001	-.001	.000	.000	.001	.000	.000	-.003
	$\hat{\sigma}$	.006	.006	.005	.007	.007	.008	.007	.009	.007
	$\sqrt{\text{MSE}}$	.006	.006	.005	.007	.007	.008	.007	.009	.008
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
Local Whittle $m = N/4$	Bias	.000	.007	.012	.016	.018	-.002	-.003	-.007	-.011
	$\hat{\sigma}$	.009	.010	.009	.012	.010	.011	.009	.010	.012
	$\sqrt{\text{MSE}}$	.009	.012	.015	.020	.021	.011	.009	.012	.016
Local Whittle $m = N/8$	Bias	.001	.002	.005	.002	.005	.001	.000	-.002	-.004
	$\hat{\sigma}$	.015	.014	.016	.015	.014	.014	.016	.013	.016
	$\sqrt{\text{MSE}}$	.015	.014	.016	.015	.015	.014	.016	.013	.016
Local Whittle $m = N/32$	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
$N = 100$										
Whittle Correct	Bias	-.001	.004	-.032	-.038	-.038	-.013	-.039	-.049	-.042
	$\hat{\sigma}$	.073	.062	.067	.076	.050	.089	.095	.090	.091
	$\sqrt{\text{MSE}}$	.073	.062	.074	.085	.062	.089	.102	.102	.099
Whittle FARIMA $(1,d,1)$	Bias	-.511	-.506	-.589	-.655	-.630	-.417	-.480	-.472	-.534
	$\hat{\sigma}$	.494	.485	.416	.406	.432	.532	.452	.418	.473
	$\sqrt{\text{MSE}}$	.707	.697	.719	.769	.761	.672	.656	.627	.710
Local Whittle $m = N/4$	Bias	-.021	.026	-.018	-.001	.037	.024	-.015	-.040	-.022
	$\hat{\sigma}$	.130	.115	.121	.134	.135	.144	.128	.128	.119
	$\sqrt{\text{MSE}}$	.130	.117	.121	.132	.138	.144	.128	.133	.120
Local Whittle $m = N/8$	Bias	-.046	-.010	-.067	-.066	-.020	-.009	-.045	-.009	.026
	$\hat{\sigma}$	.236	.205	.242	.235	.257	.220	.221	.191	.225
	$\sqrt{\text{MSE}}$	.238	.203	.249	.242	.255	.218	.224	.189	.224
Local Whittle $m = N/32$	Bias	-.051	-.184	-.300	-.328	-.068	-.105	.027	-.123	-.110
	$\hat{\sigma}$	.620	.916	.841	.888	.797	.887	.595	.847	.951
	$\sqrt{\text{MSE}}$	.616	.925	.885	.939	.792	.884	.590	.848	.948

Table 3: Estimation results for H using 50 independent copies of time series. FGN and FARIMA $(0,d,0)$ used.

the most accurate estimates of H. On the other hand, the overparametrized Whittle, where the series is assumed to be FARIMA $(1, d, 1)$, produces estimates which are still unbiased, but considerably more scattered, as shown by the standard errors. For FARIMA $(0, d, 0)$, the correctly specified and the overspecified Whittle estimator perform in essentially the same way as for FGN, that is extremely well in the first case and quite well in the second.

Note that in addition to the value of H (or d), the Whittle estimator also estimates the *values of any other parameter* present in the assumed model. When the Whittle estimator assumes a $(1, d, 1)$ series, and the series is actually either FGN or FARIMA $(0, d, 0)$, the model is unidentified and consequently the autoregressive (ϕ_1) and moving average (θ_1) coefficient estimates tend to be quite variable. However, in the cases examined here, the ϕ_1 and the θ_1 coefficients are virtually equal and thus essentially cancel each other. Thus, if an overspecification of this form occurs, it should be quite evident that the series in question is not actually FARIMA $(1, d, 1)$, but either a FGN or a FARIMA $(0, d, 0)$ series.

The *local Whittle estimator* gives good results for FGN and FARIMA $(0, d, 0)$. Consider first FGN. The choice $m = N/4$ gave the smallest value of the standard errors and, for $H \leq 0.7$, the smallest value of the MSEs. The averages of the estimated H's for this estimator can be compared with the values in Table 1, where the periodogram is replaced by the spectral density of FGN. The biases are practically identical. The estimators with $m = N/8$ and $m = N/32$ also seem to behave as expected. The standard errors, and therefore the MSEs are bigger than for $N/4$, and there is a slight bias present for $N/8$, consistent with that in Table 1.

The *local Whittle estimators* for FARIMA $(0, d, 0)$ have essentially the same standard errors as for FGN for all values of m. Also, as for FGN, there is a bias present when the estimator uses $m = N/4$, and H is relatively high. There is again a very good correspondence with the biases that we would expect for a FARIMA $(0, d, 0)$ series (see Table 1). In addition, the sample standard deviations for the local Whittle estimator are quite close to the theoretical values of $\sqrt{1/(4m)}$.

Referring now to Table 4, we examine the performance of the *aggregated Whittle* in the FGN case. FGN is the "ideal" example, in that it should

Estimation method and aggregation level		Nominal H								
		True Model FGN					True Model FARIMA$(0,d,0)$			
		.5	.6	.7	.8	.9	.6	.7	.8	.9
Whittle	Bias	.001	.001	-.001	.000	.000	-.019	-.037	-.053	-.071
$m=1$	$\hat{\sigma}$	.006	.006	.005	.007	.007	.007	.006	.008	.006
	$\sqrt{\text{MSE}}$	.006	.006	.005	.007	.007	.020	.038	.054	.071
Whittle	Bias	-.001	-.001	.003	-.004	-.003	.000	-.002	-.001	-.011
$m=10$	$\hat{\sigma}$	.022	.019	.021	.018	.017	.022	.024	.023	.018
	$\sqrt{\text{MSE}}$	.022	.019	.021	.018	.017	.022	.024	.022	.021
Whittle	Bias	-.007	-.017	-.005	-.002	-.017	-.008	-.005	-.011	-.024
$m=50$	$\hat{\sigma}$	.054	.048	.041	.042	.047	.052	.056	.046	.054
	$\sqrt{\text{MSE}}$	.054	.050	.041	.041	.050	.052	.056	.047	.058
Whittle	Bias	-.022	-.027	-.012	-.019	-.042	-.009	-.022	-.023	-.046
$m=100$	$\hat{\sigma}$	.068	.070	.056	.057	.066	.085	.079	.075	.073
	$\sqrt{\text{MSE}}$	.071	.074	.057	.060	.078	.085	.081	.078	.085

Table 4: Estimation results for H using the aggregated Whittle method on FGN and FARIMA $(0,d,0)$ series. $N = 10,000$.

scale perfectly when aggregated, and produce exact FGN no matter what the aggregation level. Since the aggregated series are shorter, however, we would expect the standard errors of the estimates to increase. In fact, they should theoretically increase by a factor of $\sqrt{m}$. We see that for $m = 1$ the standard errors are approximately 0.006, and that these errors do indeed scale with m in the expected manner. The estimates are biased for the higher values of m, and for high H's. Thus, for $m = 100$, and the nominal $H = 0.9$, the average of the estimates is 0.858. This is not an artifact of the aggregation, but it is due to the relatively short length of the aggregated series (100). In fact, comparing the biases to the biases for FGN in the bottom part of Table 3 (for $N = 100$), we see that they are almost identical.

We now turn to the *aggregated Whittle* estimator applied to FARIMA $(0,d,0)$. We see in Table 4 that the standard errors are comparable to those for FGN, for all values of m and H. But, for $m = 1$, the estimates are very definitely biased, especially for higher H's. This is to be expected since we use Whittle assuming an underlying FGN model. When $m = 10$, the bias seems to have almost disappeared, except a slight one for $H = 0.9$. As we increase m still further, the bias of the estimates returns, but this time, it is there for another reason. It comes about from the relatively short length of the series and is of the same order as for FGN with the same m. This shows that the aggregated FARIMA $(0, d, 0)$ series do indeed behave like FGN, even for m as low as 10. Since the Whittle method is biased for very short FGN series, it is better not to use very high aggregation levels.

The bottom part of Table 3 gives the information for $N = 100$. These are much shorter series, and the level of accuracy certainly suffers as a result. The correctly specified Whittle method has the smallest standard errors (0.05 to 0.10) and the smallest $\sqrt{MSE}$s (0.06 to 0.10). The overspecified Whittle estimator (assuming FARIMA $(1, d, 1)$ series) is very scattered and also biased, unlike the case for $N = 10,000$. The estimates are, in fact, meaningless. This is undoubtedly due to the short length of the series.

For the local Whittle estimator, we can again compute the theoretical standard deviations by using the formula $\sigma = 1/\sqrt{4m}$. We see from Table 3 that the standard errors $\hat{\sigma}$ are considerably larger than the theoretical values (which hold as $N \to \infty$). For such a small N, the "rule of thumb" fails. In fact, only $m = N/4$ can be used to provide an estimate, since in the other two cases, the "confidence intervals" cover virtually the entire range of H.

In conclusion, when $N = 100$, the correctly specified Whittle estimator and the local Whittle estimator with $m = N/4$ may be able to give us some information about H. A choice of $m = N/8$ or $m = N/32$ gives a totally unreliable estimate, even under the "ideal" conditions of FGN or FARIMA $(0, d, 0)$. The aggregated Whittle estimator is not applicable when N is that small.

5.2 FARIMA $(1, d, 1)$

In Table 5 we present the estimates of H for the following series of length $10,000$:

FARIMA$(1, d, 0)$	$\phi_1 = 0.5$	$d = 0.0(0.1)0.4$
FARIMA$(0, d, 1)$	$\theta_1 = 0.5$	$d = 0.0(0.1)0.4$
FARIMA$(1, d, 1)$	$\phi_1 = 0.3, \theta_1 = 0.7$	$d = 0.0(0.1)0.4$

Table 5: FARIMA $(1,d,0)$, $(0,d,1)$, and $(1,d,1)$ series estimated with the Whittle and local Whittle estimators. $N = 10,000$.

Estimation method and assumed Model		Nominal H: True Model FARIMA$(1,d,0)$ $(\phi_1 = .5)$					True Model FARIMA$(0,d,1)$ $(\theta_1 = .5)$					True Model FARIMA$(1,d,1)$ $(\phi_1 = .3, \theta_1 = .7)$				
		.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
Whittle Correct	Average	.498	.597	.695	.795	.901	.494	.598	.688	.798	.896	.496	.596	.693	.791	.894
	$\hat{\sigma}$	.023	.021	.020	.026	.025	.019	.021	.022	.026	.022	.033	.032	.029	.030	.033
	$\sqrt{\text{MSE}}$	.023	.021	.021	.026	.025	.020	.021	.025	.026	.022	.033	.032	.029	.031	.033
Whittle FARIMA $(1,d,1)$	Average	.498	.598	.695	.796	.900	.494	.598	.688	.799	.896	.496	.596	.693	.791	.894
	$\hat{\sigma}$	.023	.022	.021	.026	.026	.020	.021	.023	.026	.022	.033	.032	.029	.030	.033
	$\sqrt{\text{MSE}}$	.023	.022	.021	.026	.026	.021	.021	.026	.026	.022	.033	.032	.029	.031	.033
Whittle FARIMA $(0,d,0)$	Average	.915	1.015	1.118	1.214	1.313	.205	.307	.402	.505	.605	.204	.305	.401	.500	.600
	$\hat{\sigma}$	.010	.009	.009	.009	.009	.013	.014	.013	.014	.014	.010	.012	.013	.012	.013
	$\sqrt{\text{MSE}}$	.415	.416	.418	.414	.414	.296	.293	.297	.295	.296	.297	.295	.299	.301	.300
Local Whittle $m = N/4$	Average	.747	.842	.943	1.037	1.135	.300	.402	.495	.595	.694	.241	.343	.437	.534	.632
	$\hat{\sigma}$	.011	.010	.009	.011	.010	.010	.014	.012	.014	.013	.014	.014	.016	.014	.015
	$\sqrt{\text{MSE}}$	.247	.242	.243	.237	.236	.200	.199	.205	.205	.206	.259	.257	.264	.267	.268
Local Whittle $m = N/8$	Average	.605	.702	.801	.898	1.002	.405	.506	.600	.706	.804	.319	.421	.516	.616	.716
	$\hat{\sigma}$	.015	.014	.014	.015	.017	.013	.015	.015	.017	.015	.016	.017	.018	.019	.018
	$\sqrt{\text{MSE}}$	.106	.104	.102	.099	.103	.096	.095	.101	.096	.097	.182	.180	.185	.185	.185
Local Whittle $m = N/32$	Average	.510	.613	.707	.810	.916	.490	.592	.685	.788	.886	.468	.576	.667	.769	.870
	$\hat{\sigma}$	.036	.027	.025	.034	.031	.025	.030	.031	.032	.032	.031	.032	.030	.030	.032
	$\sqrt{\text{MSE}}$	.037	.030	.026	.035	.035	.027	.030	.034	.033	.035	.044	.039	.044	.043	.044

Table 6: FARIMA $(1,d,0)$, $(0,d,1)$, and $(1,d,1)$ series estimated with the Whittle and local Whittle estimators. $N = 100$.

Estimation method and assumed Model		Nominal H: True Model FARIMA$(1,d,0)$ $(\phi_1 = .5)$					True Model FARIMA$(0,d,1)$ $(\theta_1 = .5)$					True Model FARIMA$(1,d,1)$ $(\phi_1 = .3, \theta_1 = .7)$				
		.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
Whittle Correct	Average	.242	.339	.426	.525	.602	.207	.312	.450	.560	.619	-.020	-.009	.025	.274	.249
	$\hat{\sigma}$	.203	.154	.215	.178	.194	.202	.206	.196	.255	.185	.492	.555	.549	.500	.532
	$\sqrt{\text{MSE}}$	.328	.303	.347	.327	.355	.355	.353	.317	.348	.335	.860	.821	.866	.722	.837
Whittle FARIMA $(1,d,1)$	Average	.000	.080	.278	.313	.369	.087	.172	.367	.452	.560	-.020	-.009	.025	.274	.249
	$\hat{\sigma}$	.290	.260	.325	.265	.246	.454	.460	.389	.488	.341	.492	.555	.549	.500	.532
	$\sqrt{\text{MSE}}$	.576	.580	.530	.553	.585	.611	.625	.509	.595	.479	.860	.821	.866	.722	.837
Whittle FARIMA $(0,d,0)$	Average	.919	1.034	1.113	1.215	1.315	.056	.152	.271	.363	.433	.080	.184	.262	.357	.463
	$\hat{\sigma}$	.083	.101	.101	.104	.086	.097	.102	.082	.115	.103	.107	.102	.104	.106	.083
	$\sqrt{\text{MSE}}$	.427	.445	.425	.427	.423	.455	.460	.437	.452	.478	.433	.428	.450	.455	.445
Local Whittle $m = N/4$	Average	.800	.876	.998	1.093	1.207	.193	.308	.426	.506	.586	.137	.244	.300	.401	.507
	$\hat{\sigma}$	.098	.124	.131	.124	.122	.129	.127	.131	.155	.128	.133	.137	.122	.123	.135
	$\sqrt{\text{MSE}}$	.315	.302	.325	.317	.330	.333	.318	.303	.332	.338	.386	.381	.418	.417	.415
Local Whittle $m = N/8$	Average	.588	.693	.836	.944	1.033	.290	.385	.552	.635	.713	.181	.355	.387	.553	.596
	$\hat{\sigma}$	.200	.217	.198	.225	.233	.202	.222	.211	.234	.236	.220	.211	.209	.205	.255
	$\sqrt{\text{MSE}}$	.217	.234	.239	.265	.266	.290	.308	.256	.284	.299	.386	.322	.375	.319	.395
Local Whittle $m = N/32$	Average	.435	.465	.747	.772	.879	.213	.281	.665	.742	.483	.329	.371	.533	.706	.556
	$\hat{\sigma}$	.752	.647	.834	.792	.794	.879	.832	.728	.834	.733	.818	.798	.733	.807	.709
	$\sqrt{\text{MSE}}$	.747	.655	.827	.784	.785	.916	.883	.722	.828	.837	.828	.823	.744	.805	.782

The correctly specified *Whittle estimator* again produces good results, although increase from approximately .007 to $.020 - .033$ for the various series in Table 5. This seems due to the increased number of parameters which are used. In fact, overspecifying the model for the Whittle estimator, and using a FARIMA $(1, d, 1)$ for all of the series in Table 5 does not appreciably change the results. Both the averages of the estimates and their standard errors are virtually unchanged when using the FARIMA $(1, d, 1)$ model on the FARIMA $(1, d, 0)$ and FARIMA $(0, d, 1)$ series.

The underspecified Whittle estimator, i.e. using a FARIMA $(0, d, 0)$ model, does not give good results. As seen in Table 5, for a series which is not FARIMA $(0, d, 0)$ the estimates are very biased, although quite consistent ($\hat{\sigma}$ is very small). This consistency, however, can be very misleading since this is a non-graphical method and there may be no indication of the large bias.

When the *local Whittle estimator* with $m = N/4$ is used, problems similar to those using an underspecified Whittle estimator occur. The estimates are again quite consistent, with small standard errors (similar to those for FGN and FARIMA $(0, d, 0)$), but again very biased. The biases are almost as large as for Whittle (at least 0.2), and certainly large enough to make the estimates useless. When $m = N/8$, the results are quite similar. The standard errors are somewhat larger (by about $\sqrt{2}$, as the theory predicts), and the biases are somewhat smaller, but still very significant (at least 0.095). These biases are consistent with the theoretical ones of Table 2.

When $m = N/32$, the results are quite different. As can be seen from Table 5, the standard errors $\hat{\sigma}$ are approximately the same as for the same estimator used on the FGN and FARIMA $(0, d, 0)$ series. The 95% confidence intervals for the averages of the estimates are approximately equal to $2\hat{\sigma}\sqrt{50} = \pm .008$. The estimates are in fact slightly biased for the FARIMA $(1, d, 0)$ and FARIMA $(0, d, 1)$ cases, and somewhat more strongly biased for the FARIMA $(1, d, 1)$ case. But even in the latter case, the bias seems no larger than approximately 0.03. Again, the theoretical results of Table 2 are borne out. It seems clear that while $m = N/4$ or $N/8$ works rather well when dealing with series such as FGN or FARIMA $(0, d, 0)$, the estimator with $m = N/32$ is the better one if the series has any sort of additional short-range structure. A bias of even 0.03 would be deemed acceptable for most applications. In conclusion, the "rule of thumb" works, and simulations for $N = 1000$ suggest that it applies there as well.

Table 7 contains the results of the *aggregated Whittle* estimates for FARIMA $(1, d, 0)$, FARIMA $(0, d, 1)$, and FARIMA $(1, d, 1)$ series. We see that for $m = 1$, the estimates are very biased, because the series, while exhibiting long-range dependence, are not very close to FGN. Note that in the FGN model, H can not go beyond the stationary interval $(0, 1)$. The estimation program restricts H to the interval $[0.001, 0.999]$. (For the FARIMA $(1, d, 0)$ series, where the estimates are biased upward, the limit of 0.999 is reached for the higher H's, thus resulting in very low standard errors.) Even for $m = 10$, the biases are still quite significant, although now much smaller than for $m = 1$. When we get to $m = 50$, we see that the averages of the estimates are now quite close to the averages of the estimates for FGN. Thus, most of the bias which is now present in the estimates is not due to the fact that the series are not FGN, but to the fact that their length is now only 200. When $m = 100$, the length of the series is down to 100, and the estimates are as biased as the FGN estimates. Overall, for $N = 10,000$ and the aggregation levels discussed here, $m = 50$ seems to give the best results for the FARIMA series, as indicated by the corresponding $\sqrt{MSE}$s.

If the *local Whittle estimator* with $m = N/32$ is compared with the *Whittle estimator* using the FARIMA $(1, d, 1)$ model, two conclusions can be drawn. First, the *Whittle estimator* is, on average, less biased and more consistent. This holds for all of the series examined here, from FGN to the various FARIMA series. Second, however, the difference in accuracy is not as great for the various FARIMA series as might be expected. The $\sqrt{MSE}$s are only about 0.01 larger for the *local Whittle estimator*. Considering the parametric nature of the Whittle method, and the semi-parametric nature of the local Whittle method, this difference is not very large.

If we now compare the *aggregated Whittle* with $m = 50$ in Table 7 to the *local Whittle* with $m = N/32$ in Table 6 (note that m is not the same parameter in the different estimators), we see that their performances are fairly similar, with the *local Whittle estimator* having a slight edge. The advantage is slight enough that it may be overcome if a different m is used for the aggregated Whittle.

Table 7: FARIMA $(1, d, 0)$, $(0, d, 1)$, and $(1, d, 1)$ series estimated with the aggregated Whittle estimators, $N = 10,000$.

Estimation method and aggregation level		Nominal H: True Model FARIMA$(1,d,0)$ $(\phi_1 = .5)$					True Model FARIMA$(0,d,1)$ $(\theta_1 = .5)$					True Model FARIMA$(1,d,1)$ $(\phi_1 = .3, \theta_1 = .7)$				
		.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
Whittle $m = 1$	Bias	.346	.327	.296	.199	.099	-.251	-.273	-.296	-.309	-.323	-.244	-.268	-.290	-.308	-.321
	$\hat{\sigma}$	.008	.007	.002	.000	.000	.007	.010	.010	.012	.013	.007	.008	.010	.010	.012
	$\sqrt{\text{MSE}}$	.346	.327	.296	.199	.099	.251	.274	.296	.310	.323	.244	.268	.290	.308	.321
Whittle $m = 10$	Bias	.046	.038	.033	.030	.025	-.080	-.070	-.073	-.057	-.054	-.123	-.118	-.112	-.108	-.100
	$\hat{\sigma}$	.021	.018	.019	.024	.021	.017	.021	.023	.024	.022	.019	.018	.021	.023	.024
	$\sqrt{\text{MSE}}$	.050	.042	.038	.038	.032	.082	.073	.077	.062	.059	.125	.119	.114	.110	.103
Whittle $m = 50$	Bias	.001	.007	.005	-.009	-.021	-.032	-.007	-.024	-.024	-.027	-.040	-.024	-.015	-.026	-.031
	$\hat{\sigma}$	.045	.057	.044	.050	.046	.034	.054	.044	.048	.047	.046	.047	.048	.050	.046
	$\sqrt{\text{MSE}}$	.044	.057	.044	.051	.050	.047	.054	.049	.053	.053	.061	.053	.050	.056	.055
Whittle $m = 100$	Bias	-.009	.005	-.016	-.033	-.052	-.016	-.006	-.015	-.022	-.049	-.018	-.017	-.011	-.031	-.050
	$\hat{\sigma}$	.066	.075	.073	.074	.060	.063	.082	.067	.070	.059	.076	.075	.064	.068	.054
	$\sqrt{\text{MSE}}$	.066	.074	.074	.080	.079	.064	.081	.068	.073	.077	.076	.075	.064	.074	.073

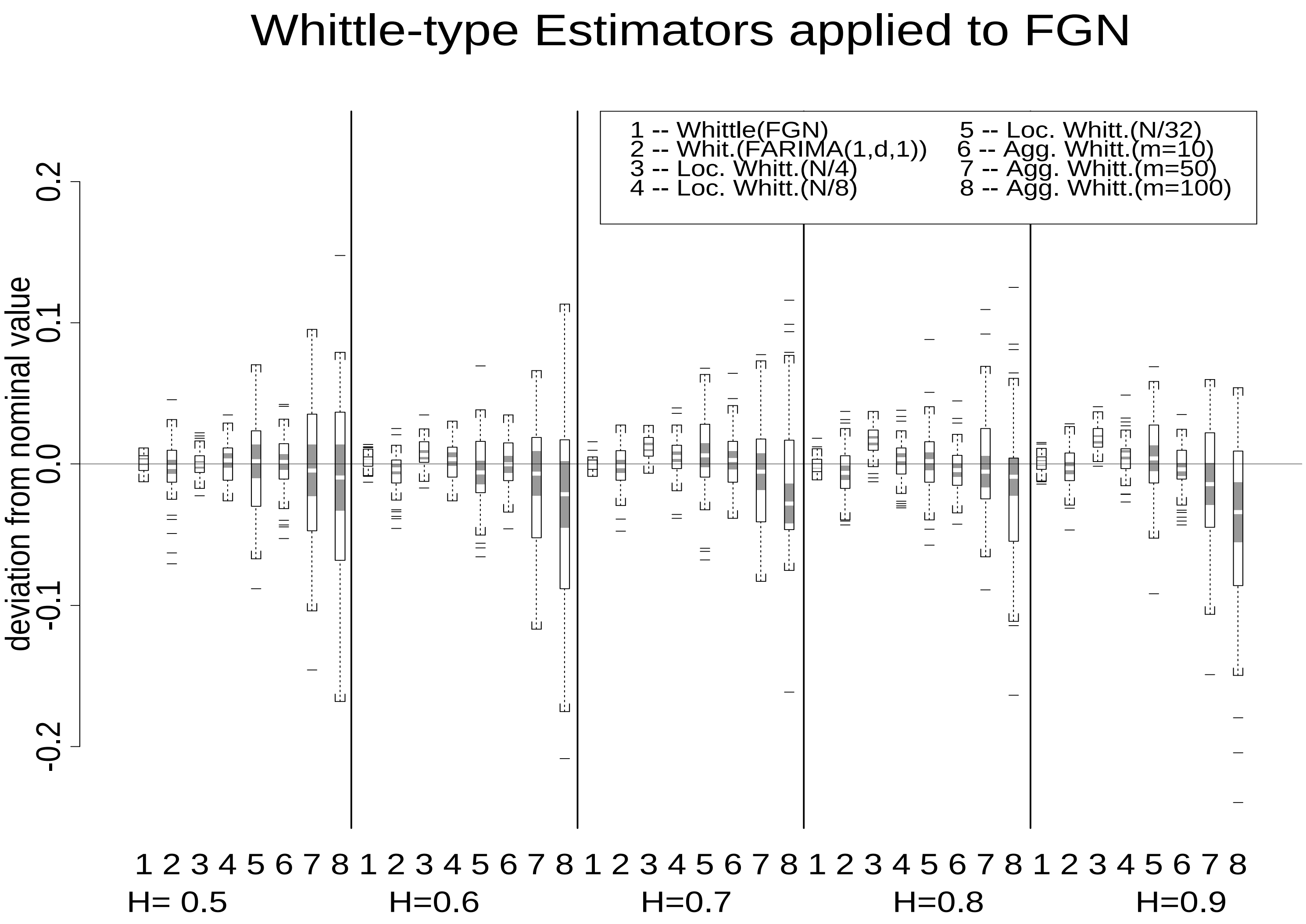
It should be noted that Whittle estimates the ϕ_1 and θ_1 coefficients wherever they are present. In general, the estimates of these coefficients have approximately the same standard deviations and MSEs as the estimates of H. This is true even when the model is overspecified, as when the actual series is FARIMA $(1, d, 0)$ or FARIMA $(0, d, 1)$ and the assumed model is FARIMA $(1, d, 1)$. In those cases, the θ_1 and the ϕ_1 coefficients are estimated, respectively, to be zero. The estimates possess approximately the same degree of accuracy as the estimates of H.

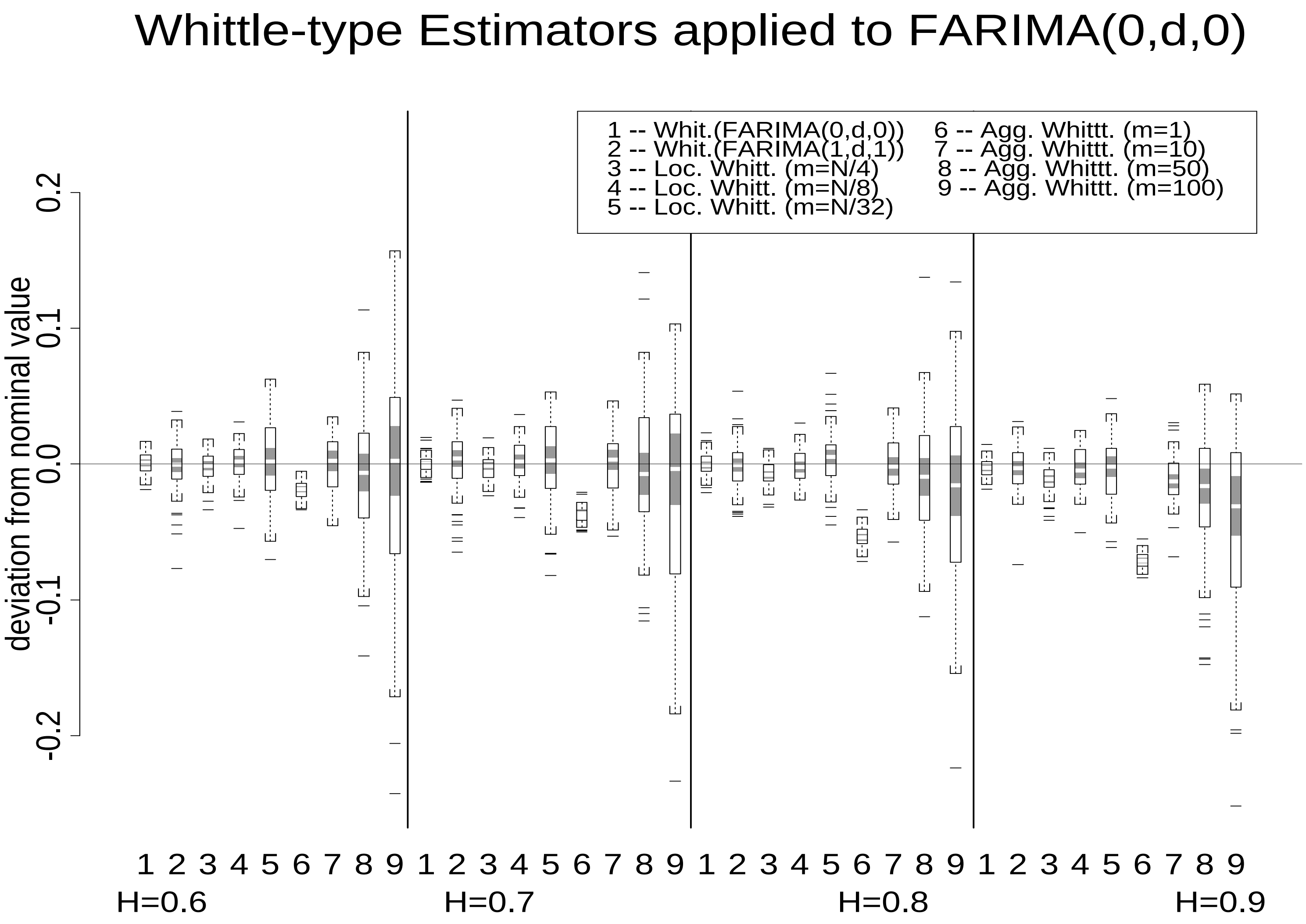
Table 6 is for $N = 100$. As in the bottom part of Table 3, the estimates produced here are very scattered, and, in addition, quite biased. Even the correctly specified *Whittle estimator*, which would be expected to give the best results has $\sqrt{MSE} \geq 0.3$. The misspecified Whittle estimators are even worse. The *local Whittle estimators*, for various m's, are just as bad. The MSEs are again very large and any confidence intervals will cover essentially the whole range of H. We see that none of the estimators discussed here can determine H with any reasonable accuracy when the series length is so small.

The box plots are the graphical representation of the results of the top part of Table 3, Table 4, Table 5 and Table 7. The vertical axis in these plots shows the deviations from the nominal values of H. For each method we have – (1) a box representing the middle 50% of the data. That box contains a) a shaded region representing an approximate 95% confidence interval about the median (see McGill, Tukey and Larsen (1978)); 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.

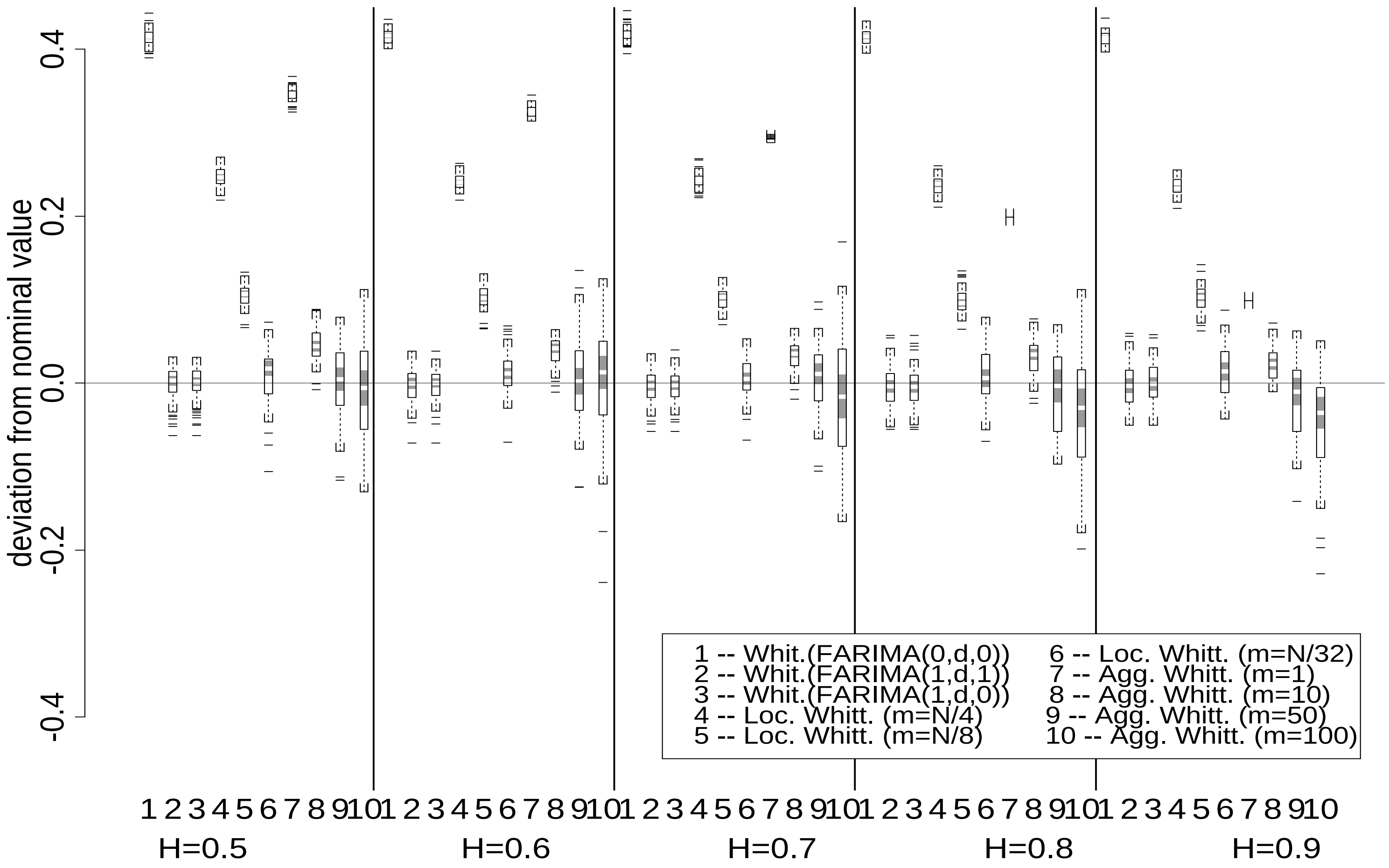
6 Application to Ethernet Data

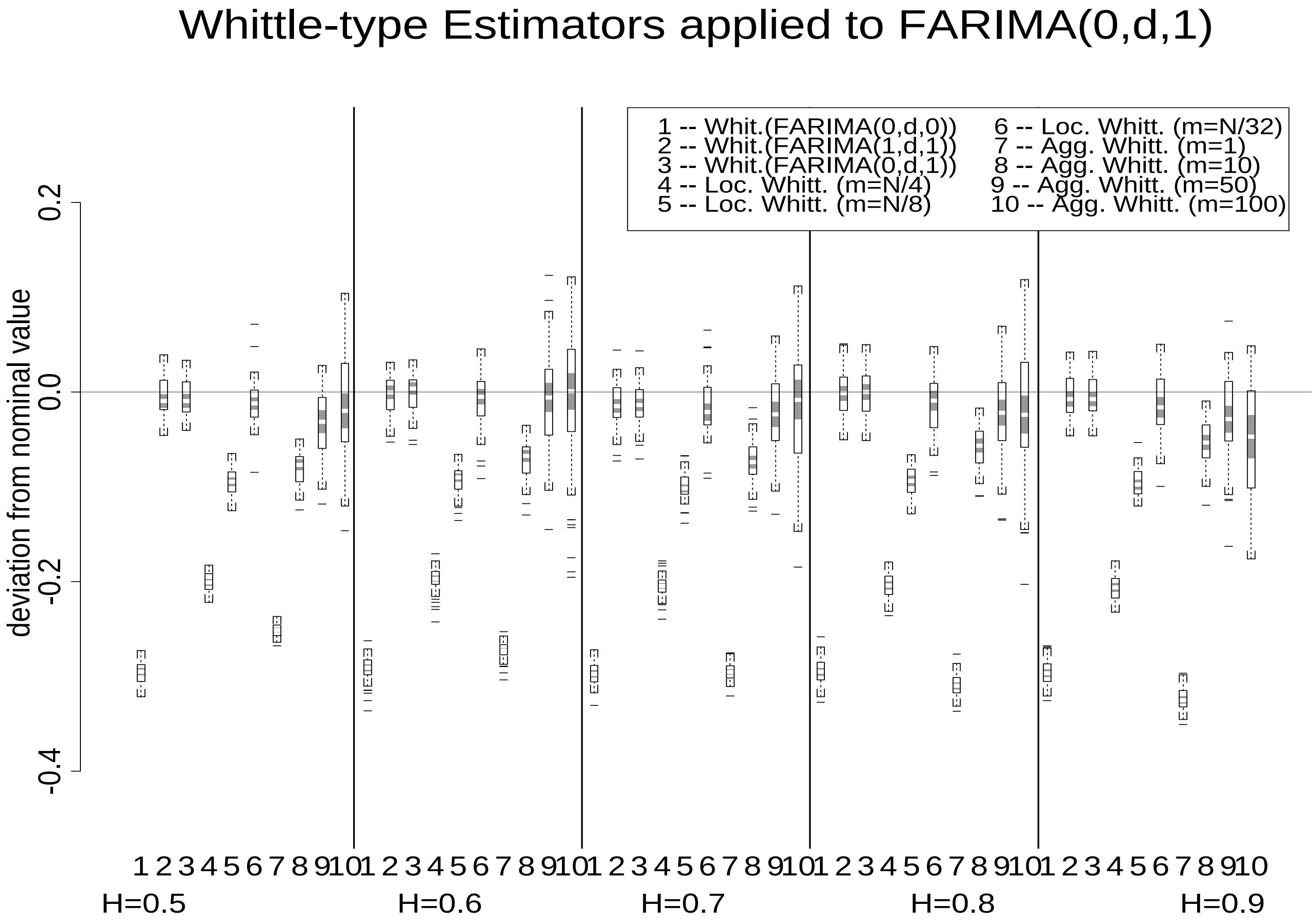
We now apply the Whittle and local Whittle estimators to real time series, obtained from a study of Ethernet traffic performed at Bellcore. Series of this type were investigated by Leland, Taqqu, Willinger and Wilson ((1994), (1995)). The first Ethernet data set gives the number of bytes per 10 milliseconds passing through a monitoring system during a "normal traffic hour" in August 1989. The second gives the number of packets sent per 10 millisec-


Whittle-type Estimators applied to FGN
1 -- Whittle(FGN)
2 -- Whit.(FARIMA(1,d,1))
3 -- Loc. Whitt.(N/4)
4 -- Loc. Whitt.(N/8)
5 -- Loc. Whitt.(N/32)
6 -- Agg. Whitt.(m=10)
7 -- Agg. Whitt.(m=50)
8 -- Agg. Whitt.(m=100)
deviation from nominal value
0.2
0.1
0.0
-0.1
-0.2
1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
H= 0.5
H=0.6
H=0.7
H=0.8
H=0.9

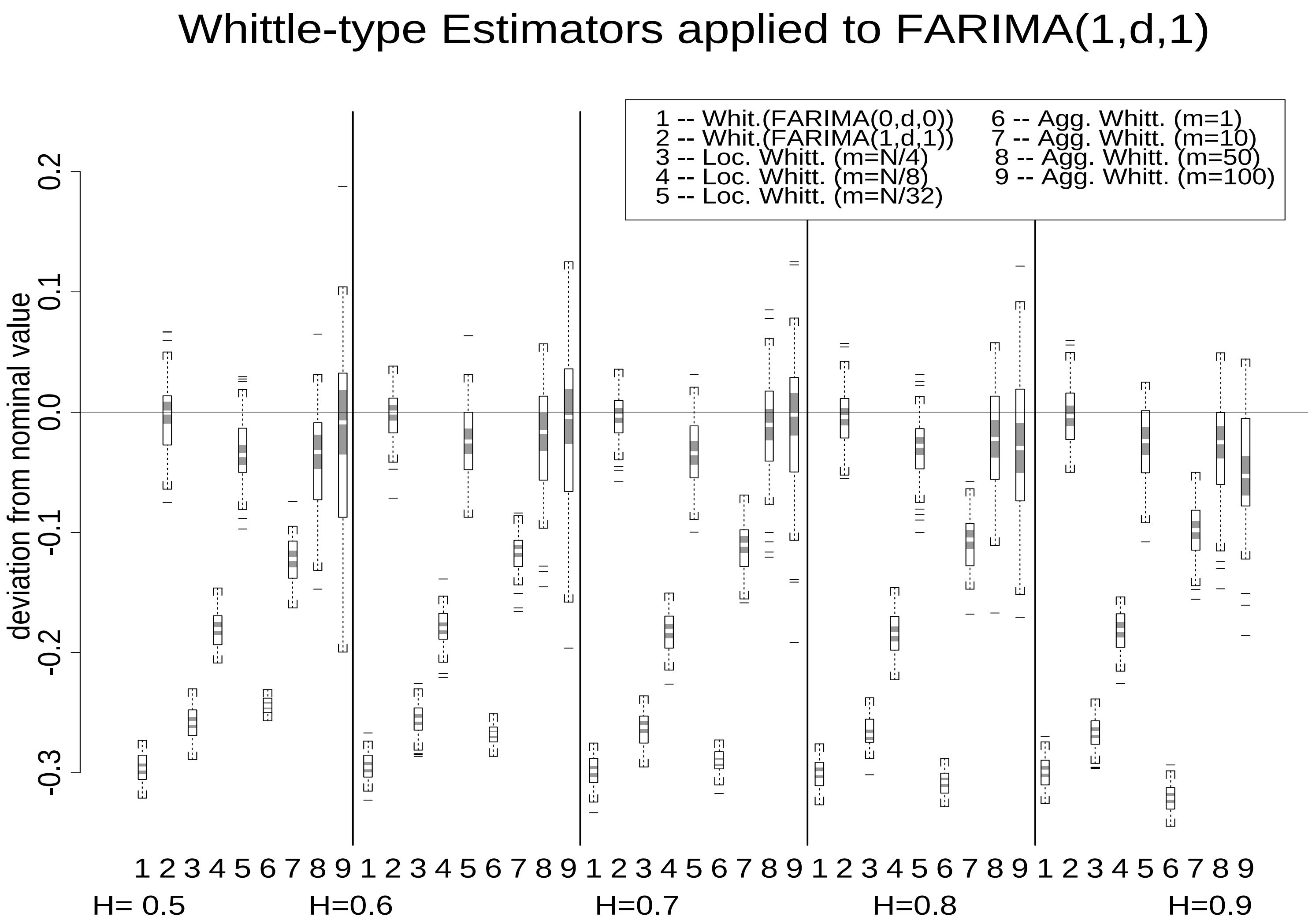

Whittle-type Estimators applied to FARIMA(0,d,0)
1 -- Whit.(FARIMA(0,d,0))
2 -- Whit.(FARIMA(1,d,1))
3 -- Loc. Whitt. (m=N/4)
4 -- Loc. Whitt. (m=N/8)
5 -- Loc. Whitt. (m=N/32)
6 -- Agg. Whittt. (m=1)
7 -- Agg. Whittt. (m=10)
8 -- Agg. Whittt. (m=50)
9 -- Agg. Whittt. (m=100)
deviation from nominal value
0.2
0.1
0.0
-0.1
-0.2
1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9
H=0.6
H=0.7
H=0.8
H=0.9

Whittle-type Estimators applied to FARIMA(1,d,0)





Whittle-type Estimators applied to FARIMA(0,d,1)
1 -- Whit.(FARIMA(0,d,0))
2 -- Whit.(FARIMA(1,d,1))
3 -- Whit.(FARIMA(0,d,1))
4 -- Loc. Whitt. (m=N/4)
5 -- Loc. Whitt. (m=N/8)
6 -- Loc. Whitt. (m=N/32)
7 -- Agg. Whitt. (m=1)
8 -- Agg. Whitt. (m=10)
9 -- Agg. Whitt. (m=50)
10 -- Agg. Whitt. (m=100)
deviation from nominal value
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7 8 9 10
H=0.5
H=0.6
H=0.7
H=0.8
H=0.9


Whittle-type Estimators applied to FARIMA(1,d,1)
1 -- Whit.(FARIMA(0,d,0))
2 -- Whit.(FARIMA(1,d,1))
3 -- Loc. Whitt. (m=N/4)
4 -- Loc. Whitt. (m=N/8)
5 -- Loc. Whitt. (m=N/32)
6 -- Agg. Whitt. (m=1)
7 -- Agg. Whitt. (m=10)
8 -- Agg. Whitt. (m=50)
9 -- Agg. Whitt. (m=100)
deviation from nominal value
0.2
0.1
0.0
-0.1
-0.2
-0.3
1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9
H= 0.5
H=0.6
H=0.7
H=0.8
H=0.9

Estimation Method and Model Specification	Byte series	Packet Series
Whittle(FGN)	.774	.830
Whittle(FARIMA$(0, d, 0)$)	.818	.889
Whittle(FARIMA$(1, d, 1)$)	.750	.853
Local Whittle ($m = N/4$)	.742	.844
Local Whittle ($m = N/8$)	.801	.898
Local Whittle ($m = N/16$)	.835	.970
Local Whittle ($m = N/32$)	.834	.885
Local Whittle ($m = N/64$)	.856	.748
Local Whittle ($m = N/128$)	.836	.932
Local Whittle ($m = N/256$)	.810	.922

Table 8: Estimation results for H for the Ethernet series.

onds during a "normal traffic hour" in August 1989. Thus, each of the series is $360,000$ long. The conclusions reached by Leland, Taqqu, Willinger and Wilson ((1994), (1995)) are that H is significantly different from 1/2 and that long-range dependence is present. The authors obtain estimates of approximately $H = 0.8$ for the *byte* data, and $H = 0.9$ for the *packet* data, by using several different methods.

One of the characteristics of real data series, is, of course, that the true model is not known. That model is very unlikely to be exactly FGN or FARIMA. Nevertheless, since the Whittle estimator requires a choice of a model, we have tested the two data series using FGN, FARIMA $(0, d, 0)$, and FARIMA $(1, d, 1)$ models. The results are presented in Table 8. We note that the estimates vary significantly depending on the model chosen, and in fact, there is no reason to assume that the correct model is one of the three examined here. It seems, nevertheless, likely that the *byte* series has an H of at least 0.75, and the *packet* series has an H of at least 0.8, but it is quite difficult to conclude anything more precise. This is an inherent limitation of *parametric* estimators: unless the parametric form is precisely known, conclusions can not easily be drawn.

We also used the local Whittle estimator with different values of m. The extreme length of the series allowed us to choose $m = N/4$, $N/8$, $N/16$, $N/32$,

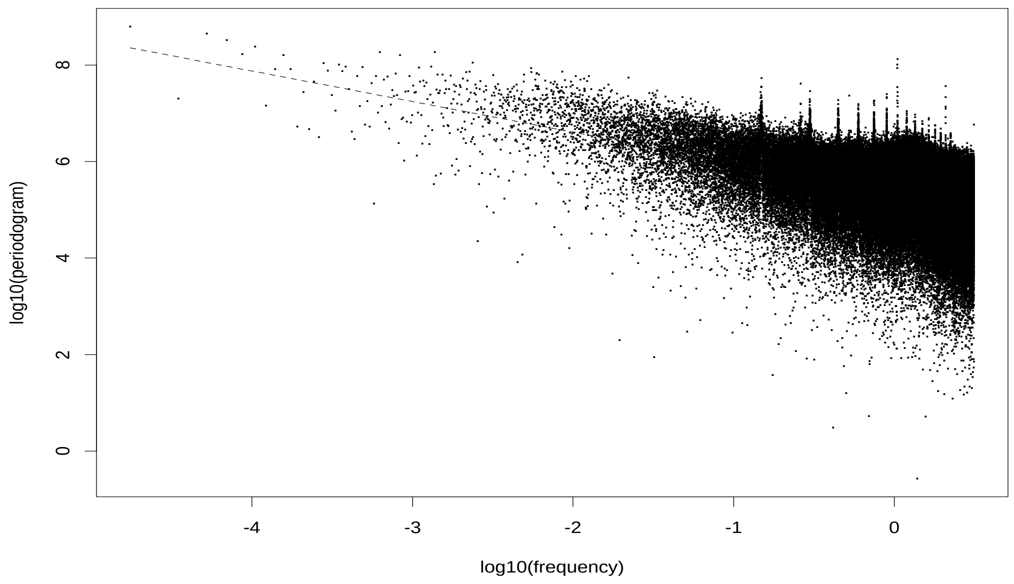


Figure 3: Log-Log plot of the periodogram of the Ethernet byte data.

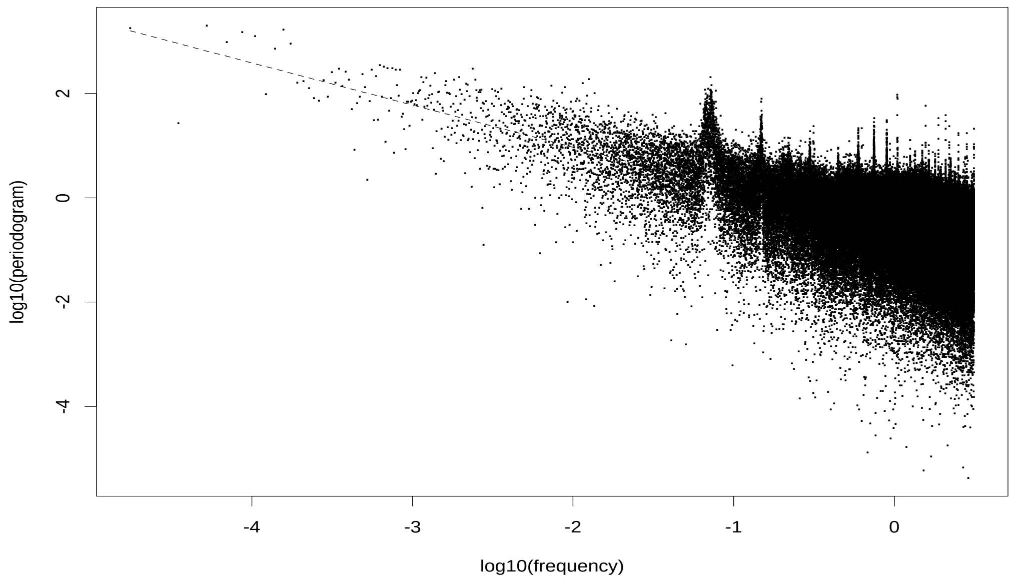


Figure 4: Log-Log plot of the periodogram of the Ethenet packet data.

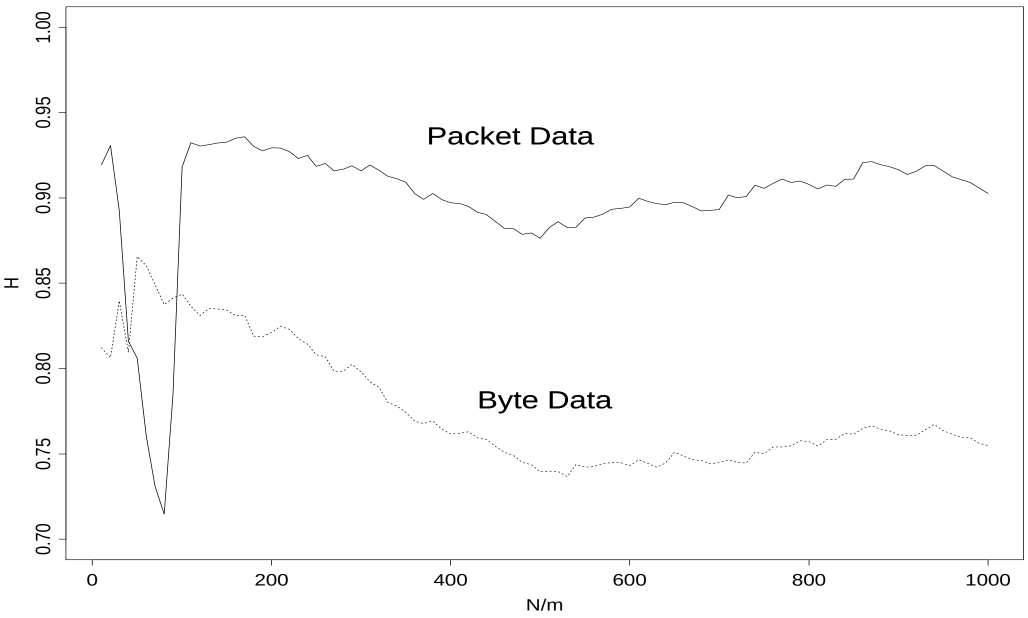


Figure 5: Log-Log plot of the local Whittle estimator for the Ethernet series

$N/64$, $N/128$, and $N/256$. Note that using $N/256$ leaves still over 1400 frequencies available for the estimation. In Taqqu and Teverovsky (1996), we plotted the local Whittle estimator in function of $N/m = 10(10)1000$, with the goal of finding a stable m-region. We reproduce this plot here (Figure 5), but as we explain below, in the Ethernet context, it is important to examine also the periodogram, because of the presence of periodicities.

The bottom part of Table 8 shows the results of using the local Whittle estimator on the Ethernet data sets. Focusing first on the *byte* series, we see that the periodogram has several very sharp peaks at various frequencies. Since we are looking for long-range behaviour of the series, it makes sense to focus on the smaller frequencies. but to avoid too small a value of m. It is difficult in this case to decide on a precise estimate, but based on Table 8 and Figure 5, $\hat{H} \cong 0.8$ seems to be a reasonable value.

The periodogram of the *packet* series displays even more pronounced peaks than for the byte series. The most prominent one is at a frequency of approximately 0.1, which corresponds to $2\pi/64$. Thus if the estimate uses $m = N/64$, the peak in question would influence the estimate unduly, and would in fact tend to reduce H by a significant amount. This is indeed what happens. One should therefore focus on smaller m's, and for this series, $\hat{H} = 0.9$ seems to be a reasonable estimate.

The aggregated Whittle method was applied to the Ethernet data in Leland, Taqqu. Willinger and Wilson ((1994), (1995)). The authors obtain $H \simeq 0.8$ for the byte data and $H \simeq 0.9$ for the packet data. (Note that Figure 4 in Willinger, Taqqu, Leland and Wilson (1995) involves aggregation for the packet data and that the first estimate ($H \simeq 0.92$) corresponds to an aggregation level of $m = 100$. As indicated in Table 8, with no aggregation, Whittle based on FGN gives $H = 0.83$.)

The aggregated Whittle estimates are fully consistent with the local Whittle estimates we obtained here and they are also close to those obtained by other methods (see Figure 7 in Leland, Taqqu. Willinger and Wilson (1994)).

7 Stable Innovations

Symmetric α-stable innovations ϵ have the characteristic function

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

where $\alpha \in (0, 2]$ and σ is the scale parameter. They possess "heavy tails",

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

implying that the variance of ϵ is infinite. In this study we use stable innovations with $\alpha = 1.5$ in FARIMA $(0, d, 0)$ and FARIMA $(1, d, 1)$ $(\phi_1 = 0.3, \theta_1 = 0.7)$ time series. These innovations were generated using the *rstab* command in S-Plus. While it is convenient to characterize the FARIMA (p, d, q) series with stable innovations by (2.6), it is more convenient to define them through the corresponding linear representation, (2.3). Since the Durbin-Levinson algorithm cannot be used – it depends heavily on the existence of the covariance function – we used instead the infinite sums in (2.3) truncated at $j = 10,000$.

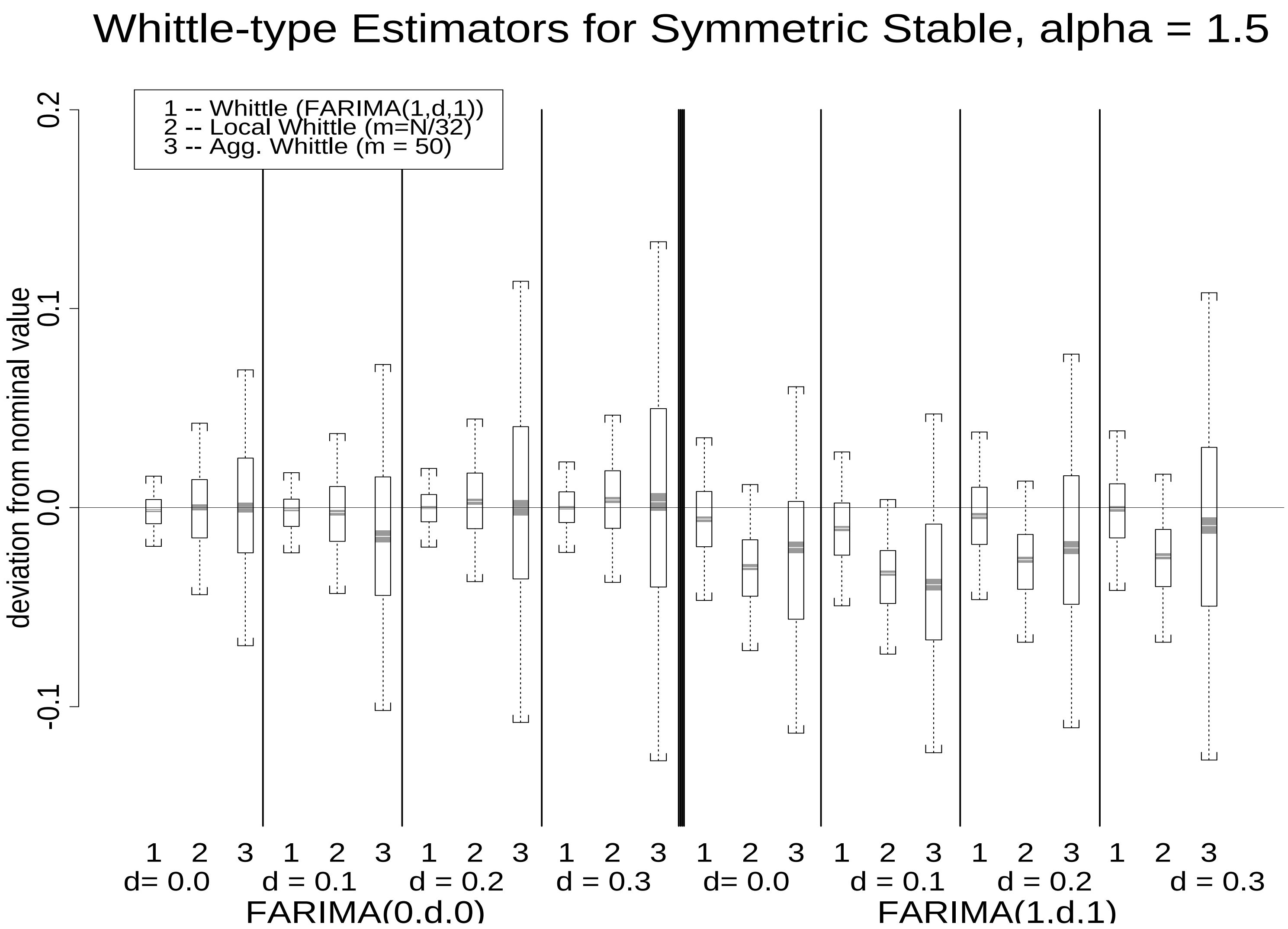
The inclusion of infinite variance innovations forces us to carefully differentiate between the two parameters d and H that are used to characterize long-range dependence. Whereas for finite variance series $H = d + 1/2$, for stable series $H = d + 1/\alpha$; see Samorodnitsky and Taqqu (1994), page 382.

		Symmetric Stable Innovations with $\alpha = 1.5$ Nominal d							
		FARIMA$(0, d, 0)$				FARIMA$(1, d, 1)$			
		0.0	0.1	0.2	0.3	0.0	0.1	0.2	0.3
Whittle	BIAS	-.003	-.004	-.001	.000	.001	-.010	-.004	.000
FARIMA	$\hat{\sigma}$	.021	.022	.023	.026	.063	.026	.030	.029
$(1, d, 1)$	$\sqrt{\text{MSE}}$	.021	.023	.023	.026	.063	.027	.030	.029
Local	BIAS	-.001	-.003	.003	.004	-.031	-.034	-.027	-.025
Whittle	$\hat{\sigma}$	.027	.027	.027	.032	.025	.026	.026	.029
$m = N/32$	$\sqrt{\text{MSE}}$	.027	.027	.027	.032	.040	.042	.037	.038
Aggreg.	BIAS	.001	-.014	.005	.005	-.030	-.041	-.017	-.008
Whittle	$\hat{\sigma}$	.045	.049	.060	.067	.053	.053	.052	.062
$m = 50$	$\sqrt{\text{MSE}}$	.045	.051	.060	.068	.061	.067	.055	.062

Table 9: Estimation results for d using 1000 independent copies of FARIMA time series of length $10,000$. The innovations are symmetric stable with $\alpha = 1.5$.

Many of the estimators of long-range dependence continue to work for infinite variance series, asymptotically returning the correct value of d or H. In particular, the Whittle estimator has been shown to converge to the correct value of d in the infinite variance case (see Klüppelberg and Mikosch (1996) and Kokoszka and Taqqu (1996)). The local Whittle estimator has not been theoretically proven to work for infinite variance innovations, although the empirical results presented here suggest that it does work reasonably well. The aggregated Whittle estimator also has not been rigorously shown to work in the infinite variance case, but the results of Kokoszka and Taqqu (1996) suggest that it can still be so used. We take d in the range $[0, 1 - 1/\alpha)$. This results in an $H < 1$, which is necessary to ensure the stationarity of the series.

Table 9 presents the results of using three different estimators on several long-range dependent series with symmetric stable innovations. We use $\alpha = 1.5$, $d = 0(.1).3$, a series length $N = 10,000$, and 1000 independent realizations. The number of realizations is increased to 1000 because infinite variance series tend to have some very large values, which means that the estimates of d or H also tend to have large outliers. Thus, to ensure meaningful results, 1000


Whittle-type Estimators for Symmetric Stable, alpha = 1.5
1 -- Whittle (FARIMA(1,d,1))
2 -- Local Whittle (m=N/32)
3 -- Agg. Whittle (m = 50)
deviation from nominal value
0.2
0.1
0.0
-0.1
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
d= 0.0
d = 0.1
d = 0.2
d = 0.3
d= 0.0
d = 0.1
d = 0.2
d = 0.3
FARIMA(0.d.0)
FARIMA(1.d.1)

realizations are used. Both FARIMA $(0, d, 0)$ and FARIMA $(1, d, 1)(\phi_1 = 0.3, \theta_1 = 0.7)$ series are simulated and the d's estimated with the following estimators: the Whittle estimator applied to an *overparametrized* FARIMA $(1, d, 1)$ model Whittle, the local Whittle estimator with $m = N/32$ and the aggregated Whittle estimator with $m = 50$. Boxplots of the results are based this time on 1000 realizations. Outliers are not displayed because with so many realizations there are a lot of them, which tends to obscure the overall boxplots.

A comparison of Table 9 with the appropriate rows of Tables 3, 4, 5 and 7 shows that the heavy-tailed nature of the innovations very rarely affects the statistical properties of the estimators. The only exception is the Whittle estimator for the FARIMA $(1, 0, 1)$ series. For that series, the standard deviation of the estimator in the stable case is much larger than in the Gaussian case. This is because, for about 1% of the series, the estimator returns the upper boundary value of $d = 0.5$ instead of the correct value of 0. (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.) This results in an increased standard deviation. Other than this, the performance of the estimators is quite robust to infinite variance innovations.

8 Conclusion

This paper has compared the Whittle, local Whittle and the aggregated Whittle estimators in time series which are commonly used to model long-range dependence. We provided benchmarks that, hopefully, will be useful to the practitioner.

The preferred estimation method depends essentially on how well the exact form of the data series is known. If the series is known to be FGN, or FARIMA of a known order, then certainly the correctly specified Whittle estimator gives the smallest biases and standard errors. The Whittle estimator produces good results for a FARIMA series whose order is not known, provided that an upper limit can be set on the number of autoregressive and moving average coefficients. If, however, this number is incorrectly chosen, then the Whittle estimator, being parametric, can give extremely biased results. Two alternatives were considered: the aggregated Whittle method, where one sums the time series over blocks of size m, thus making it closer to FGN; the local Whittle method which takes into account only the first m Fourier frequencies.

When the time series is short ($N = 100$), the aggregated Whittle estimator is unapplicable and the local Whittle estimator is generally unreliable because the standard errors are high (see Tables 3 and 6). If the series is either FGN or FARIMA $(0, d, 0)$, however, the local Whittle estimator with $m = N/4$ is acceptable. Although it cannot perform as well as the correctly specified Whittle, it offers, in this case, a significantly better estimate than the incorrectly specified Whittle.

When the time series is long ($N = 10,000$), both the aggregated Whittle and local Whittle methods were fairly successful in the cases examined in this paper, with the local Whittle estimator performing slightly better than the aggregated Whittle. A problem with aggregated Whittle is that, when the aggregation level m is large ($m = 100$), the series becomes quite short making the estimates more unreliable. An aggregation level of $m \simeq 50$ seemed to strike the best balance between bias and standard deviation. For the local Whittle method, the recommended choice is $m \simeq N/32$. In any case, both methods are much more robust than the standard Whittle. For series of length $N = 10,000$, the local Whittle and the aggregated Whittle estimators produce results only slightly worse than the *correctly* specified standard Whittle. The fact that no model needs to be specified in order to use them is a definite advantage.

The local Whittle method is computationally much faster than the non-aggregated Whittle. It is also faster than the aggregated Whittle method, although the speed gain in this case is less dramatic. A disadvantage of both the local Whittle and the aggregated Whittle methods is the delicate choice of the extra parameter m. In Taqqu and Teverovsky (1996), we suggest producing a plot of the estimator against m and choosing an m-region where the estimator is relatively stable.

For practical purposes, the aggregated Whittle is not applicable when the length N of the time series is significantly less than $10,000$ because the sample sizes would be too small. As for local Whittle, we suggest the following rule of thumb for time series of length greater than 1000: use Tables 1 and 2 (or their equivalent) to get a rough idea of the bias, and use $\sqrt{1/(4m)}$ to get a rough idea of the standard deviation.

When symmetric stable innovations are used, the estimators do not seem to be affected in a significant way. With the single exception of the Whittle estimator in the FARIMA $(1,0,1)$ case $(\mathrm{ARMA}(1,1))$ where the standard error is larger, the statistical results are almost identical for Gaussian and stable series.

The real-life data example used in this study shows the need to exercise care when using these estimators. It is often necessary to examine the periodogram or some other graphical representation of the data in order to discover any hidden structure of the series, and to make sense of the different estimates which are obtained. The local and aggregated Whittle methods lead to approximately the same estimates.

In conclusion, the original Whittle method remains by far the most accurate when the correct model **is** chosen. Judging from the simulation data, the local Whittle method with an appropriate m seems to be the preferable one when the form of the time series is not known. A value of $m \simeq N/32$ should be chosen if the length N of the time series is $10,000$. A close second choice in this case is the aggregated Whittle method.

Acknowledgement: We would like to thank Peter Robinson for his comments.

References

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edn, Springer-Verlag, New York.

Fox, R. & Taqqu, M. S. (1986), 'Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series', *The Annals of Statistics* **14**, 517–532.

Klüppelberg, C. & Mikosch, T. (1996), Self-normalized and randomly centered spectral estimates, *in* P. M. Robinson & M. Rosenblatt, eds, 'Athens Conference on Applied Probability and Time Series Analysis. Volume II: Time Series Analysis in Memory of E. J. Hannan', Springer-Verlag, New York, pp. 259–271. Lecture Notes in Statistics, **115**.

Kokoszka, P. S. & Taqqu, M. S. (1996), Parameter estimation for infinite variance fractional ARIMA, To appear in *The Annals of Statistics*.

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

McGill, R., Tukey, J. W. & Larsen, W. A. (1978), 'Variation of box plots', *The American Statistician* **32**, 12–16.

Robinson, P. M. (1995), 'Gaussian semiparametric estimation of long range dependence', *The Annals of Statistics* **23**, 1630–1661.

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

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

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

Willinger, W., Taqqu, M. S., Leland, W. E. & Wilson, V. (1995), 'Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements', *Statistical Science* **10**, 67–85.

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