Effects of Aggregation on Estimators of Long-Range Dependence

by
Michele A. Trovero

A dissertation submitted to the faculty of University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2007

Approved by:

Richard L. Smith, Professor

J. Stephen Marron, Professor

Zhengyuan Zhu, Professor

Andrew B. Nobel, Professor

F. Donelson Smith, Professor

ABSTRACT

MICHELE A. TROVERO: Effects of Aggregation on Estimators of Long-Range Dependence.
(Under the direction of Richard L. Smith)

Modern technologies have made available huge amount of data from phenomena that exhibit long-range dependence (LRD). For example, network traffic data can be sampled at intervals as small as one nanosecond. The analysis of such a big amount of data can pose practical challenges. Temporal aggregation has been employed to cope with the limitations in the storage capacity and analysis tools.

One question that arises immediately is whether the aggregated process has the same long-range dependence characteristics of the underlying process. Under mild local assumptions on the spectral density of the LRD process, we show that the fractional order d of the processes is invariant under temporal aggregation.

A second related question is whether and how the estimators of long-range dependence are affected by aggregation. We focus our attention on the log-periodogram regression estimator of Geweke and Porter-Hudak (GPH) and on the local Whittle (LW) estimator. For such estimators, a trimming parameter m needs to chosen by the user so as to balance the trade-off between bias and variance. One optimal choice of m is the value that minimizes the asymptotic mean squared error (MSE) of the estimator. We derive the asymptotic MSE of the GPH estimator for the model in consideration. We show how the MSE-optimal choice of m varies under aggregation for both the GPH and LW estimators.

We consider also the case when the sum of the LRD process and a white noise process is

observed. This model emerges from long-memory stochastic volatility models (LMSV). We derive an expression for the asymptotic MSE of LW estimator. A LW-type estimator (ELW) which accounts for the presence of noise is also considered. We derive a representation of the Hessian matrix of the ELW estimator as functionals of incomplete beta functions. We evaluate numerically the effect of aggregation on the LW and ELW estimators when the observed process is composed by a LRD process plus noise.

We perform an empirical analysis of the effects of aggregation on the UNC network data.

Contents

List of Figures

Chapter 1

Introduction

> Ireneo began by enumerating, in Latin and Spanish, the cases of prodigious memory cited in the Historia Naturalis: Cyrus, king of the Persians, who could call every soldier in his armies by name; Mithridates Eupator, who administered justice in the twenty-two languages of his empire; Simonides, inventory of mnemotechny; Metrodorus, who practised the art of repeating faithfully what he heard once. With evident good faith, Funes marveled that such things should be considered marvelous.
>
> J.L. Borges, Funes el Memorioso, 1944.

One of the assumptions of classical time series analysis is that the correlations decay exponentially over time. Or, similarly, if one looks at the spectral representation of the series, the assumption is the spectral density is bounded at the origin. There are, however, many fields in which such assumptions seem to be the exception, rather than the rule. For example, much of the recent literature on computer network traffic data, starting with the seminal paper of Leland et al. (1994), has focussed on the inadequacy of the traditional models designed for voice networks to adequately describe the characteristics of the data. Similarly, Granger (1966) in his article on "The typical spectral shape of an economic variable" argues that the typical spectral density of an economic time series has a pole at the origin. For these cases, series often show strong correlation over large lags and tend to fluctuate for long periods of time above and below a certain level, showing a pattern that is often referred to as *persistence*. Mathematically, persistence, or slowly decaying

correlations, can be parsimoniously described by the notion of *long range dependence* (LRD).

LRD occurs in a stationary time series when the autocovariance function decays so slowly that its absolute sum diverges. The concept was brought to the attention of statisticians and mathematicians by Mandelbrot and his coauthors (see Mandelbrot and Ness, 1968; Mandelbrot and Wallis, 1968, 1969a,b,c; Mandelbrot, 1977, 1983), although the phenomenon has been known for a long time, the most celebrated example being the level of the river Nile. Since ancient times the river Nile has been known for alternating long periods of dryness with long periods of flooding. This fact has been described also in a passage of the Bible (Genesis 41, 29-30) that predicts seven years of abundance followed by seven years of famine and has therefore been named the *Joseph effect* (Mandelbrot and Wallis, 1968). Reliable measurements of the yearly minimum water levels of the river Nile for the years 622-1281 A.D. (Tousson, 1925) are available and have been studied extensively (Hurst, 1951). Qualitative features of the sample path are that the series globally appears to be stationary and shows no apparent long-term trend, despite the existence of local trends and seasonal effects. The long periods of high level and low level alternate in a manner that seems random. A standard time series analysis of the series reveals that the sample autocorrelation at lag k seems to decrease according to a power law of the form $k^{-\alpha}$ for some $0 < \alpha < 1$, or, equivalently, the variance of the sample mean decays to zero at a rate that is slower than N^{-1}, where N is the sample size. The asymptotic behavior of the autocorrelation translates into the spectral domain at low frequencies, where the periodogram $I(\lambda)$ seems to have a singularity for $\lambda = 0$, and its logarithm decreases almost in linear manner with respect to the logarithm of the frequencies.

Fields in which long-range dependence behavior has been detected include, to mention a few, agricultural field trials (Fairfield Smith, 1938), textile yields (Cox, 1984), video data (Heeke, 1991;

Heyman et al., 1991), network data (Leland and Wilson, 1991), climatological data (Jones and Briffa, 1992; Smith, 1993), and Economics (Granger, 1966).

Long-range dependence sometimes goes under the alternative names of *strong dependence*, *long memory*, *slowly decaying correlation*, $1/f$ *noise*, where f is intended as frequency near the origin, *flicker noise* and *pink noise*.

In the last three decades, long-range dependence has seen an enormous increase of attention in the scientific literature, as testified in the somewhat dated bibliographical guide to long-range dependence of Willinger et al. (1996). More up-to-date accounts on the current status of studies of long-range dependence can be found in Doukhan et al. (2003) and in the eclectic database of references about $1/f$ noise literature that Wentian Li maintains at http://www.nslij-genetics.org/wli/1fnoise/.

Modern technologies of data collection have made possible to collect huge amount of data from phenomena that exhibit LRD behavior. Often, the amount that can be collected and analyzed is limited only by storage capacity and by the analysis tools. Network traffic data can be sampled by sophisticated measuring devices at intervals as small as one nanosecond. Financial data on securities traded on the stock exchange is available transaction-by-transaction (*tick-by-tick*.) There is sufficient empirical and theoretical evidence that the dynamic of packet flows in high-speed data networks exhibits LRD properties (Willinger et al., 2003). Likewise, the volatility process in stochastic volatility models for financial data is often assumed to have long memory, and persistence has been observed in the series of squared returns, log squared returns, and log absolute returns (Harvey, 1998). The analysis of such a big amount of data can pose practical challenges if one wants to look at large spans of time.

Typically, temporal aggregation has been employed as a form of data reduction to cope with the limitations in the storage capacity and analysis tools. One question that arises immediately is

whether the aggregated process has the same long-range dependence characteristics of the underlying process.

Intuitively, one can expect long-range dependence to be invariant under aggregation, since such transformation reduces the basic process to a process of lower frequency, and long memory is related to the low-frequency behavior of the process only. Another reason to expect that aggregation does not interfere with long-range dependence is the close proximity of the class of long-memory processes to the class of self-similar processes for which the autocorrelation function is invariant under aggregation (Samorodnitsky and Taqqu, 1994).

Teles et al. (1999) showed that the fractional order of a FARIMA process is invariant under aggregation. Chambers (1998) showed that the fractional order does not change for stationary processes that admit a Wold representation.

Let X_t, $t \in \mathbb{Z}$, be a stationary time series with long memory. We will assume that the process X_t has a spectral density of the form

$$f_X(\lambda) = c\lambda^{-2d}\{1 + a\lambda^b + o(\lambda^b)\} \qquad \lambda \to 0 \tag{1.1}$$

with $c > 0$, $d \in [0, 0.5)$, $b \in (0, 2]$. The parameter d is the long-range dependence parameter, or *fractional parameter*. Such a spectral density is a fairly general way to represent long memory and has been assumed, for example, by Smith (1989) and Robinson (1995a). We will show that the fractional order of the processes X_t is invariant under temporal aggregation.

A second closely related question is whether and how the estimators of long-range dependence are affected by aggregation. Several estimators of long-range dependence are now available. We will focus our attention on the class of semiparametric estimators based on the periodogram, which include the log-periodogram regression estimator of Geweke and Porter-Hudak (1983) (GPH) and

the local Whittle (LW) estimator introduced by Künsch (1987).

A common characteristics of the semiparametric estimators based on the periodogram is to make assumptions on the functional form of the spectral density only in a neighborhood of the origin. Therefore, the estimators use only the first few ordinates of the periodogram, and a trimming parameter m needs to chosen by the user so as to balance the trade off between the increase in variance that comes from having a small sample, and the bias that comes from extending the base of the estimator to portions of the periodogram where the presence of short-range components interferes with the assumption about the spectral density. One optimal choice of m is the value that minimizes the mean squared error (MSE) of the estimator.

Investigating the effect of aggregation on the estimators translates into investigating how the mean squared error and the optimal choice of the bandwidth parameter m varies under aggregation.

The asymptotic MSE of the GPH estimator was computed by Hurvich et al. (1998) for stationary Gaussian processes with a fractional spectral density. We will extend their result to Gaussian processes whose spectral density follows the model (1.1).

Since the LW estimator is defined implicitly, only a heuristic result due to Henry and Robinson (1996) is available for the asymptotic MSE.

We will show how the MSE and MSE-optimal choice of m varies under aggregation for both the GPH and LW estimators.

We will consider also the case when the process X_t may not be observed directly. Instead, a process Z_t which is the sum of X_t and white noise process is observed. This model emerges for the log of square returns from long-memory stochastic volatility models (LMSV) (Hurvich and Ray, 2001; Deo and Hurvich, 2001). For such a model, Deo and Hurvich (2001) derived asymptotic properties, including the MSE, of the GPH estimator. We will derive an expression for

the asymptotic MSE of LW estimator when the observed process is composed by a LRD process plus noise. The expression does not lend itself to finding the optimal value of m analytically. Therefore, the effect of aggregation on the estimator will be evaluated numerically.

Hurvich and Ray (2001) proposed a new LW-type estimator (ELW) which explicitly accounts for the presence of noise. Recently Hurvich et al. (2005) derived the asymptotic properties of an estimator that includes the ELW estimator as a particular case. The asymptotic MSE of the ELW estimator cannot be evaluated analytically due to the complexity of the expression of the Hessian matrix. We will derive a representation of the Hessian matrix as functionals of incomplete beta functions that can be evaluated numerically. We will evaluate the effect of aggregation on the ELW estimator numerically.

Chapter 1 introduces some notation and definitions of long range dependence in section 1.1. Section 1.2 describes the most popular probabilistic models of long-range dependence. Section 1.3 describes the first empirical findings of long-range dependence and physical models that generate it. Section 1.4 gives an account of the literature on the methods of detection and estimation of long-range dependence.

Chapter 2 describes the theoretical results. Section 2.1 describes the effect of aggregation on LRD processes. Section 2.2 analyzes the effect of aggregation on the LW and GPH estimators. Section 2.3 consider the effect of adding a noise to a LRD process and describes the analytical results that can be derived.

Chapter 3 analyzes numerically the effect of aggregation on the LW and ELW estimators when a noise component is added to a LRD process.

Chapter 4 applies the analysis to the Internet traffic data collected at the UNC main gateway, and to data generated in a lab that is supposed to mimic the behavior of the Internet network.

1.1 Notation and Definitions of Long-Range Dependence

The definition of long-range dependence given above, as non-absolute summability of the autocovariance function, turns out not to be very useful from a practical point of view. Often one of the two quantitative characteristics observed empirically for the time series of the river Nile is taken as definition. That allows us, after introducing some notation, to state the following two definitions of long-range dependence, one for the time domain, and one for the spectral domain.

Before proceeding, let us introduce some basic notation. $X = \{X_t, t \in \mathbb{N}\}$ denotes a long-memory time series observed at points $x = \{x_1, x_2, \ldots, x_N\}$. Unless otherwise specified, X is assumed to have mean zero, be stationary and have finite second moments, with autocovariance function $\gamma(k)$, such that $\sum |\gamma(k)| = \infty$. $\hat{\gamma}(k)$ is the sample autocovariance at lag k. The autocorrelation function and its corresponding sample equivalent are denoted by $\rho(k)$ and $\hat{\rho}(k)$. The spectral density of X at a frequency λ is defined by $f(\lambda) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda}\gamma(\lambda)$. The integer $m^* = [(N-1)/2]$, where $[\cdot]$ is the integer part, is the number of available frequencies for the observed series x. $L_1(\cdot)$ and $L_2(\cdot)$ indicate *slowly varying* functions at infinity and zero, respectively, that is $L_1(ax)/L_1(x) \to 1$ and $L_2(ax^{-1})/L_2(x^{-1}) \to 1$ as $x \to \infty$. The symbols "$\stackrel{d}{=}$" and "$\stackrel{d}{\to}$" indicate the equality and convergence in distribution, respectively.

We are now ready to introduce the two following definitions of long-range dependence.

Definition 1.1.1 (LRD in the Time Domain)**.** X exhibits *long-range dependence* if there exist a real number $\alpha \in (0, 1)$ and a function L_1 slowly varying at infinity such that

$$\lim_{k \to \infty} \frac{\gamma(k)}{k^{-\alpha} L_1(k)} = 1 \tag{1.2}$$

Definition 1.1.2 (LRD in the Spectral Domain)**.** X exhibits *long-range dependence* if there exist

a real number $\beta \in (0,1)$ and a function L_2 slowly varying at zero such that

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{\mid \lambda \mid^{-\beta} L_2(\mid \lambda \mid)} = 1 \tag{1.3}$$

Note that Equation (1.3) is also valid the case of short-range, in which case $\beta = 0$, while this is not true for Equation (1.2).

The two definitions are not equivalent in general, however, they are under the conditions of the following Theorem (Taqqu, 2003; Bingham et al., 1987, Theorems 1.7.2 and 4.10.1).

Theorem 1.1.1 (Abelian-Tauberian Theorem for LRD)**.** *If there exists a value* $0 < a < \infty$ *such that* $\gamma(k)$ *is monotone for all* $k > a$*, then Definition 1.1.1 and Definition 1.1.2 are equivalent with* $\alpha = 1 - \beta$ *and*

$$L_2(x) = \frac{1}{2\pi} \Gamma(\alpha + 1) \sin \frac{\pi(1-\alpha)}{2} L_1(1/x) \tag{1.4}$$

The value $H = 2 - 2\alpha$ is known as *Hurst parameter* and is taken as a measure of long-range dependence. The closer H is to 1, the stronger the dependence is. The case $H = 1/2$ corresponds to *short-range dependence*, while for $0 < H < 1/2$, X is said *invertible*, or *negative dependent*, or *antipersistent*

Long-range dependence can be generated by both physical and probabilistic models. The following sections will illustrated some examples taken from different fields.

1.2 Probabilistic Models of Long-Range Dependence

1.2.1 Self-Similar Processes

Self-similar processes were introduced by Kolmogorov (1941), but were popularized among statisticians mainly thanks to the works of Benoit Mandelbrot and his co-authors (Mandelbrot and Ness, 1968; Mandelbrot and Wallis, 1968, 1969a,b,c).

The idea of self-similarity is that the process exhibits a similar probabilistic structure independently of the "resolution" at which one looks at it. They are important in probability because of their connection to limit theorems.

The exposition of this section will follow loosely Samorodnitsky and Taqqu (1994) and Taqqu (2003).

Throughout this section, $Z = \{Z(t),\ t \in \mathbb{R}\}$ indicates a finite variance continuous time process.

Definition 1.2.1 (Stationary Process)**.** If, for any choice of $\{t_1, t_2, \cdots, t_k\} \in \mathbb{R}^k$, $k \in \mathbb{N}$, and any constant $c \in \mathbb{R}$, the distribution of the vector $(Z(t_1 + c), Z(t_2 + c), \cdots, Z(t_k + c))'$ does not depend on c, the process Z is said to be *stationary*.

Definition 1.2.2 (Self Similar Process)**.** A continuous time process Z is said to be *self-similar* with self-similarity parameter H (H-ss) if, for any rescaling constant $c > 0$, the finite-dimensional distributions of the process $\{c^{-H} Z(ct),\ t \in \mathbb{R}\}$ are the same as those of Z.

The definition states that a change in the time scale is equivalent to a change in the state space scale. Self-similar processes are also known as *self affine* processes, a term coined by Mandelbrot.

The limiting behavior of a self-similar process Z varies according to the values of the self

similarity parameter H (Vervaat, 1987; Beran, 1994). The equality

$$Z(t) \stackrel{d}{=} t^H Z(1)$$

implies that, for t that diverges to infinity,

1. if $H < 0$, then $Z(t) \stackrel{d}{\to} 0$
2. if $H = 0$, then $Z(t) \stackrel{d}{=} Z(1)$
3. if $H > 0$ and $Z(t) \neq 0$, then $|Z(t)| \stackrel{d}{\to} \infty$.

And, for t that converges to zero,

1. if $H < 0$ and $Z(t) \neq 0$, then $|Z(t)| \stackrel{d}{\to} \infty$
2. if $H = 0$, then $Z(t) \stackrel{d}{=} Z(1)$
3. if $H > 0$, then $Z(t) \stackrel{d}{\to} 0$

If one ignores the trivial case of $Z(1) \equiv 0$, the above properties imply that a self-similar process can be stationary only for $H = 0$.

There are, nevertheless, two important relations between self-similar and stationary processes. The first establishes a one-to-one correspondence between the class of stationary processes and the class of self-similar processes.

Theorem 1.2.1. *If Z is H-ss, then*

$$Y(t) = e^{-tH} Z(e^t), \qquad -\infty < t < \infty$$

is stationary. Conversely, if Y is stationary, then

$$Z(t) = t^H Y(\ln t), \qquad 0 < t < \infty$$

is H-ss.

For a proof, see Samorodnitsky and Taqqu (1994), p. 313.

The second, due to Lamperti (1962), is a central limit-type theorem which establishes a relation between self-similar processes and normalized partial sums of stationary processes.

Theorem 1.2.2. *(i) Suppose that $\{U(t)\}_{t\geq 0}$ is a discrete or continuous time process continuous in probability, and such that*

$$\frac{U(\xi t) + g(\xi)}{f(\xi)} \xrightarrow{d} Z(t), \qquad \text{as } \xi \to \infty \tag{1.5}$$

with $0 < f(\xi) \to \infty$ as $\xi \to \infty$. Then there exists $H > 0$, $w(\xi)$ such that $w(\xi) \to w$, $w \in \mathbb{R}$, as $\xi \to \infty$, and a slowly varying function at infinity L for which

$$f(\xi) = \xi^H L(\xi) \tag{1.6}$$

$$g(\xi) = w(\xi)\xi^H L(\xi) \tag{1.7}$$

and the process $\{Z(t) - w\}_{t\geq 0}$ is H-ss

(ii) Conversely, every process Z such that $\{Z(t) - w\}_{t\geq 0}$ is an H-ss process, with $H > 0$ and $w \in \mathbb{R}$, admits the representation (1.5).

If one takes $\xi = n$, $w = 0$, $U(0) = 0$ and $U(nt) = \sum_{k=1}^{[nt]} X_k$, $t > 0$, for some stationary sequence $\{X_k,\ k \in \mathbb{N}\}$, Theorem 1.2.2 says that every limit of normalized partial sums of a stationary series

is a self-similar process, and viceversa. In other words, self-similar processes play an analogous role for stationary processes that stable distributions play for distributions. If the X_k are Gaussian with long-range dependence, then the limit is fractional Brownian motion (see Definition 1.2.4).

For the purpose of modeling data that looks stationary, one only needs to model processes with stationary increments.

Definition 1.2.3 (Stationary Increment Process)**.** Z has *stationary increments* if, for all $c \in \mathbb{R}$ and fixed $t \in \mathbb{R}$, the distribution of

$$Z(t+c) - Z(t) \tag{1.8}$$

does not depend on t. The process Z is *H-sssi* if it is self-similar with stationary increments. For integer j, the first difference process $X_j = \Delta Z(j) := Z(j) - Z(j-1)$ is a stationary time series.

Example 1.2.1 (Brownian Motion)**.** A Brownian motion $\{W(t),\ t \in \mathbb{R}\}$ is a Gaussian process with mean 0 and autocovariance function $\gamma(t_1, t_2) = \min(|t_1|, |t_2|)$. It is H-sssi with $H = \frac{1}{2}$ because, for $c > 0$, $-\infty < t_1, t_2 < \infty$, $EW\{(ct_1)W(ct_2)\} = \min(ct_1, ct_2) = c\min(t_1, t_2) = E\{[c^{\frac{1}{2}}W(t_1)][c^{\frac{1}{2}}W(t_2)]\}$

Proposition 1.2.3. *An H-sssi process Z, with $EZ^2(1) = \sigma^2$, has the following properties:*

1. $Z(0) = 0$ *a.s.*

2. *If* $H \neq 1$, $EZ(t) = 0$ *for all* $t \in \mathbb{R}$.

3. $Z(t) \stackrel{d}{=} Z(-t)$.

4. $E[Z^2(t)] = |t|^{2H}\sigma^2$.

5. The autocovariance function is

$$\begin{aligned} \gamma(s,t) &= EZ(s)Z(t) \\ &= \frac{\sigma^2}{2}\left\{|t|^{2H} - |t-s|^{2H} + |s|^{2H}\right\} \end{aligned} \tag{1.9}$$

and it is non-negative definite.

6. $H \leq 1$.

7. If $H = 1$, $Z(t) = tZ(1)$ *a.s. for all* t.

For a proof, see Taqqu (2003), p.8.

The distribution of a Gaussian process is entirely determined by the structure of the first two moments. Since, for a H-sssi process, these are defined in 2 and 5, there is exactly one Gaussian H-sssi process, up to a constant.

Definition 1.2.4 (Fractional Brownian Motion)**.** A Gaussian H-sssi process, $0 < H \leq 1$ is called *fractional Brownian motion* (FBM) and is denoted by $\{B_H(t),\ t \in \mathbb{R}\}$. It is called *standard fractional Brownian motion* if $\sigma^2 \equiv 1$.

FBM admits several different representations. We refer the reader to Embrechts and Maejima (2002) and Samorodnitsky and Taqqu (1994) for more details.

We now shift our attention to the increments of H-sssi processes, with $0 < H < 1$.

Proposition 1.2.4. *Let* $\{Z_t, t \in \mathbb{R}\}$ *be an* H*-sssi process. The increment process* $\{X_j\} = \{Z(j) - Z(j-1)\}, j \in \mathbb{Z}$ *has the following properties:*

1. X_j *is stationary*

2. $EX_j = 0$ *for all* $j \in \mathbb{Z}$.

3. The variance is

$$E[Z(j) - Z(j-1)]^2 = \sigma^2 \tag{1.10}$$

4. The autocovariance function is

$$\begin{aligned} \gamma(k) &= \frac{\sigma^2}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \\ &= \frac{\sigma^2}{2}\Delta^2|k|^{2H} \qquad k \in \mathbb{N} \end{aligned} \tag{1.11}$$

and $\gamma(k) = \gamma(-k)$ *for* $k \in \mathbb{Z}^-$.

where Δ is the difference operator.

Increments of H-sssi processes are interesting because, for certain values of H, they form a class of stationary processes with long memory. In fact, it follows from (1.10) and (1.11) that the autocorrelation function of X_j is

$$\rho(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \tag{1.12}$$

One can find the asymptotic behavior of $\rho(k)$ through a Taylor expansion. It can be shown (Beran, 1994) that, for $0 < H < 1$ and $H \neq 1/2$,

$$\rho(k) \sim H(2H-1)k^{2H-2} \qquad \text{as } k \to \infty \tag{1.13}$$

Hence, the autocorrelation function decays asymptotically as a power law. For $1/2 < H < 1$ its sum diverges, and, therefore, X_j has long-range dependence. For $0 < H < 1/2$, $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ and, in particular, $\sum_{k=-\infty}^{\infty} \rho(k) = 0$ due to the telescoping nature of $\rho(k)$. This latter case happens

rarely in practice, although it may be generated by overdifferencing a series. For $H = 1/2$, the autocorrelation function is trivially zero, for $k \neq 0$, and there is no dependence. That is, for example, the case of the increments of Brownian motion.

The Gaussian case, as usual, deserves a name by itself.

Definition 1.2.5 (Fractional Gaussian Noise)**.** The first difference process of a fractional Brownian Motion is named *fractional Gaussian Noise* (FGN).

The dependence characteristics of the first difference of a H-sssi process translate, as expected, into the frequency domain.

Theorem 1.2.5. *The spectral density of the first difference of an H-sssi is given by*

$$f(\lambda) = 2c_f(1 - \cos\lambda) \sum_{l=-\infty}^{\infty} |2\pi l + \lambda|^{-2H-1} \tag{1.14}$$

with $c_f = \frac{\sigma^2}{2\pi} \sin(\pi H)\Gamma(2H + 1)$.

It can be shown (Beran, 1994), by a Taylor expansion, that

$$f(\lambda) \sim c_f|\lambda|^{1-2H} + O(|\lambda|^{\min(3-2H,2)}) \qquad \text{as } \lambda \to 0 \tag{1.15}$$

1.2.2 FARIMA and FEXP Processes

Fractional Gaussian noise is an idealized model whose covariance structure depends uniquely on the parameter H which determines both the long- and short-range dependence features. When dealing with real data, it is often necessary to allow for greater flexibility to capture the dependence

structure adequately. One possibility is to assume a spectral density of the form

$$f(\lambda) = \mid 1 - e^{i\lambda} \mid^{-(2H-1)} f^*(\lambda) \tag{1.16}$$

where f^* is usually a smooth bounded function on $[-\pi, \pi]$, nonvanishing at zero which captures the short-range dependence structure. For $\lambda \to 0$, $f(\lambda) \sim |\lambda|^{-(2H-1)}$, hence the model (1.16) has long memory.

One such type of processes is the well known class of FARIMA models, which stands for fractional autoregressive integrated moving averages. FARIMA models were introduced by Adenstat (1974) and studied by Granger (1980b) and Hosking (1981).

FARIMA models are like ARIMA models of Box and Jenkins (1971), but the differencing parameter d can take any real value.

Definition 1.2.6 (FARIMA model)**.** Let B be the lag operator, and ε_t a stationary sequence of *iid* innovations with mean zero and variance σ^2. For integers p and q, assume that the polynomials $\phi(x) = 1 + \sum_{j=1}^{p} \phi_j x^j$ and $\psi(B) = 1 + \sum_{j=1}^{q} \psi_j x^j$ admit distinct roots outside the unit circle. The process X is a FARIMA(p, d, q) if

$$\phi(B)(1 - B)^d X_t = \psi(B)\varepsilon_t \quad d \in \mathbb{R}^+ \tag{1.17}$$

where

$$(1 - B)^d := \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \tag{1.18}$$

and

$$\binom{d}{k} := \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} \tag{1.19}$$

It can be shown (Brockwell and Davis, 1991, p. 525, Theorem 13.2.2) that the spectral density of a stationary FARIMA(p, d, q) process follows Equation (1.16), with $H = d + 1/2$ and

$$f^*(\lambda) = \frac{\sigma^2}{2\pi} \frac{\mid \psi(e^{-i\lambda}) \mid^2}{\mid \phi(e^{-i\lambda}) \mid^2} \tag{1.20}$$

which is the spectral density of an ARMA(p, q) model.

Another class of models for which Equation (1.16) holds is the FEXP (Fractionally integrated Exponential) class, which generalizes the so-called Bloomfield exponential class of models (Bloomfield, 1973) and was studied by Janacek (1987) and Beran (1993).

Definition 1.2.7 (FEXP model)**.** A *FEXP process* X is characterized by a spectral density (1.16), with

$$f^*(\lambda) = \exp\left\{\sum_{j=0}^{p-1} \theta_j f_j(\lambda)\right\} \tag{1.21}$$

where $f_0 \equiv 0$ and $f_1, f_2, \ldots, f_{q-1}$ are smooth functions in $[-\pi, \pi]$ such that $f_j(\lambda) = f_j(-\lambda)$ and, for any n, the $n \times (p-1)$ matrix Ω with column vectors $[f_j(2\pi/N), f_j(2\pi 2/N), \ldots, f_j(2\pi n/N)]'$ $(j = 1, 2, \ldots, p-1)$ is nonsingular.

In particular, for

$$f_j = \frac{\cos(j\lambda)}{\sqrt{\pi}} \qquad j = 0, 1, \ldots, p-1$$

we have

$$f(\lambda) = \mid 1 - e^{i\lambda} \mid^{-(2H-1)} \exp\left\{\sum_{j=0}^{p-1} \theta_j \frac{\cos(j\lambda)}{\sqrt{\pi}}\right\} \tag{1.22}$$

We will indicate by FEXP(p) a process that has spectral density (1.22). The coefficients θ_j are known as *cepstrum* coefficients in the time series literature.

1.3 Empirical Findings of Long-Range Dependence

1.3.1 Uniformity Trials

Fairfield Smith (1938) was among the first to give a detailed account of evidence of long-range dependence. He analyzed the results from uniformity trials, that are experiments commonly conducted in agronomy to determine the optimal plot size that provides, roughly speaking, the best information at the minimum cost. An experimental area was planted uniformly with wheat, in 38 rows, 6 inches apart. Four rows at each side were disregarded to avoid border effects. The remaining area was sectioned in 1080 elementary plots of $\frac{1}{2}$ square foot and the yield of each one recorded. Adjacent elementary plots were then combined in non-overlapping rectangular plots of size k, for k varying from 1 to 120. For each fixed k, the average yield and its sample variance $V(k)$ over all rectangles of size k were computed. For long-range dependent series, the variance of the sample mean is of order k^{2H-2}, for large k (see also Section 1.4.1). Smith found that $V(k)$ follows, approximately, the following relation:

$$\log V(k) = a + b \log(k) \tag{1.23}$$

with b around -0.749. This implies that the variance converges to zero at a much slower rate that if the observations were independent.

Fairfield Smith found similar results in all the uniformity trial data that he analyzed.

Whittle (1956, 1962) proposed an explanation to the phenomenon observed by Smith using stochastic partial differential equations. He considered a process Y_ν with $\nu = (\mathbf{x}, t) = (x_1, \ldots, x_m, t)$, $m \geq 1$, where t is a time index, and $\mathbf{x}$ is a position index. In the case of uniformity trials, Y_ν may represent soil fertility, and $m = 3$. He considered two opposite effects: on one side a uniformizing

effect due to diffusion of salts in the soils, on the other side, a random disturbance due to weather, artificial fertilization and other effects. He assumed that, when the two effects are in equilibrium, the process follows the stochastic partial differential equation

$$\frac{\partial Y_\nu}{\partial t} + \alpha Y_\nu = \frac{1}{2}\nabla^2 Y_\nu + \epsilon_\nu \tag{1.24}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_m^2}$$

and ϵ_ν is a random field.

Under suitable condition on the random component ϵ_ν, Whittle showed that if the spatial covariance of the process that is solution to the above stochastic partial differential equation depends only on the Euclidean distance between two points x and x' in space, and decays asymptotically like $d_E(x, x')^{4H-4}$, for some H between $\frac{1}{2}$ and 1, then $V(k)$ converges to zero like a constant times k^{2H-2}. If the trials were independent, one would expect $V(k)$ to decrease like k^{-1} instead.

1.3.2 The Hurst Effect

The first attempt to measure long-range dependence that lead to a specific estimator is due to the hydrologist Hurst (1951) in his studies of the level of the river Nile. Let X_t be the flux of a reservoir at time t, and $Y_T = \sum_{t=1}^{T} X_t$ be the cumulative influx up to time T. The ideal capacity of the reservoir between time t and $t + k$ can be shown to be equal to the *adjusted range*

$$R(t,k) = \max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] \tag{1.25}$$

Let

$$S(t,k) = \sqrt{k^{-1} \sum_{i=t+1}^{t+k} (X_i - \bar{X}_{t,k})^2} \tag{1.26}$$

be the sample standard deviation of the flux between time t and k, where $\bar{X}_{t,k} = k^{-1} \sum_{i=t+1}^{t+k} X_i$ is its sample mean.

The standardized ratio

$$Q(t,k) = \frac{R(t,k)}{S(t,k)} \tag{1.27}$$

is known as *rescaled adjusted range* or *R/S-statistic*.

For the River Nile data, Hurst observed that, for large k,

$$\log E[R/S] \approx a + H \log k \tag{1.28}$$

with $H > \frac{1}{2}$. This is in contrast to what is expected from a short-memory processes, for which the R/S ratio, for large k, should be approximately equivalent to $k^{\frac{1}{2}}$. This empirical finding is known as *Hurst effect*, and the parameter H is known as *Hurst parameter*.

1.3.3 Critical Phenomena

Cassandro and Jona-Lasinio (1978) describe how long-range dependence arises in physics in the contest of critical phenomena. Suppose a thermodynamics system can be described by a random field on a finite d-dimensional lattice $\Lambda \in \mathbb{Z}^d$. For instance, let S_i be the spin at the position $i \in \Lambda$ of a ferromagnetic substance. The quantity $M_\Lambda(S) = \sum_{i \in \Lambda} S_i$ is called *total magnetization*. The properties of the system are studied considering the limit as $\Lambda \to \mathbb{Z}^d$. In particular, one seeks a

normalizing value $\sigma(\Lambda)$ such that the ratio

$$M_\Lambda^\star(S) = \frac{M_\Lambda(S) - E[M_\Lambda(S)]}{\sigma(\Lambda)} \tag{1.29}$$

converges to a nondegenerate random variable.

Under normal conditions, the spins S_i are weakly dependent and

$$\sigma(\Lambda) \approx c|\Lambda|^{\frac{1}{2}}$$

where $|\Lambda|$ is the number of elements in Λ and $c > 0$.

There are however critical temperatures at which the spins show long-range correlation, for example in the case of spontaneous magnetization of a ferromagnetic substance, or the transition from liquid form to gaseous form of a gas. In such cases,

$$\sigma(\Lambda) \approx c|\Lambda|^{\alpha}$$

for some $\frac{1}{2} < \alpha < 1$.

1.3.4 Aggregated Series

Suppose the series S_t aggregates individual time series $X_t^{(j)}$ $(j = 1, 2, 3, \ldots)$, where each $X_t^{(j)}$ follows an AR(1) process defined by

$$X_t^{(j)} = \alpha_j X_{t-1}^{(j)} + \epsilon_t^{(j)} \qquad\qquad j = 1, 2, 3, \ldots \tag{1.30}$$

with $-1 < \alpha_j < 1$, and $\epsilon_t^{(j)}$'s serially independent random variables with zero mean and variance σ_j^2.

Suppose, furthermore, that the parameters α_j and σ_j^2 are randomly drawn from populations with distributions F_α and F_σ and means α and σ_ϵ, respectively.

Then the spectral density of the partial sum

$$S_t^{(N)} = \sum_{j=1}^{N} X_t^{(j)} \tag{1.31}$$

for large N, is approximately equal to

$$\begin{aligned} f^{(N)}(\lambda) &= \sum_{j=1}^{N} f_j(\lambda) \\ &= \sum_{j=1}^{N} \frac{\sigma_j^2}{2\pi} \frac{1}{|1-\alpha_j e^{-i\lambda}|^2} \\ &\approx \frac{N}{2\pi}\sigma_\epsilon^2 \int \frac{1}{|1-ue^{-i\lambda}|^2} dF_\alpha(u) \end{aligned} \tag{1.32}$$

where f_j is the spectral density of $\{X_t^{(j)}\}$.

Granger (1980a) showed that if F_α is a beta distribution with suitable parameters, then $f^{(N)}(\lambda)$ has a singularity at 0, and thus $\{S_t^{(N)}\}$ approximates a series with long memory.

1.4 Detection and Estimation of Long-Range Dependence

This section presents an overview of different methods for detecting and measuring long-range dependence. Following Definitions 1.1.1 and 1.1.2, the methods are classified according to the domain of application. Usually, time domains methods are based on the sample autocorrelation function or second moments of moving sample averages over sub-blocks of observations, while

spectral-domain methods are based on the periodogram. A separate section is devoted to the Wavelet estimator, which does not fall in either of above two categories, although it can be thought of as a different form of spectral decomposition.

An account of the recent developments in estimation theory can be found in Giraitis and Robinson (2003), Moulines and Soulier (2003) and Abry et al. (2003). A comparison of several of these estimators is found in Taqqu et al. (1997), Taqqu and Teverovsky (1998) and Bardet et al. (2003). An excellent description of several estimation methods, which includes examples and Splus code, is available from Murad Taqqu's web page at http://math.bu.edu/people/murad/methods.

1.4.1 Time-Domain Methods

Several methods of detection of long-range dependence in the time domain rely on graphical procedures. Most of them, while providing an estimate for H, do not offer an easy way to compute confidence intervals, making the interpretation of the estimates difficult. Hence, they should be regarded only as useful exploratory tools. Nevertheless, they are widely used for estimation as well, partly due to the fact that they appeared first, and thus are well known, and partly because they are easy to implement.

Sample ACF and PACF

One obvious way to detect long-range dependence is to plot the sample autocorrelation $\hat{\rho}(k)$ of the series. If the series has long-range dependence, according to Definition 1.1.1, $\hat{\rho}(k)$ will decay slowly like a power function, proportional to k^{2H-2}.

Under the null hypothesis of zero correlation, and mild regularity conditions (see, e.g., Brockwell and Davis, 1991), $\sqrt{n}\hat{\rho}(k)$ are asymptotically standard normal *iid* random variables. It is common to plots the approximate 95% confidence interval lines corresponding to $\pm 2/\sqrt{n}$ and check for

departures from the null hypothesis, together with a slow power-like decay. Such behavior is often difficult to detect from the plots only. The definition implies a slow asymptotic decay, but nothing about the absolute value of the ACF. It is indeed possible that the sample ACF of a LRD series will lie within the $\pm 2/\sqrt{n}$ bands for the entire observed range (see Beran, 1994, pp. 89–90). In addition, correlations at high lags cannot be estimated reliably.

A more suitable visual display is plotting $\log|\hat{\rho}(k)|$ vs $\log k$. If the decay is hyperbolic, the points will be scattered around a line with slope equal to $2H - 2$. Fitting a linear regression will also offer a rough estimate of H.

Analogous comments also apply to the partial autocorrelation function, which decays at a rate of $k^{-H-1/2}$.

The R/S approach

The R/S approach is historically the first and still the best known of methods for measuring long-range dependence. It is discussed in detail in Mandelbrot and Wallis (1969b) and Mandelbrot (1975).

Based on Hurst's empirical findings, one heuristic approach to estimate H is as follows:

1. Divide the series into K block of size N/K

2. Compute the R/S statistics $Q(t_i, k)$, as defined in Equation (1.27), with starting values $t_i = iN/K + 1$ for all possible k such that $t_i + k < N$

3. Plot its logarithm against the logarithm of k. This is sometimes called the *pox plot* for the R/S statistics.

4. The estimated slope from the regression will be then the estimate of H.

The theoretical foundations of this method rests on the following theorem of Mandelbrot (1975).

Theorem 1.4.1. *Let X_t be such that X_t^2 is ergodic and $t^{-H}\sum_{s=1}^{t} X_s$ converges weakly to a fractional Brownian motion as t tends to infinity. Then, as $k \to \infty$,*

$$\frac{k^{-H}R(t,k)}{S(t,k)} \xrightarrow{d} \xi$$

where ξ is a nondegenerate random variable.

The R/S method has several limitations. First it is not clear what cut-off value k to use in the regression for the asymptotic result to apply, nor can R/S be computed for large values of k. Second, the distribution of R/S is neither normal nor symmetric and its values are correlated at different time points t, which raises questions about the use of linear regression. Lastly, Bhattacharya et al. (1983) have shown that the methodology is not robust with respect to slight non-stationarity. In other words, it is possible to construct a short-memory process with slowly decaying trend for which the R/S method produces an estimate of H greater than $\frac{1}{2}$.

Lo (1991) provided a test for the null hypothesis of absence of long-range dependence based on the asymptotic distribution of a modified version of the R/S statistic. He considered R as above, but suggested using the entire length of the series N, and instead of the standard deviation S, a weighted sum of the autocovariances, namely,

$$S_q(N) = \sqrt{S^2(1,N) + 2\sum_{j=1}^{q} w_j(q)\hat{\gamma}(j)} \tag{1.33}$$

where

$$w_j(q) := 1 - \frac{j}{q+1}, \qquad q < N$$

are weights chosen so that S_q is the sample autocovariance of an aggregated or averaged time series.

If the series has no long-range dependence, Lo showed that, given the right choice of q, the distribution of the statistic

$$V_q(N) := N^{-1/2} R(1, N)/S_q(N) \tag{1.34}$$

is asymptotic to

$$W_1 := \max_{0 \le t \le 1} W_0(t) - \min_{0 \le t \le 1} W_0(t) \tag{1.35}$$

where W_0 is the standard Brownian bridge.

This fact allows to compute the 95% confidence interval for W_1

$$P(W_1 \in [.809, 1.862]) = 0.95$$

which can be used to set up a test for the null hypothesis of absence of long-range dependence.

Lo's result is asymptotic for $N,\ q(N) \to \infty$. For finite N, $q(N)$ influences both the power and the size of the test, and it's not clear what the right choice is. Additionally, it indicates only whether there is long-range dependence or not, it does not provide an estimate for H.

The Aggregated Variance and Absolute Moment Methods

A LRD stationary time series of length N with finite variance is characterized by the variance of the sample mean being of order N^{2H-2} (Beran, 1994). This suggests the following method of estimation.

1. For an integer m between 2 and $N/2$, divide the series in blocks of length m and compute

the sample average over each kth block.

$$\bar{X}_k^{(m)} := \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t, \qquad k = 1, 2, \ldots, [N/m], \tag{1.36}$$

2. For each m, compute the sample variance of $\bar{X}_k^{(m)}$

$$s_m^2 := \frac{1}{([N/m]-1)} \sum_{k=1}^{[N/m]} (\bar{X}_k^{(m)} - \bar{X})^2$$

3. Plot $\log s_m^2$ against $\log m$.

For sufficiently large values of m, the points should be scattered around a straight line with slope $2H - 2$. In the case of short-range dependence ($H = .5$,) the slope is equal to -1. It is often convenient to draw such a line as reference, however, small departures from short-range dependence are not easy to detect visually.

The estimate of H is found fitting a least square line to the points of the plot. In practice, neither the left nor the right end points should be used in the estimation. On the left end, the low number of observations in each block introduces bias due to short-range effects. On the right end, the low value of $[N/m]$ makes the estimate of s_m^2 unstable. The two thresholds that restricts the estimation range are left to the discretion of the researcher.

One disadvantage of the Aggregated Variance (AV) method is that it does not provide an explicit estimation of the variance of the estimator. Another disadvantage is that is has been found not to be very robust to departures from standard Gaussian assumptions (Taqqu and Teverovsky, 1998).

A generalization of the AV method is the Absolute Moments (AM) method.

Consider the series of the averages defined in (1.36), and compute its nth absolute moment

$$AM_n^{(m)} = \frac{1}{[N/m]} \sum_{k=1}^{[N/m]} \left| \bar{X}_k^{(m)} - \bar{X} \right|^n \tag{1.37}$$

$AM_n^{(m)}$ is asymptotically proportional to $m^{n(H-1)}$. To find an estimate for H, compute $AM_n^{(m)}$ for different values of m, and plot it in a log-log plot against m. The point should be scattered along a line with slope $n(H-1)$.

For $n = 2$ the AM method reduces to the AV method.

A further generalization of the case $m = 1$ was introduced by Higuchi (1988), who suggested to use a sliding window to compute the averages, instead of non-overlapping blocks, thus providing better accuracy for shorter time series, at the cost of being more computationally intensive than the AV or AM methods. He considered the functional

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^{m} \left[\frac{N-1}{n} \right]^{-1} \sum_{k=1}^{(N-i)/m} \left| \sum_{j=i+(k-1)m+1}^{i+km} X_j \right| \tag{1.38}$$

and showed that the expected value of $L(m)$ is asymptotic to C^{mH-2}, for a constant C.

Gaussian Maximum Likelihood

Let X be causal and invertible, or, in other terms, let X admit the representations

$$X_t = \sum_{s=1}^{\infty} \phi_s X_{t-s} + \varepsilon_t \tag{1.39}$$

and

$$X_t = \sum_{s=0}^{\infty} \psi_s \varepsilon_{t-s} \tag{1.40}$$

with innovations ε_t uncorrelated with mean 0 and variance σ_ε^2.

Suppose that the autocovariance and spectral density of X can written in parametric form as $\gamma(k;\theta^0)$ and $f(\lambda;\theta^0)$, where $\theta^{0\prime} = [\theta_1^0, \theta_2^0, \ldots, \theta_M^0]' \in \mathbb{R}^M$ is a vector of parameters. θ_1 typically represents a scale parameter which can be σ^2 in (1.11), or, more generally, the variance of the innovation in (1.40). Often it is only a nuisance parameter that varies freely from the other parameters, and, therefore can be eliminated from the estimation and written as a function of the other parameters. $\theta_2 = H$ is the usual Hurst parameter which measures long-range dependence, and $(\theta_3, \ldots, \theta_M)$ is a set of parameters of the model that typically describe the short-range behavior of X. For example, it can be the MA and AR polynomials in a FARIMA model.

An interesting fact (see Brockwell and Davis, 1991, p. 168, Corollary 5.1.1) is that the best one-step linear predictor of X is

$$\hat{X}_t = \sum_{s=1}^{\infty} \phi_s X_{t-s} \tag{1.41}$$

and the mean squared prediction error corresponding to $\hat{X}_t$ is given by Kolmogorov's Formula

$$\text{MSPE} = \sigma_\varepsilon^2 = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda;\theta) d\lambda \right\} \tag{1.42}$$

In particular, if $f(\lambda;\theta) = \theta_1 f^*(\lambda;\eta)$, with $\theta' = [\theta_1 \ \eta']'$, then we have the following *standard representation*:

$$\int_{-\pi}^{\pi} \log f(\lambda;\theta) d\lambda = 2\pi \log \theta_1 \tag{1.43}$$

Let $\Sigma_N(\theta)$ be the variance-covariance matrix of x, and $\mid \Sigma_N(\theta) \mid$ its determinant. If X is Gaussian, the MLE is found maximizing the log-likelihood, given by

$$\mathscr{L}_N(x;\theta) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log \mid \Sigma_N(\theta) \mid - \frac{1}{2} x' \Sigma_N(\theta)^{-1} x \tag{1.44}$$

The strong consistency of the Gaussian MLE is covered under the proof of Hannan (1973) for the short-range dependence case. In fact, Hannan's basic assumptions are that the series is ergodic and linear in innovations that are martingale differences with existing second moment. Hannan's proof of asymptotic normality, instead, does not extend to the LRD case, since it assumes the square integrability of the spectral density.

Dahlhaus (1989) extended the Central Limit Theorem of Hannan to the case of LRD, and proved that

$$\sqrt{N}(\hat{\theta} - \theta^0) \xrightarrow{d} \mathcal{N}\left(0, \left[\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}\log f(\lambda;\theta)\right\}\left\{\frac{\partial}{\partial\theta}\log f(\lambda;\theta)\right\}' d\lambda\right]^{-1}\right) \tag{1.45}$$

Note that the rate of convergence $\sqrt{N}$ is the same as for the short memory case. Dahlhaus showed also that the MLE is still asymptotically efficient in the case of a LRD Gaussian series, in the sense that it reaches the Cramer-Rao lower bound.

1.4.2 Frequency-Domain Methods

Frequency domain estimators of H are based on Definition 1.1.2 of long-range dependence.

These methods rely on the periodogram of X, which is defined by

$$I(\lambda) = \frac{1}{2\pi N}\left|\sum_{t=1}^{N} X_t e^{i\lambda t}\right|^2 \tag{1.46}$$

Usually, it is evaluated at the Fourier frequencies $\lambda_{k,N} = \frac{2\pi k}{N}$, $0 \leq k \leq m^*$.

Let λ_k be a generic frequency in $[-\pi, \pi]$, not necessarily a Fourier frequency. For linear short-range dependence processes, it is a well known fact (see, e.g., Brockwell and Davis, 1991, Proposition 10.3.1 and Proposition 10.3.2) that the periodogram has the following properties:

1. The periodogram is an asymptotically unbiased estimator of the spectral density, i.e.:

$$E(I(\lambda_k)) \longrightarrow f(\lambda_k) \tag{1.47}$$

   as $N \to \infty$, uniformly on $[-\pi, \pi]$.

2. Under suitable conditions, the periodogram ordinates are asymptotically uncorrelated.

3. For any given finite s-tuple of distinct frequencies $\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_s}$, such that $\lambda_{k_j} \to \lambda_j$, as $k_j \to \infty$, the periodogram ordinates $I(\lambda_{k_1}), I(\lambda_{k_2}), \ldots, I(\lambda_{k_s})$ are asymptotically independent exponential random variables with means
   $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_s)$.

In general, these properties do not hold anymore for stationary LRD series (Hurvich and Beltrao, 1993).

For stationary invertible fractional processes satisfying (1.16) and some conditions on the cumulants of ε_t in the MA representation (1.40), Yajima (1989) showed that property 3 still holds if $\lambda_j \pm \lambda_{j'} \neq 2\pi l$, for $l \in \mathbb{Z}$.

It has to be noted that this result is true only for a finite number of frequencies converging to distinct limits away from zero, while in the SRD case the limit may be the same, and, in particular, zero, as long as the sequences are distinct. Most frequency-domain estimators of LRD are instead functions of the periodogram at an increasing number of Fourier frequencies in a neighborhood of the origin, for which this result does not apply.

We restrict, therefore, our attention to the periodogram of fractional linear processes satisfying (1.16) and computed at Fourier frequencies $\lambda_{k,N} = \frac{2\pi k}{N}$ tending to zero. Let $I^*(\lambda) = \frac{I(\lambda)}{f(\lambda)}$ be the rescaled periodogram. It can be shown (Hurvich and Beltrao, 1993) that, for processes satisfying

(1.16),

$$\lim_{N\to\infty} \mid E(I^*(\lambda_{k,N}) - 1) \mid \neq 0$$

and

$$\lim_{N\to\infty} \mid Cov(I^*(\lambda_{k,N}), I^*(\lambda_{l,N})) \mid \neq 0$$

However, under appropriate conditions (see Robinson, 1995a), there exists a sequence $r(f;k)$ with $\lim_{k\to\infty} r(f;k) = 0$, such that, for any N,

$$\mid E(I^*(\lambda_{k,N}) - 1) \mid \leq r(f;k) \tag{1.48}$$

for $1 \leq k \leq m^*$. In other words, the bias is small for frequencies away from zero.

Similarly, there exists a sequence $r(f;k,l)$, with $\sum_{1\leq k<l\leq m^*} r(f;k,l) = O(\log^r(n))$, for some $r > 0$, such that

$$\mid Cov(I^*(\lambda_{k,N}), I^*(\lambda_{l,N})) \mid \leq r(f;k,l) \tag{1.49}$$

for $1 \leq k \leq m^*$.

Recently, Lahiri (2003) provided necessary and sufficient conditions for the asymptotic independence of the periodogram ordinates under both short- and long-range dependence. He showed that $I(\lambda_{j,N})$ and $I(\lambda_{k,N})$ are asymptotically independent if $\lambda_{j,N}$ and $\lambda_{k,N}$ are "sufficiently asymptotically distant", even if both converge to zero, that is

$$\mid N(\lambda_{j,N} - \lambda_{k,N}) \mid \to \infty \qquad \text{as } N \to \infty \tag{1.50}$$

The necessity of this condition depends on the taper function used and on the value of H.

Finally, Deo (1997) showed that for any given integer s, and a s-uple of distinct Fourier fre-

quencies $\lambda_{k_1,N}, \lambda_{k_2,N}, \ldots, \lambda_{k_s,N}$,

$$(I^*(\lambda_{k_1,N}), I^*(\lambda_{k_2,N}), \ldots, I^*(\lambda_{k_s,N})) \quad \xrightarrow{d} \tag{1.51}$$

$$\big((Z_1^2 + Z_2^2)/2, (Z_3^2 + Z_4^2)/2, \ldots, (Z_{2s-1}^2 + Z_{2s}^2)/2\big) \tag{1.52}$$

where $Z = (Z_1, \ldots, Z_{2s})'$ is a vector that has multivariate normal distribution with non-diagonal covariance matrix. See Deo (1997, Corollary 3) for the expression of such a matrix.

The Whittle Approximate MLE

The MLE method discussed in section 1.4.1 seems to solve the problem of likelihood estimation for H. Nevertheless, it poses some practical issues. First, the dimension of the $\Sigma_N(\theta)$ is $N \times N$. The inversion can take a great amount of resources if the series is long. Second, $\Sigma_N(\theta)$ can be nearly singular for high values of H, due to the slowly decaying correlations. Furthermore, if X has mean different from zero, x_i is usually replaced by the centered value $(x_i - \bar{x})$. For long-memory series, the rate of convergence of $\bar{x}$ is N^{2-2H}, which can be lower than the rate $\sqrt{N}$ of the convergence for $\hat{H}_{MLE}$. This fact may impair the efficiency of the estimates of H. Indeed, empirical studies (Cheung and Diebold, 1994) suggest that the approximation (1.60) which depends only on the periodogram at the Fourier frequencies $1, \ldots, m*$ and, therefore, is not affected by $\bar{x}$, may perform better than other forms presented here.

For these reasons, equation (1.44) is often replaced by other target functions that imply the same first order distribution for the estimators, but have computational advantages. The idea arose first in the context of short-range dependence with the work of Whittle (1953).

To approximate equation (1.44), first note that, under regularity conditions,

$$\lim_{N\to\infty} \frac{1}{N} \log \mid \Sigma(\theta) \mid \to \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda;\theta) d\lambda \tag{1.53}$$

Therefore, $\log \mid \Sigma(\theta) \mid$ can be replaced by

$$\frac{N}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda;\theta) d\lambda \tag{1.54}$$

MA Approximation

Consider the representation (1.40) of X, and write $\theta = \left(\frac{\sigma_\varepsilon^2}{2\pi}, \eta'\right)'$. Then $\Sigma = \Sigma^*(\eta)/\theta_1$ for some matrix Σ^*. Assume $f(\lambda;\theta) = f^*(\lambda;\eta)/\theta_1$. The standard representation (1.43) holds, so that (1.44) can be replaced by

$$-\frac{1}{2} \log \theta_1 - \frac{1}{2NB\theta_1} x' \Sigma^*(\eta)^{-1} x \tag{1.55}$$

The estimate for η is found minimizing the quadratic form

$$x' \Sigma^*(\eta)^{-1} x$$

and θ_1 is estimated by (1.55) at its minimum.

AR Approximation

If, instead of the MA representation, we use the AR representation (1.39) of X, and approximate x_t by the finite truncation

$$x_t = \sum_{s=1}^{t-1} \psi(s;\theta) x_{t-s} + \varepsilon_t^*(\eta) \qquad\qquad t = 1, \ldots, N \tag{1.56}$$

Then equation (1.55) can be replaced by

$$-\frac{1}{2N}\log\theta_1 - \frac{1}{2N}\sum_{t=1}^{N}\frac{\varepsilon_t^*(\eta)^2}{\theta_1} \tag{1.57}$$

This procedure was suggested by Box and Jenkins (1971) for short-memory time series, but the truncation (1.56) makes it less attractive for LRD series. On the other hand, Beran (1994) asserts that the asymptotic results for this approximation hold also in the case $H \geq 1$, which makes it suitable for non-stationary series.

Whittle Approximation

A further approximation to $\Sigma_N(\theta)$ is due to Whittle (1953), who discovered it in the context of shot-memory series. Let $A(\theta) = [\alpha(j-l)]_{j,l=1,\ldots,N}$, and

$$\alpha(j-l) = (2\pi)^{-2}\int_{-\pi}^{\pi}\frac{1}{f(\lambda;\theta)}e^{i(j-l)\lambda}d\lambda \tag{1.58}$$

$A(\theta)$ is asymptotically equivalent to the inverse of $\Sigma_N(\theta)$ so that minimizing (1.44) is approximately equivalent to minimizing

$$\mathscr{L}_W(\theta) = \frac{1}{2\pi}\left\{\int_{-\pi}^{\pi}\log f(\lambda;\theta)d\lambda + \int_{-\pi}^{\pi}\frac{I(\lambda)}{f(\lambda;\theta)}d\lambda\right\} \tag{1.59}$$

We can further approximate (1.59) by a Riemann sum at the Fourier frequencies $\lambda_j = \frac{2\pi j}{N}$. The approximate MLE can be found minimizing the following equation:

$$\mathscr{L}_W^*(\theta) = \sum_{j=1}^{m^*}\log f(\lambda_j;\theta) + \sum_{j=1}^{m^*}\frac{I(\lambda_j)}{f(\lambda_j;\theta)} \tag{1.60}$$

where $m^* = \left[\frac{N-1}{2}\right]$. In particular, if the standard representation (1.43) holds, (1.44) is approximately equivalent to

$$\log \theta_1 + \frac{1}{N\theta_1} \sum_{j=1}^{m^*} \frac{I(\lambda_j)}{f^*(\lambda_j; \eta)} \tag{1.61}$$

The approximation (1.61) was first suggested by Graf (1983) for FGN. His argument made the simplifying assumption that

$$I(\lambda; \theta) = f(\lambda; \theta)\xi_j \tag{1.62}$$

where ξ_j are independent exponential random variables. This assumption, as we have seen, is not formally correct, as it holds only for fixed frequencies away from zero.

Asymptotic Results and Extensions

In a seminal paper, Fox and Taqqu (1986) extended the asymptotic normality of the Whittle estimator to the LRD case. Their basic intuition was that the vanishing $f^*(\lambda; \theta)^{-1}$ in (1.59) near $\lambda = 0$ compensate for the blowing off of $I(\lambda)$, so that square integrability is not required.

Giraitis and Surgailis (1990) generalized the result of Fox and Taqqu to series that are linear in *iid* innovations with finite fourth moment. Hosoya (1997) allowed for martingale differences innovations and other more general models. Heyde and Gay (1993) consider also multivariate models.

In the case of short-range dependence, the asymptotic normality of the Whittle estimate is achieved under a variety of hypotheses about the series X. However, in the LRD case, the class of limiting distributions is richer, and Gaussianity seems to be the exception rather than the rule. Giraitis and Taqqu (1999) considered a process $X = P(Y)$ with spectral density $\theta_1 f(\lambda; \eta)$, where Y is a Gaussian LRD process, and $P(\cdot)$ is a polynomial function. They showed that the Whittle

estimate of H is consistent, but does not converge with rate $\sqrt{N}$, and its asymptotic distribution may not be Gaussian.

The GPH estimators

While the Whittle estimator attains asymptotic efficiency, it requires, as any parametric procedure, global assumptions on the spectral density, introducing model bias if the model chosen is not correct. Short-range features have no influence at low frequencies. If the only interest is the estimation of the long-range dependence parameter, one can disregard modeling the spectral density at high frequencies, and base the estimation only on the lower tail of the periodogram. The estimates will then be valid across several classes of models.

If one assumes that the spectral density, in a neighborhood of the origin, can be approximated by

$$f_{c,H}(\lambda) := c\lambda^{1-2H} \tag{1.63}$$

By talking logarithm, one obtains the following relation

$$\log I(\lambda_j) = \log c + (1-2H)\log \lambda_j + u_j \tag{1.64}$$

A natural estimator of H can be obtained using least squares on the regression model (1.64) restricted to m frequencies in the lower tail of the periodogram, with $m(N) \to \infty$ and $m/N \to 0$ as $N \to \infty$.

This approach was first suggested by Geweke and Porter-Hudak (1983) (GPH), but they considered the more restrictive fractional model (1.16), essentially replacing $\log \lambda_j$ with $\log \mid 1 - e^{i\lambda_j} \mid$.

They argued that

$$m^{1/2}(\hat{H}_{GPH} - H) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{24}\right) \tag{1.65}$$

However, their proof is not formally accurate because they considered the u_j asymptotically uncorrelated and homoscedastic.

Robinson (1995a) gave a formally correct proof of the above result for Gaussian X, and indicated that it holds also in the setting of GPH. Velasco (2000) extended the result to linear non-Gaussian processes.

In the same paper, Robinson showed that the variance in (1.65) can be reduced by "pooling" adjacent periodogram ordinates prior to logging and advocated deleting some of the smallest Fourier frequencies. Likewise, Hurvich and Beltrao (1994); Andrews and Guggenberger (2000); Robinson and Henry (2000) propose estimators that reduce the bias.

One issue with the GPH estimator is how to choose the number of frequencies to use in the regression. GPH, based on empirical observations, proposed $m = \sqrt{N}$. Hurvich and Beltrao (1994) proposed cross-validation, and showed that trimming to the left can adversely affect the GPH estimator. Hurvich et al. (1998) provided a formula for the asymptotic MSE for Gaussian series, and showed that the optimal number of frequencies m_{opt} that minimizes the MSE is equal to $CN^{4/5}$, for some constant C that depends on the spectral density f. Smith (1989), in un unpublished paper, achieved heuristically a similar result. Hurvich and Deo (1999) suggested a plug-in rule for computing C and m_{opt} that results in a consistent estimator and yields to better results than the suggestion of GPH even in small samples.

The main advantages of the GPH estimator is that it is graphical, and it provides a closed form expression for the estimate, and therefore it can be implemented easily. The asymptotic theory, however, is somewhat complicated because of the nonlinearity implied by the logarithms.

The Local Whittle Estimator

The local Whittle estimator (LWE), or Gaussian semiparametric estimator (GSE), is another local semiparametric estimator based on the lower tail of the periodogram. It was initially suggested by Künsch (1987) and later studied by Robinson (1995b). The idea is to combine the efficiency of the Whittle estimator with the flexibility of the GPH procedure.

Like the GPH estimator, it assumes that the spectral density $f(\lambda)$ of the process can be approximated by the function

$$f_{c,H}(\lambda) = c\lambda^{1-2H} \tag{1.66}$$

for frequencies λ in a neighborhood of the origin.

The LWE of the Hurst parameter, $\widehat{H}_{LW}(m)$, is implicitly defined by minimizing the Whittle contrast

$$\sum_{j=1}^{m} \left\{ \log f_{c,H}(\lambda_j) + \frac{I_N(\lambda_j)}{f_{c,H}(\lambda_j)} \right\} \tag{1.67}$$

with respect to c and H, limited to a subset of m frequencies in the lower tail of the periodogram.

Robinson (1995b) showed that, if the innovations in the Wold representation (1.40) are martingales with finite second moments, and under some mild regularity conditions on f in a neighborhood of the origin, the LW estimator is consistent for H in a compact subset of $(0, 1)$, as m diverges at a slower rate than N.

Under slightly stronger conditions, which include finite fourth moments of the innovations, he proved also that

$$\sqrt{m}(\hat{H}_{LW} - H) \to \mathscr{N}\left(0, \frac{1}{4}\right) \tag{1.68}$$

Like the GPH estimator, the LWE is only $\sqrt{m}$ consistent, and, thus, it is less efficient than the Whittle approximate MLE when the correct model is chosen. On the other hand, since it relies

only on local assumptions, it can be applied to a wider range of models, and it's asymptotically more efficient than the GPH estimator.

Again, like the GPH estimator, the LWE depends on the number of frequencies m over which the summation is performed. It should be chosen so as to balance the trade-off between adding more bias as m increases, due to the fact that the approximation (1.66) holds only in a neighborhood of 0, and increasing the variance as m decreases. Henry and Robinson (1996) computed heuristically the asymptotic bias and variance for this estimator and suggested a consistent plug-in rule for choosing the optimal bandwidth m_{opt} that minimizes the asymptotic MSE.

The LWE has been proved to be fairly robust to deviation from standard assumptions and seems to be one of the best performing estimators (Taqqu and Teverovsky, 1998; Bardet et al., 2003).

The FEXP and FAR Estimators

Another estimator based on the log periodogram is the FEXP estimator, initially proposed by Janacek (1987), and studied by Moulines and Soulier (1999) and Hurvich and Brodsky (2001). Unlike the GPH estimator which makes only local assumptions, the FEXP estimation is a *global*, or *broadband*, method, in that it makes assumption on the spectral density over the whole set of Fourier frequencies. The spectral density is assumed to admit a representation on the cosine basis in the form of an FEXP(∞) model, and is approximated by a finite truncation FEXP(p), defined in Equation (1.22).

An ordinary least square regression is performed for the model

$$\log I(\lambda_j) = (1-2H)\log|1-e^{i\lambda_j}| + \sum_{k=0}^{p-1} \frac{\cos(k\lambda)}{\sqrt{(}\pi)} \tag{1.69}$$

Equation (1.69) can also be justified in term of generalized linear models (see Beran, 1993).

As for the GPH case, this method allows for a closed expression of the estimate.

Hurvich et al. (2002) considered the model (1.69) applied to a tapered and pooled version of the periodogram, with X either Gaussian or linear. They showed that, under some regularity conditions on f, if p diverged at a rate slower than $\frac{N}{\log^5 N}$, the FEXP estimator is consistent and asymptotically normal with convergence rate $\sqrt{\frac{N}{p}}$. The FEXP estimator can, therefore, attain a faster convergence that the LWE, but it requires assumptions on the spectral density over the whole spectrum.

The FAR estimator is another global semiparametric estimator. Like the FEXP estimator, it is based on a truncated expansion of the spectral density. In this case, a FAR(p) model is used to approximate f, with p increasing with the sample size. The estimate of H is found minimizing the Whittle contrast, where f is replaced by the spectral density of the finite-order FAR(p) model.

Kokoszka and Bhansali (1999), using an invertible LRD process, proved the consistency and conjectured the asymptotic normality of the FAR estimator if the coefficients of the AR polynomial decay exponentially fast and p diverges at a rate slower than N.

1.4.3 Wavelet Method

The Wavelet estimator differs somewhat from the semiparametric methods that we have presented so far in that it does not use directly the properties implied by either definition of long-range dependence. It does, however, utilize, in an indirect way, properties of the spectral density.

Wavelet transforms have been used to detect features of a signal at different locations and scales. Therefore, they are a natural tool for investigating the scaling properties that are associated with self-similar processes, or long-range dependent processes that are asymptotically self-similar. The

use of wavelet for analyzing LRD was first suggested by Flandrin (1989), and later developed by Abry et al. (1998) and Veitch and Abry (1999).

Let $X := \{X(t) : t \in \mathbb{R}\}$ be a continuous time stationary $\mathbb{L}^2$ process with spectral density (1.16), for $\frac{1}{2} < H < 1$. Assume that X is observed at every point in the interval $[0, N]$. The *coefficient*, or *detail*, of the *discrete wavelet transform of* X *at scale* $n = 2^j$ *and location* k is defined as

$$\begin{aligned} d_X(j,k) &:= 2^{-j/2} \int_{\mathbb{R}} \psi(2^{-j}t - k) X(t) dt \\ &= 2^{j/2} \int_{\mathbb{R}} \psi(t) X(2^j[t+k]) dt \end{aligned} \tag{1.70}$$

where $\psi \in \mathbb{L}^1 \cap \mathbb{L}^2$ is called the *mother wavelet* and is supposed to have Q *vanishing moments*, also called Q *zero moments*, that is

$$\int_{\mathbb{R}} t^k \psi(t) dt = 0 \qquad k = 0, \ldots, Q-1 \tag{1.71}$$

and

$$\int_{\mathbb{R}} t^Q \psi(t) dt \neq 0 \tag{1.72}$$

The value j is called *octave*.

The wavelet coefficients have the following properties.

1. If X is stationary, so is $\{d_X(j,k),\ k \in \mathbb{N}\}$, for any $j \in \mathbb{N}$.

2. If X is Gaussian, so is $\{d_X(j,k),\ k \in \mathbb{N}, j \in \mathbb{N}\}$.

3. If X is LRD, then the Q-moment property of ψ implies that $\{d_X(j,k),\ k \in \mathbb{N}\}$ is a short-range

dependent process for any $j \in \mathbb{N}$. In particular, it can be shown that, for $k \neq k'$ and $j \in \mathbb{N}$,

$$Ed_X(j,k)d_X(j,k') = O\left(|k-k'|^{2(H-Q)-2}\right) \tag{1.73}$$

so that the correlation of the wavelet coefficient decreases as Q increases.

It can be shown (Abry et al., 1998) that

$$\nu_j := Var[d_X(j,k)] \approx 2^{j(2H-1)} f^*(0)\Phi(1-2H) \tag{1.74}$$

where

$$\Phi(\gamma) := \int \lambda^{-\gamma} |\hat{\psi}(\lambda)|^2 d\lambda \tag{1.75}$$

and $\hat{\psi}$ is the Fourier transform of ψ.

This fact suggests to estimate H from the log-regression of $v_j := \log_2(\nu_j)$ on j, since

$$v_j \approx \log_2[f^*(0)\Phi(1-2H)] + 2(H-1)j \tag{1.76}$$

We need then to estimate the variance of the wavelet coefficients. Let $n_j \approx [N2^{-j}]$ be the number of coefficients available at the octave j. The numbers of coefficients roughly decreases by half as the scale doubles. Let

$$\hat{\nu}_{j,N} := \frac{1}{n_j} \sum_{k=1}^{n_j} d_X(j,k)^2 \tag{1.77}$$

be the sample second moment of the wavelet coefficients at octave j. Under some regularity conditions on f^* and ψ (Bardet et al., 2000), $\hat{\nu}_{j,N}$ is a consistent estimator of ν_j. Hence, one can regress $\hat{v}_{j,N} := \log_2(\hat{\nu}_{j,N})$ on $j_1, \ldots, j_p$, where j_1 and j_p are user-chosen trimming numbers such

that j_1 is the lowest octave for which the diagram is linear, and j_p is the largest octave for which v_j can be estimated reliably.

Due to the decreasing number of available coefficients, as j increases, the variance of $\hat{\nu}_{j,N}$ is not constant. It is, therefore, more appropriate to use weighted regression rather than ordinary least squares. Let $\hat{\sigma}_j^2$ be the estimate of the variance of $\hat{\nu}_{j,N}$. Define

$$S_0 := \sum_{i=1}^{p} \hat{\sigma}_{j_i}^{-2} \tag{1.78}$$

$$S_1 := \sum_{i=1}^{p} j_i \hat{\sigma}_{j_i}^{-2} \tag{1.79}$$

$$S_2 := \sum_{i=1}^{p} j_i^2 \hat{\sigma}_{j_i}^{-2} \tag{1.80}$$

Abry et al. (1995) suggested the following estimator for H:

$$\hat{H}_w := \sum_{i=1}^{p} w_i \hat{\nu}_{j_i,N} - \frac{1}{2} \tag{1.81}$$

where

$$w_i := \frac{j_i S_0 - S_1}{2\hat{\sigma}_i^2 (S_2 S_0 - S_1^2)} \tag{1.82}$$

Abry et al. (1998) improved the estimator with some modification to estimate jointly $f^*(0)$ and H, and to improve the finite-sample properties. A heuristical asymptotic theory is presented in Abry et al. (2000), but it is based on the assumption that the wavelet coefficients are iid, which is not verified in the case of LRD processes. Bardet et al. (2000) proved that the wavelet estimator is consistent for a Gaussian process X, under some regularity conditions on f^* and ψ, when $j_1(N)$ diverges at a slower rate than N, and $|j_i(N) - j_1(N)| \leq k$, $i = 2, \ldots, p$, for some finite k. They showed that it is asymptotically normal under the additional condition that

$\lim_{N\to\infty} 2^{-j_1(N)(1+2\beta)}N = 0$, where $\beta > 0$ is a parameter that determines some regularity features of f^* in a non-vanishing neighborhood of the origin. They also proposed a plug-in rule for j_1 that minimizes the asymptotical MSE.

Remarks

- The asymptotic variance of $\hat{H}_w$ depends on H, unlike for the semiparametric estimators based in the periodogram. This implies that confidence intervals or tests based on the asymptotic distribution of $\hat{H}_w$ will be more complicated because they need some estimate of H.

- The scale 2^j is chosen so that one can use the pyramidal algorithm to compute the wavelet coefficients. In reality, the theory holds for any choice of scales that verifies the conditions stated (see Bardet et al., 2003, for details).

- The algorithm is of order $O(N)$, and it has a finite number of steps, unlike the local Whittle estimator whose minimization algorithm step has unknown order.

- The estimator is invariant with respect to non-stationarity induced by polynomial trends $P(t)$ of order less than $Q-1$, since the Q vanishing moments of ψ make so that the wavelet coefficients of $X(t)$ and $X(t)+P(t)$ are equal.

- Empirical studies (Bardet et al., 2003) show that it performs well when compared to other estimators.

Chapter 2

Theoretical Results

In several fields where long range dependence arises the amount of data collected can be such that it is often convenient to aggregate it in bins for analysis purposes. Equivalently, the sampling frequency of a cumulative variable can be decided arbitrarily by the analyst. One such example is network data for which the source is practically endless and the sampling resolution is only limited below by the current technology of the networking monitoring systems. The UNC network data discussed in Chapter 4 were sampled at 1ms intervals. `Splus` could not compute the Fourier transform of the original data because it exhausted the memory resources. We had to resort to memory-efficient programs written in C language to able to handle data sets of such dimension. Failing that, the only other option would have been to aggregate the data or, equivalently, increase the length of the sampling interval.

One question that arises immediately is whether the aggregated process has the same long-range dependence characteristics of the underlying process.

Let X_t, $t \in \mathbb{Z}$, be a stationary time series with long memory. Denote by $Y_t^{(k)}$ the time series

derived by aggregating X_t at bins of size k:

$$\begin{aligned} Y_t^{(k)} &:= X_{(t-1)k+1} + X_{(t-1)k+2} + \cdots + X_{tk} \\ &= \sum_{j=0}^{k-1} B^j X_{kt} \end{aligned} \tag{2.1}$$

where B is the lag operator. Denote by N the length of the observed series X_t, and assume for simplicity that N is a multiple of k, then $N_k = \frac{N}{k}$ is the length of $Y_t^{(k)}$.

The original process X_t will be called the *basic* process.

One wants to know what is the relationship between the long-memory parameter of the basic process X_t and that of Y_t. Intuitively, one can expect long-range dependence to be invariant under aggregation, since such transformation reduces the basic process to a process of lower frequency, and long memory is related to the low-frequency behavior of the process only. Another reason to expect that aggregation does not interfere with long-range dependence is the close proximity of the class of long-memory processes and the class of self-similar processes for which the autocorrelation function is invariant under aggregation.

A second closely related question is whether and how the estimators of long-range dependence are affected by aggregation. Since we will be concerned only with the Geweke and Porter-Hudak (henceforth GPH) and Local Whittle (henceforth LW) estimators, that translate to investigate how the mean squared errors (MSE's) and the optimal choice of the bandwidth parameter m varies under aggregation.

Section 2.1 will present results that answer the first question, while Section 2.2 will present results that answer the second question for the GPH and LW estimators. Section 2.3 considers the case when the observed process is composed of a LRD process plus a white noise.

2.1 Aggregation and Long-Range Dependent Processes

In order to understand how aggregation affects long-range dependence we must first see how the long-memory characteristics of the aggregated process Y_t defined in (2.1) relate to the long-range dependence of the basic process X_t.

We will assume that the process X_t has a spectral density of the form

$$f_X(\lambda) = c\lambda^{-2d}\{1 + a\lambda^b + o(\lambda^b)\} \qquad \lambda \to 0 \tag{2.2}$$

with $c > 0$, $d \in [0, 0.5)$, $b \in (0, 2]$. Such a spectral density is a fairly general way to represent long memory and has been assumed, for example, by Smith (1989) and Robinson (1995a).

Likewise, we will assume that the aggregated process $Y_t^{(k)}$ has a spectral density of the form

$$f_k(\lambda) = c_k\lambda^{-2d_k}\{1 + a_k\lambda^{b_k} + o(\lambda^{b_k})\} \tag{2.3}$$

The problem becomes then to find the relationship between the parameters (a, b, c, d) of (2.2) and the parameters (a_k, b_k, c_k, d_k) of (2.3).

Teles et al. (1999) considered the case when X_t is a stationary and invertible FARIMA(p, d, q) process and proved that Y_t is a FARIMA(p, d, ∞) process. The proof is based on earlier results of Wei (1990) about ARIMA processes that extend without modification to the case of fractional d.

Chambers (1998) investigated the more general case when the basic process X_t admits the Wold representation

$$(1 - B)^{(\delta+d)}X_t = \sum_{h=0}^{\infty} \psi_h e_{t-h} \tag{2.4}$$

where $\psi_0 = 1$, $\sum |\psi_h| < \infty$, e_t is a white noise sequence with variance σ_e^2, $\delta \in \mathbb{Z}$ and $-0.5 < d < 0.5$

is the usual long-range dependence parameter. Chambers considered two cases:

1. $Y_t^{(k)}$ is a *stock* variable, that is a variable that is observed once every k intervals, so that $Y_t^{(k)} = X_{tk}$

2. $Y_t^{(k)}$ is a *flow* variable, that is $Y_t^{(k)}$ is the sum of the values of X_t in the previous $k-1$ subperiods, as defined in Equation (2.1).

He showed that for a stock variable, the spectral density of $Y_t^{(k)}$ is given by

$$f^S(\lambda) = \frac{\sigma_e^2}{2\pi}|1 - e^{-i\lambda/k}|^{-2(\delta+d)}\psi(e^{-i\lambda/k})\psi(e^{i\lambda/k}) \qquad -\pi < \lambda < \pi \tag{2.5}$$

where $\psi(z) = \sum_{h=0}^{\infty} \psi^h z^h$, and satisfies

$$f^S(\lambda) \sim \frac{\sigma_e^2}{2\pi}\psi(1)^2|\lambda|^{-2(\delta+d)}k^{2(\delta+d)} \qquad \text{as } \lambda \to 0. \tag{2.6}$$

For a flow variable, the spectral density of $Y_t^{(k)}$ is

$$f^F(\lambda) = \frac{\sigma_e^2}{2\pi}|1 - e^{-i\lambda}|^{-2(\delta+d)}\psi(e^{-i\lambda})\psi(e^{i\lambda})|\sum_{j=0}^{k} e^{ij\lambda}|^2 \qquad -\pi < \lambda < \pi \tag{2.7}$$

and satisfies

$$f^F(\lambda) \sim \frac{\sigma_e^2}{2\pi}\psi(1)^2|\lambda|^{-2(\delta+d)}k^{2+2(\delta+d)} \qquad \text{as } \lambda \to 0. \tag{2.8}$$

As already seen for the FARIMA models, Chambers's results imply that the fractional order of the aggregated process is the same as that of the basic process.

A related result was derived by Hannig et al. (2001). They considered a stationary time series with sample autocorrelation function $\rho_X(l)$, $l \in \mathbb{Z}$, and binning intervals of size $k = 2$. They showed

that for a smooth $\rho_X(l)$ such that $\rho_X(2l-1)+\rho_X(2l+1) \approx 2\rho_X(l)$,

$$\rho_2(l) \approx \frac{2\rho_X(l)}{1+\rho_X(1)} \tag{2.9}$$

where ρ_2 is the autocorrelation function of $Y_t^{(2)}$.

Similarly, for the periodogram, they showed that

$$I_2(\lambda) \approx 2I_x\left(\frac{\lambda}{2}\right) \qquad -\pi < \lambda < \pi \tag{2.10}$$

A similar but more general result for the spectral density has been derived by Cao in an unpublished note (see Appendix B). He showed that for a stationary time series X_t, $t \in \mathbb{Z}$, with spectral density $f_X(\lambda)$, $\lambda \in [0, 2\pi)$ and the aggregated process $Y_t^{(k)}$ defined in Equation (2.1), the spectral density of $Y_t^{(k)}$ is given by

$$f_k(\lambda) = \frac{1}{k}\sum_{j=0}^{k-1} \frac{\sin^2\frac{\lambda}{2}}{\sin^2\frac{\lambda+2\pi j}{2k}} f_X\left(\frac{\lambda+2\pi j}{k}\right) \tag{2.11}$$

Since, to our knowledge, this result has not been published, we are reporting the full proof in Appendix B.

We can use the above result to find the relationship between the spectral density of X_t and $Y_t^{(k)}$.

First note that if X_t is a white noise process, then

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}, \quad f_k(\lambda) = \frac{k\sigma^2}{2\pi} \tag{2.12}$$

Substituting in (2.11), it gives the following identity for any λ.

$$k^2 = \sum_{j=0}^{k-1} \frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda+2\pi j}{2k}} \tag{2.13}$$

In a neighborhood of the origin, we have

$$\lim_{\lambda \to 0} \frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda+2\pi j}{2k}} = \begin{cases} k^2 & \text{if } j = 0 \\ 0 & \text{if } j > 0 \end{cases} \tag{2.14}$$

Hence, for small λ

$$f_k(\lambda) \approx k f_X\left(\frac{\lambda}{k}\right) \tag{2.15}$$

which is similar to the result (2.10) that Hannig et al. (2001) developed for the periodogram.

When $j = 0$, a Taylor series expansion around the origin of the numerator and denominator in the left-hand side of (2.14) gives

$$\frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda}{2k}} = k^2\left(1 - \frac{\lambda^2}{12}\right) + o(\lambda^2) \tag{2.16}$$

while, when $j > 0$,

$$\frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda+2\pi j}{2k}} = \frac{\lambda^2}{4 \sin^2 \frac{\pi j}{k}} + o(\lambda^2) \tag{2.17}$$

Combining the leading terms of Equation (2.16) and Equation (2.17) with Cao's result (2.14), we

have

$$\begin{aligned} f_k(\lambda) &= \frac{k^2}{k}\left(1-\frac{\lambda^2}{12}\right) f_X\left(\frac{\lambda}{k}\right) + \frac{1}{k}\sum_{j=1}^{k-1} \frac{\lambda^2}{4\sin^2\frac{\pi j}{k}} f_X\left(\frac{\lambda+2\pi j}{k}\right) \\ &= k\left(1-\frac{\lambda^2}{12}\right) c\left(\frac{\lambda}{k}\right)^{-2d}\left\{1+a\left(\frac{\lambda}{k}\right)^b\right\} \\ &\quad + \frac{1}{k}\sum_{j=1}^{k-1} \frac{\lambda^2}{4\sin^2\frac{\pi j}{k}}\left\{f_X\left(\frac{2\pi j}{k}\right)+O(\lambda)\right\} \\ &\approx ck^{1+2d}\lambda^{-2d}\left[\left(1-\frac{\lambda^2}{12}\right)\left\{1+a\left(\frac{\lambda}{k}\right)^b\right\}\right. \\ &\quad \left. + \frac{\lambda^{2d+2}}{4ck^{2d+2}}\sum_{j=1}^{k-1}\frac{1}{\sin^2\frac{\pi j}{k}} f_X\left(\frac{2\pi j}{k}\right)\right] \end{aligned}$$

The errors are or order $o(\lambda^2)+o(\lambda^b)$ in the the first term inside the square brackets, and of order $O(\lambda^{2d+3})$ in the second term. It follows immediately that

$$c_k = k^{1+2d}c \tag{2.18}$$

$$d_k = d \tag{2.19}$$

$$b_k = b \tag{2.20}$$

and

$$a_k = \begin{cases} ak^{-b} & \text{if } b<2 \\ ak^{-b}-\frac{1}{12} & \text{if } b=2 \end{cases} \tag{2.21}$$

We can state the following Proposition.

Proposition 2.1.1 (Invariance of Long Memory)**.** *Let X_t, $t\in\mathbb{Z}$, be a long-memory stationary time series characterized by the spectral density* (2.2)*. Let $Y_t^{(k)}$ be the time series defined in* (2.1)

obtained by aggregating X_t at bins of size k. Then $Y_t^{(k)}$ and X_t have the same fractional order.

2.2 Aggregation and Estimators of Long-Range Dependence

To understand how aggregation affects the performance of the estimators of long-range dependence we need to be able to compare the estimators' mean squared error (MSE) at different aggregation levels. As mentioned before, we will focus our attention on the GPH and on the LW estimators.

2.2.1 Aggregation and the GPH Estimator

Hurvich et al. (1998) considered a stationary Gaussian long-memory time series X_t, $t \in \mathbb{Z}$, with spectral density

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) \tag{2.22}$$

where $d \in (-0.5, 0.5)$ and $f^*(\cdot)$ is an even, positive, continuous function on $[-\pi, \pi]$, bounded above and bounded away from zero with first derivative at zero $f^{*'}(0) = 0$ and second and third derivatives bounded in a neighborhood of the origin.

The function f^* models the short-term correlation structure of the process. For example, for a FARIMA(p, d, q) series, f^* is the spectral density of an ARMA(p, q) process.

They showed that, for $N \to \infty$, $m \to \infty$ at a slower rate than N such that $m \log m/N \to 0$, the asymptotic MSE of the GPH estimator of d is

$$\text{MSE}(\hat{d}_{GPH}) = \frac{4\pi^4}{81} \left\{ \frac{f^{*''}(0)}{f^*(0)} \right\} \frac{m^4}{N^4} + \frac{\pi^2}{24m} + O\left\{ \frac{m(\log^3 m)}{N^2} \right\} + o\left(\frac{m^4}{N^4}\right) + o\left(\frac{1}{m}\right) \tag{2.23}$$

which, ignoring the remainders, leads to the optimal choice of m

$$m_{opt} = CN^{4/5} \tag{2.24}$$

where

$$C := \left(\frac{27}{128\pi}\right)^{1/5} K^{2/5} \tag{2.25}$$

and $K := \frac{f^{*''}(0)}{f^*(0)}$.

In a follow-up paper, Hurvich and Deo (1999) suggested a plug-in estimator for C. They considered a Taylor expansion of $\log f^*$ at the origin which yields the relation

$$\log I(\lambda_j) = \log f^*(0) - \gamma - 2d \log|2\sin(\lambda_j/2)| + \frac{\lambda_j^2}{2}K + \frac{\lambda_j^3}{6}R_j + \varepsilon_j \tag{2.26}$$

where $\varepsilon_j = \log I(\lambda_j)/f(\lambda_j) + \gamma$, $\gamma = 0.577216\ldots$ is Euler's constant, and R_j is uniformly bounded in a neighborhood of the origin. Hence, they suggested to estimate K from the second slope coefficient of the the regression of $\log I(\lambda_j)$ on $\log|2\sin(\lambda_j/2)|$ and $\lambda_j^2/2$, for $j = 1, 2, \ldots, L$ where L is an arbitrary number of frequencies. They showed that, if $L \propto n^6$, the plug-in estimator of C obtained by replacing K with $\hat{K}$ in Equation (2.25) is a consistent estimator for C.

In an unpublished work Smith (1989) considered a spectral density of the form (2.2). The spectral density of Hurvich et al. (1998) implies that $b = 2$ in (2.2).

Under this assumption, Smith showed heuristically that the asymptotic mean squared error of the GPH estimator is

$$MSE_{GPH} = A^2 \left(\frac{m}{N}\right)^{2b} + \frac{B}{m} \tag{2.27}$$

where $A = ab(2\pi)^b/2(b+1)^2$ and $B = \pi^2/24$. This leads to the optimal values of m

$$m_{opt} = \left(\frac{\pi^2}{12A^2b}\right)^{\frac{1}{2b+1}} N^{\frac{2b}{2b+1}} \tag{2.28}$$

corresponding to the MSE

$$MSE_{opt} = \frac{2b+1}{2b}(2bA^2)^{\frac{1}{2b+1}} \left(\frac{\pi^2}{6N}\right)^{\frac{2b}{2b+1}} \tag{2.29}$$

We provide a proof of Smith's results in Appendix A. The proof is based on results in Robinson (1995b) and Hurvich et al. (1998).

If we replace in Equations (2.28) and (2.29) the expression of the parameters of the spectral density of $Y_t^{(k)}$ in terms of the parameters of the basic process given in Equations (2.18) through (2.20), and replace N by $N_k = \frac{N}{k}$, we can immediately verify that m_{opt} and MSE are unchanged by the aggregation. This allows us to state the following Proposition.

Proposition 2.2.1 (Invariance of the GPH estimator)**.** *Let X_t, $t \in \mathbb{Z}$, be a long-memory stationary time series characterized by the spectral density*

$$f_X(\lambda) = c\lambda^{-2d}\{1 + a\lambda^b + o(\lambda^b)\} \tag{2.30}$$

with $c > 0$, $d \in [0, 0.5)$, $a > 0$, $0 < b < 2$. Denote by $Y_t^{(k)}$ the time series derived by aggregating X_t at bins of size k:

$$Y_t^{(k)} := \sum_{j=0}^{k-1} B^j X_{kt} \tag{2.31}$$

where B is the lag operator. Then the asymptotic efficiency of the GPH estimator of the long-memory parameter d is invariant under aggregation. In particular, the optimal values of the band-

width parameter m at which the minimum of the asymptotic MSE is attained is also invariant under aggregation.

If $b = 2$, then the order of $m_{opt}^{(k)}$ will still be the same as m_{opt}, but its value will depend on k and a. More precisely, $m_{opt}^{(k)} = \frac{a^2}{\left(a - \frac{k^2}{12}\right)^2} m_{opt}$, which implies that $m_{opt}^{(k)} < m_{opt}$ if $k^2 > 24a$.

2.2.2 Aggregation and the Local Whittle Estimator

Henry and Robinson (1996) computed heuristically the asymptotic bias and variance for this estimator and suggested a consistent plug-in rule for choosing the optimal bandwidth m_{opt} that minimizes the asymptotic MSE. They considered the same hypotheses that lead to the asymptotic normality in Robinson (1995b), but extended the model to have spectral density similar to the one in Smith (1989):

$$f(\lambda) = c\lambda^{-2d}[1 + a\lambda^b + o(\lambda^b)] \tag{2.32}$$

as $\lambda \to 0^+$, for some $b \in (0, 2]$.

They found that the MSE can be approximated by

$$\frac{1}{4}\left[\frac{1}{m} + a^2 \frac{b^2}{(b+1)^4} \lambda_m^{2b}\right] \tag{2.33}$$

which leads to

$$m_{opt} = \left[\frac{(b+1)^4}{2b^3 a^2 (2\pi)^{2b}}\right]^{\frac{1}{1+2b}} N^{\frac{2b}{1+2b}} \tag{2.34}$$

In particular, for $b = 2$,

$$m_{opt} = \left(\frac{3N}{4\pi}\right)^{4/5} |a|^{-2/5} \tag{2.35}$$

For the model (2.32), with $b = 2$, they suggested the following iterative plug-in rule:

1. Set $\hat{m}^{(0)} = N^{4/5}$.

2. Find $\hat{c}^{(s)}(\hat{m}^{(s)})$ and $\hat{d}^{(s)}(\hat{m}^{(s)})$ using the LWE.

3. Perform the regression

$$\frac{I(\lambda_j)}{\hat{c}^{(s)}\lambda^{-2\hat{d}^{(s)}}} = \alpha^{(s)} + \beta^{(s)}\lambda_j^2 j = 1, 2, \ldots, \hat{m}^{(s)} \tag{2.36}$$

   and set $a(\hat{d}^{(s)}) = \hat{\beta}^{(s)}$.

4. Let

$$\hat{m}^{(s+1)} = \left(\frac{3N}{4\pi}\right)^{4/5} \left|a(\hat{d}^{(s)})\right|^{-2/5} \tag{2.37}$$

   Then repeat steps 1-4 until convergence is attained.

A proof of (2.33) would require an evaluation of the bias $E(\hat{d}_{LW} - d)$. Since the LW estimator is defined implicitly, a direct computation of the bias, as it is done by Deo and Hurvich (2003) for the GPH estimator, is not possible. We attempted to bound the expected value of the bias by some quantities that would converge to zero or to some negligible error term, but failed to do so. In fact, a direct proof may very well be impossible, as suggested by Deo and Hurvich (2001), and then again by Cliff Hurvich in a private conversation.

An indirect proof of (2.33) was given recently by Andrews and Sun (2004) while introducing the class of *adaptive local polynomial Whittle estimators* (ALPW) that generalizes the LW estimator. They considered a spectral density of the form

$$f(\lambda) = |\lambda|^{-2d}\varphi(\lambda) \tag{2.38}$$

where $d \in [d_1, d_2], -1/2 < d_1 < d_2 < 1/2$ and $\varphi(\lambda)$ is a positive bounded at zero function that

characterizes the SRD behavior. They assumed regularity condition in a neighborhood of zero for $\varphi(\lambda)$ such that its logarithm can be approximated by an even polynomial of degree $2r$, viz. $\log G - \sum_{k=1}^{r} \theta_k \lambda^{2k}$. This approximation is used to specify a *local polynomial Whittle* (LPW) likelihood function. They provide an adaptive estimator of d (ALPW) that uses the data to select r and m, up to a constant. Under suitable smoothness conditions on $\varphi(\lambda)$, this estimator is shown to be $N^{1/2-\delta}$-consistent for all $\delta > 0$ and, hence, attaining a rate of convergence that is arbitrarily close to the parametric rate. Their proof of the asymptotic properties of the ALPW estimator differs from the typical method for LW estimators pioneered by Robinson (1995b) which consists in establishing consistency first, and, then, assume consistency when proving the normality of the estimator. Such an approach would be problematic for the ALPW estimator because the LPW log-likelihood becomes flat as a function of θ as $n \to \infty$, and the rate at which it flattens differs for each element of θ. In their proof, Andrews and Sun establish consistency and asymptotic normality simultaneously. First they show that there exists a solution to the first order conditions with probability that goes to one as n diverges, and this solution is consistent and asymptotically normal. Then they show that the negative LPW log-likelihood is a strictly convex function of the parameters. This implies that, if there is a solution to the first order conditions, then it is unique and equals the minimizing values. Therefore the solution must be consistent and asymptotically normal. Andrews and Sun's approach has two advantages. First, as mentioned, it eliminates the need of a separate proof for consistency. Second, if one lets m diverge to infinity at what is found to be the asymptotically MSE-optimal rate, the asymptotic mean of the estimator can be interpreted as the asymptotic bias. In their Theorem 2 and Corollary 1, Andrews and Sun show that the asymptotic bias of the LW estimator, as defined above, is equal to the expression found by Henry and Robinson (1996).

Similarly to what we did for the GPH estimator, we can replace in Equations (2.33) and (2.34) the parameters of spectral density of X_t with the the expression of the parameters of the spectral density of $Y_t^{(k)}$ in terms of the parameters of the basic process given in Equations (2.18) through (2.20), and replace N by $N_k = \frac{N}{k}$. We can immediately verify that MSE and m_{opt} are unchanged by the aggregation. This allows us to state a Lemma similar to the one we stated the GPH estimator.

Proposition 2.2.2 (Invariance of the LW estimator)**.** *Let X_t, $t \in \mathbb{Z}$, be a long-memory stationary time series characterized by the spectral density*

$$f_X(\lambda) = c\lambda^{-2d}\{1 + a\lambda^b + o(\lambda^b)\} \tag{2.39}$$

with $c > 0$, $d \in [0, 0.5)$, $a > 0$, $0 < b < 2$. Denote by $Y_t^{(k)}$ the time series derived by aggregating X_t at bins of size k:

$$Y_t^{(k)} := \sum_{j=0}^{k-1} B^j X_{kt} \tag{2.40}$$

where B is the lag operator. Then the asymptotic efficiency of the Local Whittle estimator of the long-memory parameter d is invariant under aggregation. In particular, the optimal values of the bandwidth parameter m at which the minimum of the asymptotic MSE is attained is also invariant under aggregation.

If $b = 2$, then the order of $m_{opt}^{(k)}$ will still be the same as m_{opt}, but its value will depend on k and a. More precisely, $m_{opt}^{(k)} = \frac{a^2}{\left(a - \frac{k^2}{12}\right)^2} m_{opt}$, which implies that $m_{opt}^{(k)} < m_{opt}$ if $k^2 > 24a$.

2.3 Long Memory with Added Noise

In this section we will consider the estimation of the long-memory parameter d when X_t is not observed directly, but, instead, the process

$$Z_t = \mu + X_t + \varepsilon_t \tag{2.41}$$

is observed, where μ is the mean and ε_t is a serially uncorrelated process with mean zero and variance σ_ε^2, independent of X_t.

This case emerges naturally from the Long Memory Stochastic Volatility (LMSV) model in which the process of returns r_t follows the equation

$$r_t = \sigma_t \zeta_t \qquad \sigma_t = \sigma e^{\frac{X_t}{2}} \tag{2.42}$$

where ζ_t is a white noise process with unit variance and X_t is a long-memory process, usually a FARIMA process. Taking the log of the squared returns one has

$$\begin{aligned} Z_t &= \log r_t^2 \\ &= \log \sigma^2 + E(\log \zeta_t^2) + X_t + [\log \zeta_t^2 - E(\log \zeta_t^2)] \\ &= \mu + X_t + \varepsilon_t \end{aligned}$$

See Crato and Ray (2002) and Deo and Hurvich (2001) and references therein for more details on the LMSV model.

The ratio $\frac{\sigma_\varepsilon^2}{\sigma_x^2}$ is named *noise-to-signal ratio* (ns). The higher ns, the more difficult it is to estimate the parameters of the signal process X_t.

2.3.1 The GPH Estimator with Added Noise

Deo and Hurvich (2001) assumed that X_t has spectral density (2.22) so that the spectral density of Z_t has an added component due to the white noise

$$f_Z(\lambda) = |2\sin\left(\frac{\lambda}{2}\right)|^{-2d} f^*(\lambda) + \frac{\sigma_\varepsilon^2}{2\pi} \tag{2.43}$$

They derived the MSE of the GPH estimator and showed that, for $n \to \infty$, $m \to \infty$, and $N^{-2d} m^{2d} \log m \to \infty$, then

$$E(\hat{d}_{GPH} - d) = -(2\pi)^{2d} \frac{\sigma_\varepsilon^2}{2\pi f^*(0)} \frac{d}{(2d+1)^2} \left(\frac{m^{2d}}{2^{2d}}\right) + O\left(\frac{\log^3 m}{m}\right) + o\left(\frac{m^{2d}}{n^{2d}}\right) \tag{2.44}$$

and

$$Var(\hat{d}_{GPH}) = \frac{\pi}{24m} + o(m^{-1}) + O\left(\frac{m^{4d}}{n^{4d}\log^2 m}\right) \tag{2.45}$$

which imply that $\hat{d}_{GPH}$ is consistent for d if $m = Kn^\delta$, for some constant K and $0 < \delta < 1$. The first term in the bias is due to ε_t and is dominant only if $\delta > (2d+1)^{-1}2d$. Hence $\hat{d}_{GPH}$ will have increasingly negative bias as m and as the ns ratio increase.

Furthermore, the showed that asymptotic normality still holds. Under the stronger conditions that $n^{-4d} m^{4d+1} \log^2 m \to 0$ and $\log^2 n = o(m)$,

$$m^{1/2}(\hat{d}_{GPH} - d) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{24}\right) \tag{2.46}$$

That is, the limiting distribution is unchanged compared to the one found by Robinson (1995a) and Hurvich et al. (1998).

2.3.2 The Extended LW Estimator

Hurvich and Ray (2001) considered the following parametrization for the spectral density of Z_t

$$f_\theta(\lambda) = b_0(1 + b_1\lambda^{-2d}) \qquad \lambda \downarrow 0 \tag{2.47}$$

where $\theta = (b_0, b_1, d)$ are the parameters. They proposed an extension to the Local Whittle estimator that consists in minimizing the contrast

$$L(\theta) = \sum_{j=1}^{m} \left[\log f_\theta(\lambda_j) + \frac{I_Z(\lambda_j)}{f_\theta(\lambda_j)}\right] \tag{2.48}$$

with respect to θ. They conjectured that the estimate of $\hat{d}$ is still asymptotically normal with asymptotic variance $(4m \log^2 \lambda_m)$ when $m^{-1} + m^5/n \to 0$. In a Monte Carlo study, they showed that the extended LW (ELW) estimator performs better than the GPH or LW estimator with two parameters when the noise-to-signal ratio is high.

Recently, Hurvich et al. (2005) considered a more general model that encompasses both the LMSV model and the fractionally integrated exponential GARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996). Their spectral density has the form

$$G\lambda^{-2d}\left(1 + h(d, \theta, \lambda)\right) \tag{2.49}$$

with

$$h(d, \theta, \lambda) = \theta_0\lambda^{2d} + \sum_{j=1}^{u} \theta_j \lambda^{\beta_j} \tag{2.50}$$

and β_j, $j = 1, \ldots, u$ is a known sequence of positive numbers.

They showed that the ELW estimator obtained by minimizing the Whittle contrast (2.48) is

consistent for $d \in (0,1)$ and asymptotically normal for $d \in (0,3/4)$. In particular, if the spectral density of the short memory component of the signal is sufficiently smooth, a convergence rate of $N^{2/5-\delta}$ can be attained, where $\delta > 0$ is arbitrarily small.

We considered the following parametrization of the LMSV model

$$f_\theta(\lambda_i) = \alpha \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\} \tag{2.51}$$

where $\theta = (\alpha, \beta, \gamma)$ and considered the ELW estimator that minimizes the contrast (2.48). We heuristically computed the expected value of the Hessian matrix of $\hat{\theta}$ as a first step to determine the MSE. The details of the computations are in Appendix A.

If we were able to compute analytically the asymptotic MSE for such a model, we could compare the MSE of the model at different aggregation levels, in a similar fashion to what we already did for the GPH and LWE. In particular, we could to determine if there exists an optimal aggregation level at which the MSE is minimized. However, the complexity of the expected value of the Hessian matrix in Appendix A does not allow us to find an analytical solution. We will have to resort to numerical evaluations of the MSE, which is the topic of the next chapter.

2.3.3 The LW estimator with of Added Noise

If the underlying model contains added noise, and the LW estimator is used to estimate d, the MSE of the estimator will likely be greater because of the added bias introduced by the noise component which makes estimating the signal a more difficult task. One can expect that the additional bias be greater as the noise-to-signal ratio increases or d decreases.

Similarly to what Deo and Hurvich (2003) did for the GPH estimator, we can compute the MSE of the LW estimator when noise is added to a LRD process. In Appendix A we show that the bias

of the LW estimator can be approximated by the following expression.

$$\frac{\gamma\delta}{(\delta+1)^2}\left(\frac{2\pi m}{N}\right)^{\delta} - \frac{\alpha\beta_1}{(1-\beta_1)^2}\left(\frac{2\pi m}{N}\right)^{-\beta_1} \tag{2.52}$$

where $\alpha = \frac{\sigma_\varepsilon^2}{2\pi c}$, and $\beta_1 = -2d$.

The first term of equation (2.52) is the bias of the LW estimator when no noise is present. The second term represents the additional bias introduced by adding the noise component. As expected, the second component of the bias increases as ns increases.

The optimal value of the MSE, and the corresponding values of m, can only be found numerically, and will depend on the value of d. Similarly to the ELW, the effect of aggregation on the MSE of the LW estimator when the model is misspecified will be addressed with a numerical analysis in the next chapter.

Chapter 3

Numerical Analysis of Local Whittle Estimators in Presence of Noise

As mentioned in section 2.3, and again in Appendix A, when an error term is added to the underlying model, the MSE of the LW estimator, and, consequentially, the optimal value of the bandwidth parameter m, cannot be evaluated analytically.

In this chapter we present a numerical analysis of the effect of adding noise and aggregating the original process on the MSE of LW estimators, and how these factor influence the optimal choice of m. We considered two estimators: the Local Whittle (LW) and the Extended Local Whittle (ELW) estimator which explicitly accounts for the presence of the noise term.

We simulated 100 FARIMA(0,d,0) series of length $2^{19} \approx 5 * 10^5$ for each value of d in $(0.1, 0.2, 0.3, 0.4)$. The algorithm we used for the simulation is based on the circular embedding of the covariance matrix and it is described in Dietrich and Newsam (1997). The variance of the innovation was set to 1.

Each series $\{x_t\}$ was added a white noise component ε_t. The noise-to-signal ratio ns for the resulting series $z_t = x_t + \varepsilon_t$ takes values $(0, 0.1, 0.5, 1, 1.5)$.

The series z_t was aggregated nine times in bins of two observations. For each level of aggregation

$k = 1, \ldots, 9$, a new series $z_t^{(k)} = z_{2t-1}^{(k-1)} + z_{2t-1}^{(k-1)}$ was created, with $z_t^{(0)} := z_t$. The series $\{z_t^{(k)}\}$ has length $N^{(k)} = 2^{19-k}$, with $N^{(k)}$ ranging from $N^{(1)} = 2^{18} = 262,144$ at the first level of aggregation, to $N^{(9)} = 2^{10} = 1024$ at the highest level of aggregation.

3.1 The LW Estimator

For each value of d, ns, k, and a fixed value of the bandwidth parameter m, the performance of the LW estimator was evaluated by comparing the MSE computed in three different ways:

1. The actual MSE of the LW estimator across all series
2. The asymptotic MSE according to equation (2.33)(HRMSE), as computed by Henry and Robinson (1996)
3. The asymptotic MSE of the LW estimator corrected by the presence of the noise component according to equation (A.16)(CHRMSE).

The optimal value of the bandwidth parameter m which minimizes the MSE was computed for each combination of d, ns, and k and each one of the MSE's expression described above. All computations were done using the R statistical package. The minimization problem for the LW estimator uses the `nlm` function. The scale parameter c was concentrated out of the Whittle contrast so that a faster one-dimensional problem could be solved. When the spectral density of the LRD process has the form $f(\lambda) = |1 - \exp(i\lambda)|^{-2d} f^*(\lambda)$, the constants γ and a in equations (A.16) and (2.33), respectively, can be approximated by $\frac{f^{*''}(0)}{2f^*(0)} + \frac{d}{12}$, as shown in Delgado and Robinson (1996).

Since the LW estimator is defined implicitly, and an explicit expression for the MSE is not available, one cannot use standard minimization algorithms to find the value $m^{(opt)}$ that minimizes the

MSE. We found that iterative plug-in procedures, such as the one proposed by Henry and Robinson (1996), did not converge in most cases. Besides, such a procedure would not be appropriate in this context because it is based on an asymptotic approximation of the MSE, whereas we are trying to compute the minimal value of the actual MSE. Therefore, the value of $m^{(opt)}$ was computed by "brute force". Each MSE expression was evaluated over a range of values of m, and $m^{(opt)}$ was chosen as the values that minimizes the MSE over such range. Unfortunately, this resulted in a heavy computational burden because of the large number on non-linear optimizations required to find the LW estimators. At first, we attempted to use a fixed step between the values of m. The estimated completion time was about 500 days. Noticing that the values of the MSE stabilizes as m becomes progressively large, a variable step which increases as m increases can be used without prejudice for the analysis. By trial and error, we settled on the following formula:

$$m_{i+1} - m_i = \left[\frac{\log(\text{lastm})}{\log(2^9)} * m_i^{1/3} + 10 \right] \tag{3.1}$$

where $\text{lastm} = \min(\left[\frac{N}{2}\right], N^{5/6})$. The value $N^{5/6}$ was chosen because the optimal m, according to the analysis of Henry and Robinson (1996), is of order $O(N^{4/5})$. The values of lastm are 58385, 32768, 18390, 10321, 5792, 3250, 1824, 1024, 574, 322, for aggregation levels $k = 0, 1, \ldots, 9$, respectively. Equation (3.1) provides a reasonable compromise between the need to have a good resolution of m at all aggregations levels and an acceptable computational time. The whole process still took about 30 days to complete on a personal computer with two 2.6GHz Pentium Xeon processors.

Figure 3.1 shows the step interval of m when no aggregation is applied, that is, for $N^0 = 2^{19}$.

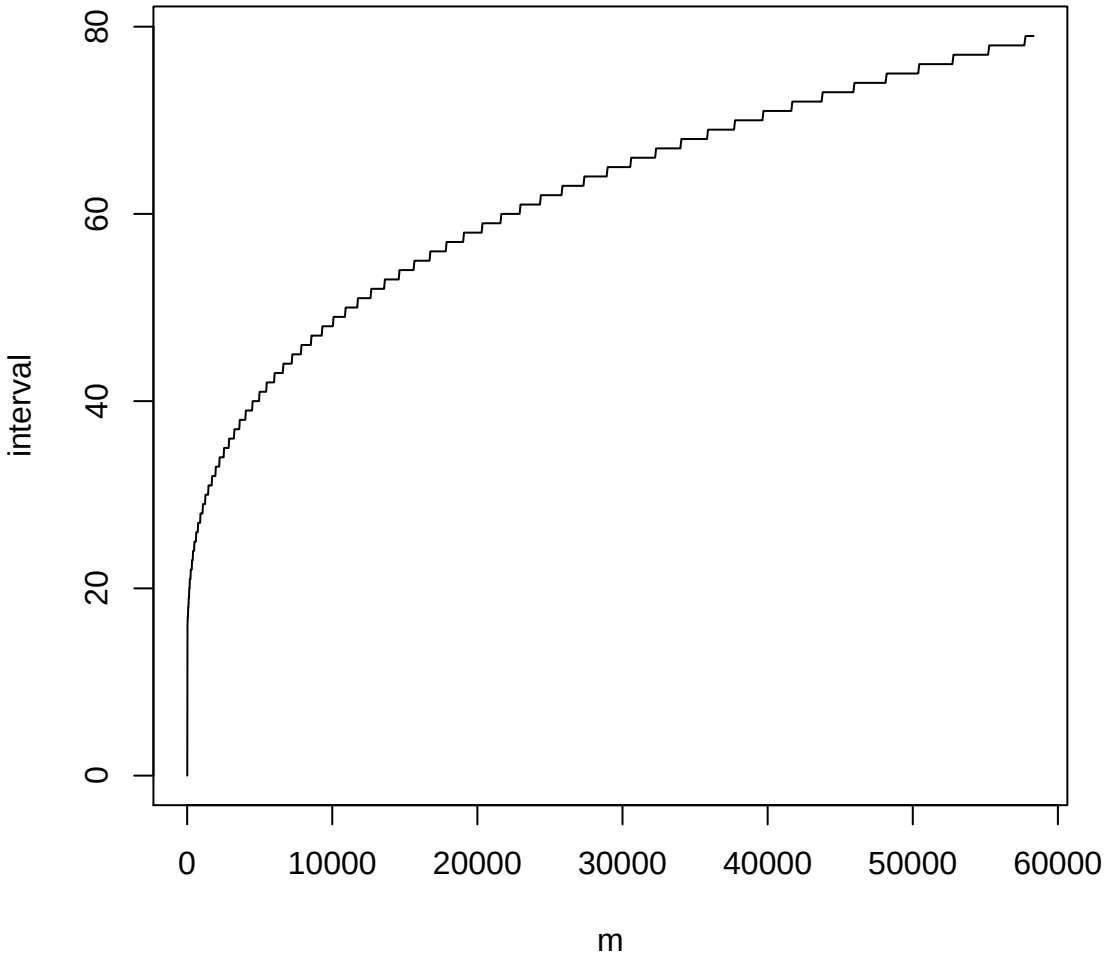

Figure 3.1: Step of m, N=2^{19}

3.1.1 MSEs Comparisons

Figure 3.2 through Figure 3.21 compare, for a given value of long-range dependence parameter d, noise-to-signal ratio ns, and aggregation level k, the actual MSE of the Local Whittle estimator (MSE) of d obtained from the simulated series to the asymptotic expression of the MSE of Henry and Robinson (1996)(HRMSE), equation (2.33), and the similar asymptotic expression (A.16) which accounts for the presence of the noise component (CHRMSE). The plot on the right-hand side display the three estimates of the MSE. The plots on the left-hand side, display the difference between the two asymptotic expression (HRMSE and CHRMSE) and the actual MSE.

Displaying plots for all 200 combinations of values of (d, ns, k) is impractical. Therefore, we are displaying here only the plots relative to $d = 0.1$ and $d = 0.4$, and to $k = 0$, that is, no aggregation is performed, and $k = 9$, the highest level of aggregation. The display is sufficient for illustrating the general behavior. The remaining plots are available on request.

We can observe the following facts.

- When there is no noise ($ns = 0$), HRMSE=CHRMSE. The second term of the bias component of CHRMSE in equation (A.15) is equal to zero because $\alpha = \frac{\sigma_\varepsilon^2}{2\pi c} = 0$. Therefore HRMSE and CHRMSE are equivalent.

- As the number of frequencies increase, both HRMSE and CHRMSE overestimate the actual MSE. This is not entirely unexpected, because both equations account for a bias term due to short-range dependency that is not present in the FARIMA$(0, d, 0)$ processes used in the simulation.

- When the long-range dependence is mild ($d = 0.1$) and N is high, HRMSE and CHRMSE are virtually indistinguishable. This is again because the second term in equation (A.16) is very

close to zero because of the high value of N, and the small value of $\beta_1 = -2d$. Vice versa, as N decreases and d increases, the difference between the two asymptotic MSE expressions becomes more pronounced.

- Similarly, as the noise-to-signal ratio ns increases, so does the difference between HRMSE and CHRMSE.

- As the aggregation level increases, the difference between the asymptotic expressions and the actual MSE increases.

Both asymptotic expressions tend to underestimate the actual MSE when m is low, and to overestimate it when m is high. The overestimation at large frequencies is more pronounced for the CHRMSE and when d and ns are high, and N is low, for the reasons discussed above. The actual MSE of the LW estimator is almost unchanged when the noise is added. The performance of the LW estimator when applied to the purely LRD processes FARIMA$(0, d, 0)$, does not seem to be affected by the addition of the noise component as much as the theoretical results would suggest.

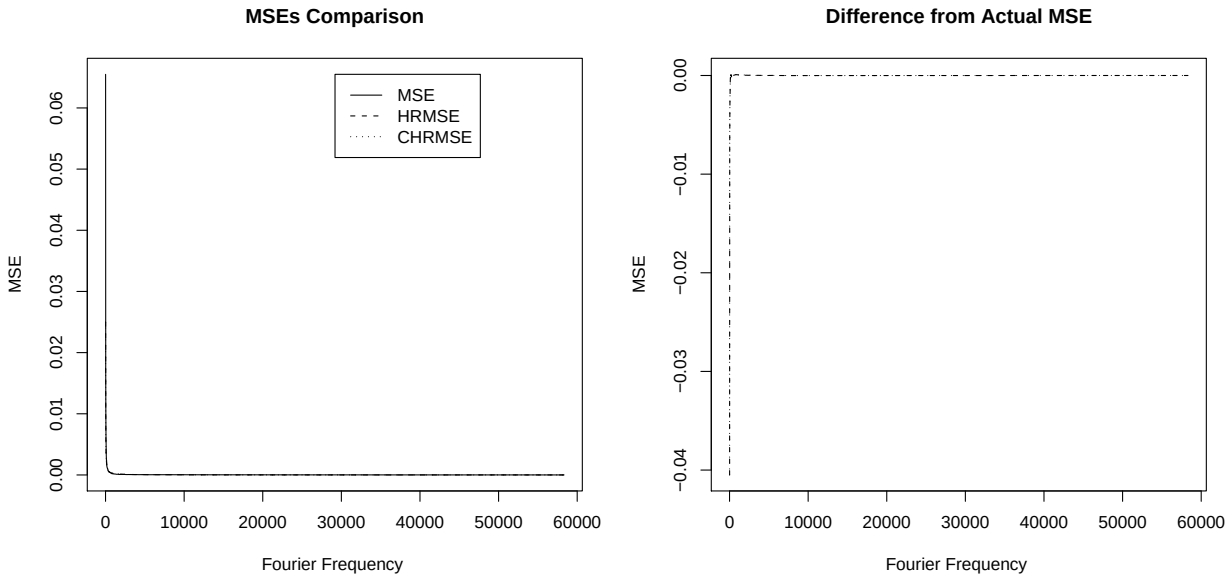

Figure 3.2: Comparison of MSEs, $d = 0.1$, $ns = 0$, $N = 2^{19}$

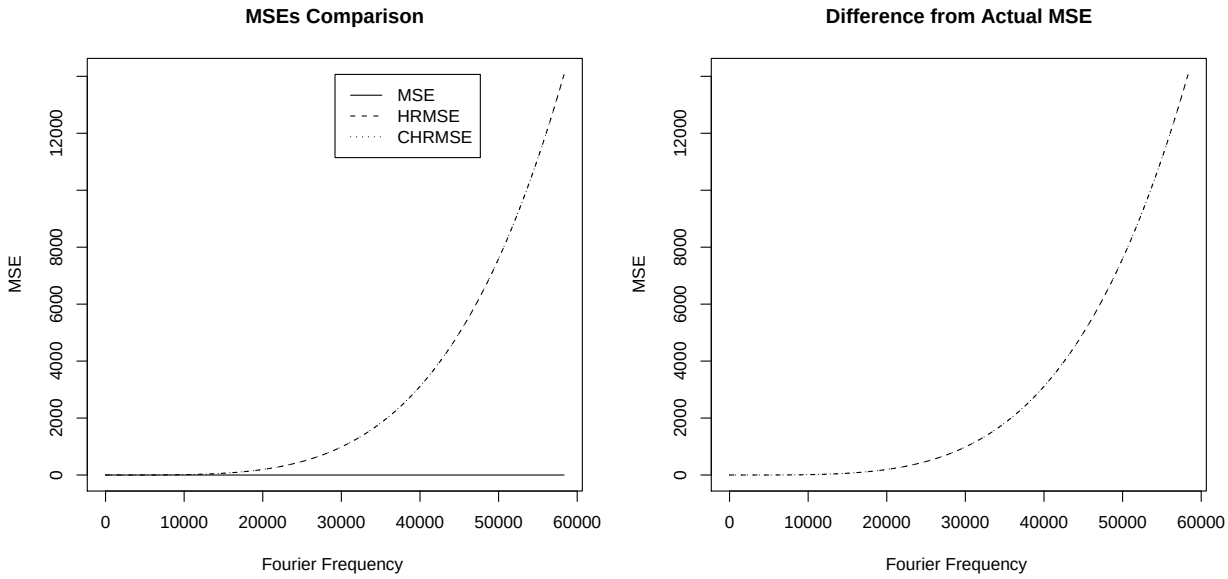

Figure 3.3: Comparison of MSEs, $d = 0.1$, $ns = 0.1$, $N = 2^{19}$

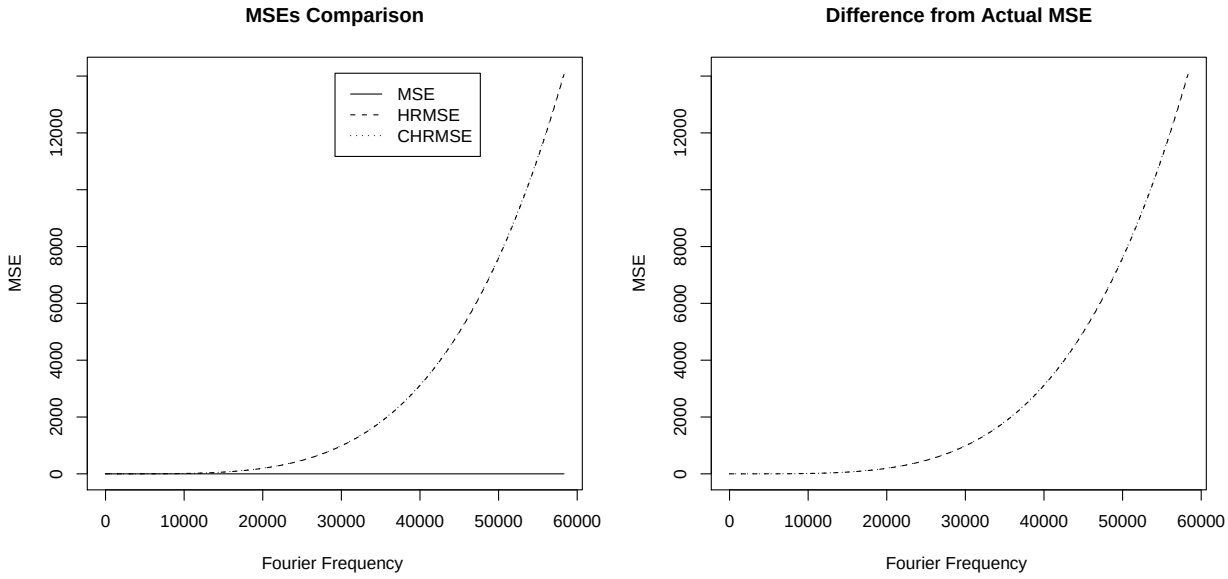

Figure 3.4: Comparison of MSEs, $d = 0.1$, $ns = 0.5$, $N = 2^{19}$

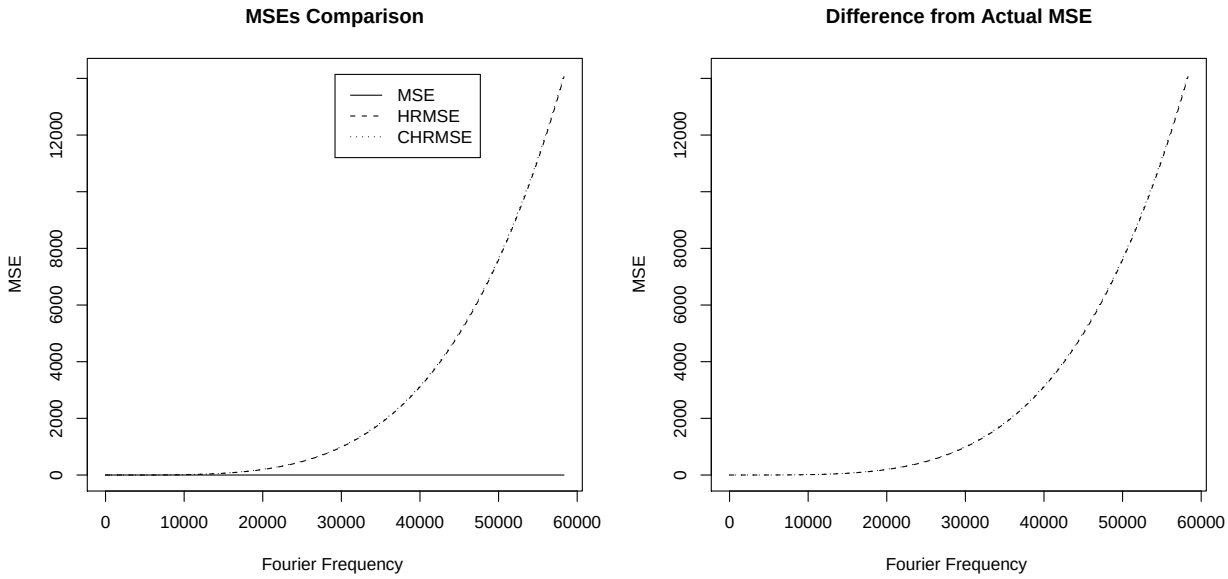

Figure 3.5: Comparison of MSEs, $d = 0.1$, $ns = 1$, $N = 2^{19}$

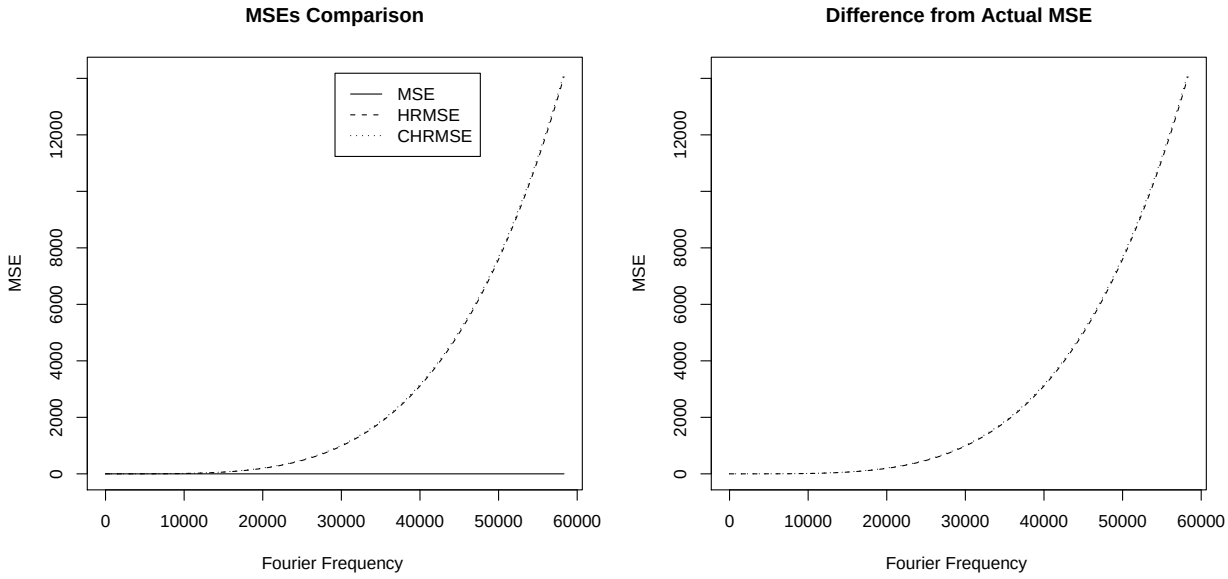

Figure 3.6: Comparison of MSEs, $d = 0.1$, $ns = 1.5$, $N = 2^{19}$

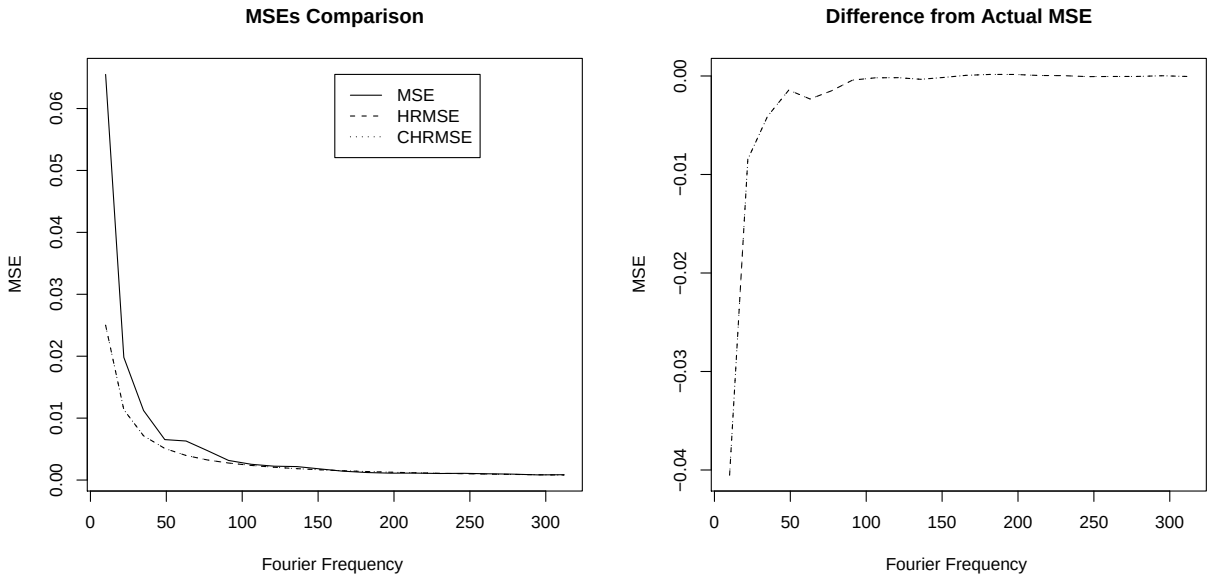

Figure 3.7: Comparison of MSEs, $d = 0.1$, $ns = 0$, $N = 2^{10}$

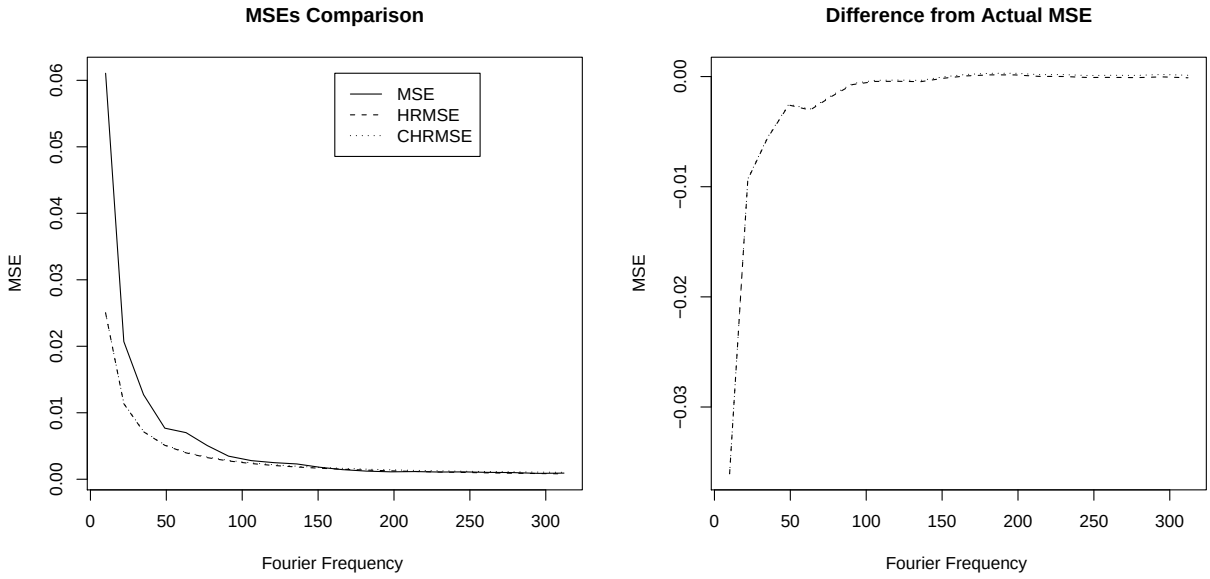

Figure 3.8: Comparison of MSEs, $d = 0.1$, $ns = 0.1$, $N = 2^{10}$

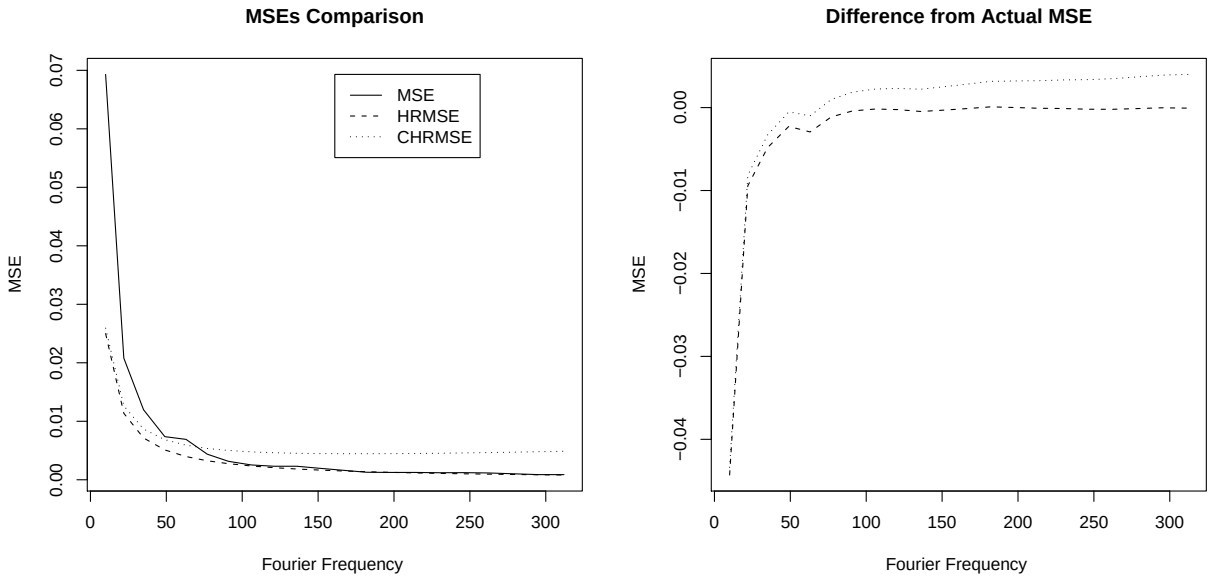

Figure 3.9: Comparison of MSEs, $d = 0.1$, $ns = 0.5$, $N = 2^{10}$

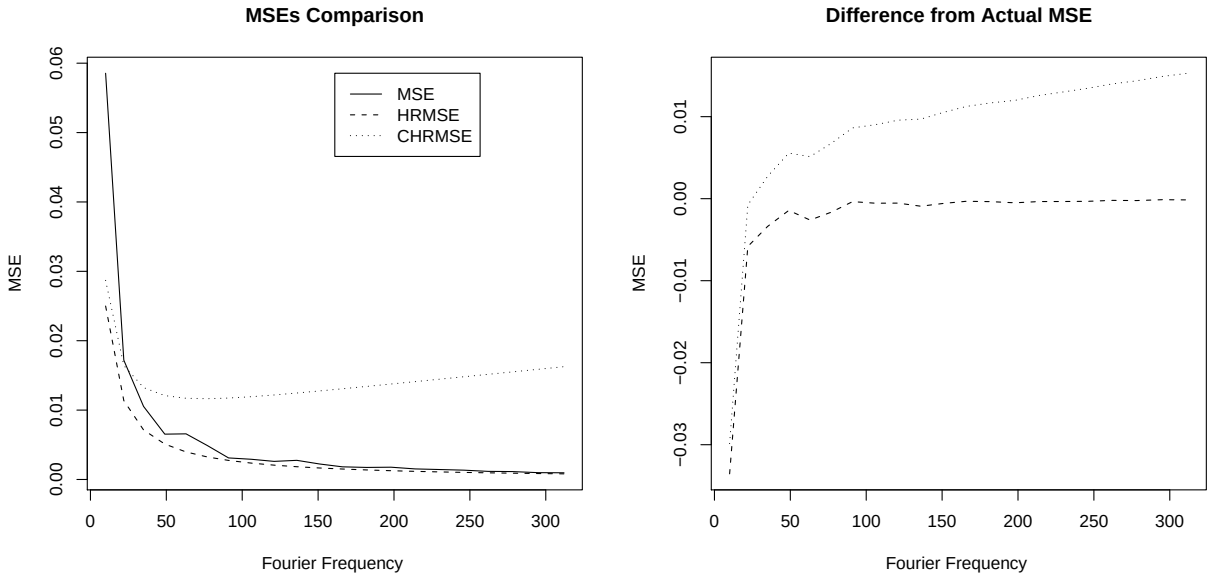

Figure 3.10: Comparison of MSEs, $d = 0.1$, $ns = 1$, $N = 2^{10}$

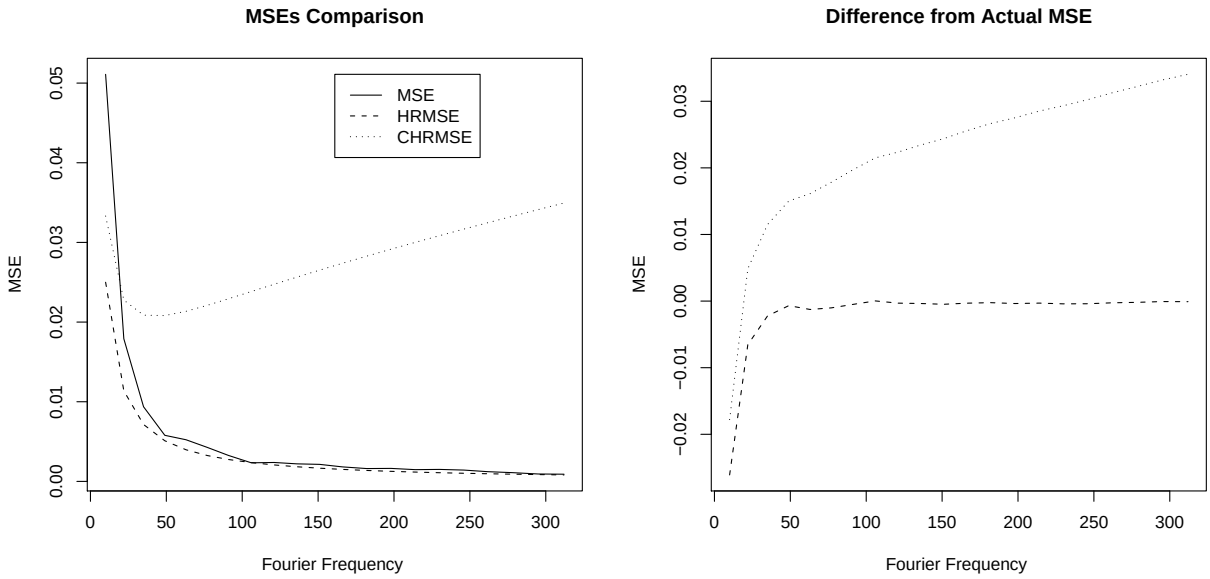

Figure 3.11: Comparison of MSEs, $d = 0.1$, $ns = 1$, $N = 2^{10}$

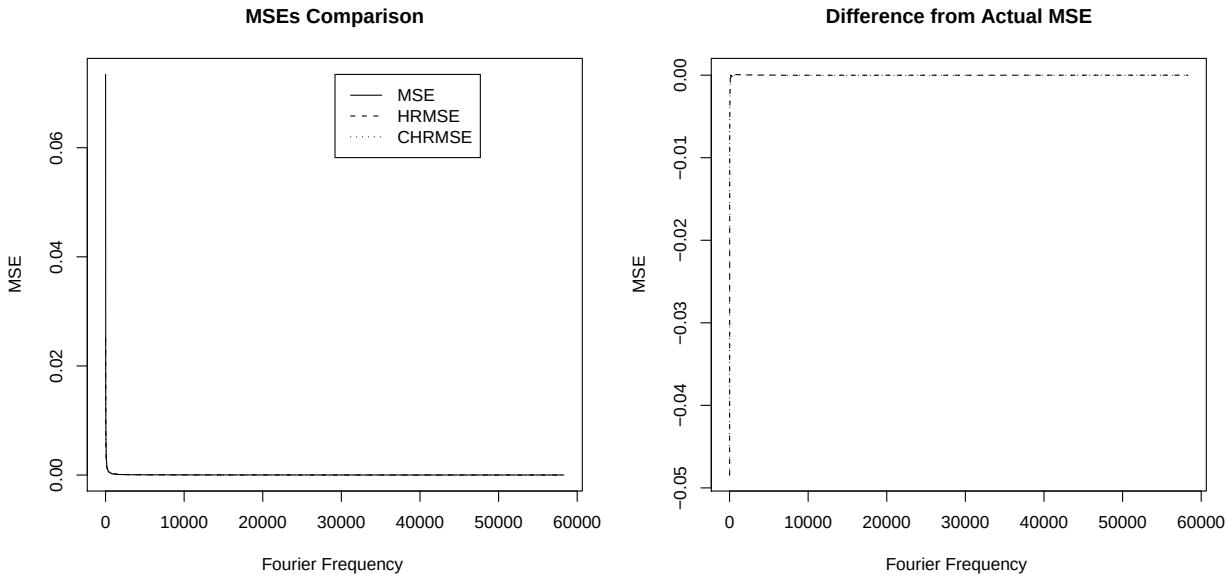

Figure 3.12: Comparison of MSEs, $d = 0.4$, $ns = 0$, $N = 2^{19}$

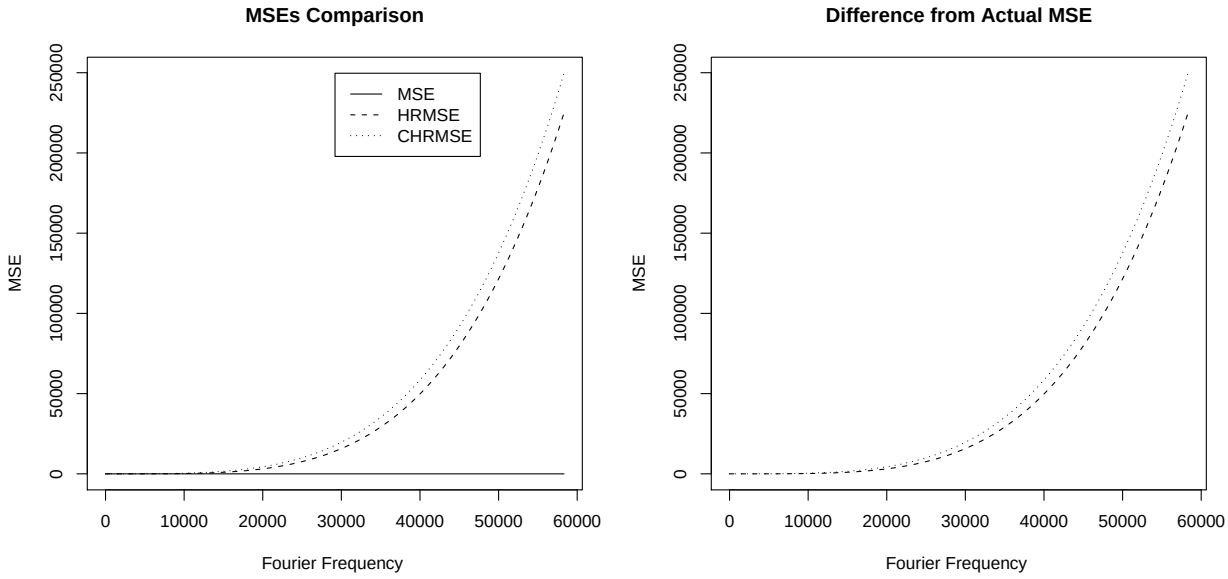

Figure 3.13: Comparison of MSEs, $d = 0.4$, $ns = 0.1$, $N = 2^{19}$

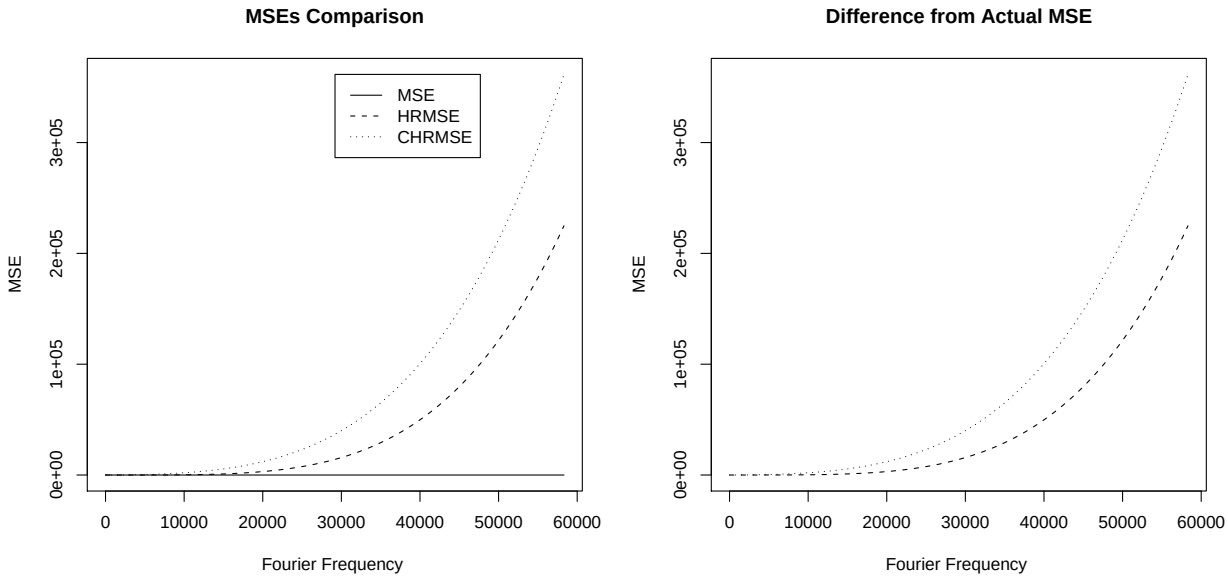

Figure 3.14: Comparison of MSEs, $d = 0.4$, $ns = 0.5$, $N = 2^{19}$

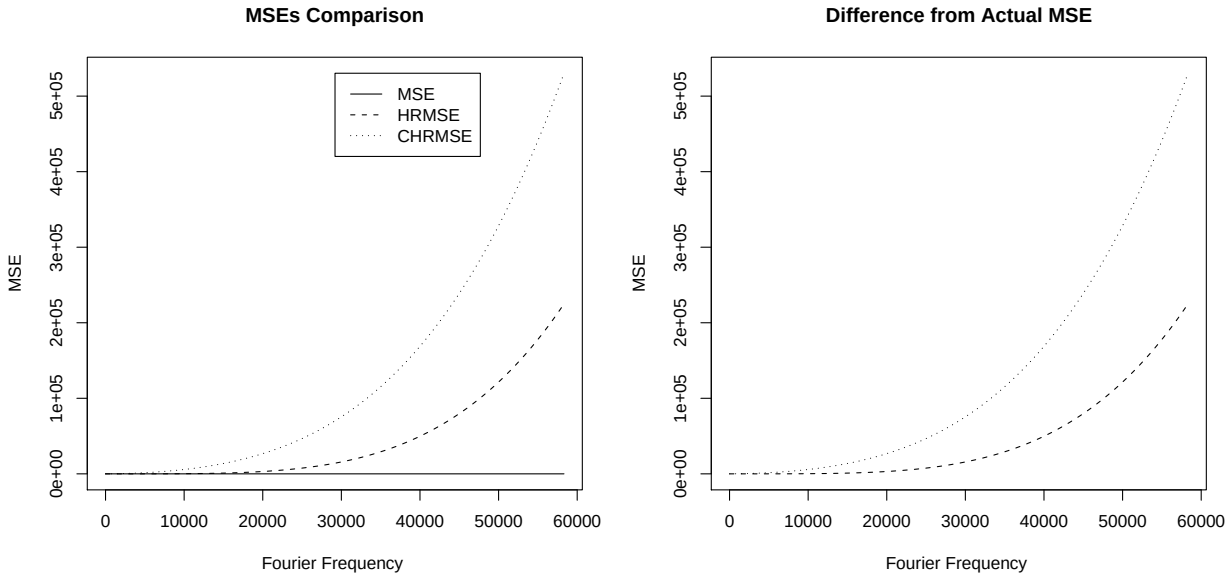

Figure 3.15: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{19}$

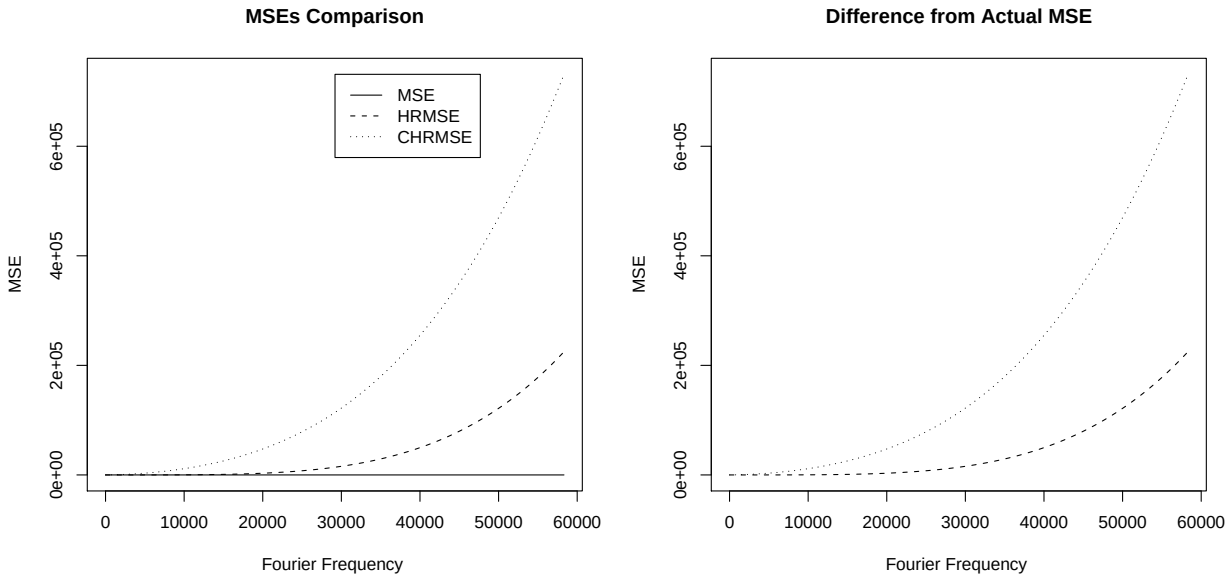

Figure 3.16: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{19}$

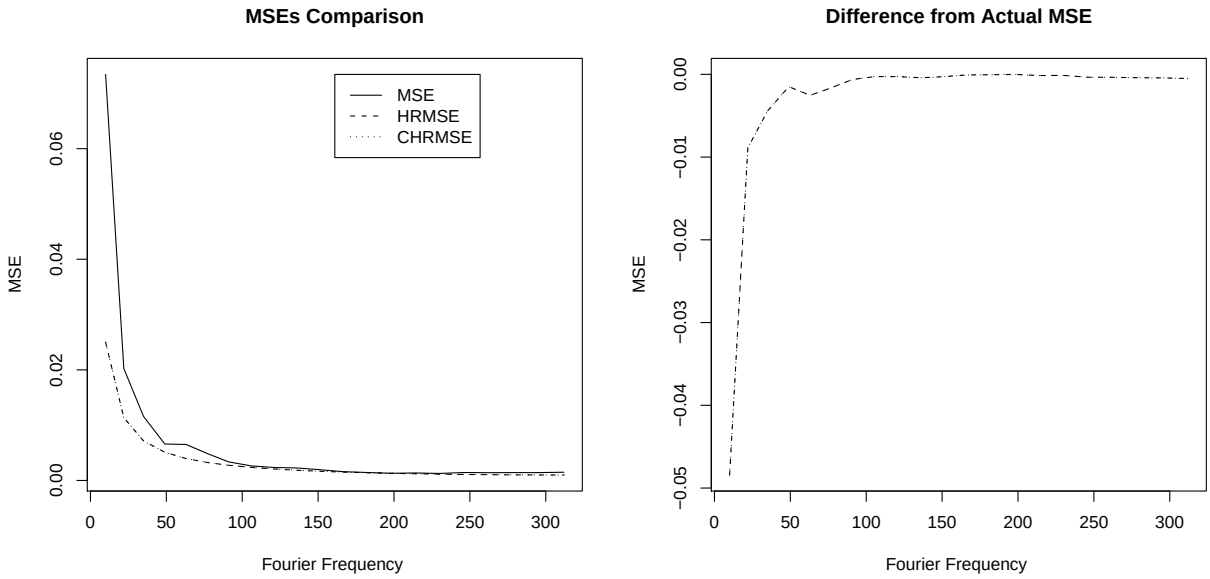

Figure 3.17: Comparison of MSEs, $d = 0.4$, $ns = 0$, $N = 2^{10}$

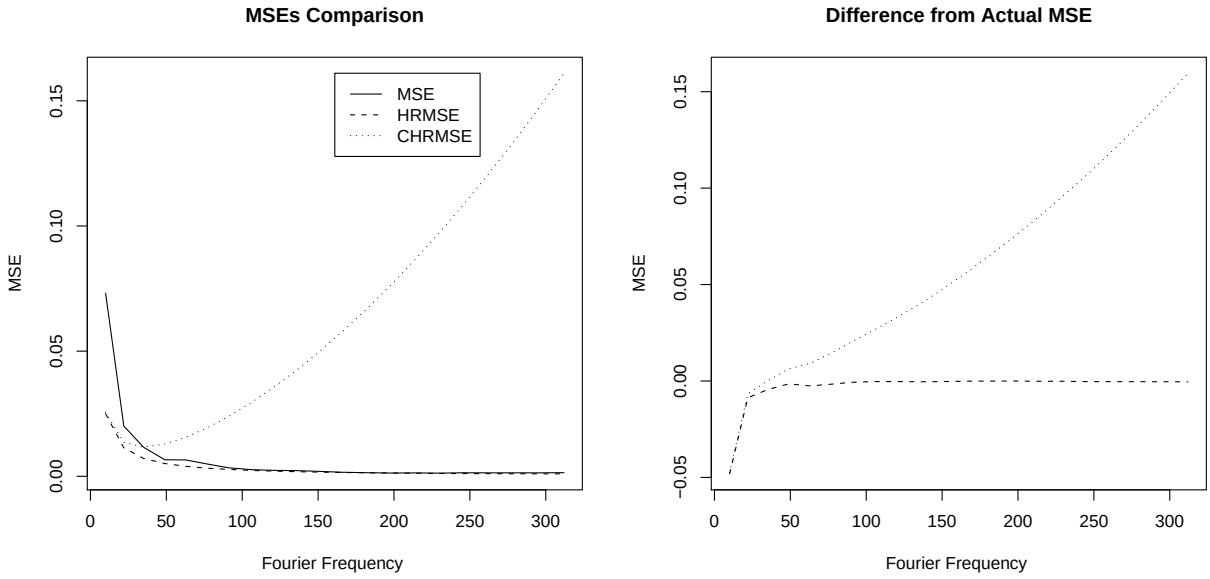

Figure 3.18: Comparison of MSEs, $d = 0.4$, $ns = 0.1$, $N = 2^{10}$

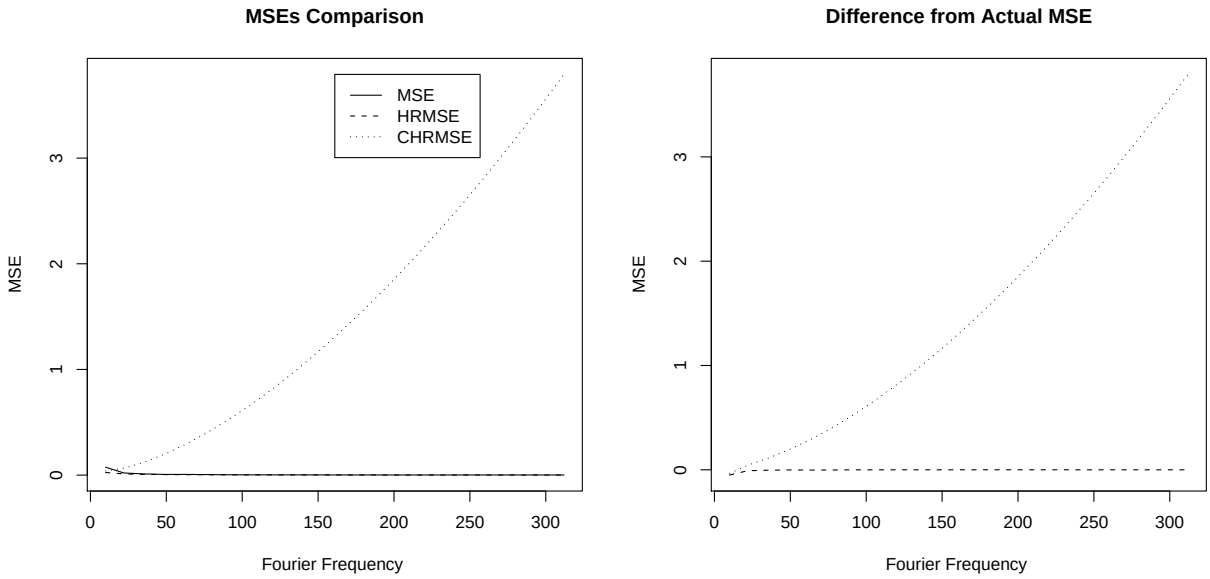

Figure 3.19: Comparison of MSEs, $d = 0.4$, $ns = 0.5$, $N = 2^{10}$

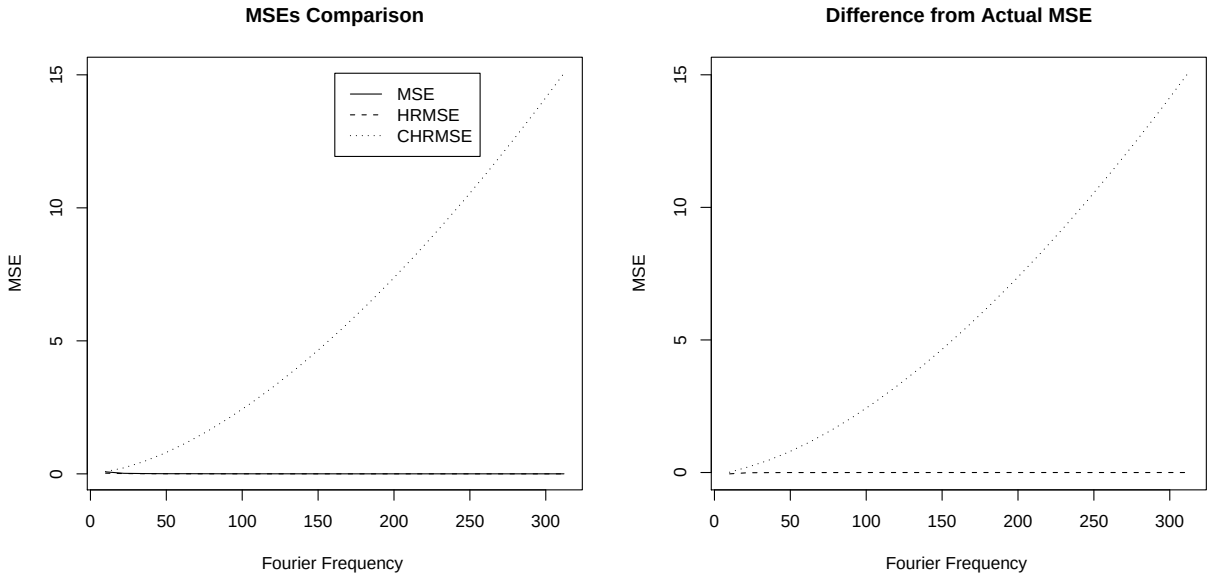

Figure 3.20: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{10}$

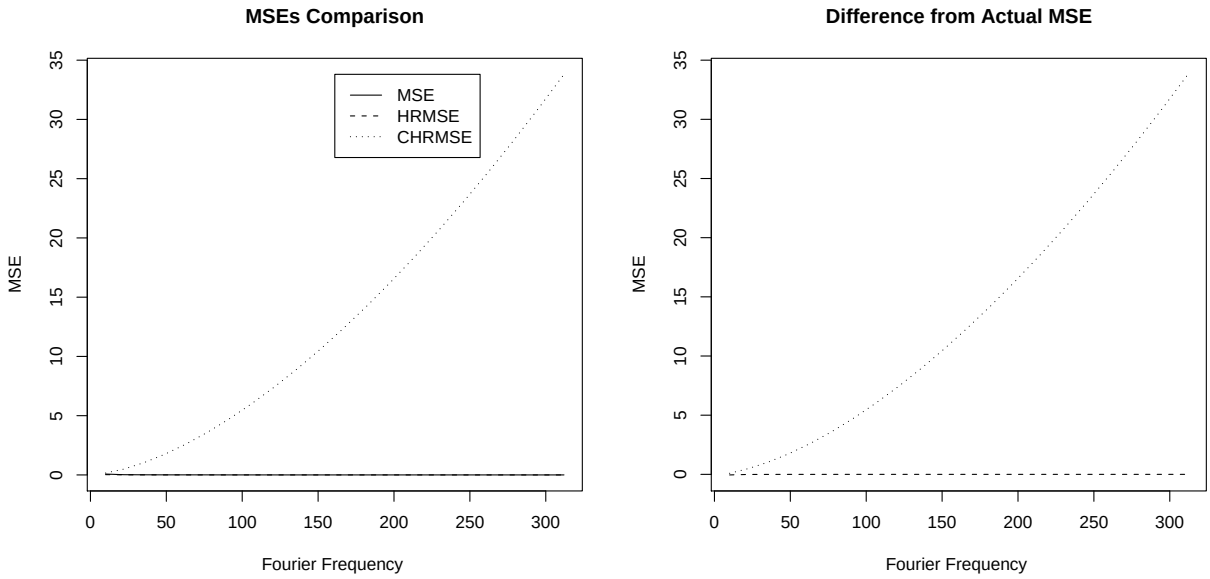

Figure 3.21: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{10}$

3.1.2 Effects of Aggregation

Crato and Ray (2002) advocate that aggregation reduces the ns ratio and, thus, reduces the bias of the estimation. On the other hand, for a finite sample, the reduction in the number of observations implied by aggregation would increase the variance of the estimator.

Figures 3.22 through 3.41 display how the minimal MSE and optimal choice of m vary with aggregation, for given d and ns, $d \in (0.1, 0.2, 0.3, 0.4)$ and $ns \in (0.1, 0.5, 1, 1.5)$. The left-hand side plots display the minimal MSE. The right-hand side plots display the value of the bandwidth parameter m corresponding to the minimal MSE. The x-axis displays the logarithm in base 2 of the number of observations in each aggregation bin. For example, on the extreme right of the x-axis, corresponding to the highest level of aggregation, $k = 9$, each observation of the series $\{z^{(9)}\}$ is the sum of 2^9 contiguous observations of the series $\{z^{(0)}\}$. To appropriately interpret the plots, recall that the highest value of m available for each aggregation level is 58385, 32768, 18390, 10321, 5792, 3250, 1824, 1024, 574, 322, for aggregation levels $k = 0, 1, \ldots, 9$, respectively. For comparison purposes, also the minimal HRMSE and CHRMSE, and relative optimal values of m are displayed. However, as noted in the previous section, these grossly overestimate the MSE in most cases.

We can observe the following facts.

- When there is no added noise ($ns = 0$), the series are nearly pure LRD processes. That implies that there is no trade-off between bias and variance, since the bias is null due to the lack of short-range dependence components. This is reflected in the plots by the fact that, when $ns = 0$, the minimal MSE keeps increasing with aggregation, and that the optimal m is essentially the last value of m available. The variance is inversely proportional to m. Therefore, as m increase, the MSE decrease.

- If even a small amount of noise is introduced, e.g. $ns = 1$, moderate aggregation can reduce

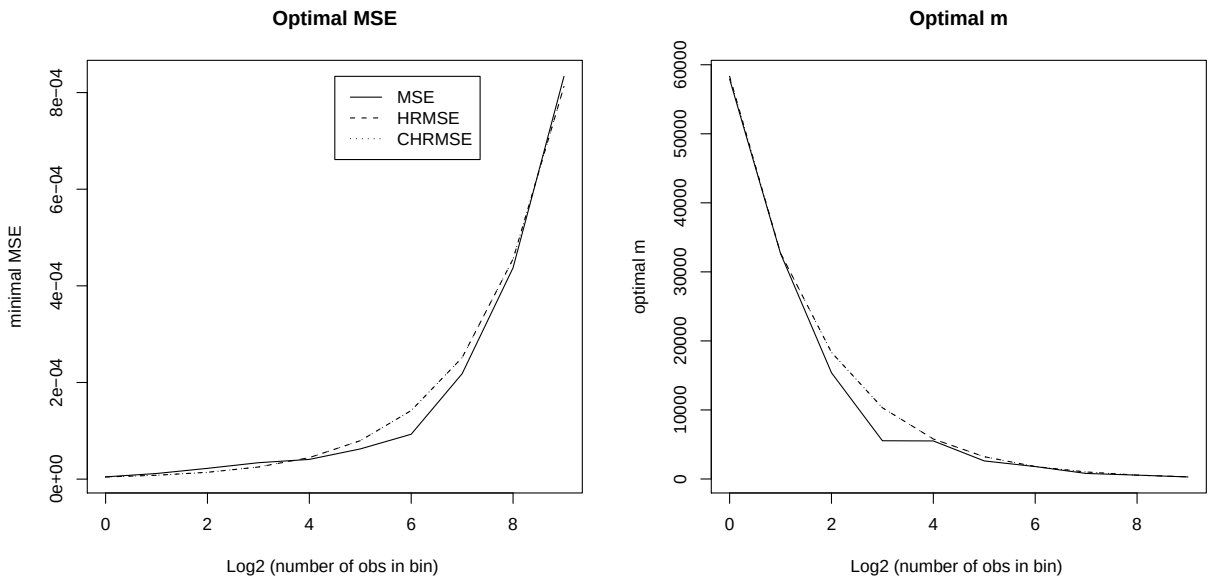

Figure 3.22: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0$

the MSE.

- The amount of aggregation that can be used before the MSE increases again is dependent on the size of the noise-to-signal ratio ns. After a certain level of aggregation, the series behave again as a pure LRD processes, that is, the MSE decreases monotonically with m.

The above facts suggest that, if the series is suspected to be the sum of noise component and LRD component, the MSE of the LW estimator can be improved by a small amount of aggregation, provided that the data available is long enough not to run into an increase of the variance component of the MSE.

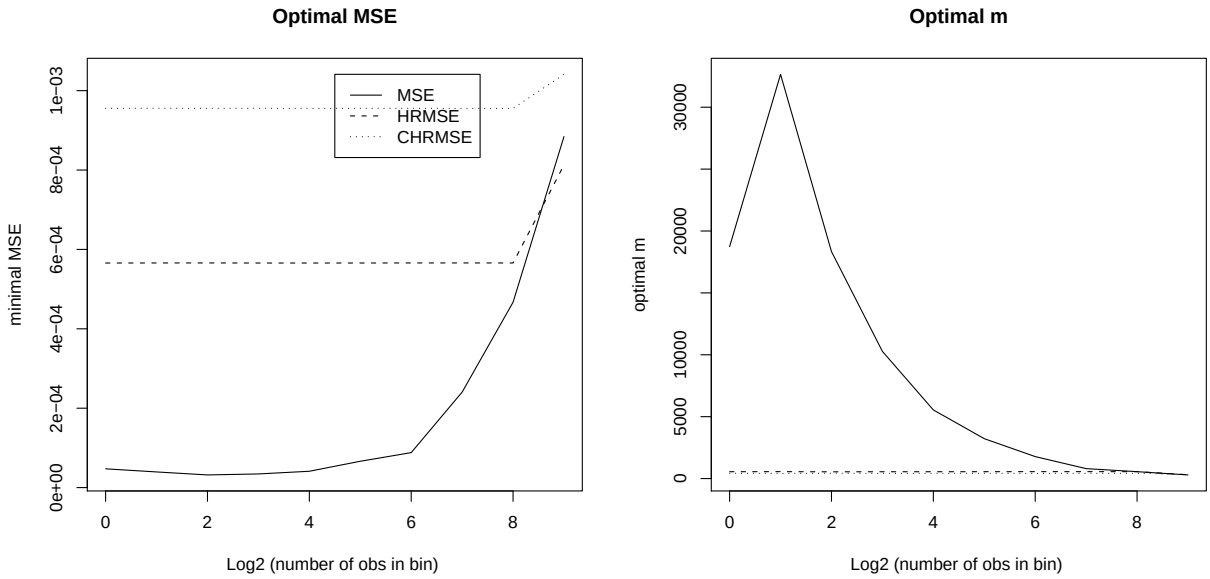

Figure 3.23: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0.1$

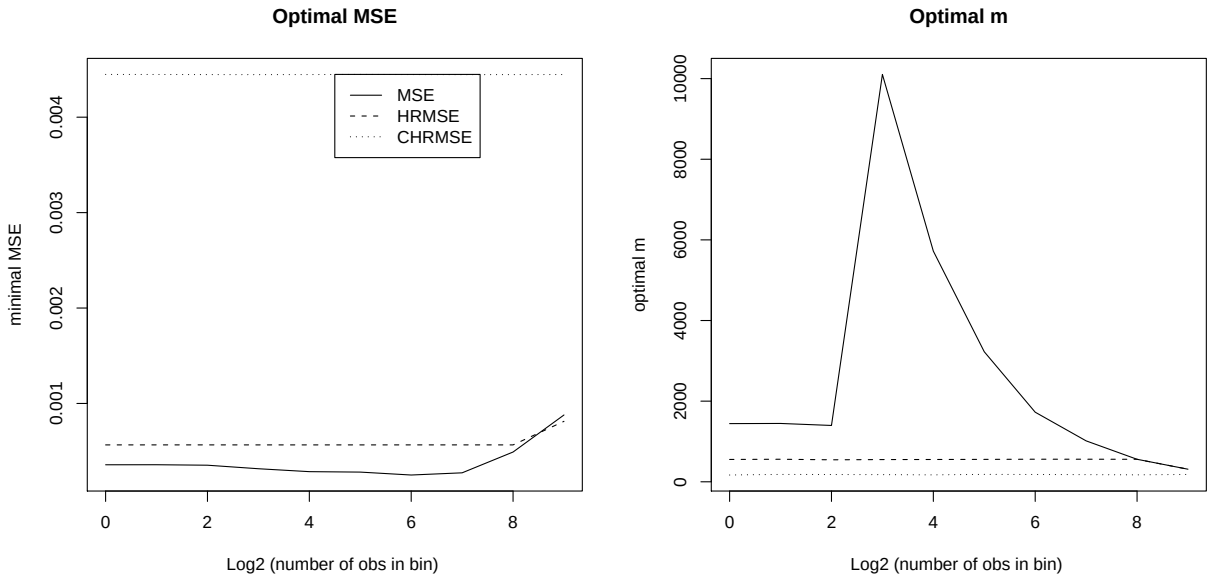

Figure 3.24: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0.5$

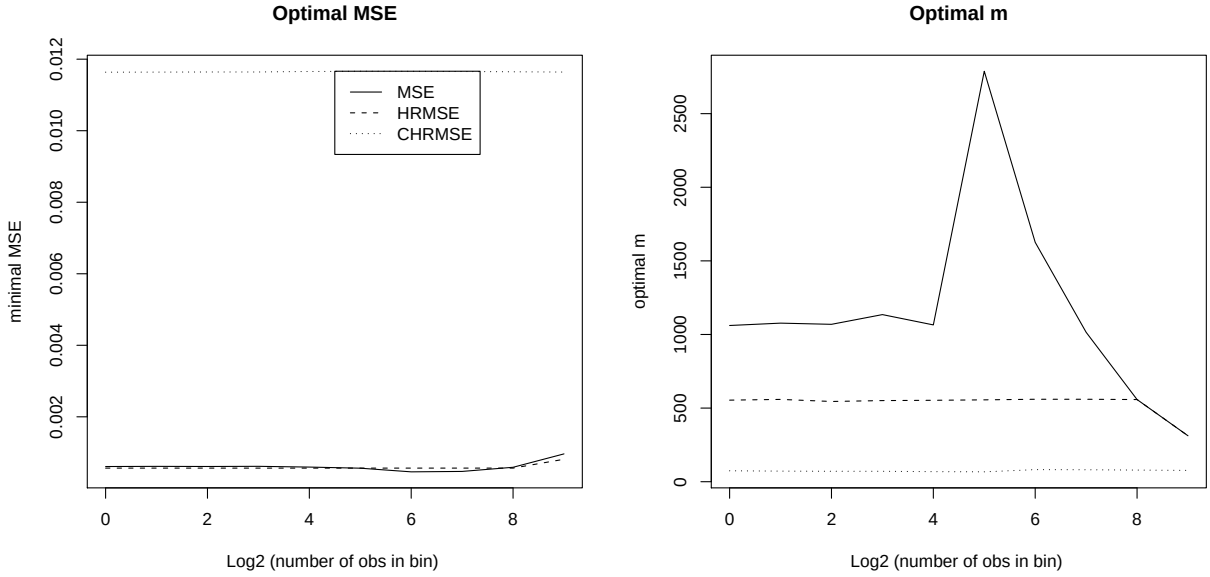

Figure 3.25: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 1$

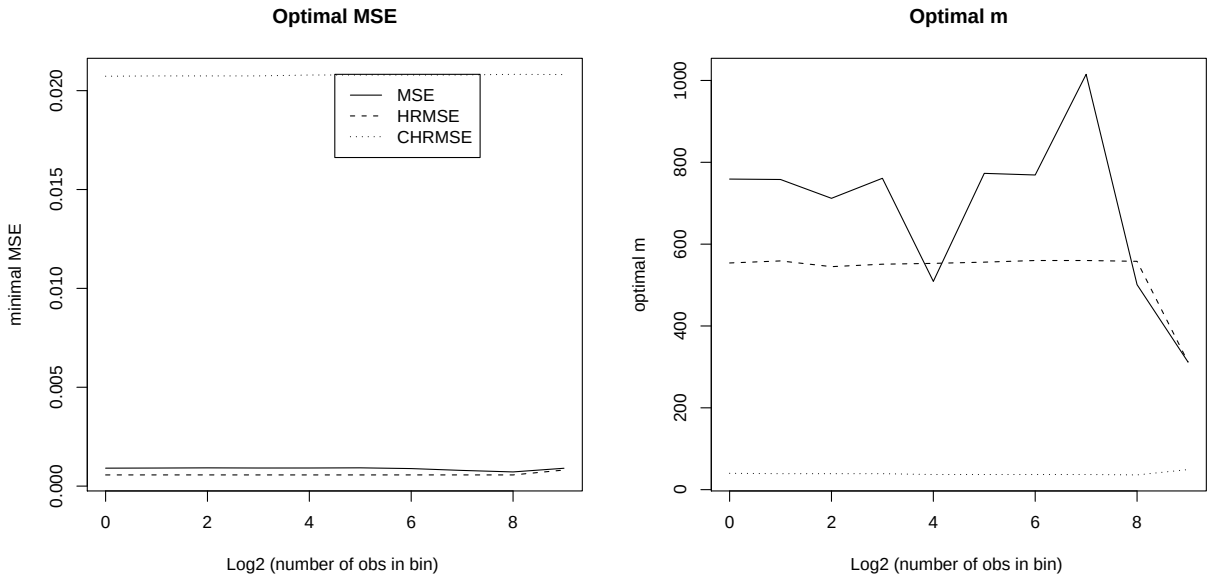

Figure 3.26: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 1.5$

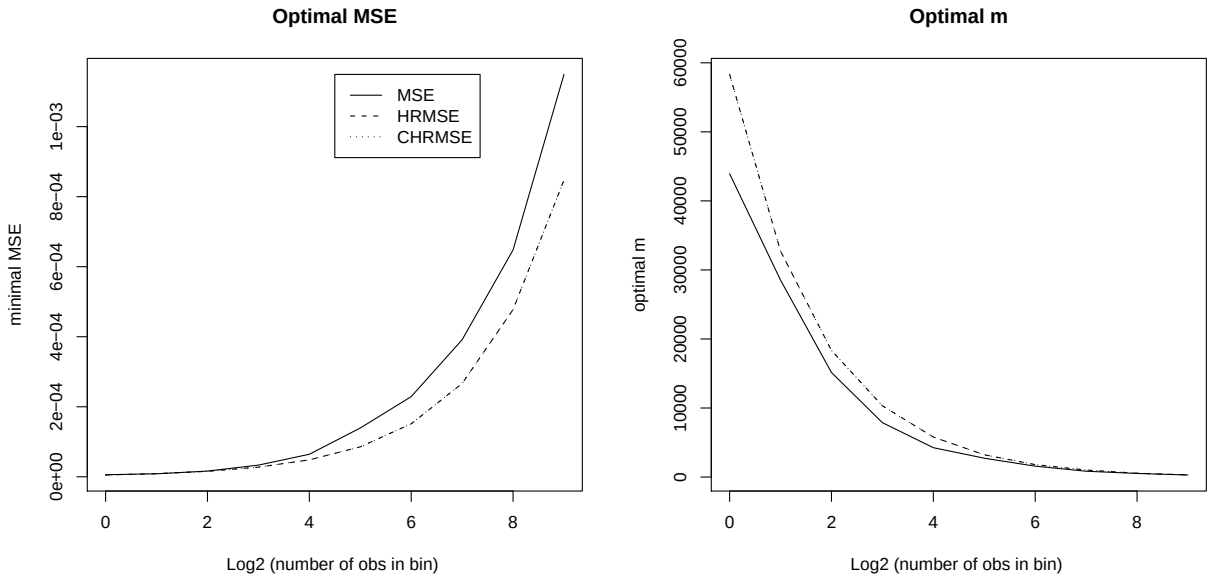

Figure 3.27: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0$

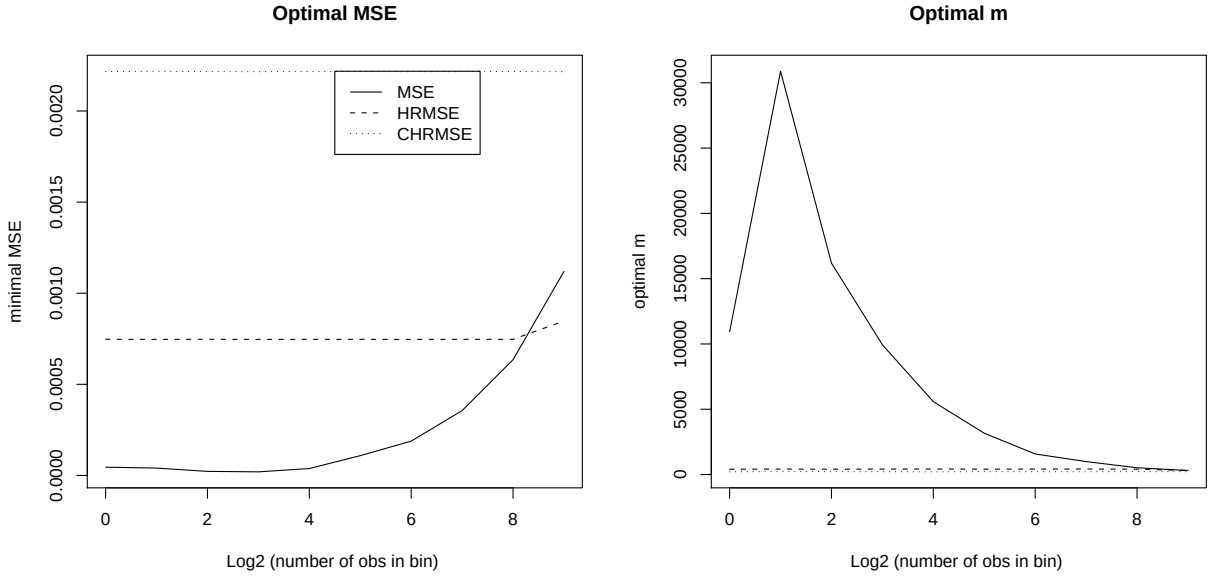

Figure 3.28: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0.1$

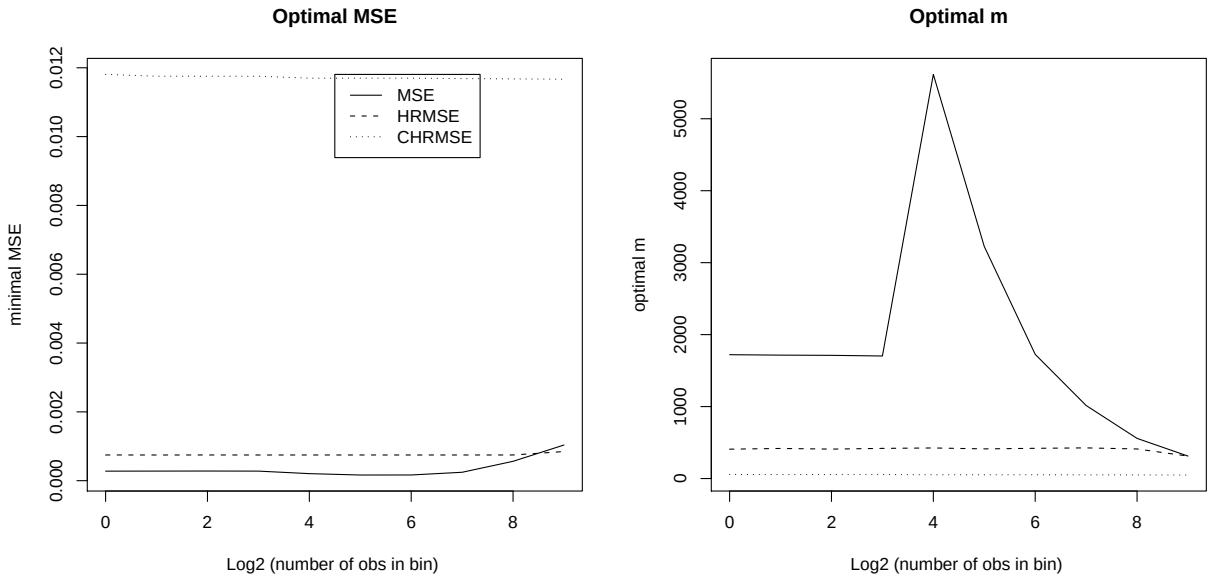

Figure 3.29: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0.5$

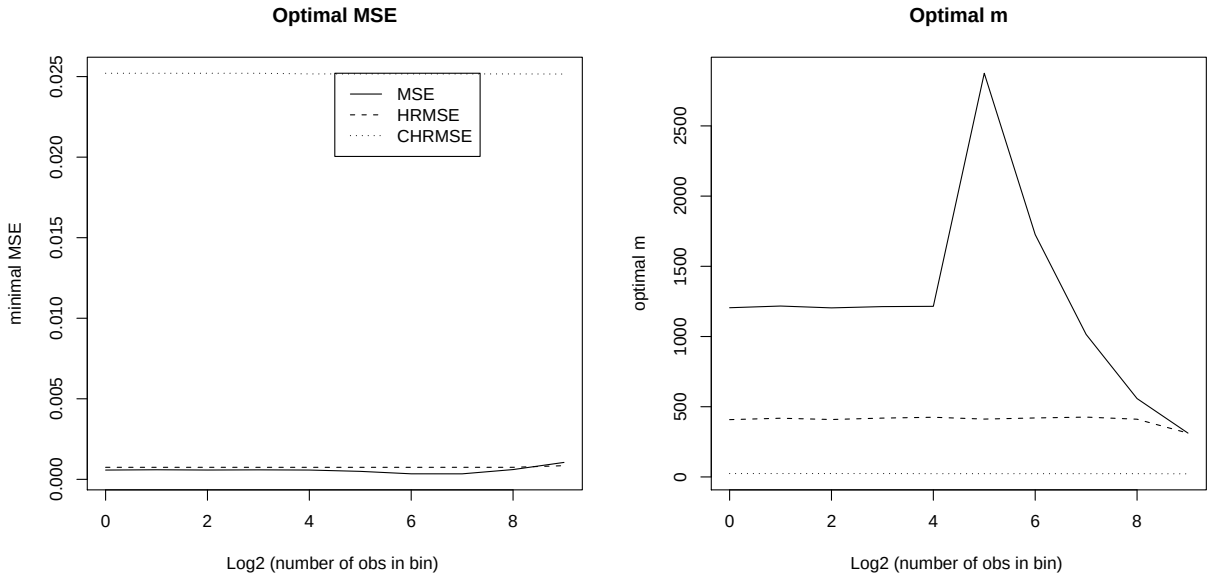

Figure 3.30: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 1$

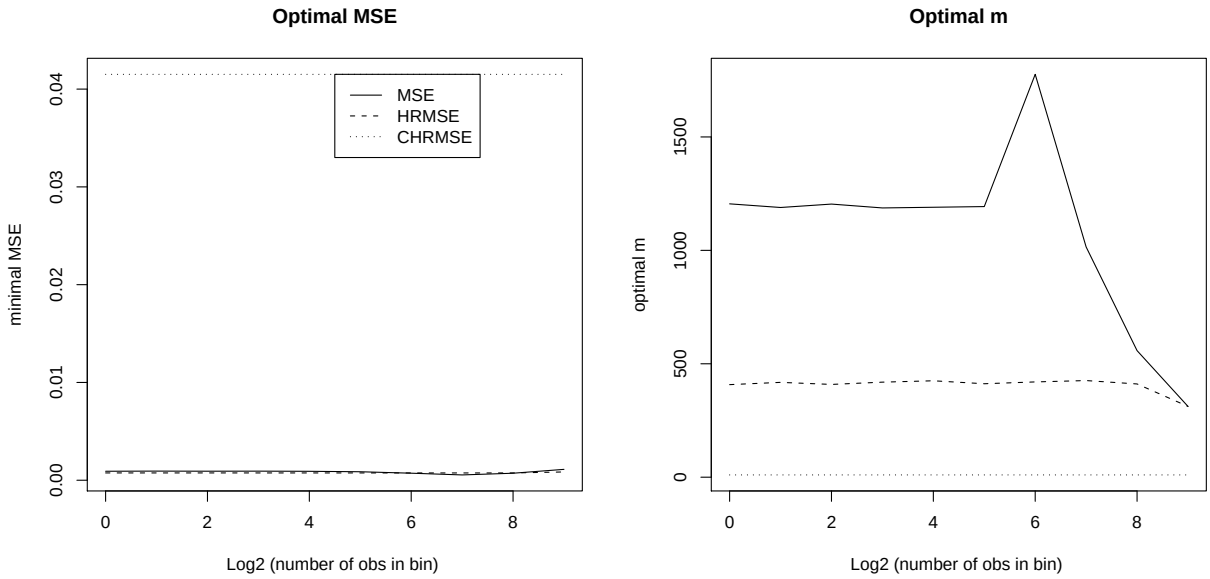

Figure 3.31: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 1.5$

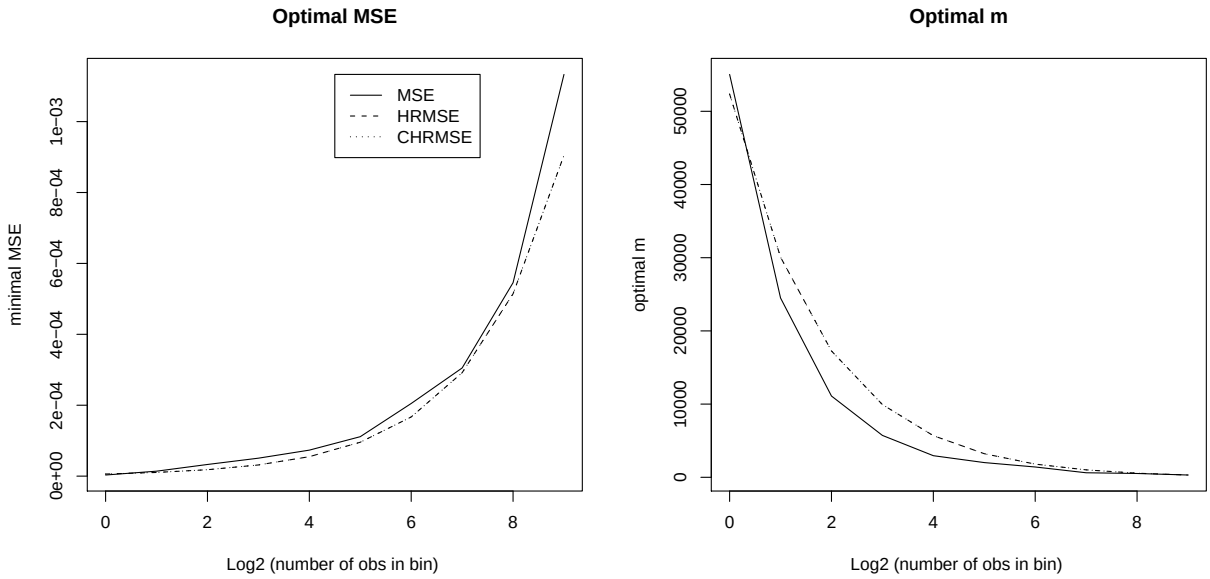

Figure 3.32: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0$

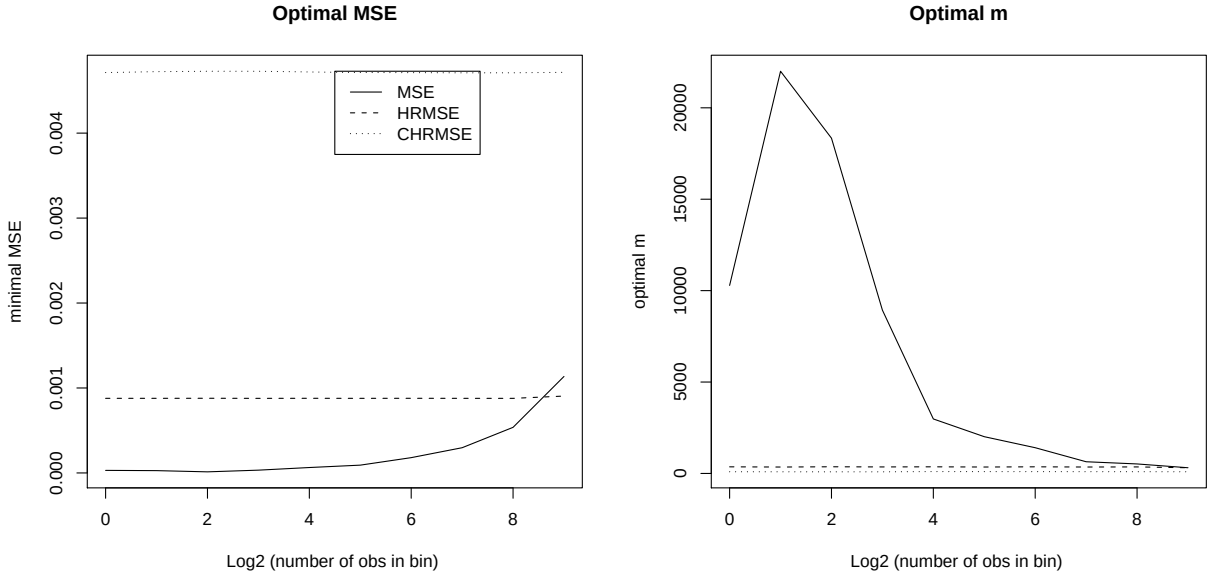

Figure 3.33: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0.1$

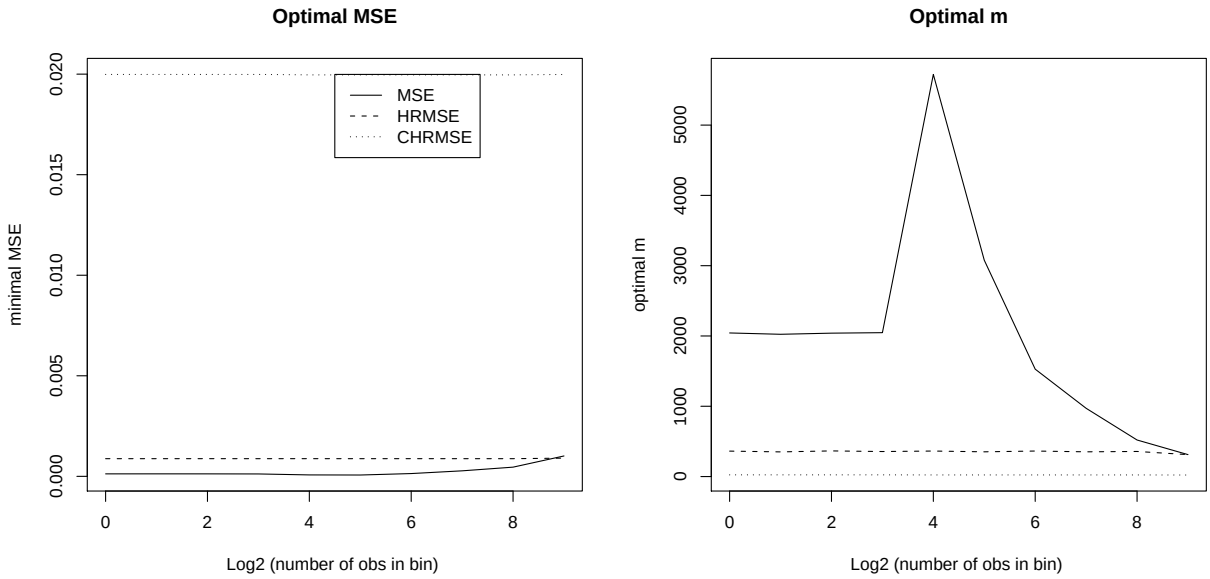

Figure 3.34: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0.5$

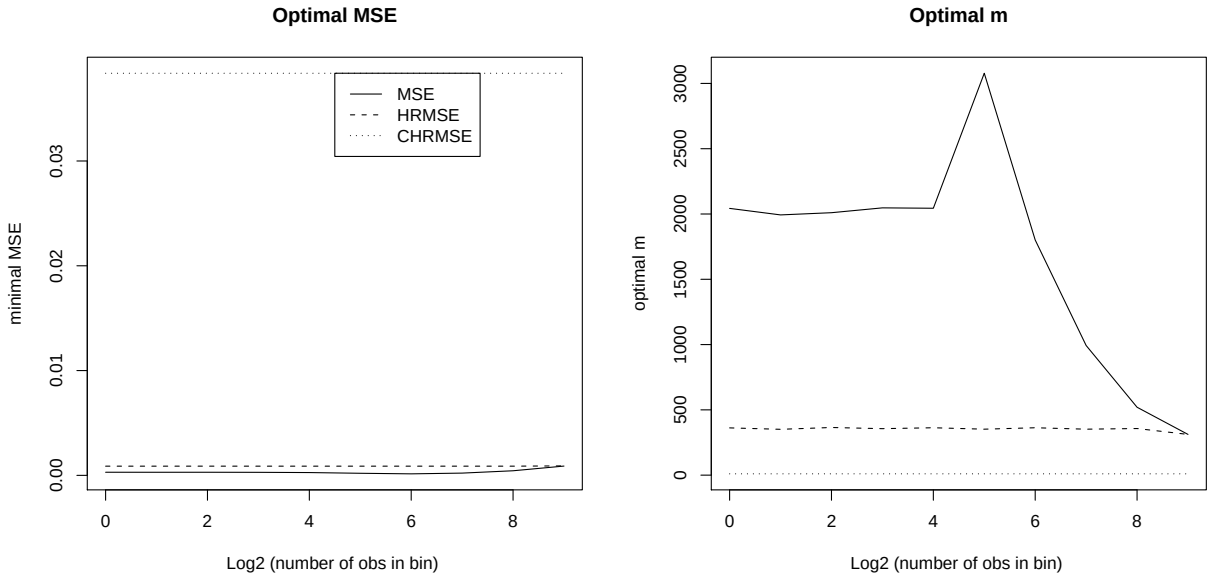

Figure 3.35: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 1$

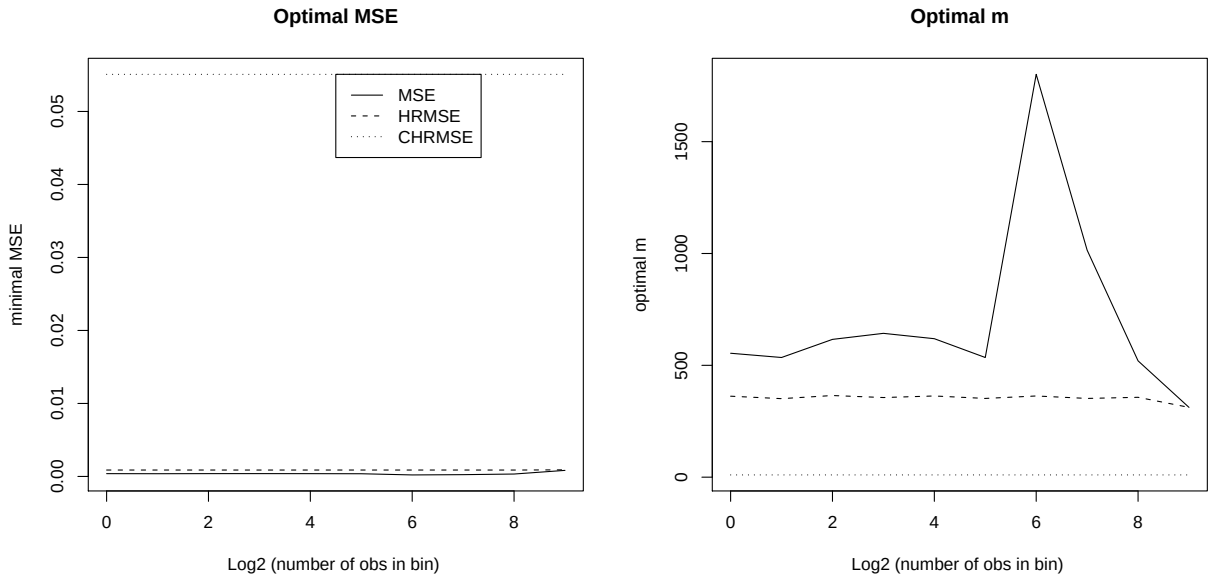

Figure 3.36: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 1.5$

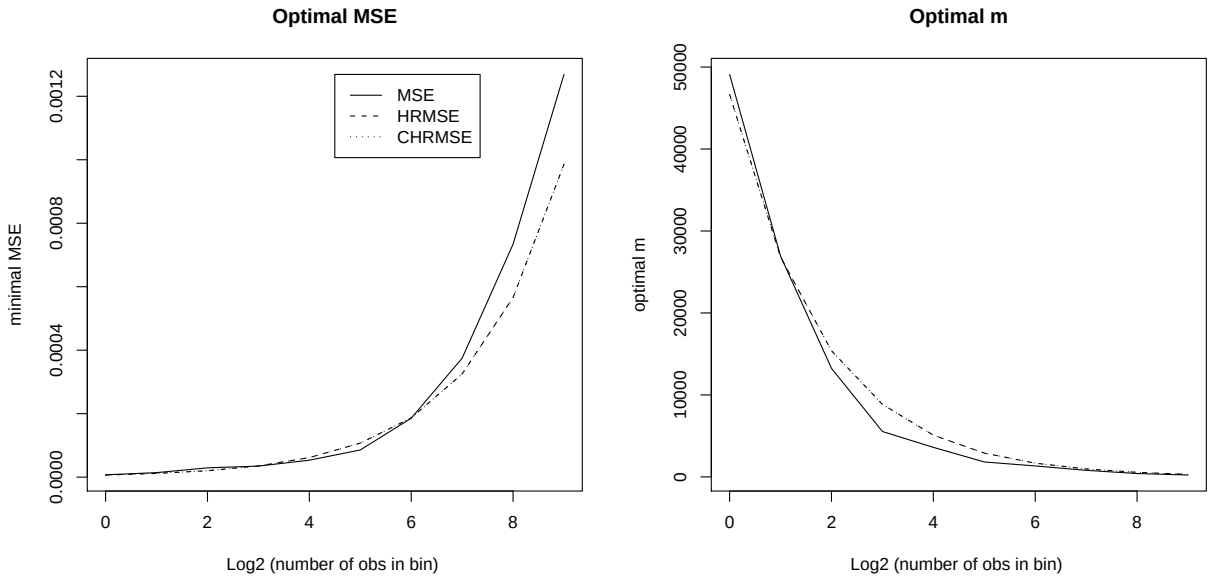

Figure 3.37: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0$

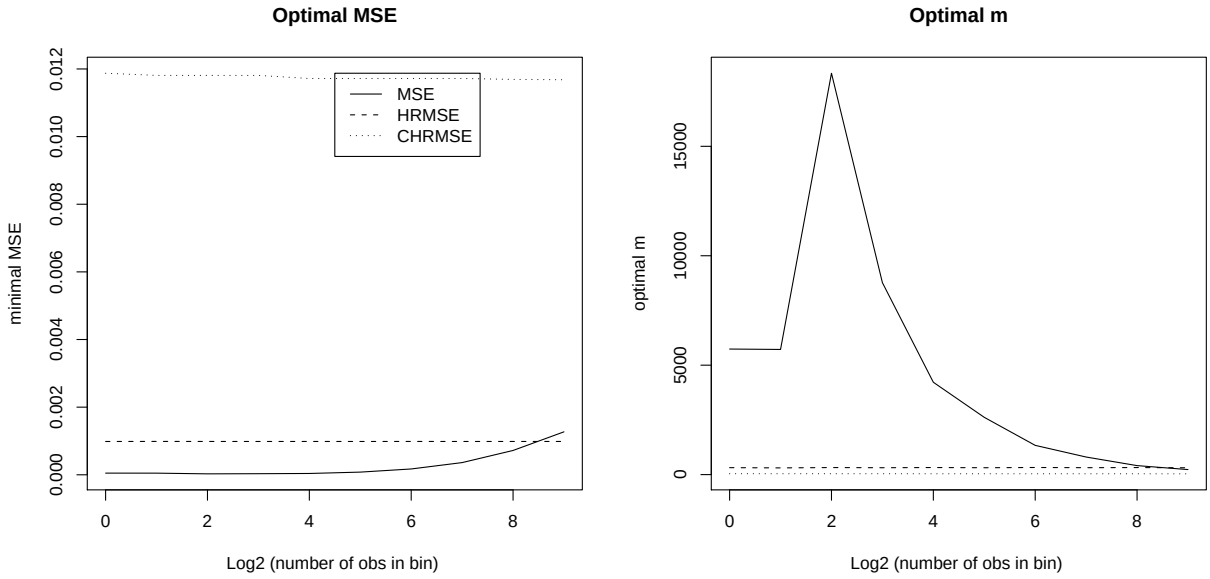

Figure 3.38: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0.1$

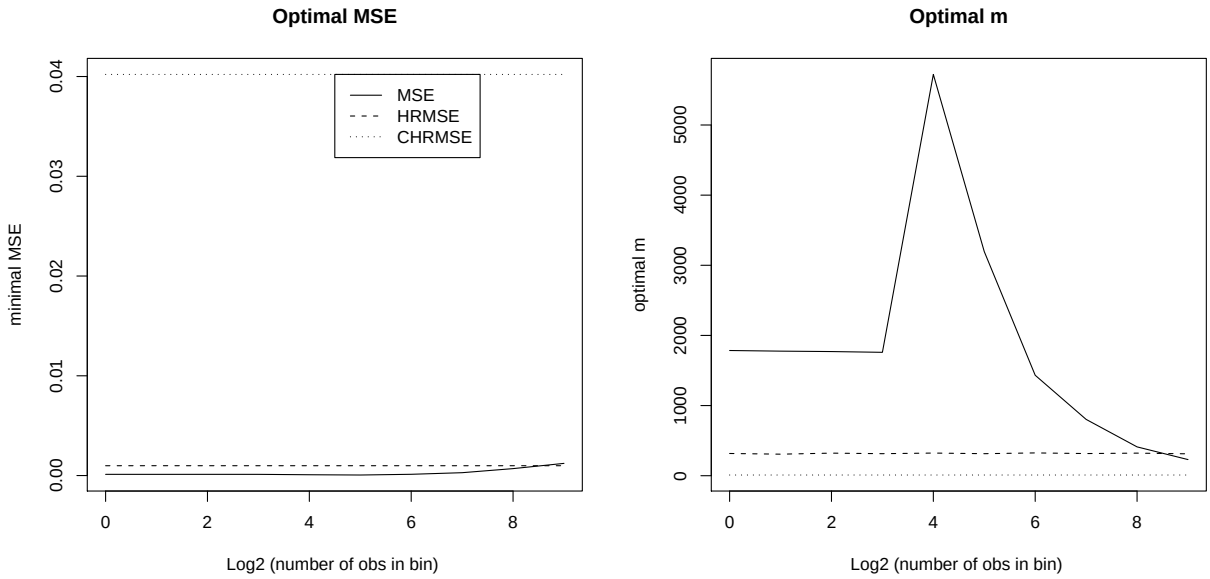

Figure 3.39: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0.5$

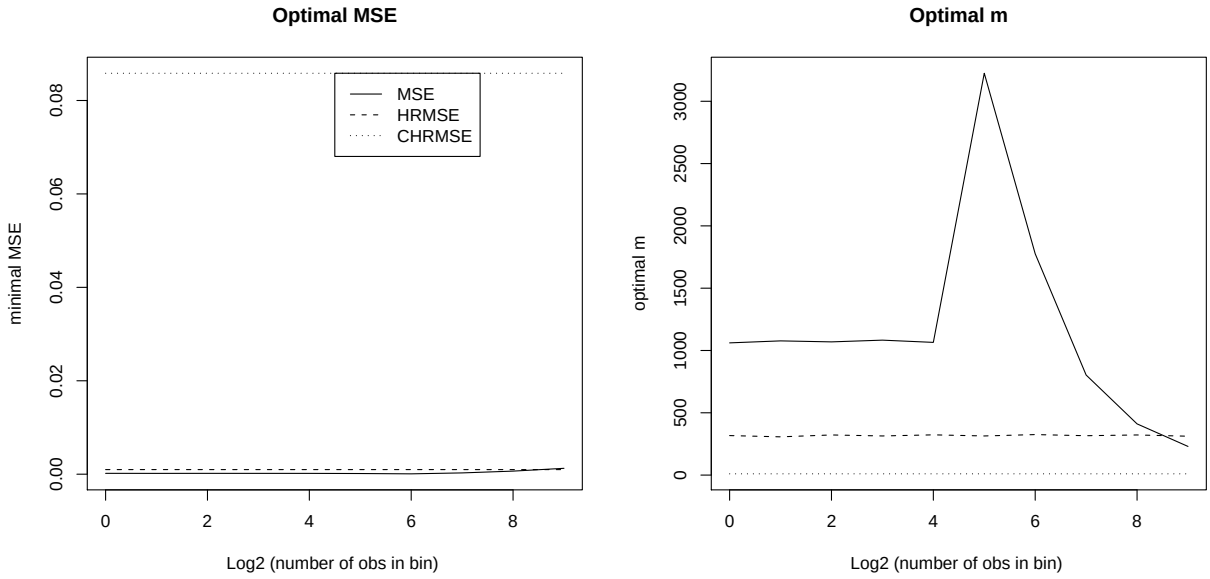

Figure 3.40: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 1$

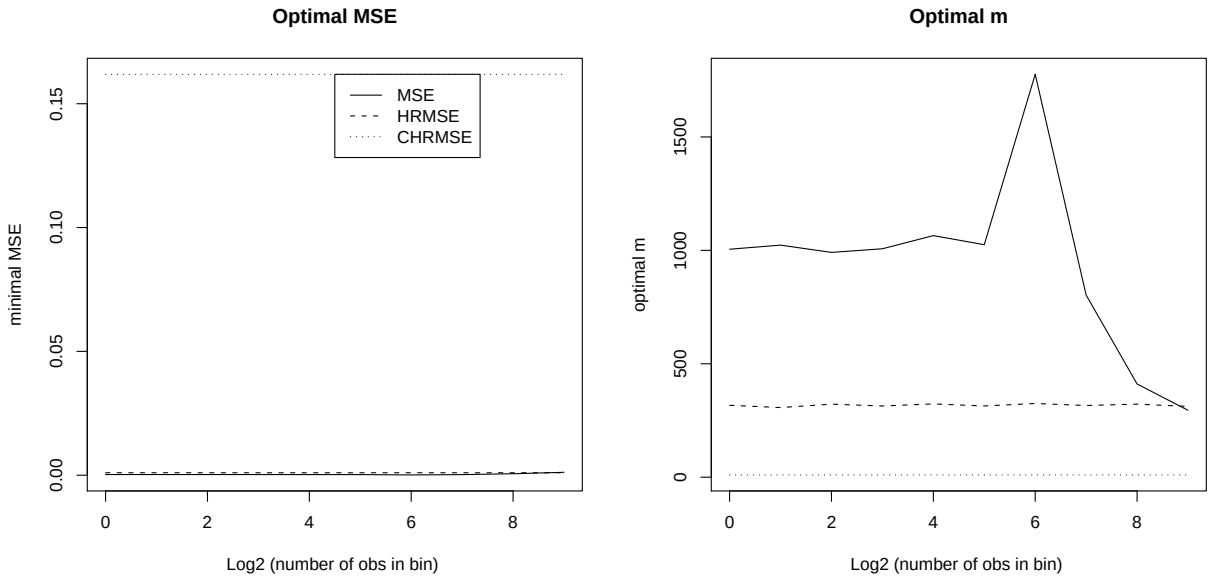

Figure 3.41: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 1.5$

3.2 The Extended LW Estimator

As explained in Chapter 2, an analytical expression for the MSE of the ELW estimator is not available. Therefore, the MSE can only be evaluated numerically, and so the effect of aggregation on the estimator.

In this section we will analyze numerically the MSE of the ELW estimator; how it is affected by aggregation and how it compares to the MSE of the LW estimator. The framework of the analysis is similar to Section 3.1. Therefore, we will describe only where it differs.

A short-range dependent component was introduced by applying an AR(1) filter with parameter 0.5 to the simulated FARIMA(0, d, 0) used in the previous section, thus generating FARIMA(1, d, 0) series. Due to time constraints, the series length was truncated to 2^{17} observations, allowing for 7 levels of aggregations. The number of series used for computing the MSE was restricted to 10. Additionally, the fix portion of the step of m was increased from 10 to 20.

The log-likelihood of the ELW flattens as a function of the parameters as N diverges, and the rate at which it flattens differs for each parameter (Hurvich et al., 2005). This makes the analysis of the ELW estimator challenging not only analytically, but numerically too. The estimation is highly sensitive to the starting values, it needs more data point (higher m) than the LW estimator to attain convergence, often does not converge or, especially when N is large, it is common that the parameters run into regions that cause the optimizer to overflow. We found the most stable parametrization to be the one suggested by Hurvich and Ray (2001), that is, the spectral density in a neighborhood of the origin has the form

$$f(\lambda) = b_0(1 + b_1\lambda^{-2d}) \tag{3.2}$$

The parameter b_0 can be concentrated out the log-likelihood so that a two-dimensional problem can be solved. We can rewrite the local Whittle contrast as

$$L(\tilde{\theta}) = \sum_{1}^{m} \left\{ \log \tilde{f}(\lambda_j) + \frac{I(\lambda_j)}{\tilde{f}(\lambda_j)} \right\} \tag{3.3}$$

where $\tilde{\theta} = (b_1, d)$,

$$\tilde{f}(\lambda) = \tilde{b}_0(1 + b_1\lambda^{-2d}) \tag{3.4}$$

and

$$\tilde{b}_0 = \frac{1}{m} \sum_{1}^{m} \frac{I(\lambda_j)}{1 + b_1\lambda_j^{-2d}} \tag{3.5}$$

The optimization is performed initially using the `nlm` function of `R`, where the step of the algorithm has been limited to try and prevent that the parameters run into regions that overflow the target function. The true values of the parameters are used as starting value. If the optimization does not converge, the GPH estimates with the variance of the error component assumed to be 1 are used as starting values. Should the estimation still not converge, or should it converge to values that are not admissible, the `nlmin` function is used, which enables to constrain the parameters to the admissible region. The constrained optimization slows down the process considerably and is avoided when possible.

3.2.1 MSEs Comparisons

Figure 3.42 through Figure 3.61 compare, for a given value of long-range dependence parameter d, noise-to-signal ratio ns, and aggregation level k, the MSE of the ELW estimator (ELW) of d to the MSE of the LW estimator (LW), and the asymptotic expression of the MSE of Henry and Robinson (1996)(HRMSE), equation (2.33), and the similar asymptotic expression (A.16) which accounts

for the presence of the noise component (CHRMSE). The plot on the left-hand side display the estimates of the MSE. The plots on the right-hand side, display the difference from the LW MSE.

We computed also an asymptotic expression for the MSE of ELW estimator based on the approximation

$$E(\hat{\theta} - \theta_0) \approx -H^{*-1}(\theta_0) E \nabla h(\theta_0) \tag{3.6}$$

$$Var(\hat{\theta} - \theta_0) \approx H^{*-1}(\theta_0) \tag{3.7}$$

$$\tag{3.8}$$

where H^* is the approximation to the expected value of the Hessian matrix given by equations (A.30)–(A.30), and $E\nabla h(\theta_0)$ is the expected value of the gradient. This last approximation of the MSE is not displayed in the plots because it is on a widely different scale than the other estimates for any value of ns different from zero, presumably because the Hessian matrix is nearly singular.

Again, we are displaying here only the plots relative to $d = 0.1$ and $d = 0.4$, and to $k = 0$, that is, no aggregation, and $k = 7$, the highest level of aggregation. The display is sufficient for illustrating the general behavior. The remaining plots are available on request.

We can observe the following facts.

- The variance of the ELW is higher at low frequencies. This is not unexpected, because the ELW tries to estimates the variance of the error component as well.

- For all values of m, d, ns, k, the MSE of the ELW is almost uniformly greater than the MSE of the LW. This is true also when the noise-to-signal ratio is high. It seems that the better asymptotic properties of the ELW estimator do no translate into better finite sample

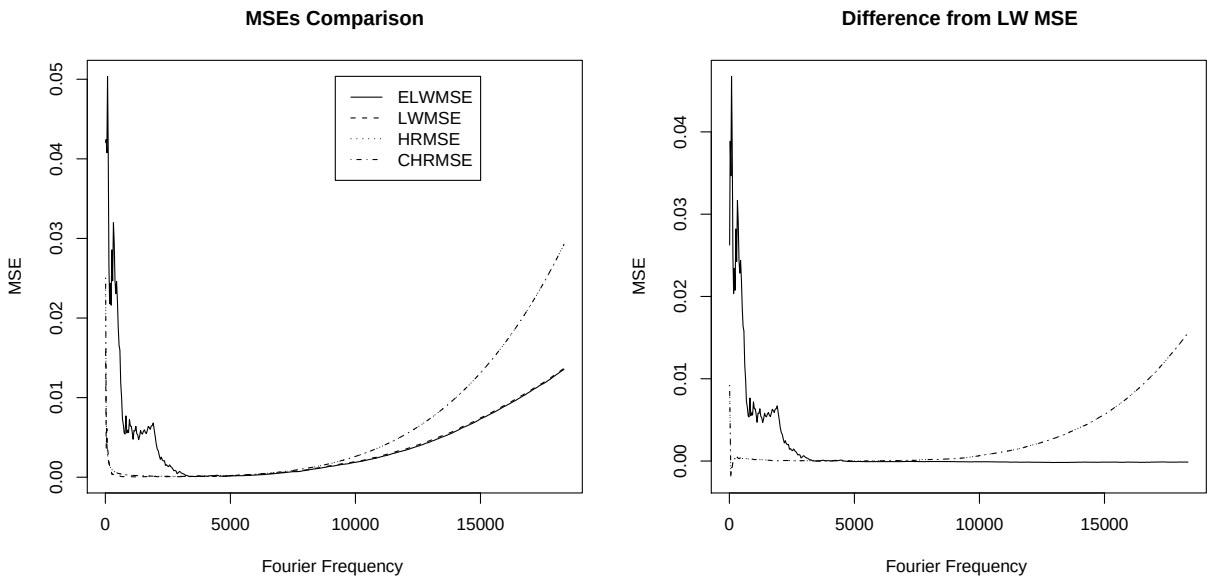

Figure 3.42: Comparison of MSEs, $d = 0.1$, $ns = 0$, $N = 2^{17}$

properties. In fact, in most cases, the LW proves to be superior.

- Even with the added SRD AR component, both the HRMSE and CHRMSE approximations to the MSE of the LW estimator keep grossly overestimating the actual MSE. However, the CHRMSE estimator, which explicitly accounts for the error term, gives a better approximation to the MSE for moderate values of ns. For large values of ns, the value of d becomes more important. When d is high and m is low, the HRMSE performs better. Eventually, though, for large m, the CHRMSE performs better for all values of d. Unfortunately, the CHRMSE requires the knowledge of the variance of the noise, which is not estimated by the LW estimator.

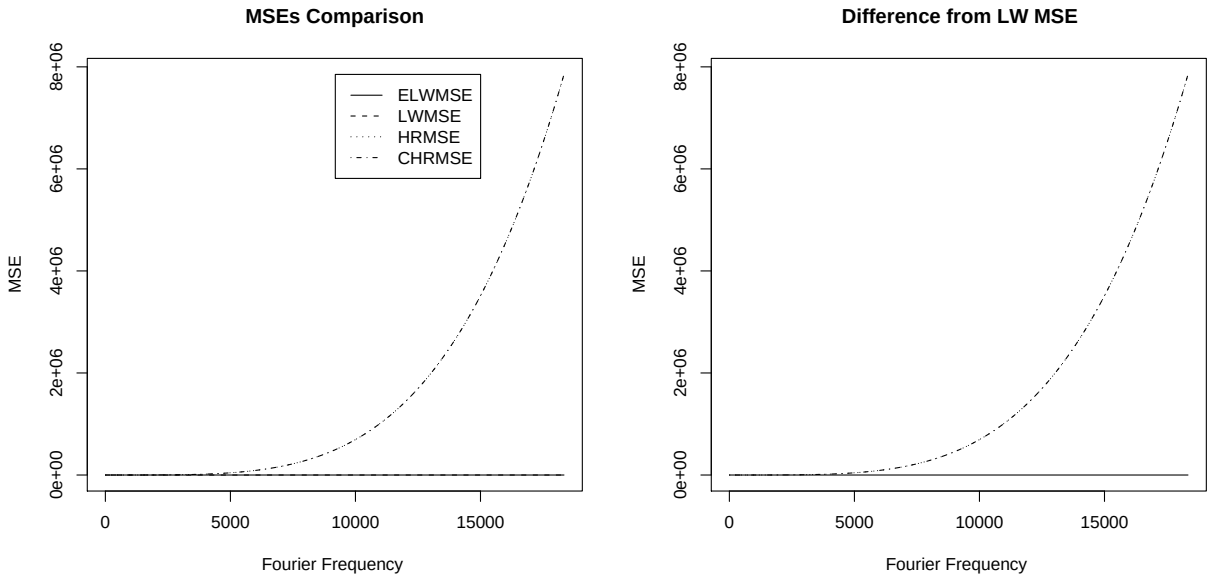

Figure 3.43: Comparison of MSEs, $d = 0.1$, $ns = 0.1$, $N = 2^{17}$

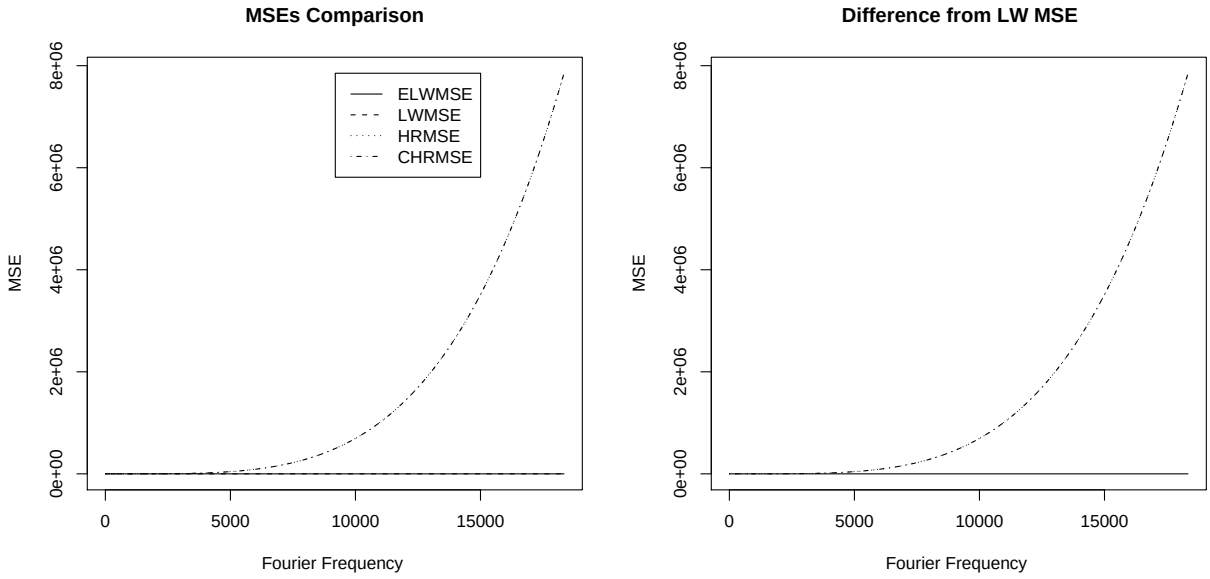

Figure 3.44: Comparison of MSEs, $d = 0.1$, $ns = 0.5$, $N = 2^{17}$

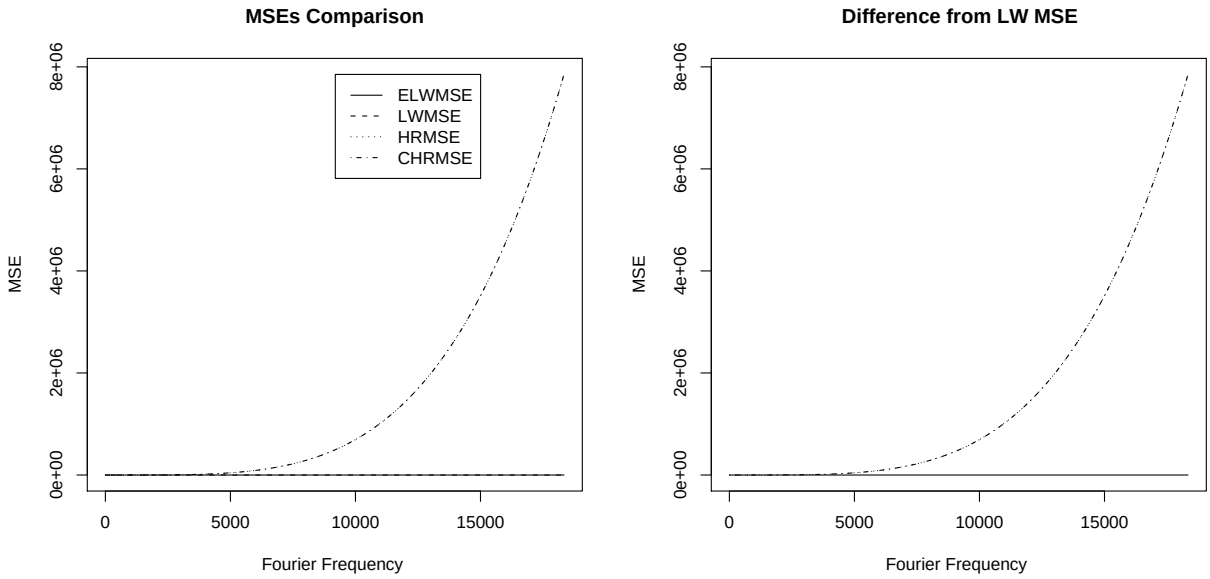

Figure 3.45: Comparison of MSEs, $d = 0.1$, $ns = 1$, $N = 2^{17}$

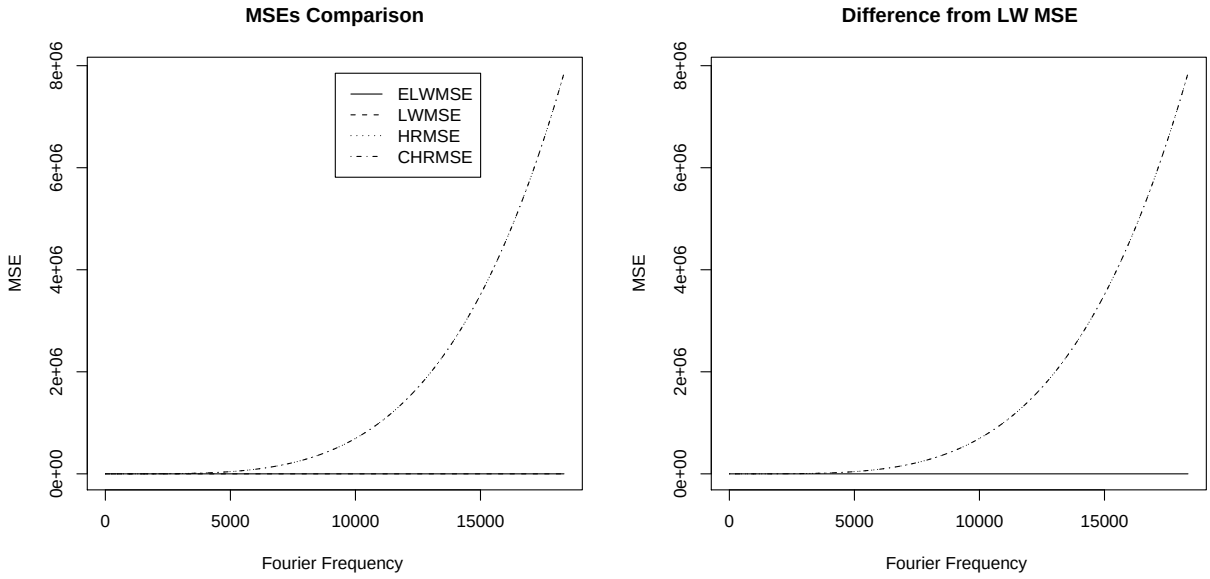

Figure 3.46: Comparison of MSEs, $d = 0.1$, $ns = 1.5$, $N = 2^{17}$

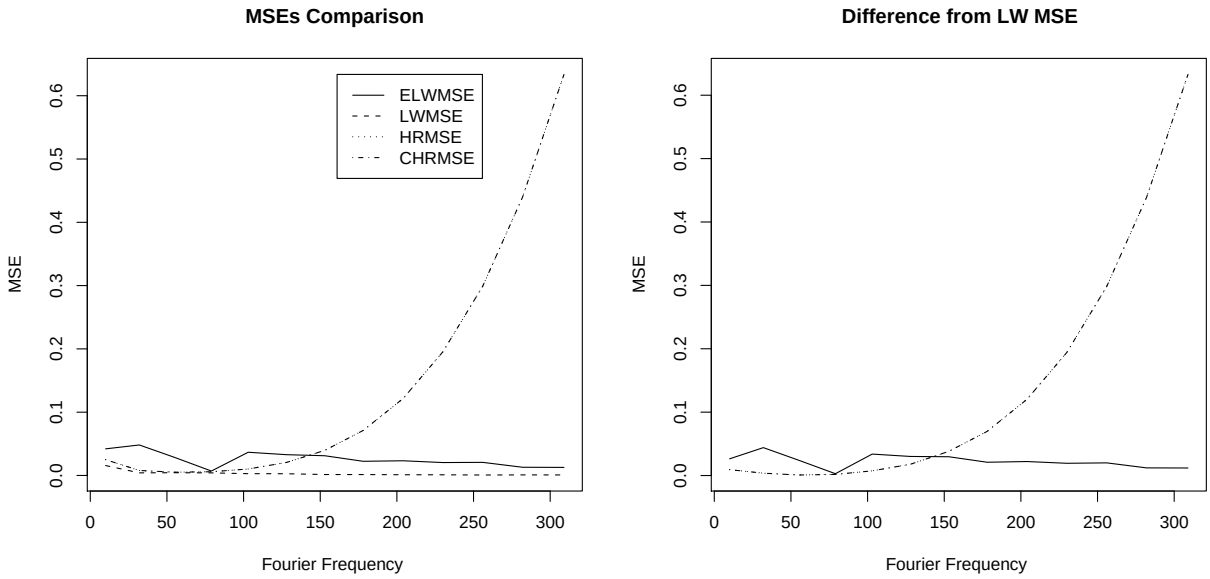

Figure 3.47: Comparison of MSEs, $d = 0.1$, $ns = 0$, $N = 2^{10}$

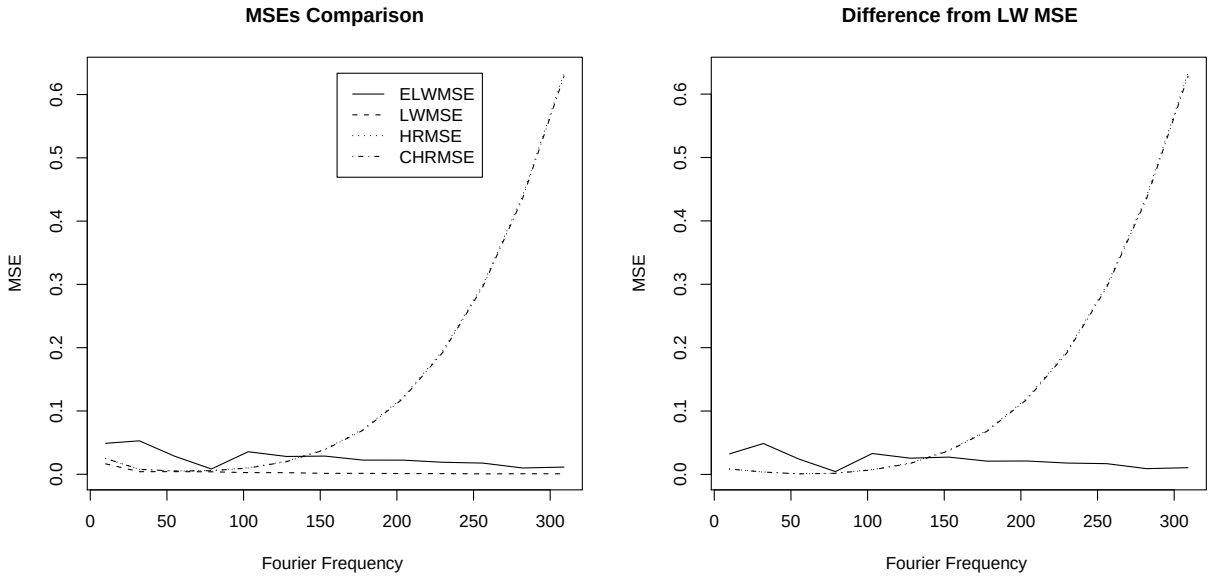

Figure 3.48: Comparison of MSEs, $d = 0.1$, $ns = 0.1$, $N = 2^{10}$

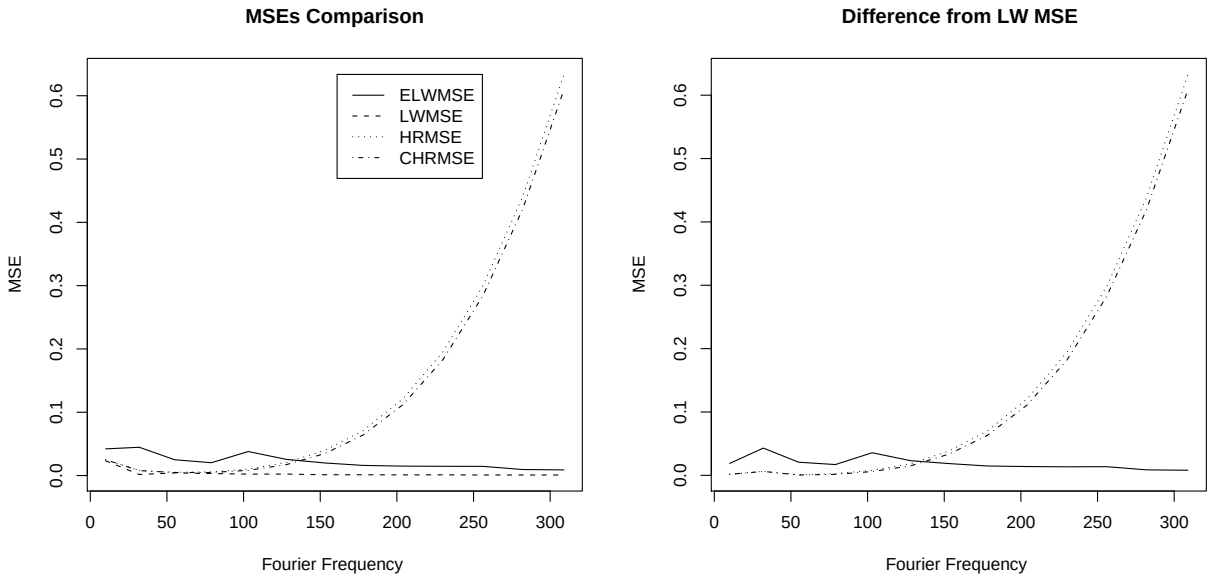

Figure 3.49: Comparison of MSEs, $d = 0.1$, $ns = 0.5$, $N = 2^{10}$

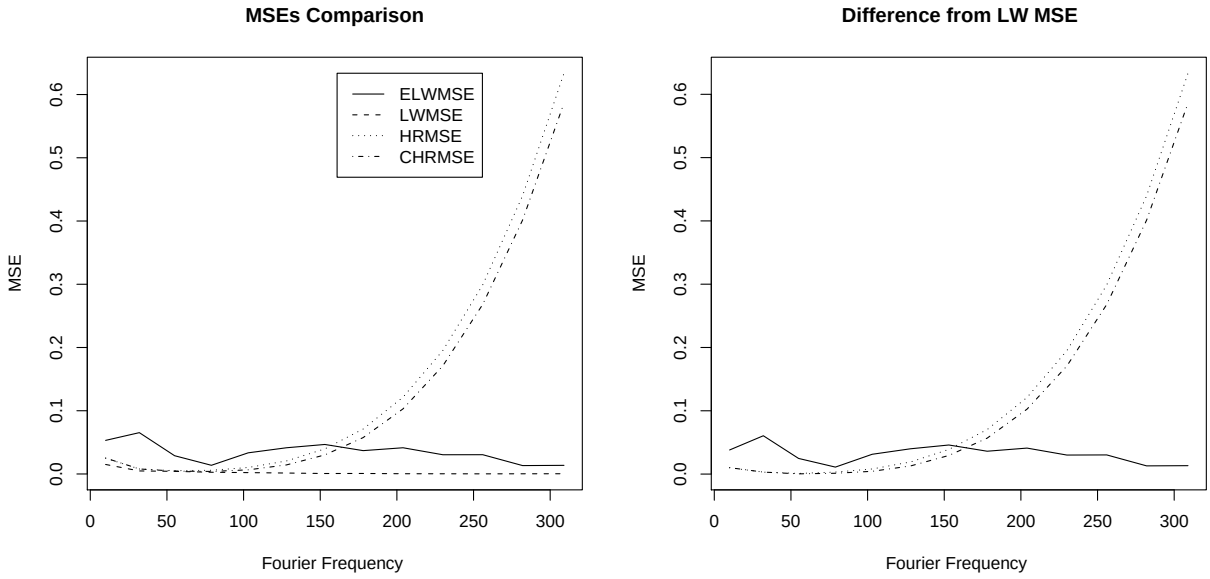

Figure 3.50: Comparison of MSEs, $d = 0.1$, $ns = 1$, $N = 2^{10}$

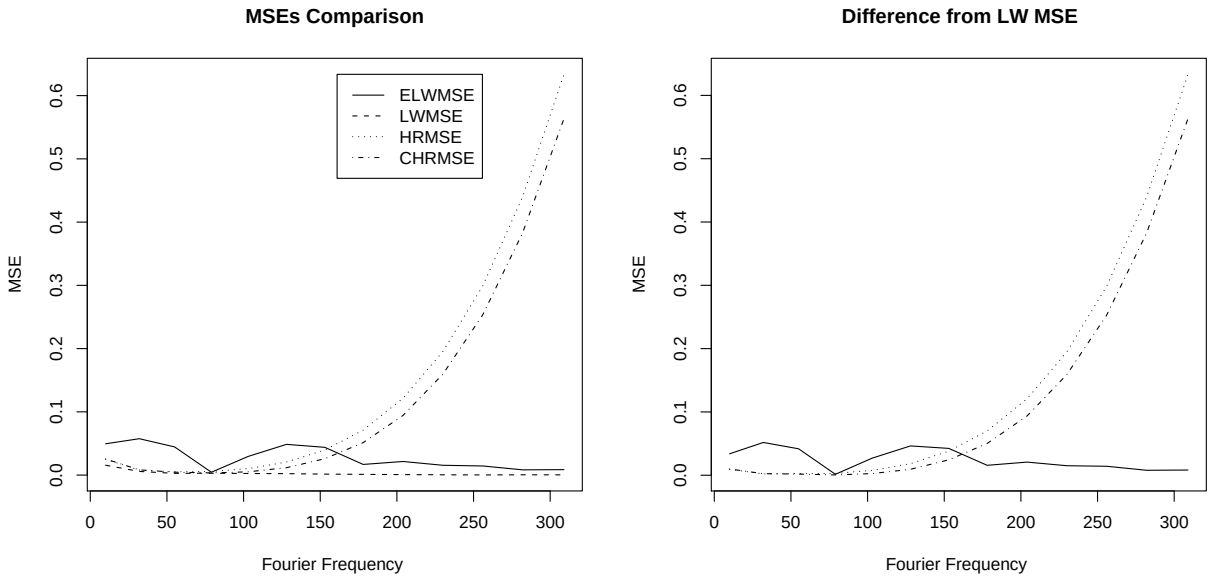

Figure 3.51: Comparison of MSEs, $d = 0.1$, $ns = 1.5$, $N = 2^{10}$

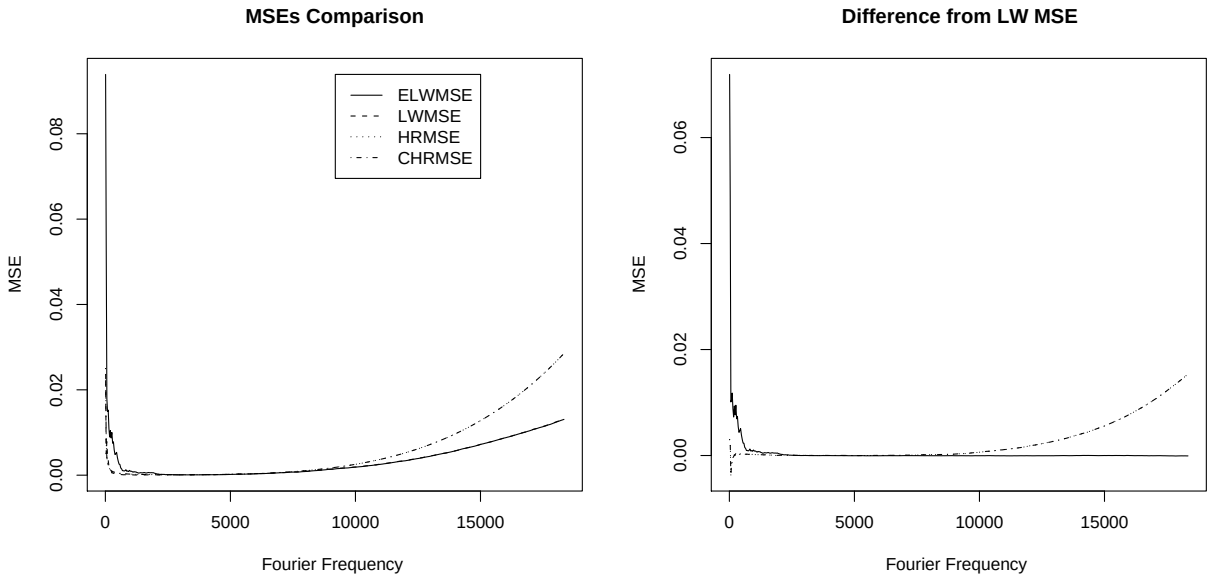

Figure 3.52: Comparison of MSEs, $d = 0.4$, $ns = 0$, $N = 2^{17}$

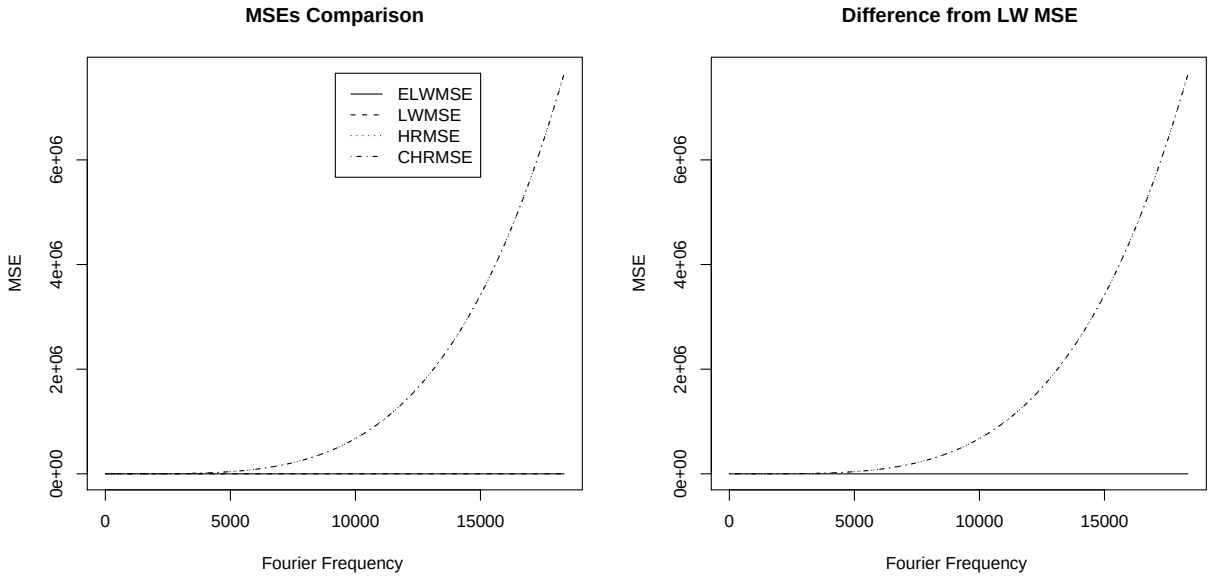

Figure 3.53: Comparison of MSEs, $d = 0.4$, $ns = 0.1$, $N = 2^{17}$

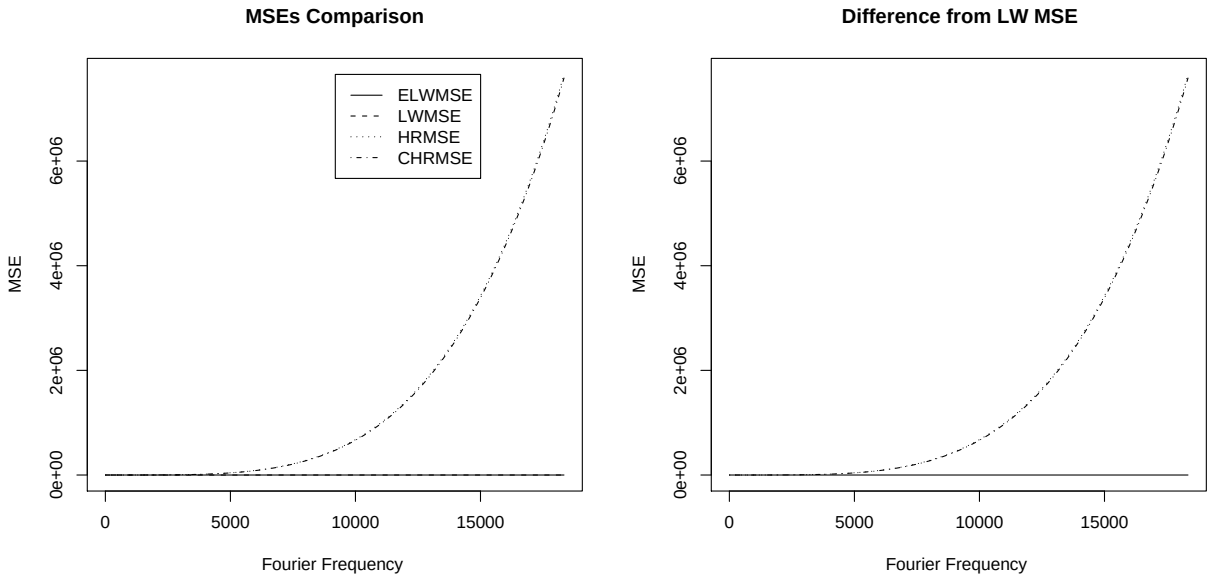

Figure 3.54: Comparison of MSEs, $d = 0.4$, $ns = 0.5$, $N = 2^{17}$

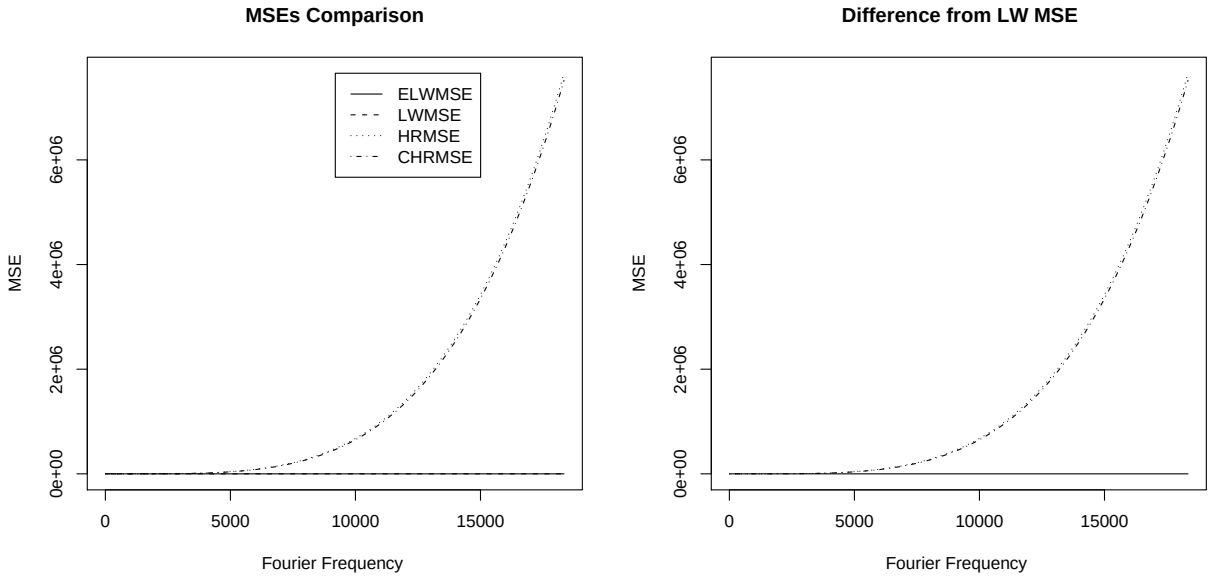

Figure 3.55: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{17}$

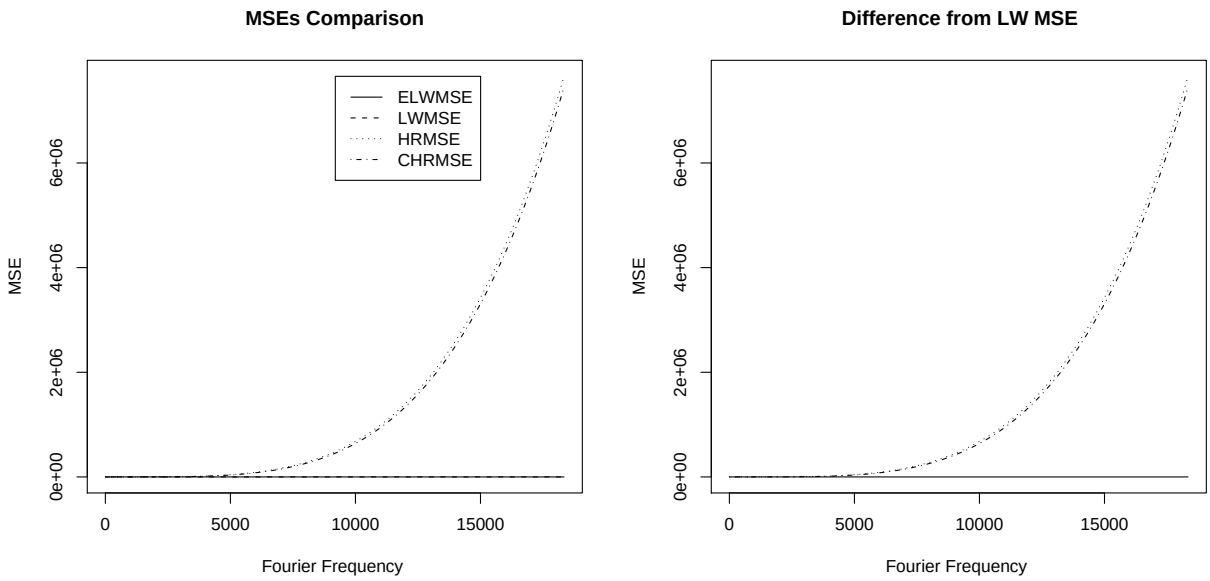

Figure 3.56: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{17}$

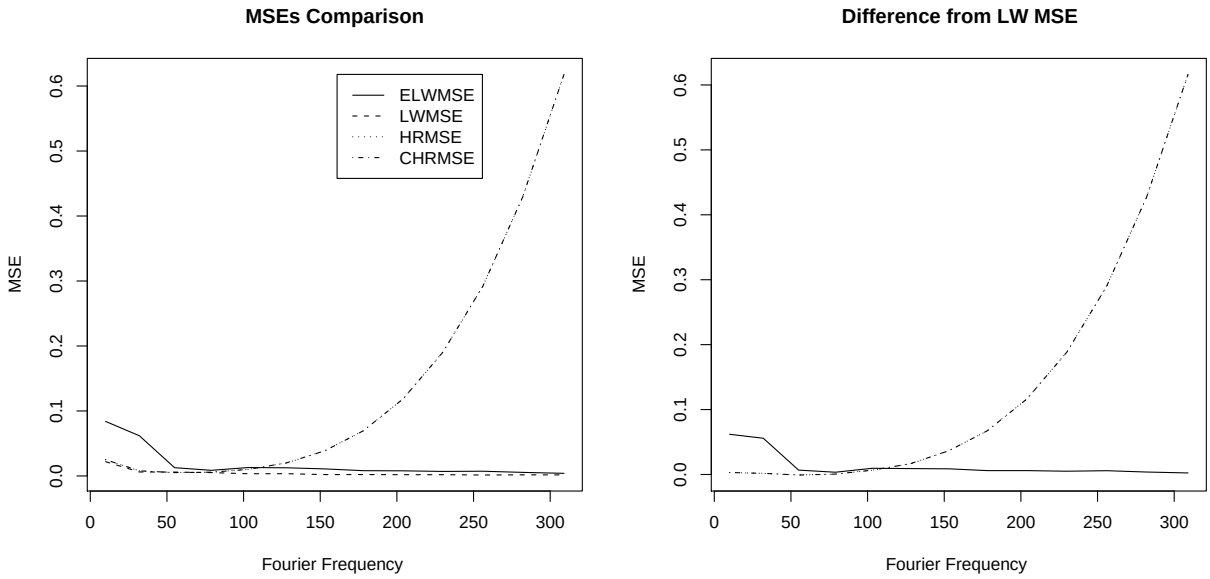

Figure 3.57: Comparison of MSEs, $d = 0.4$, $ns = 0$, $N = 2^{10}$

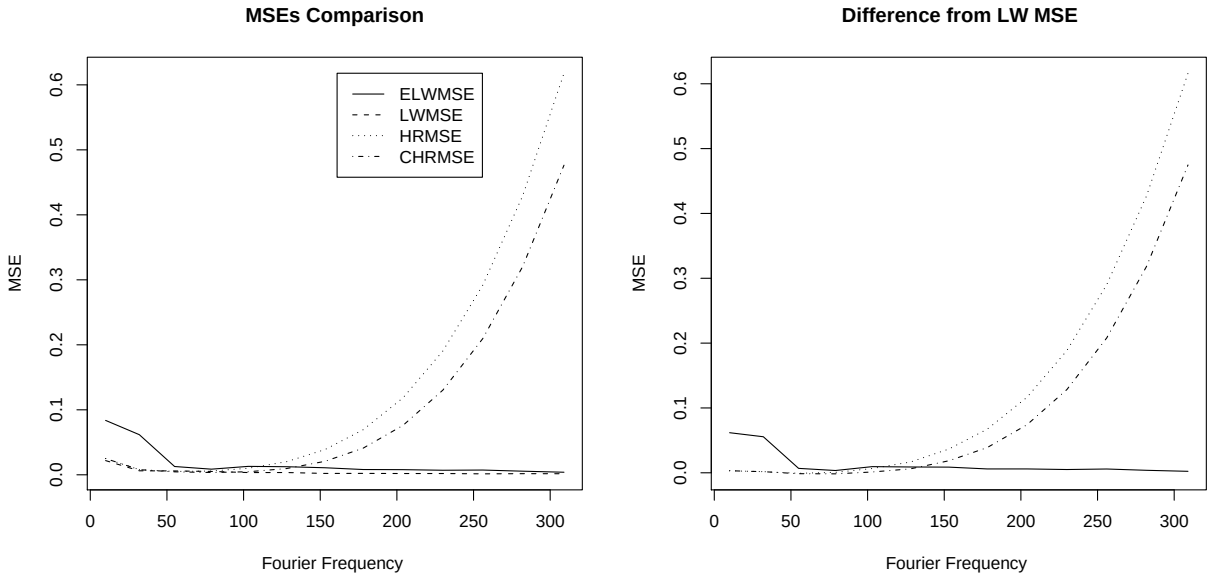

Figure 3.58: Comparison of MSEs, $d = 0.4$, $ns = 0.1$, $N = 2^{10}$

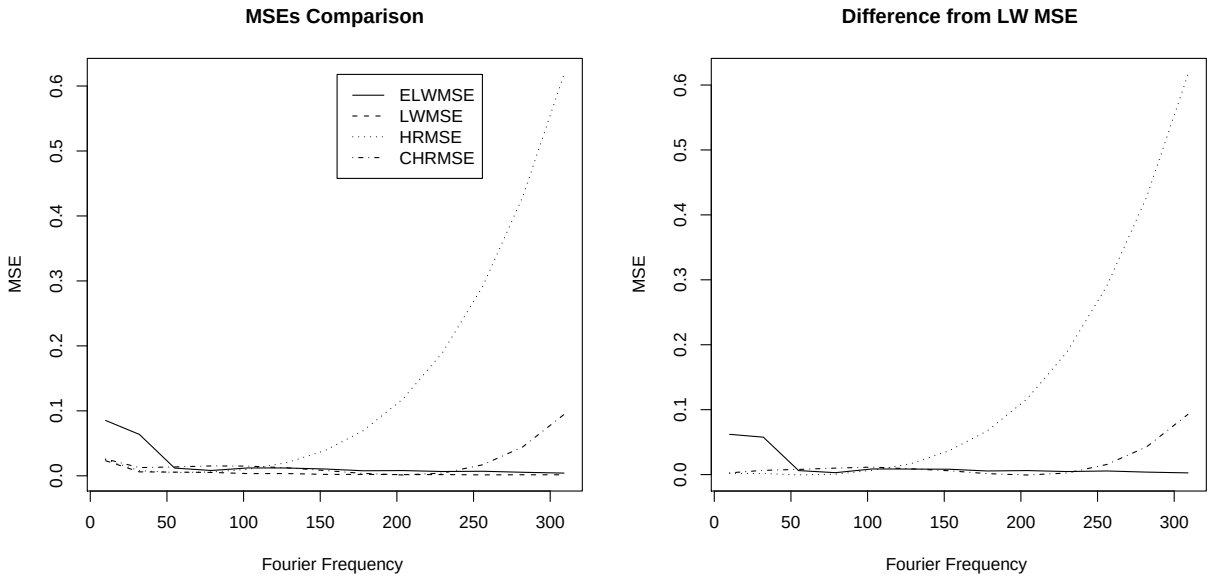

Figure 3.59: Comparison of MSEs, $d = 0.4$, $ns = 0.5$, $N = 2^{10}$

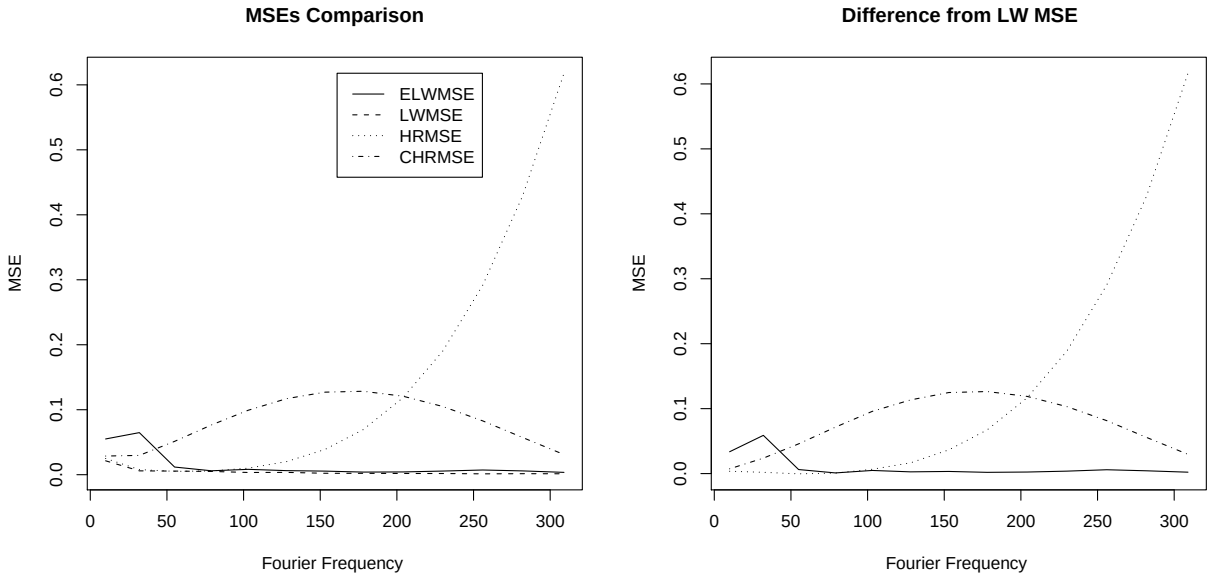

Figure 3.60: Comparison of MSEs, $d = 0.4$, $ns = 1$, $N = 2^{10}$

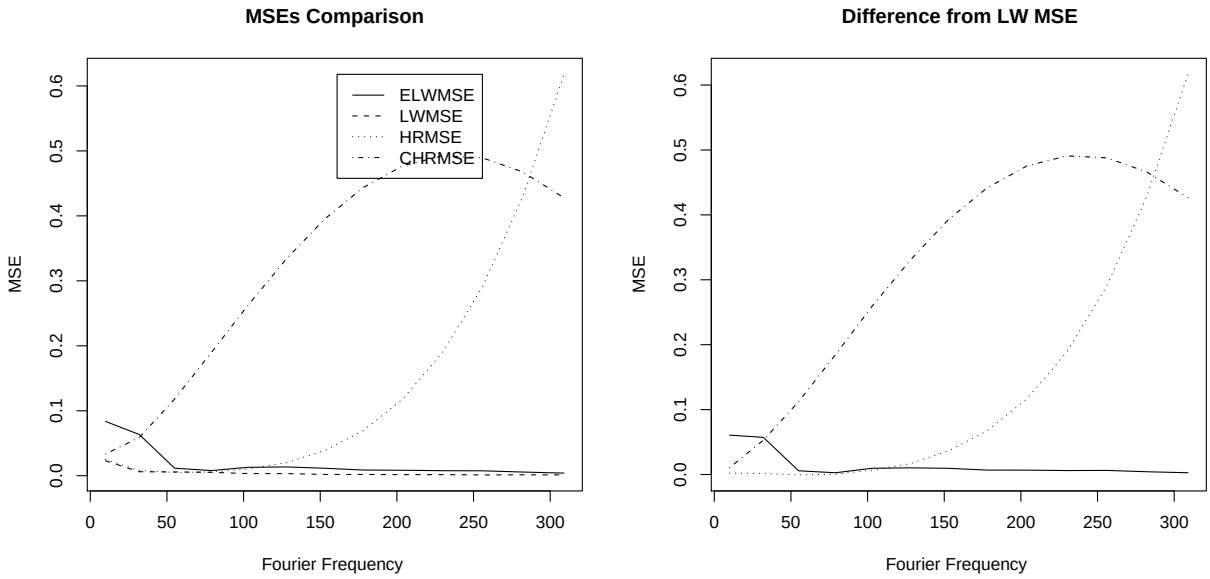

Figure 3.61: Comparison of MSEs, $d = 0.4$, $ns = 1.5$, $N = 2^{10}$

3.2.2 Effects of Aggregation

Figures 3.62 through 3.81 display how the minimal MSE and optimal choice of m vary with aggregation, for given d and ns, $d \in (0.1, 0.2, 0.3, 0.4)$ and $ns \in (0.1, 0.5, 1, 1.5)$. The left-hand side plots display the minimal MSE. The right-hand side plots display the value of the bandwidth parameter m corresponding to the minimal MSE. The x-axis displays the logarithm in base 2 of the number of observations in each aggregation bin. Again, we considered the MSE of the LW estimator, the MSE of the ELW estimator, and the HRMSE and CHRMSE approximations.

We can observe the following facts.

- The MSE of the ELW estimator is almost always monotonically increasing with aggregation.
- The high choice of m for the ELW estimator indicates that the variance component largely dominates the MSE.
- The LW estimator performs almost uniformly better than the ELW estimator, even when the noise-to-signal ratio is high. The difference is more pronounced the lower the value of d.

Given the observations above, and the fact that the ELW is numerically unstable, it seems hard to recommend its use in practice for estimating LRD, even when the presence of a noise component is suspected.

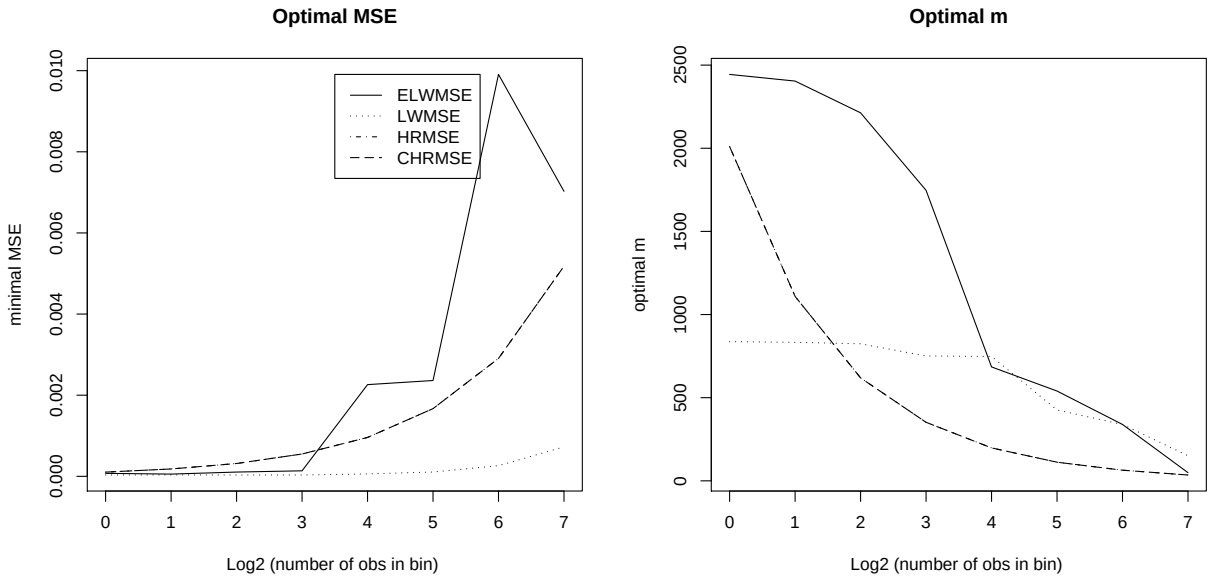

Figure 3.62: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0$

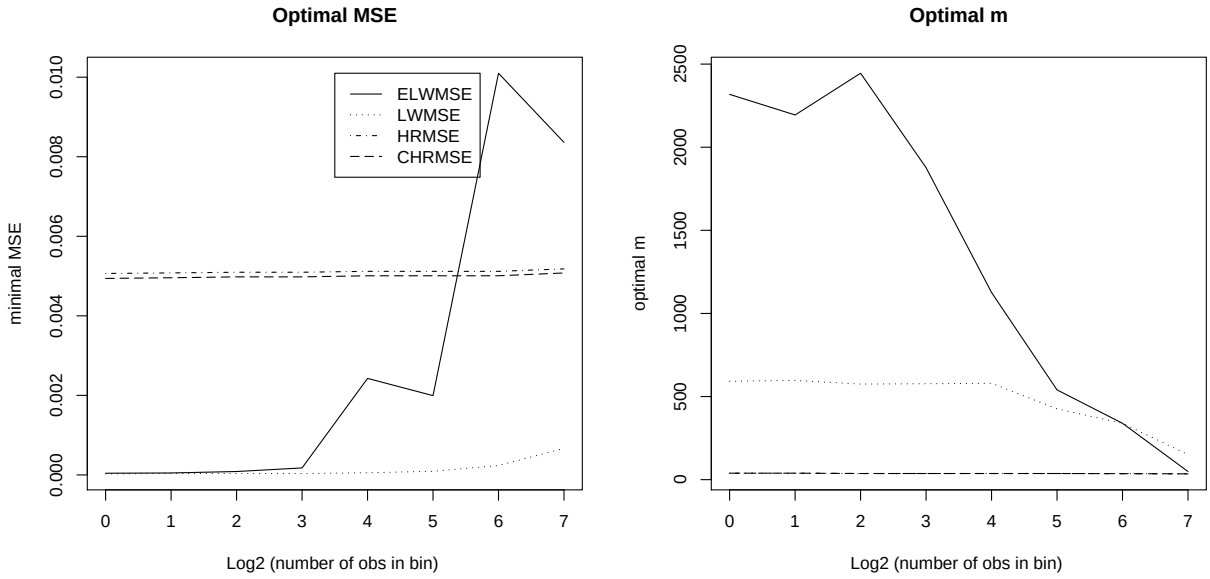

Figure 3.63: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0.1$

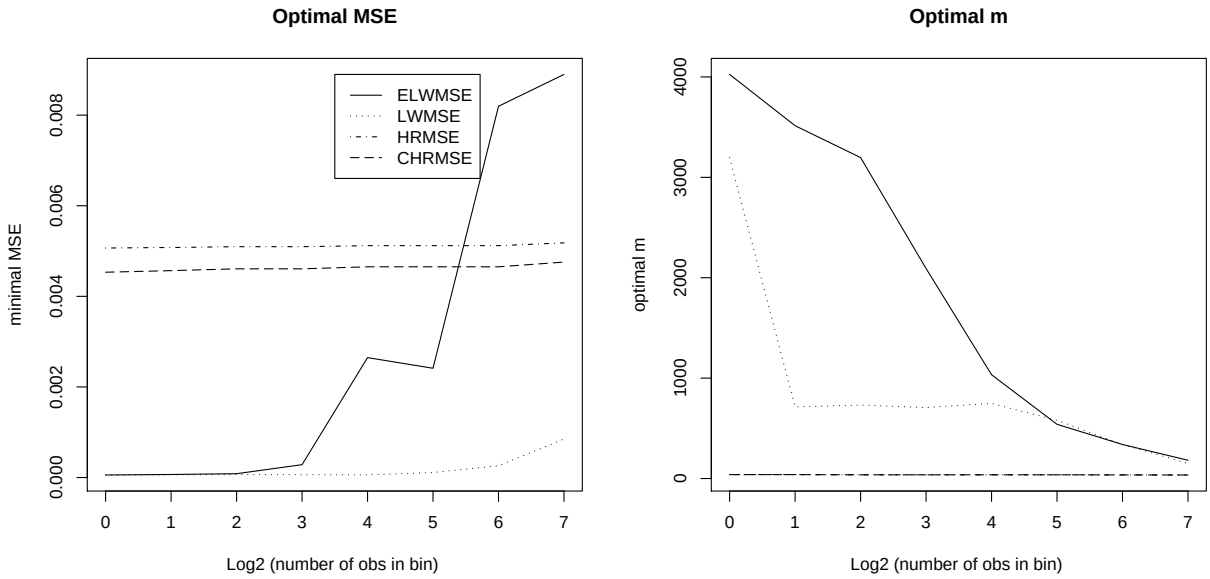

Figure 3.64: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 0.5$

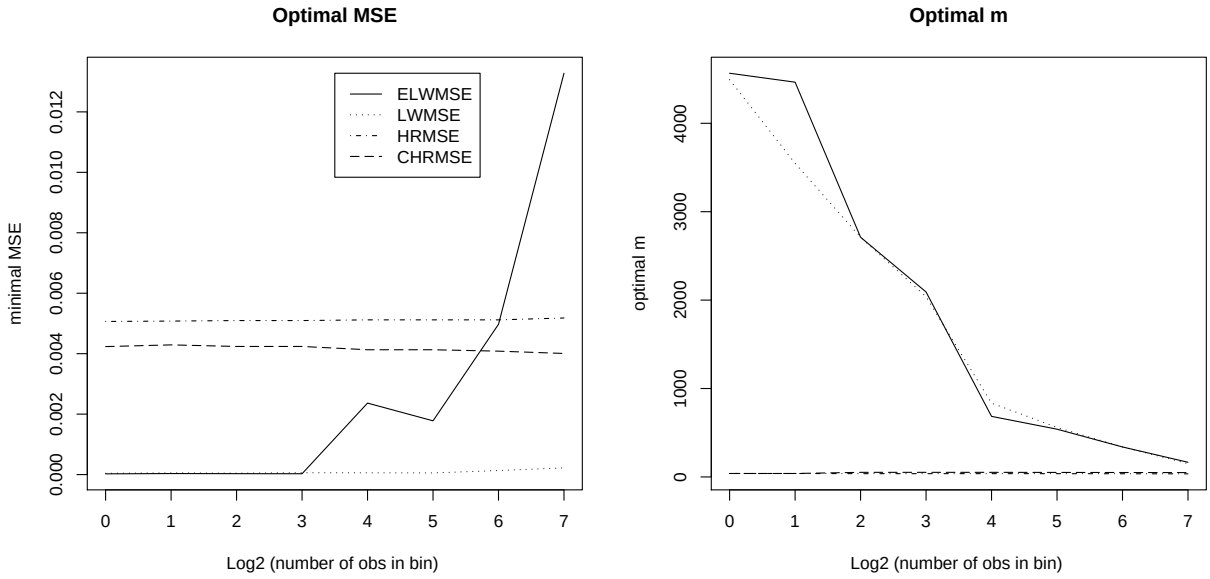

Figure 3.65: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 1$

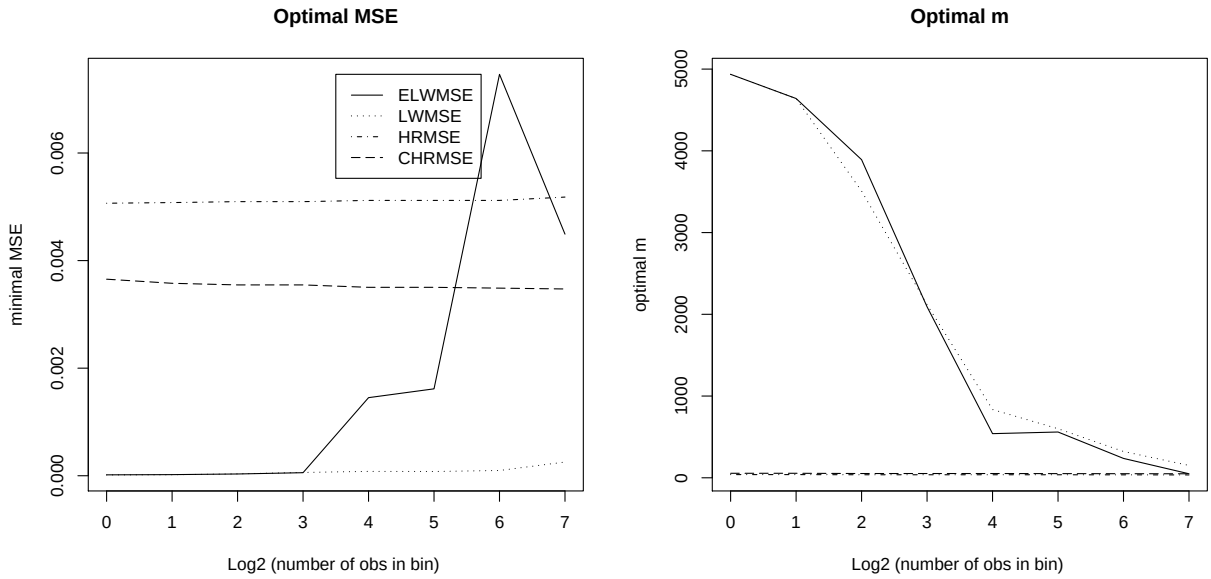

Figure 3.66: MSEs and Optimal m vs. Aggregation, $d = 0.1$, $ns = 1.5$

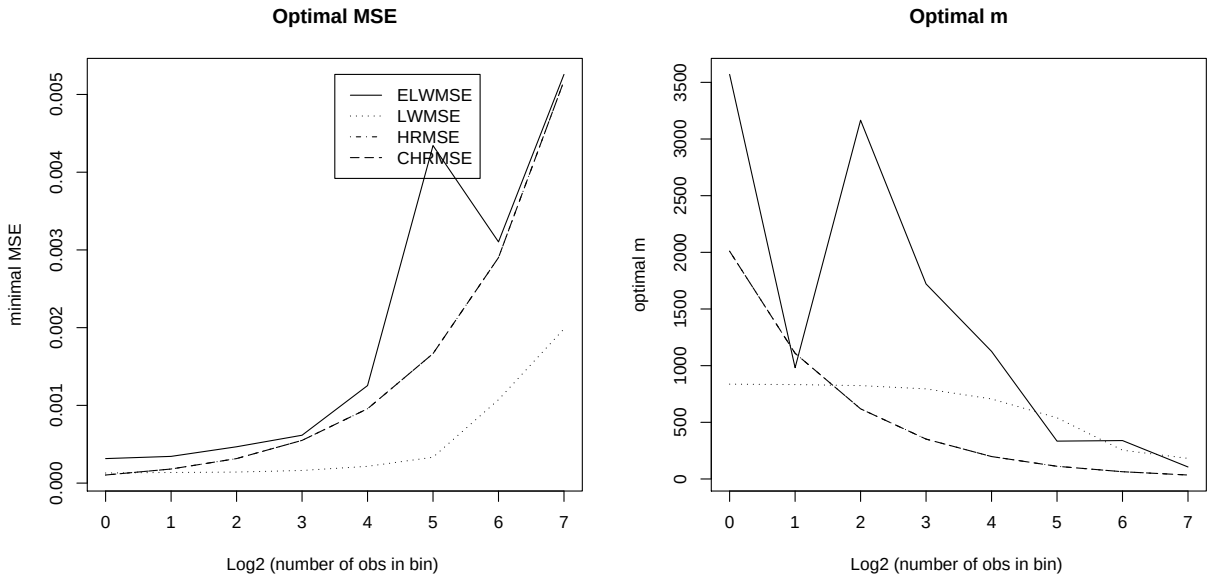

Figure 3.67: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0$

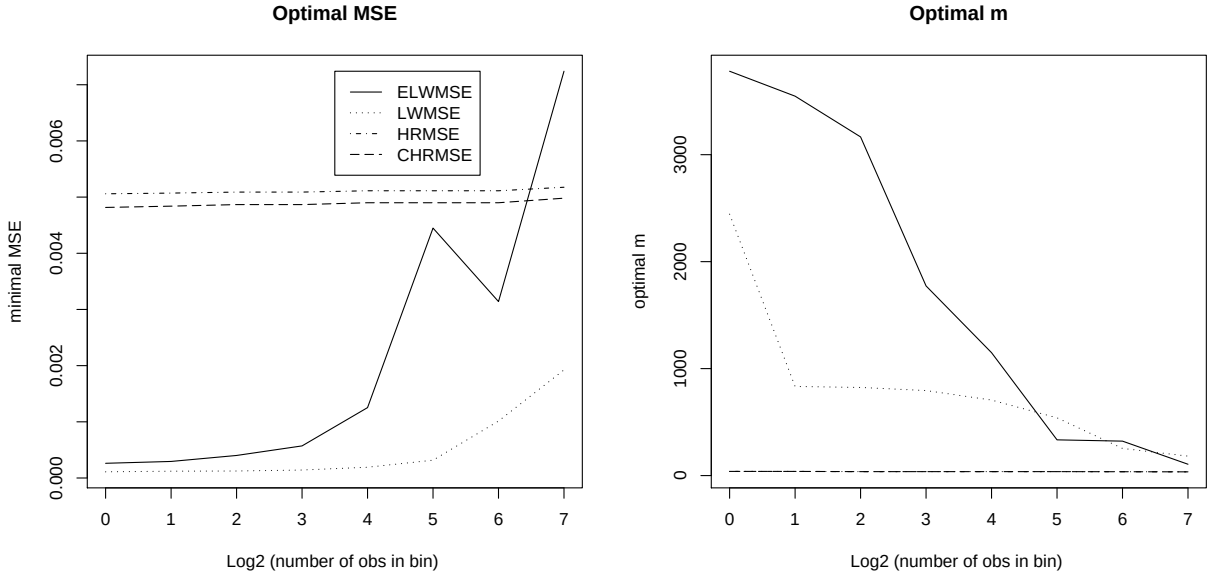

Figure 3.68: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0.1$

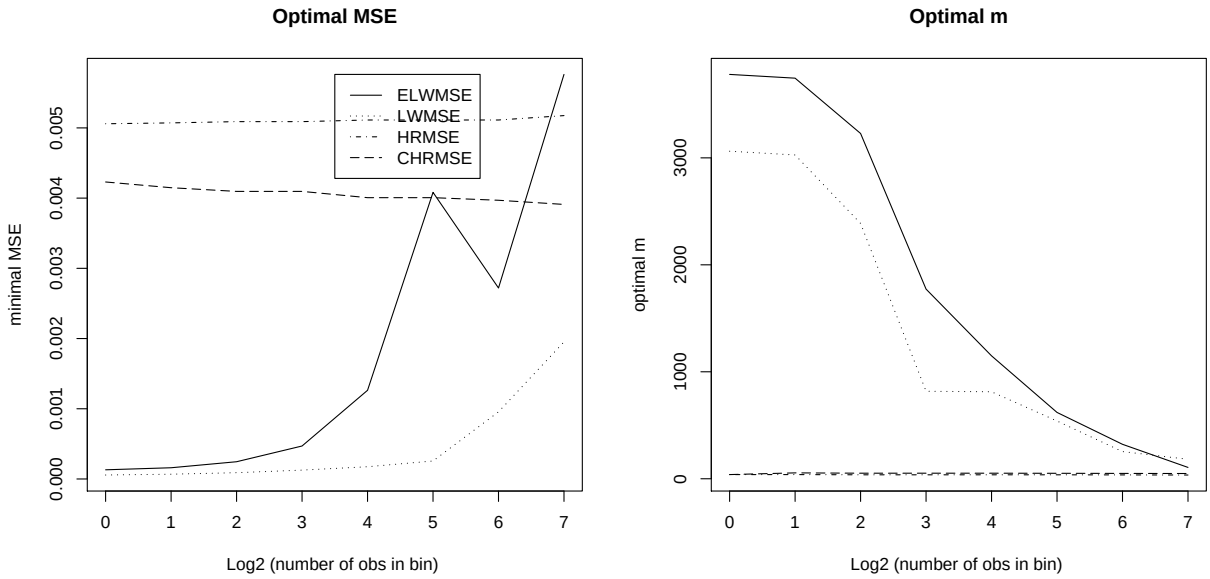

Figure 3.69: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 0.5$

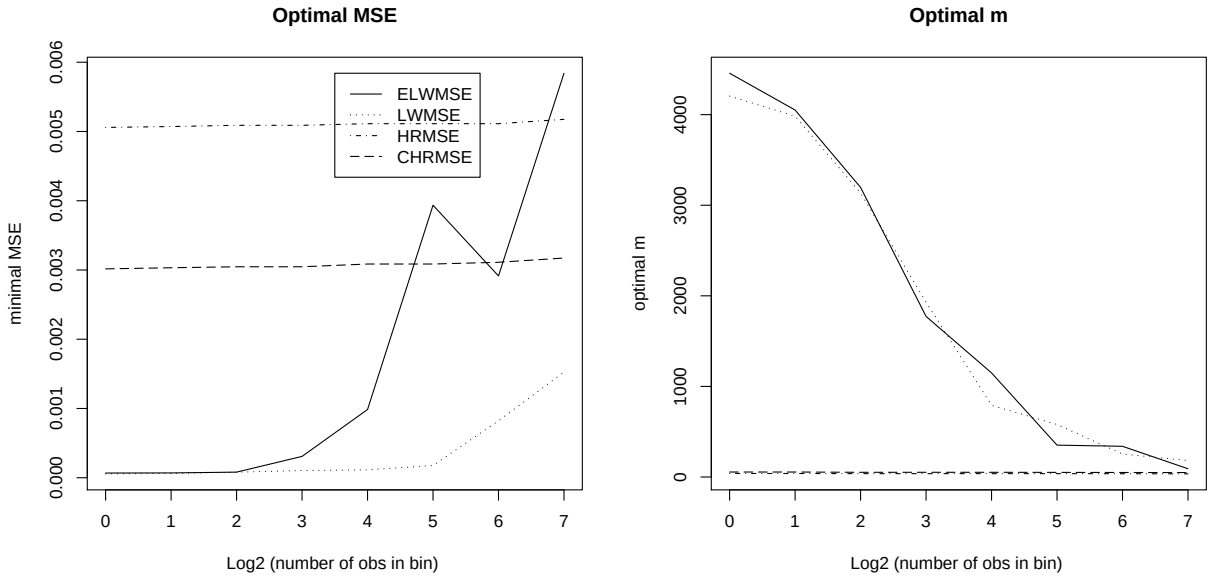

Figure 3.70: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 1$

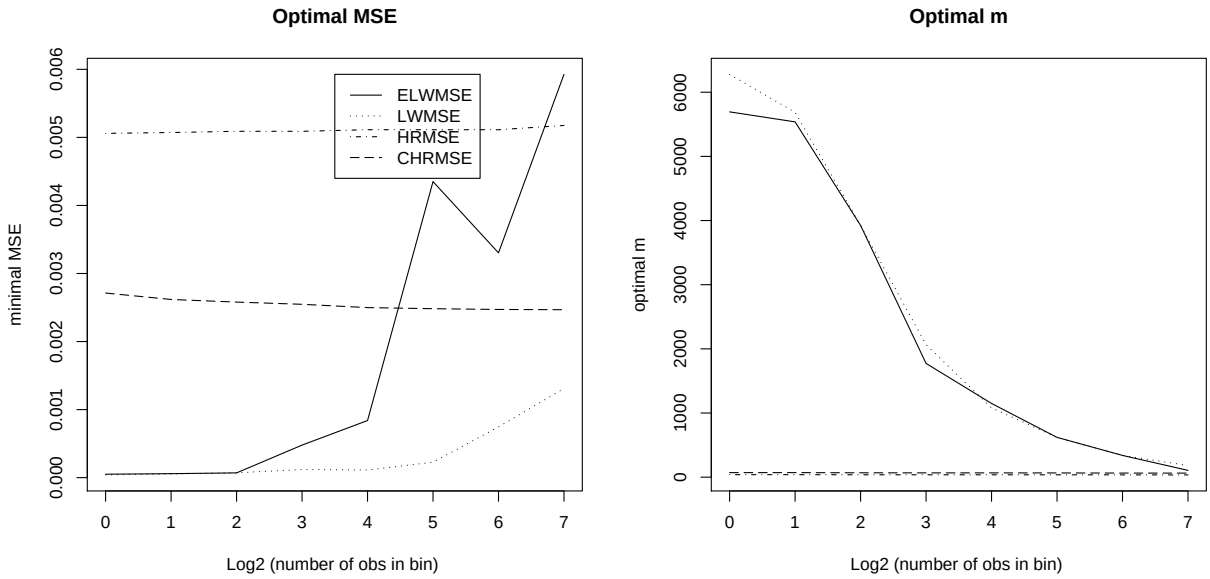

Figure 3.71: MSEs and Optimal m vs. Aggregation, $d = 0.2$, $ns = 1.5$

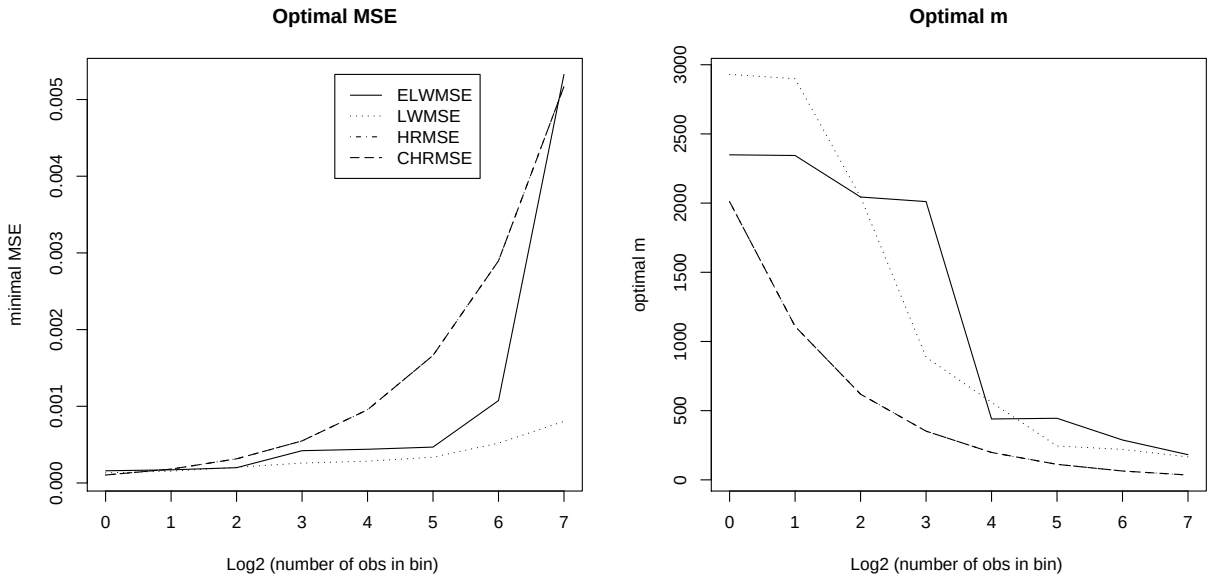

Figure 3.72: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0$

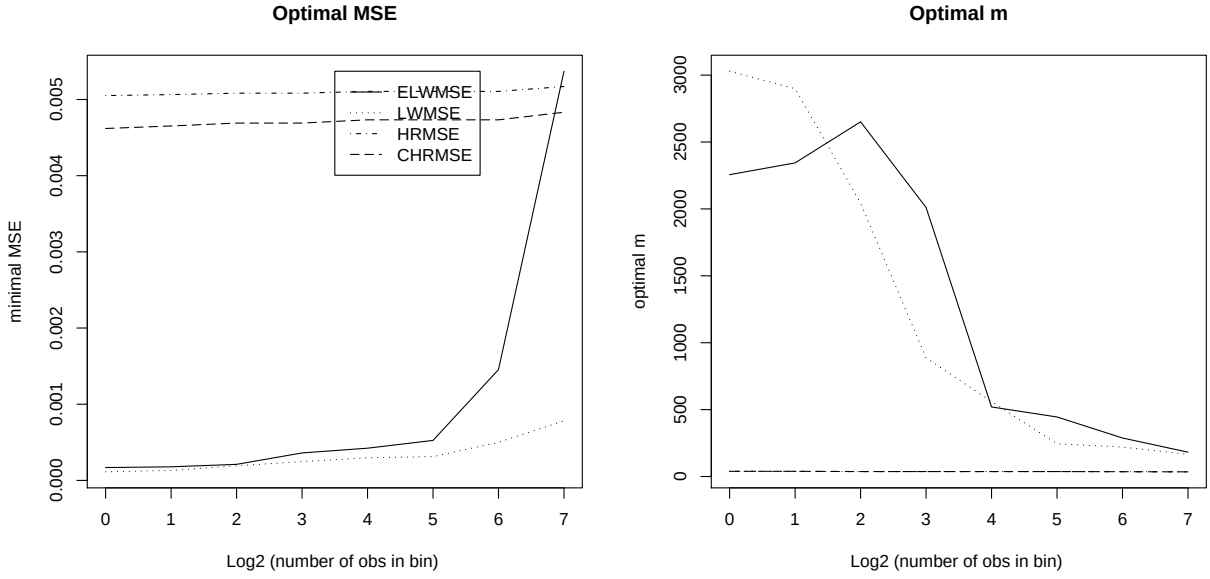

Figure 3.73: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0.1$

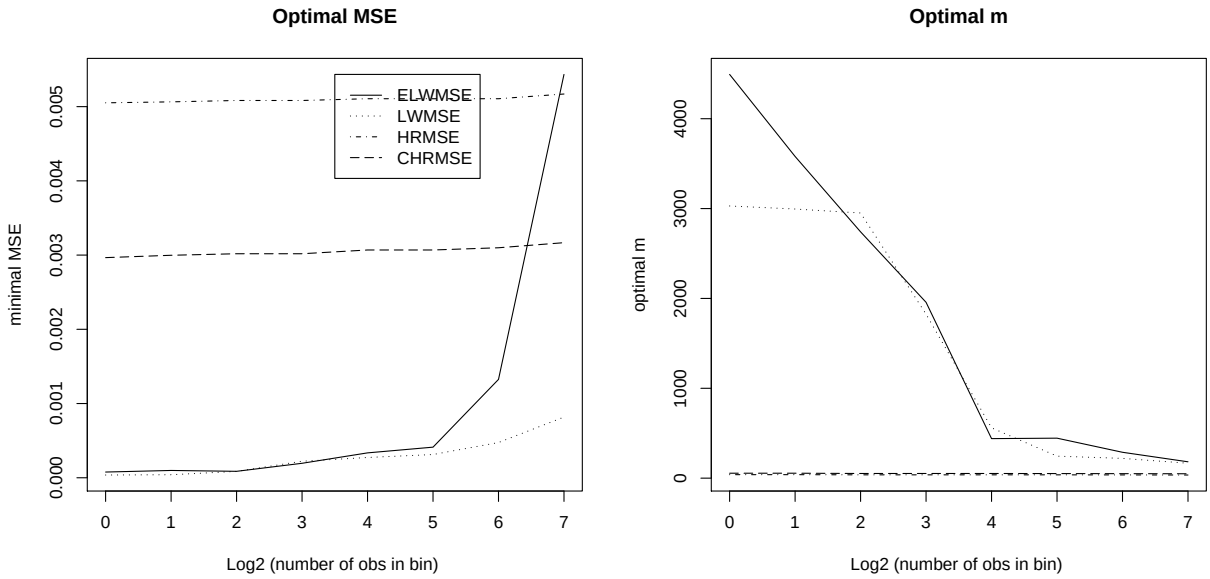

Figure 3.74: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 0.5$

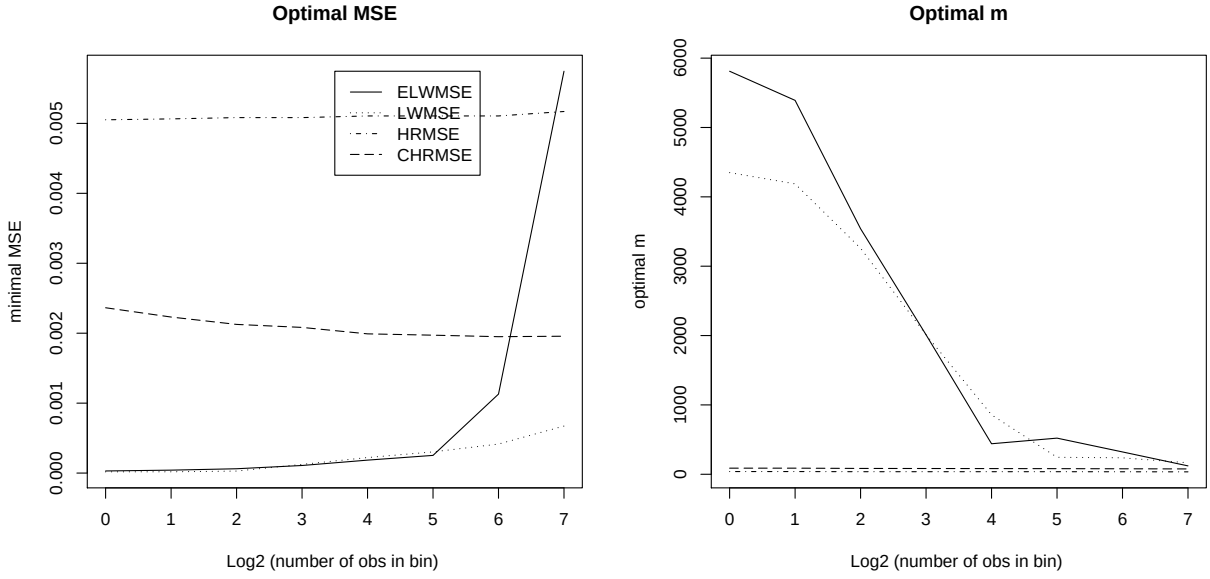

Figure 3.75: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 1$

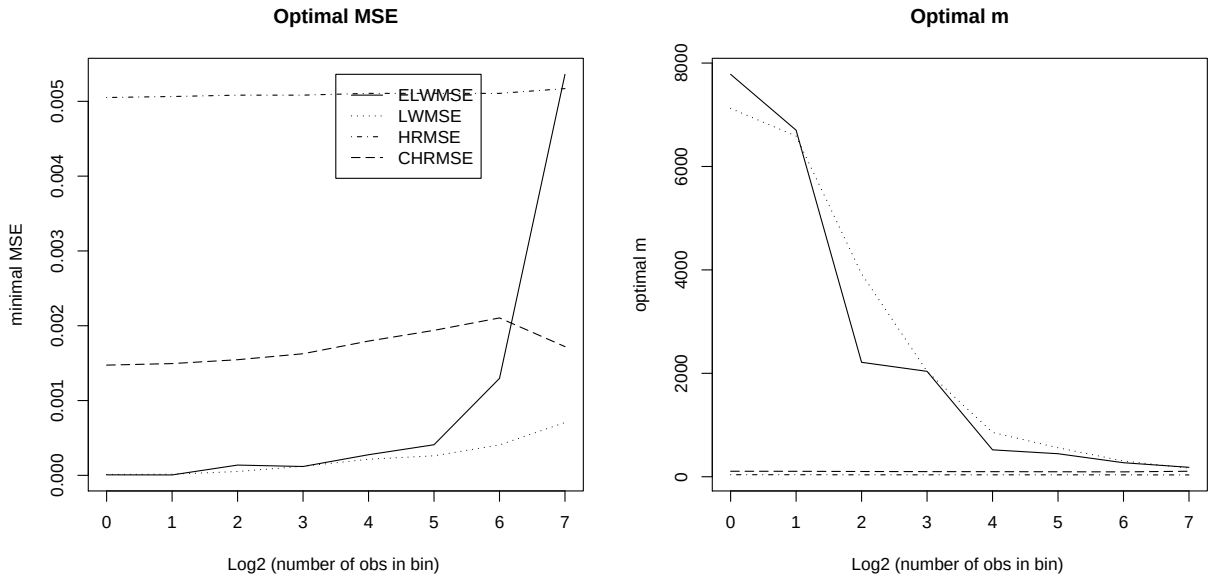

Figure 3.76: MSEs and Optimal m vs. Aggregation, $d = 0.3$, $ns = 1.5$

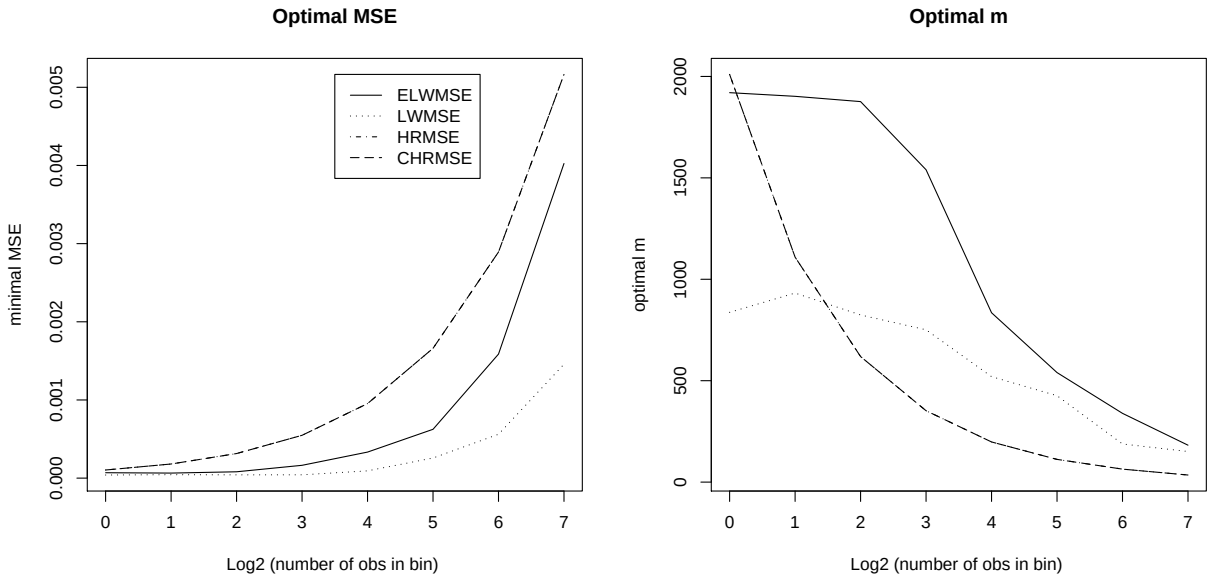

Figure 3.77: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0$

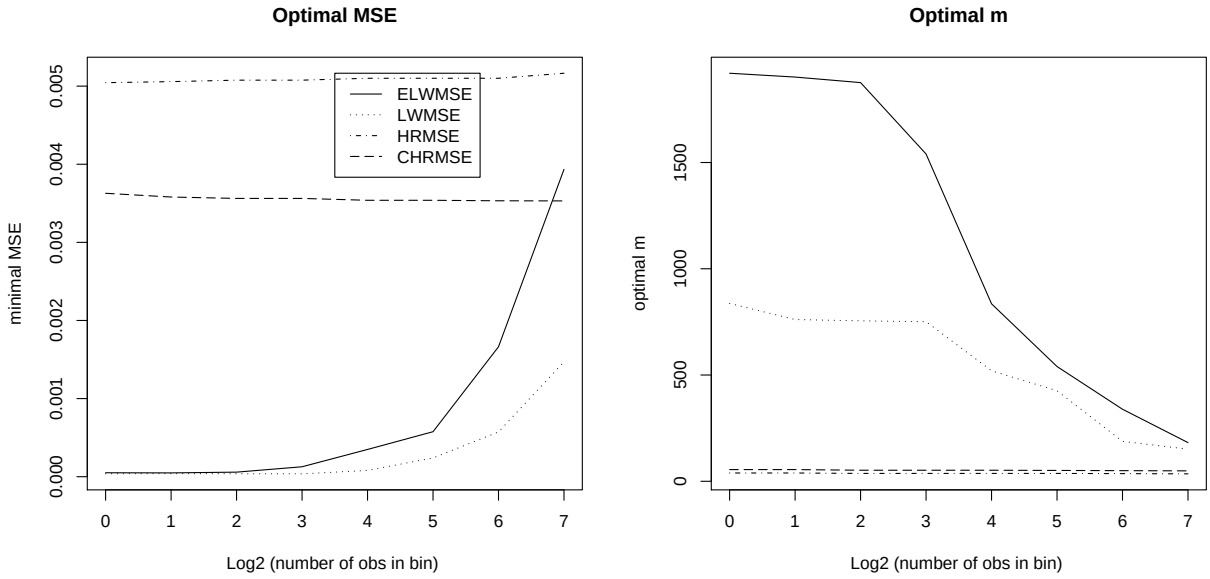

Figure 3.78: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0.1$

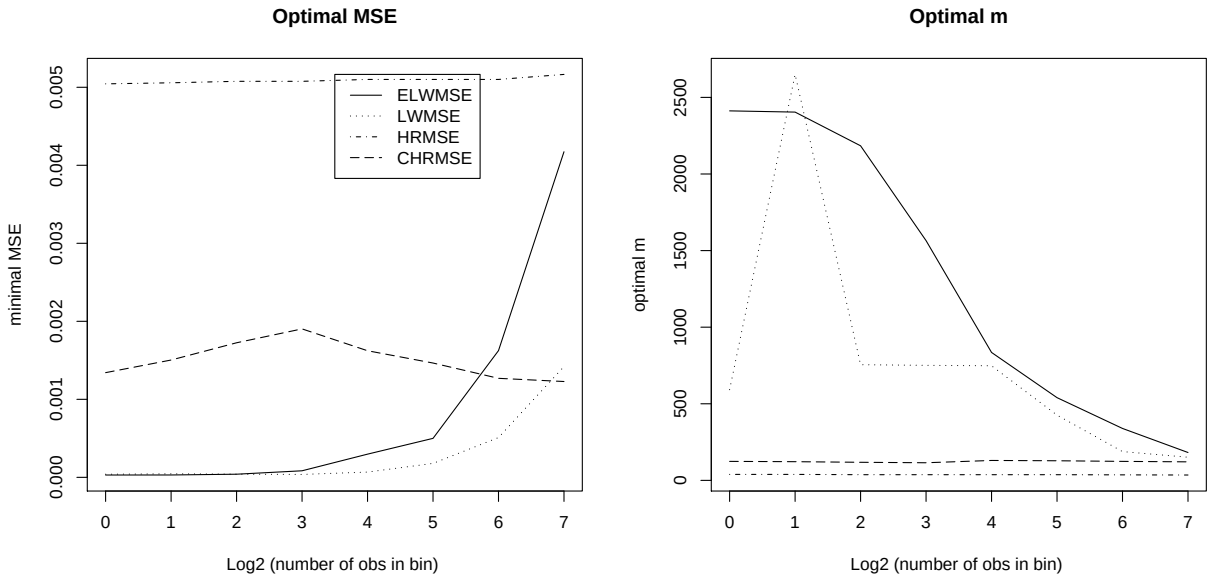

Figure 3.79: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 0.5$

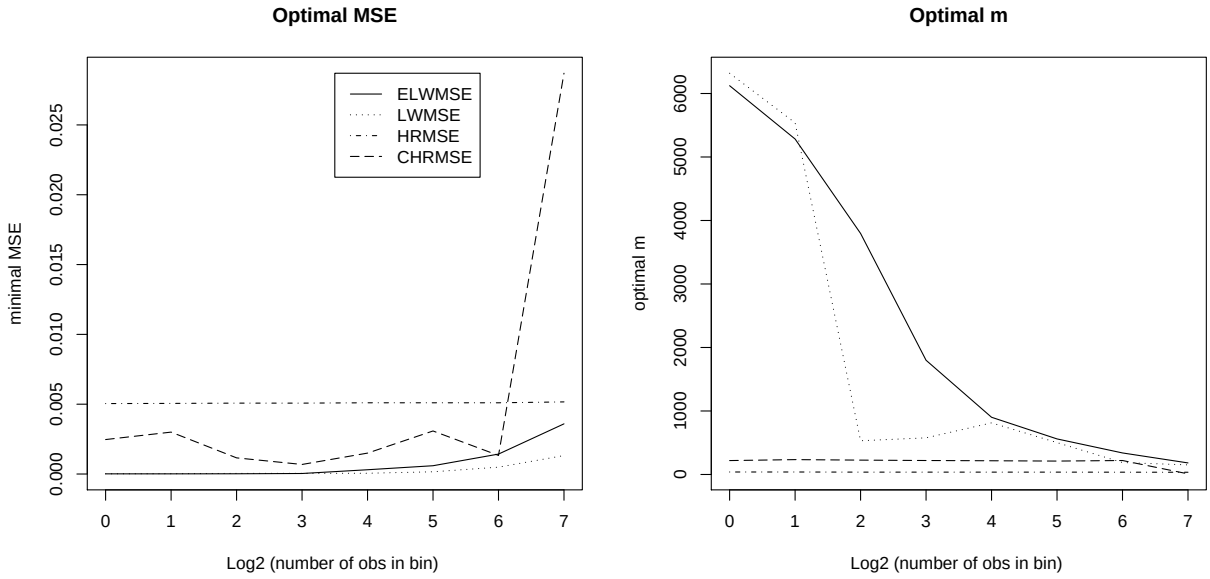

Figure 3.80: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 1$

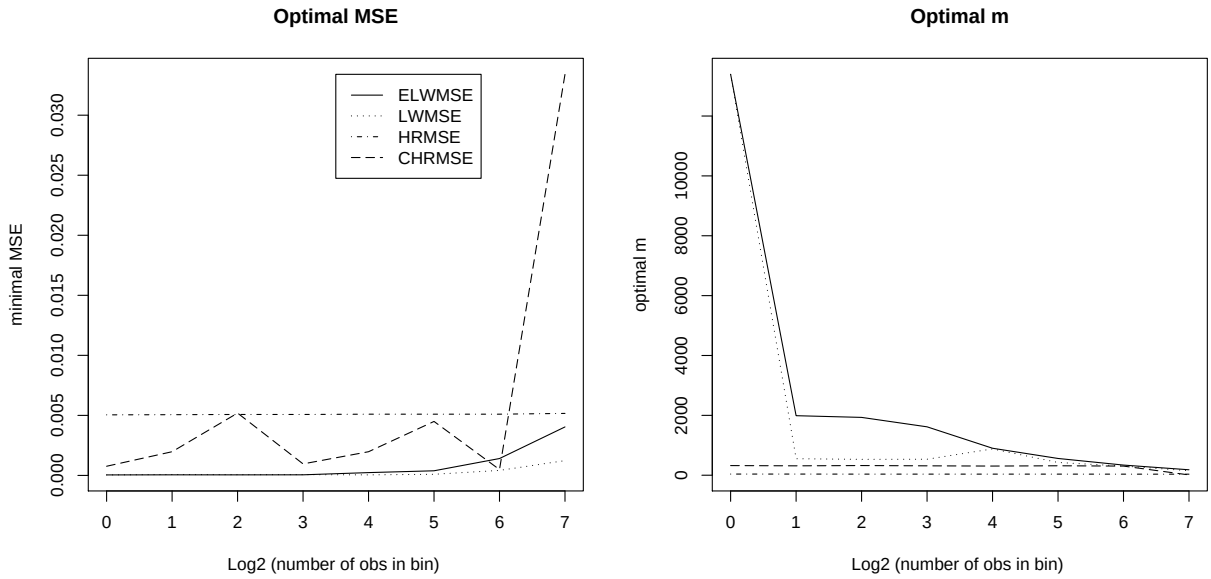

Figure 3.81: MSEs and Optimal m vs. Aggregation, $d = 0.4$, $ns = 1.5$

Chapter 4

UNC Internet Data Analysis

In this chapter we will investigate how aggregation influences the Local Whittle (LW) and the Geweke and Porter-Hudak (GPH) estimators in applications, and whether our preliminary findings in Chapter 2 represent a reasonable approximation of the behavior of the estimates in a finite sample.

While the definition of long-range dependence based on the periodogram involves a slow-varying function (see Definition 1.1.2), for estimation purposes it is assumed that the spectral density of a LRD process in a neighborhood of the origin can be approximated by

$$f(\lambda) \approx c\lambda^{-2d} \tag{4.1}$$

where $c > 0$ is a constant.

We will focus our attention on how the estimates of the parameters c and d vary across aggregation levels, and, in particular, how that affects the choice of the optimal bandwidth parameter m_{opt}.

Section 4.1 describes in details the data that were utilized for this purpose. In Section 4.2 we illustrate the results of the analysis for a selection of the data sets considered. Finally, in

Section 4.2.3 we draw our conclusions and suggest possible developments.

4.1 Data Description

4.1.1 UNC Internet Data

The data consist of 20 separate data sets collected by the DIstributed and Real-Time systems (DIRT) group (http://www.cs.unc.edu/Research/dirt/) of the UNC Computer Science Department during the second week of April 2002. A monitor of one microsecond resolution and approximately one millisecond accuracy was placed on the main router that connects the UNC campus network to the Internet. Packet headers and arrival times were recorded for TCP/IP packets traversing the router, and the traces were subsequently processed to create time series of packet arrival counts per each millisecond interval. All data traversed a 1 Gbps (Giga bites per second) Ethernet link from the ISP router to the campus aggregation switch. The UNC campus itself is on switched 1 Gbps Ethernet network and includes the UNC hospitals, for a total of over 50,000 computers and 35,000 users, giving rise to a very heterogeneous usage which includes business, education, entertainment, and a variety of protocols such as http, ftp, ssh, etc. Each data set spans approximately a 2 hour period in a way to represent the different usage times during a normal week. The 2-hour time interval was selected somewhat arbitrarily in order to maintain the traffic level approximately stationary within each data set. The number of observations in each data set is of the order of 72 million. In aggregate, the traces represent 40 hours of network entry traffic into the UNC campus and include information about 3.55 billion packets that carried 1.17 Terabytes of data. The average load of utilization of the link during the duration of the data collection precess ranged from a low of 2.7% on Tuesday at 5:00 AM, to a high of 9.1% on Thursday at 3:00 PM. More details on the data can be found in Park et al. (2005).

4.1.2 Lab Data

The synthetic data was generated in a laboratory testbed network designed to emulate a network with characteristics similar to the UNC campus network. At one end of the link there were 18 machines ("clients") that emulated the behavior of hundreds of Web users. On the other end of the link, there were 10 machines ("servers") that acted as web servers. The traffic load is controlled by a parameter that specifies a fixed size population of users browsing the web and was set to 2%, 5%, 8%, 11%, and 14% of capacity to emulate the real load recorded on the UNC main router link. More details on the data collection can be found in Park et al. (2005), and on the data generation process in Le et al. (2003).

4.1.3 Simulated data

Three different traces that simulate FGN have been analyzed. To be consistent with the findings about the real data, all have Hurst parameter $H = 0.9$, equivalent to $d = 0.4$, but differ in the algorithms used to generate them. The first and last data sets (`fgn_pipiras` and `fgn_long`) have 72 million observations to replicate the length of the real data. The second data set (`fgn_zhu`) instead has 524288 observations due to memory limitations in the software package used to generate it. The variance has been set to 20, given the fact that the mean number of packet per millisecond for the UNC data is approximately 20 and that, for the Poisson model, the variance is equal to the mean. As with the previous data sets, these series have been discussed in details in Park et al. (2005).

Details of the wavelet method, corresponding to the data set `fgn_pipiras`, can be found in Abry and Sellan (1996) and Pipiras (2005). It is based on a fast orthogonal wavelet transform that simulates a sequence of two discrete times FARIMA series. The last series is then subsampled to

obtain an approximation to a fractional Brownian motion whose first difference series constitutes the FGN.

Details about the first spectral method, corresponding to the data set `fgn_zhu`, can be found in Wood and Chan (1996) and Dietrich and Newsam (1997). This method is an exact simulation of the FGN and relies on the fact that the periodic repetition of the covariance function of the FGN in the interval $[0, 1]$ is a circulant matrix.

Details about the second spectral method, corresponding to the data set `fgn_long`, can be found in Paxson (1995) and Danzig et al. (2000). It is based on generating sample paths that have the same power spectrum as the FGN.

Three FARIMA series were also simulated using the `arima.fracdiff.sim` function of `Splus`. The function is based on the algorithm described in Haslett and Raftery (1989). Due to the amount of time needed to generate the series (over 36 hours each on the UNC statistical server), the number of observations is limited to one million per each series. The series are as follows: FARIMA$(1, .3, 0)$, with $ar(1) = 0.2$; FARIMA(1, .3, 0) with $ar(1) = 0.9$; FARIMA(1, .3, 1) with $ar(1) = 0.9$ and $ma(1) = 0.5$.

4.2 Analysis

4.2.1 UNC Internet Data

We chose to display only the results for the subset of data collected on Wednesday April, 10 2002 at 9:30 pm. This data set was selected because it yielded consistent estimates of the long range dependence parameter across different estimators and it does not show clear visual evidence of nonstationarity (see Park et al. (2005) for details.)

Since the algorithm that we used to compute the FFT is most efficient when the number of

observations can be factored in small primes (less or equal to 13), we used only the first 7,200,000 observations.

The data were aggregated in bins of one hundredth of a second (1cs), one tenth of a second (1ds) and one second (1s). The local Whittle and GPH estimates were computed for the original 1ms data and all aggregation levels for a number of frequencies m ranging from 10 to the minimum of 10,000 and the number of frequencies available. That means that the maximum number of frequencies used is 10,000 for all data sets, except for the 1s aggregation level, for which it is 3,600 due to the limited numbers of observations available at the highest aggregation level.

Figure 4.1 through Figure 4.8 illustrate the estimates for each fixed aggregation level. The last portion of the plot titles indicate the aggregation level, while the subscript `_detr` indicates that the data has been detrended prior to estimation.

The two plots in the first row from top show the local Whittle and the GPH estimates of d versus $m/10$.

We used two methods for selecting the bandwidth parameter m and uniquely determining the estimates. First we looked at the plot of $\hat{d}$ versus m and found a common value of m where the estimates seemed to stabilize. We picked such a value as choice of m for all data sets. We named this *automatic method*. The vertical line in the plots corresponding to 200 (i.e. m=2,000) indicates such a choice. For this particular data set, $m = 2,000$ seems excessive since the estimates stabilize well before that. This fact is confirmed also by the plots on the third row from top, which show the estimates of $\hat{d}$ versus the inverse of the periodogram ordinates, according to the visual method for selecting m suggested by Taqqu and Teverovsky (1996). However, one must remember that the choice of $m = 2,000$ was made to keep the values uniform across all UNC internet data sets.

The second method to select m, which we named *tuned*, consists in finding an estimate of the

asymptotic MSE of $\hat{d}$, and choosing the value of m that minimizes it. We attempted to use the plug-in algorithm suggested by Henry and Robinson (1996) but found that it did not converge to a unique values in most cases. Therefore we resorted to compute estimate the asymptotic MSE for each value of m for which the parameter was estimated and then plotting it against m to find its minimum. For the local Whittle estimator we used the formula for asymptotic MSE in Henry and Robinson (1996), while for the GPH estimator, we used the formula in Smith (1989).

The plots on the second row show the estimated MSE against m. The second vertical line in the plots indicate the ordinate at which the MSE attains its minimum. In the case of the LW estimator, the variance dominates the estimated bias so that the optimal choice of m (m_{opt}) is always the maximum m available. This happens regardless whether the variance was estimated using the Hessian matrix or using the asymptotic expression $Var(\hat{d}_{LW}) \approx \frac{1}{4m}$. Vice versa, for the GPH estimator, the bias dominates the variance by several orders of magnitude so that the m_{opt} is the value that minimizes the asymptotic bias.

Figure 4.10 show the estimated parameters across aggregation levels for the tuned and automatic method. The first column of the plots refers to the estimates of the long range dependence parameter d, while the the second to the estimates of the logarithm of the scale parameter $\log c$.

The first striking feature is that the estimates of d do not vary much across aggregation levels for both estimators and both estimation methods for a given number m of frequencies used in the estimation. The only exception is the tuned methods for the LW estimator at 1s aggregation presumably because the number of frequencies available is lower, and, therefore, forces a lower choice of m.

The second striking feature is that the logarithm of the scale parameter c increases linearly with aggregation. According to formula (2.18) $c_k = k^{1+2d}c$, where k is the number of observation

in each bin. In our case, $k = 10^j$, $j = 0, 1, 2, 3$. It follows that

$$\log c_k = \log c + j(1 + 2d) \log 10 \tag{4.2}$$

The estimated coefficient of the regression of $\log c$ on j is displayed on the bottom of the plots, along with the theoretical value computed using Formula (4.2) where for d we used the estimated $\hat{d}$ at 1ms level. In all cases, the estimate of the coefficient is very close to the theoretical values.

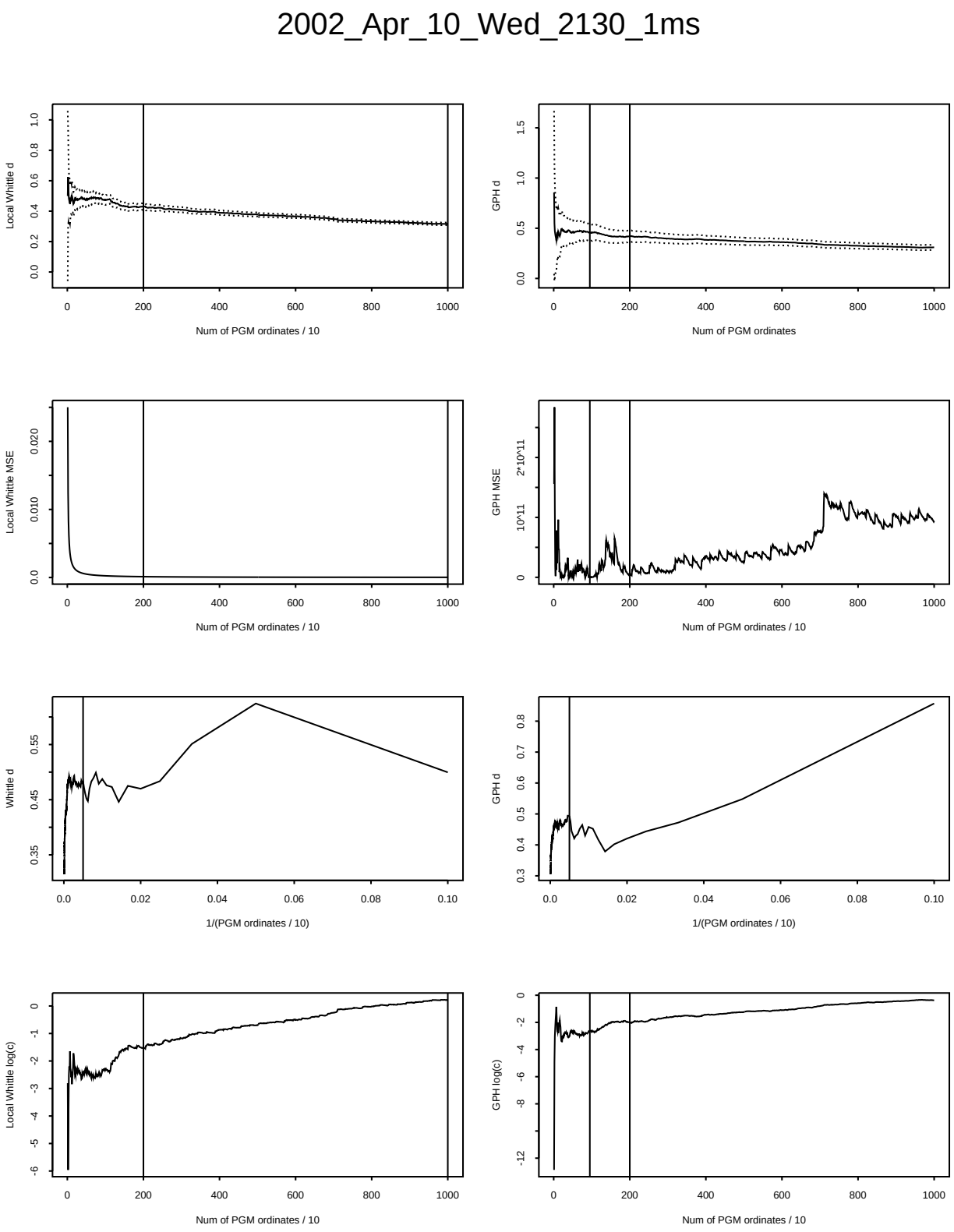

Figure 4.1: Apr 10 21:30, 1ms. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

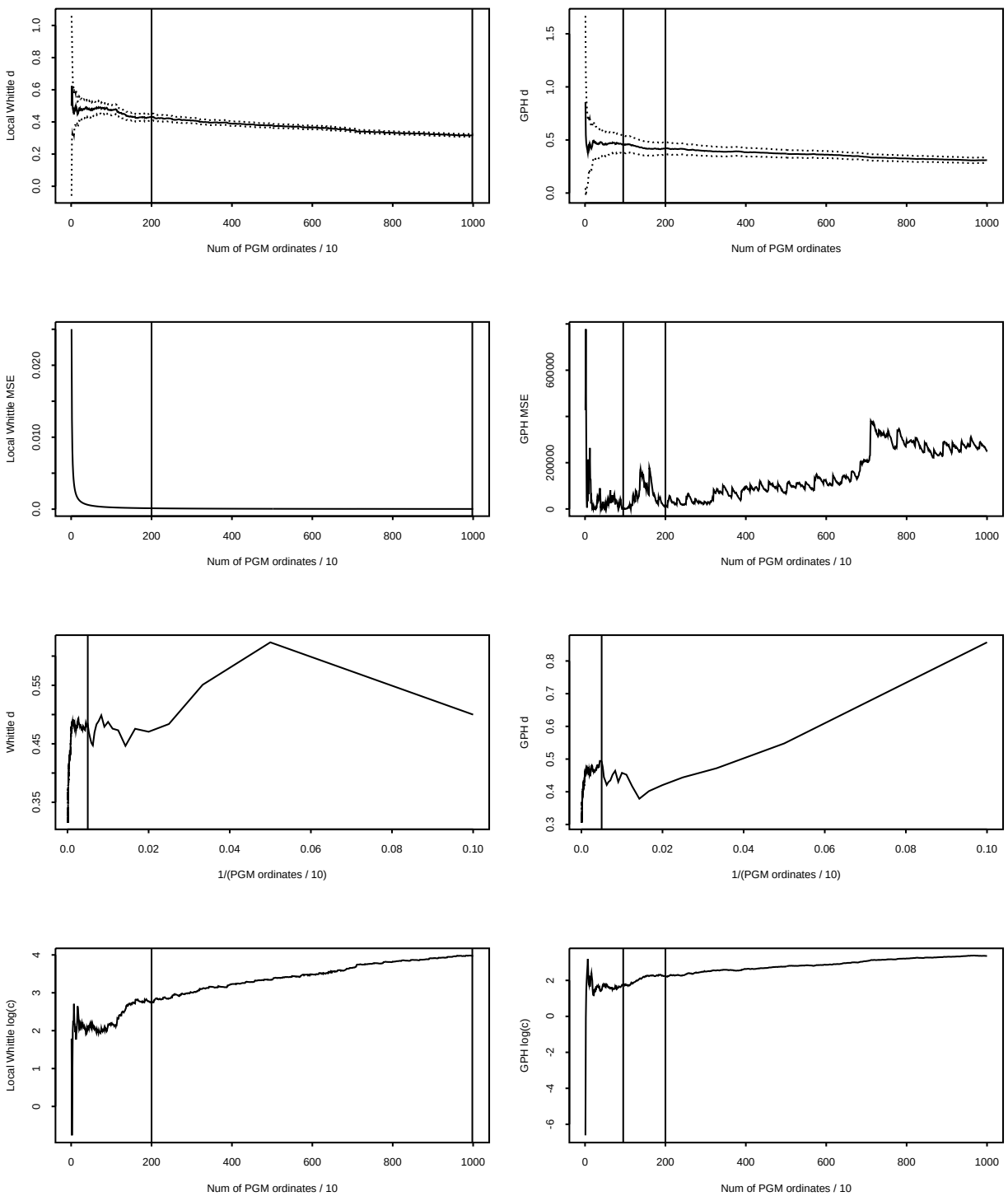

Figure 4.2: Apr 10 21:30, 1/100-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

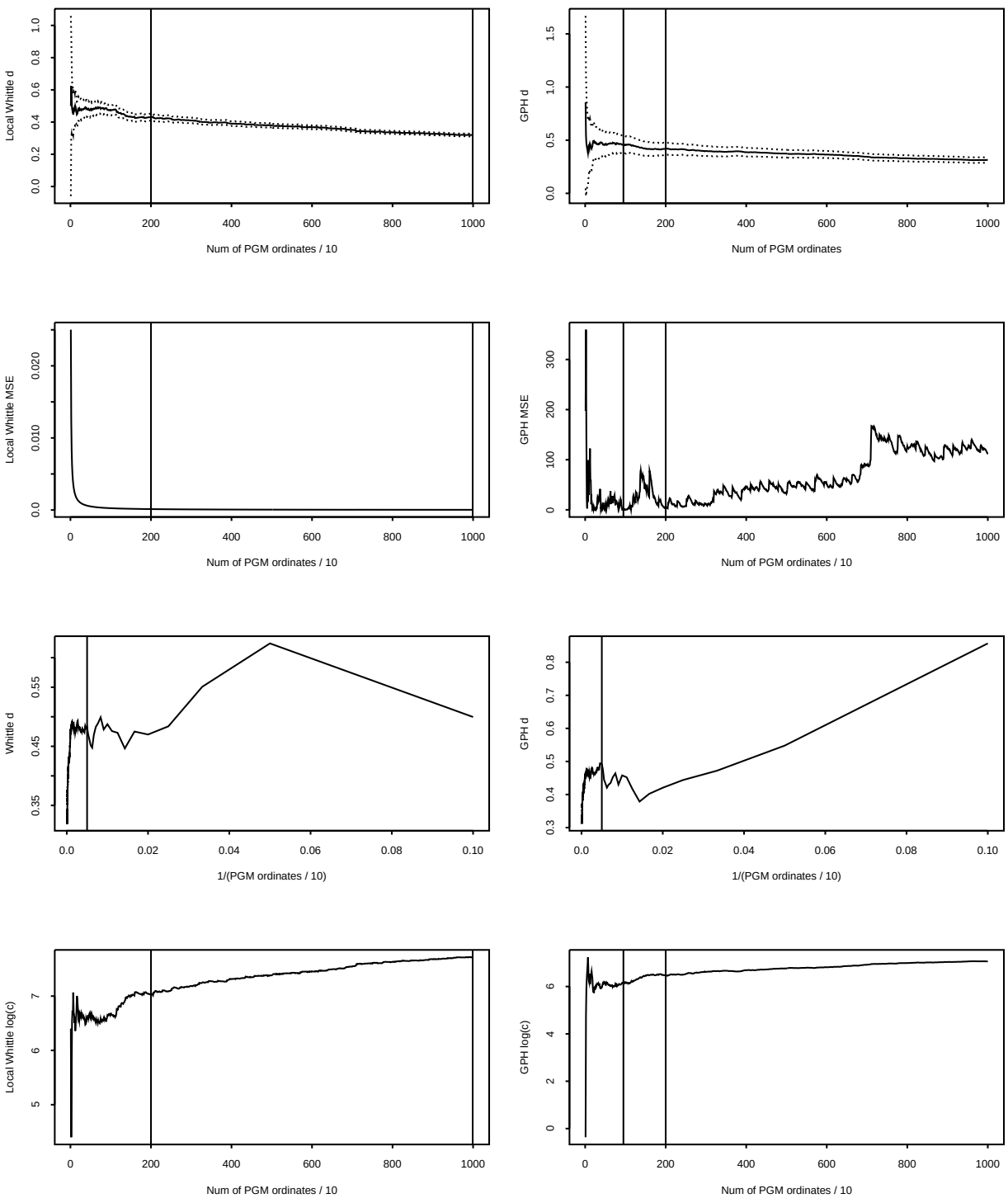

Figure 4.3: Apr 10 21:30, 1/10-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

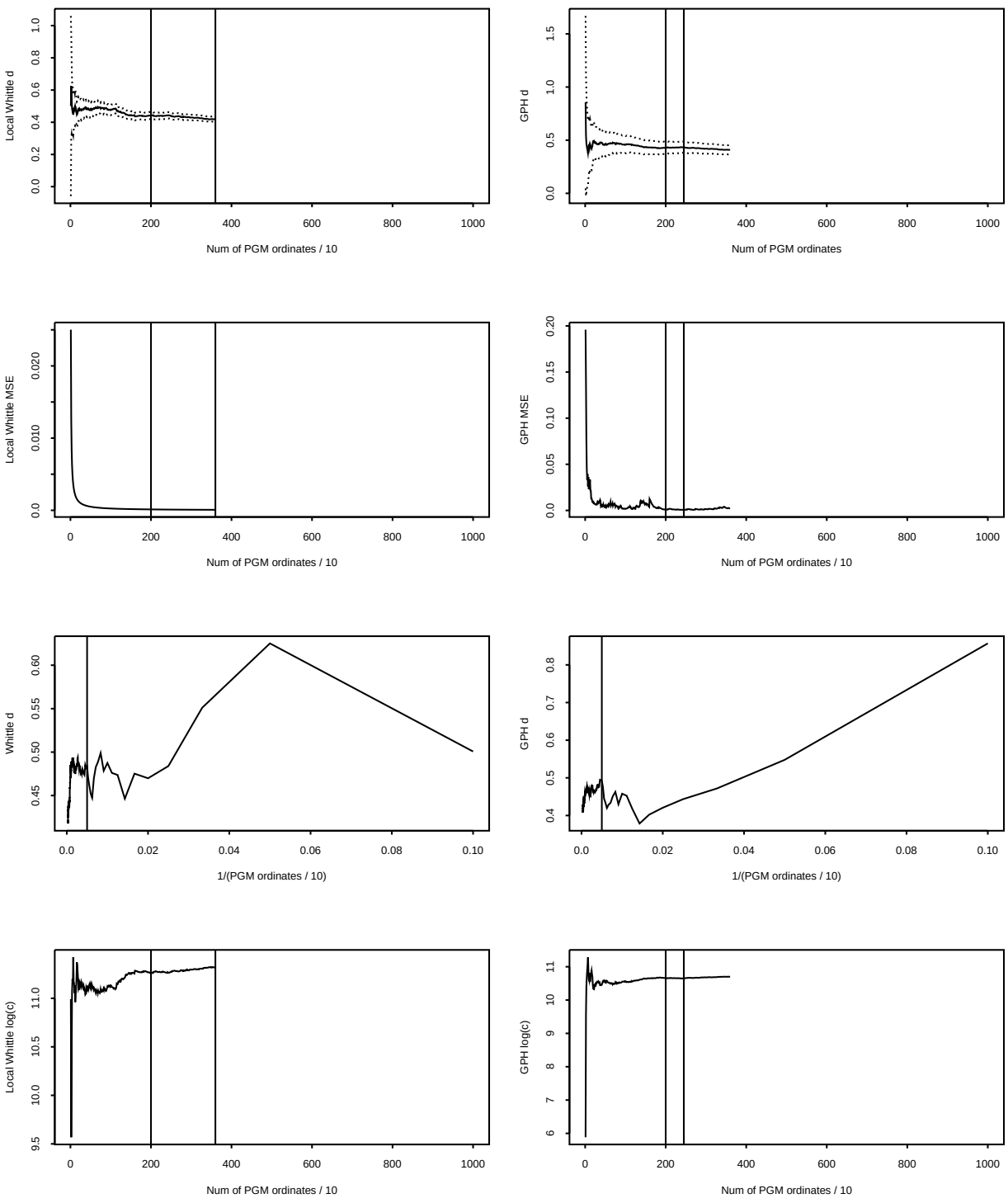

Figure 4.4: Apr 10 21:30, 1 Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

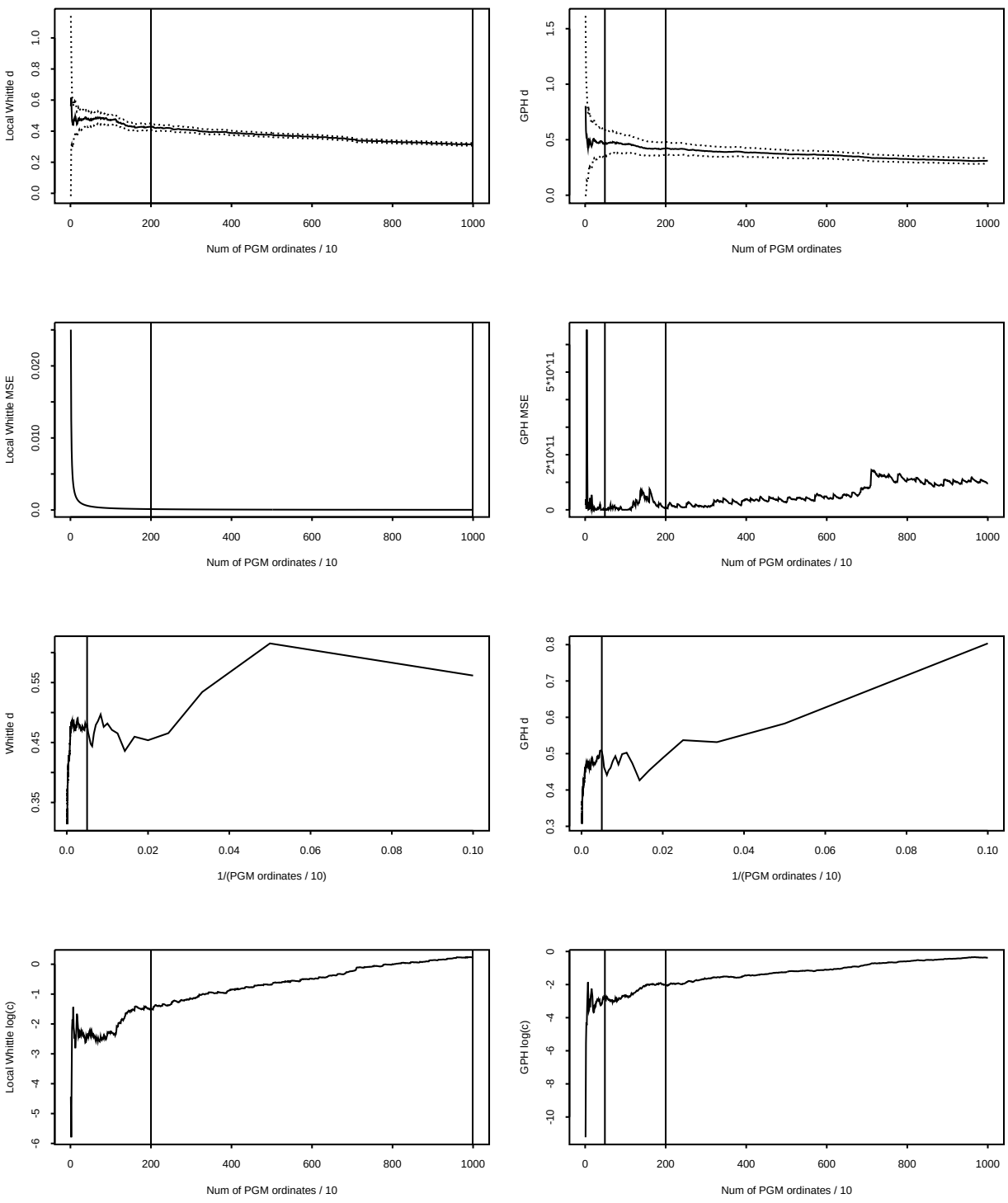

Figure 4.5: Apr 10 21:30, 1ms Detrended. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

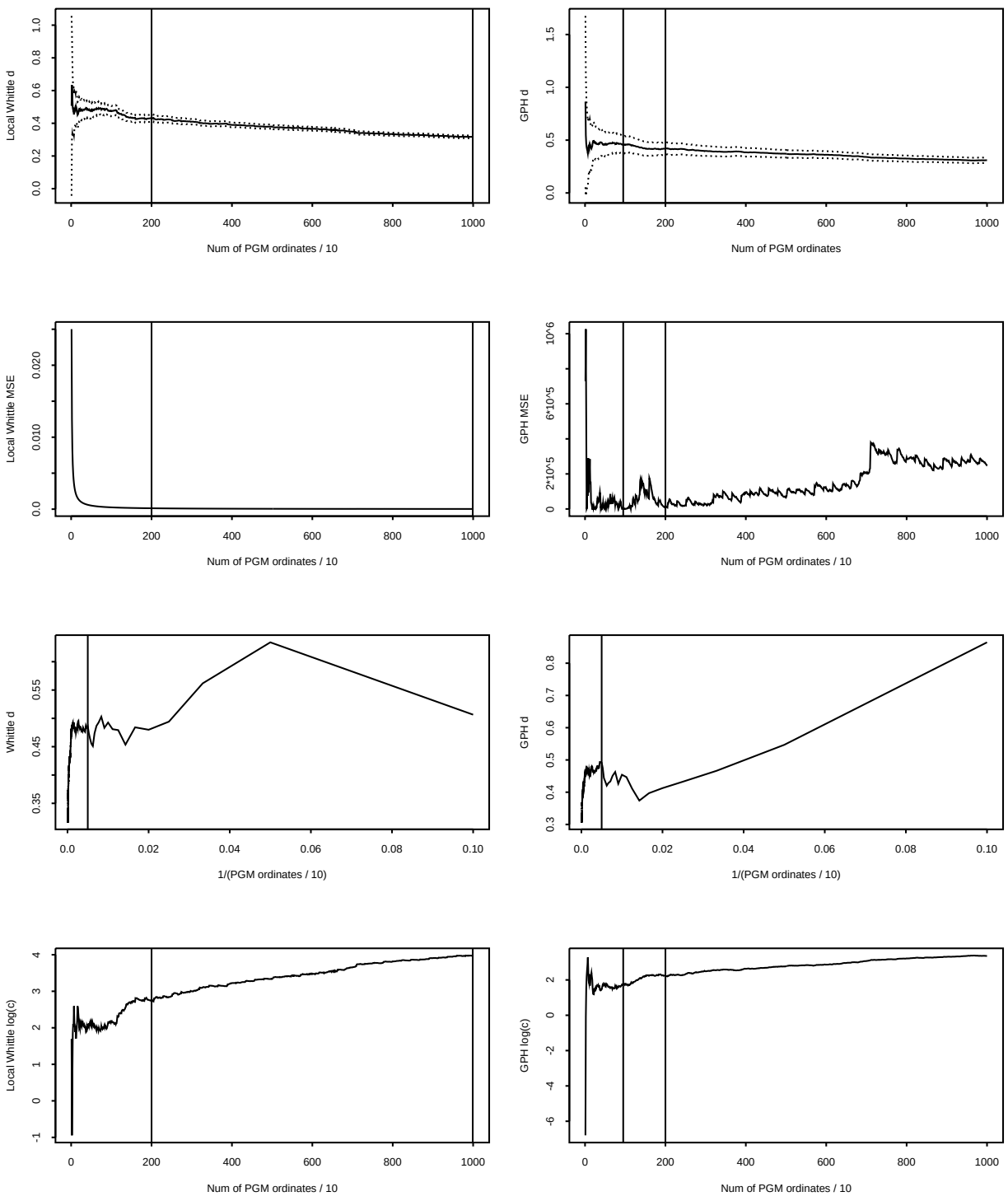

Figure 4.6: Apr 10 21:30, 1/100-th Sec Detrended. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

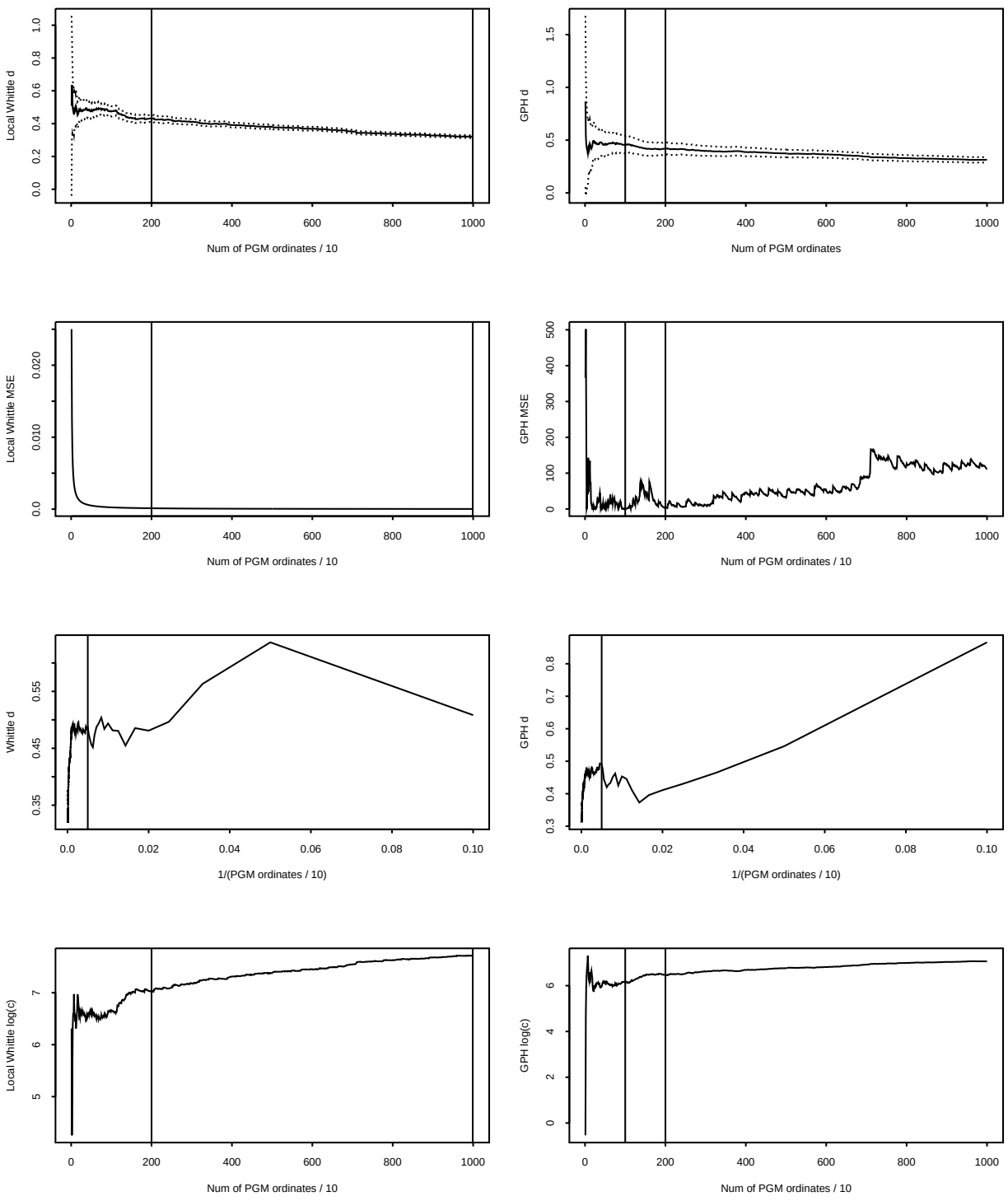

Figure 4.7: Apr 10 21:30, 1/10-th Sec Detrended. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

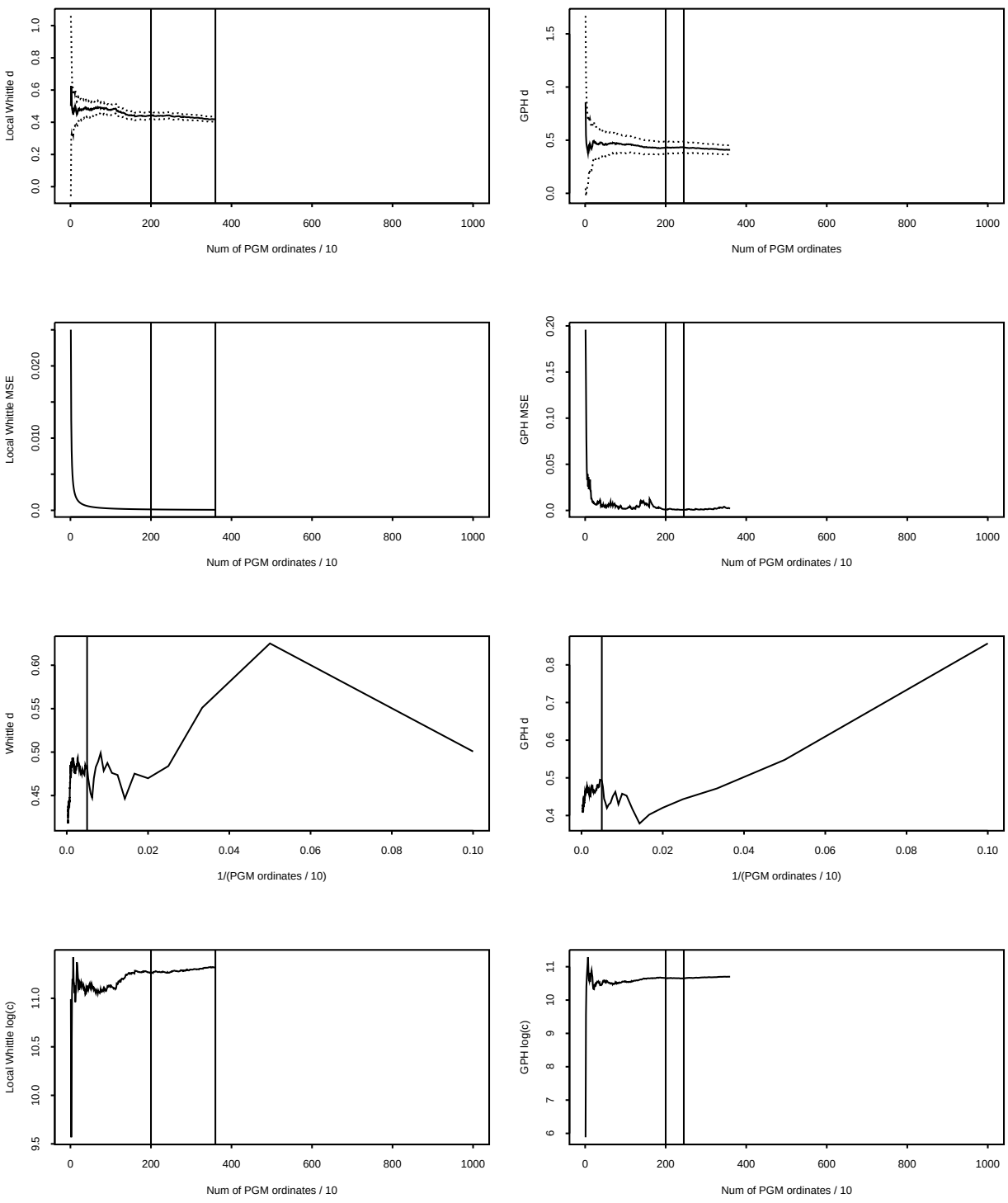

Figure 4.8: Apr 10 21:30, 1 Sec Detrended. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

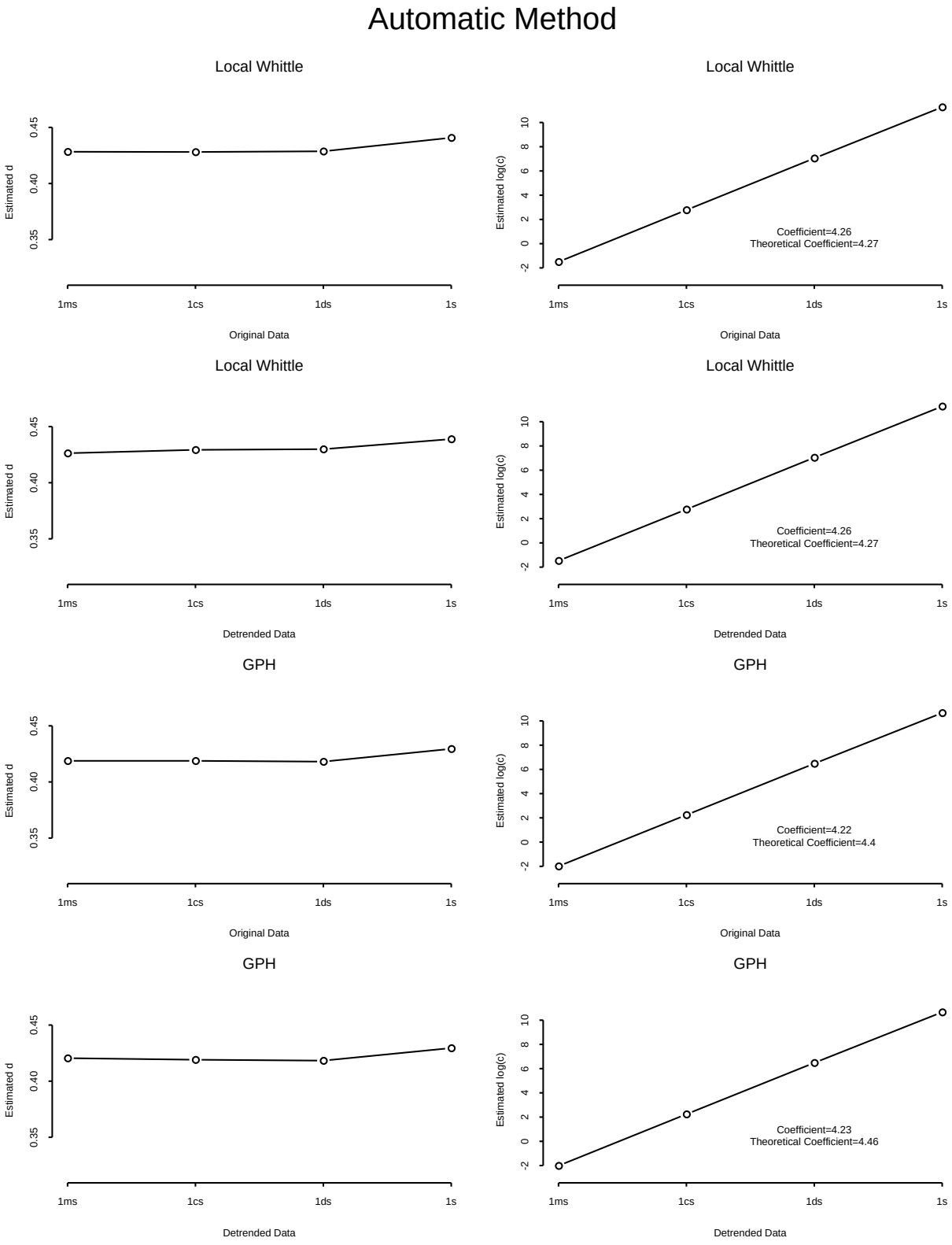

Figure 4.9: Estimates vs Aggregation, Automatic Method, Apr 10 21:30. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

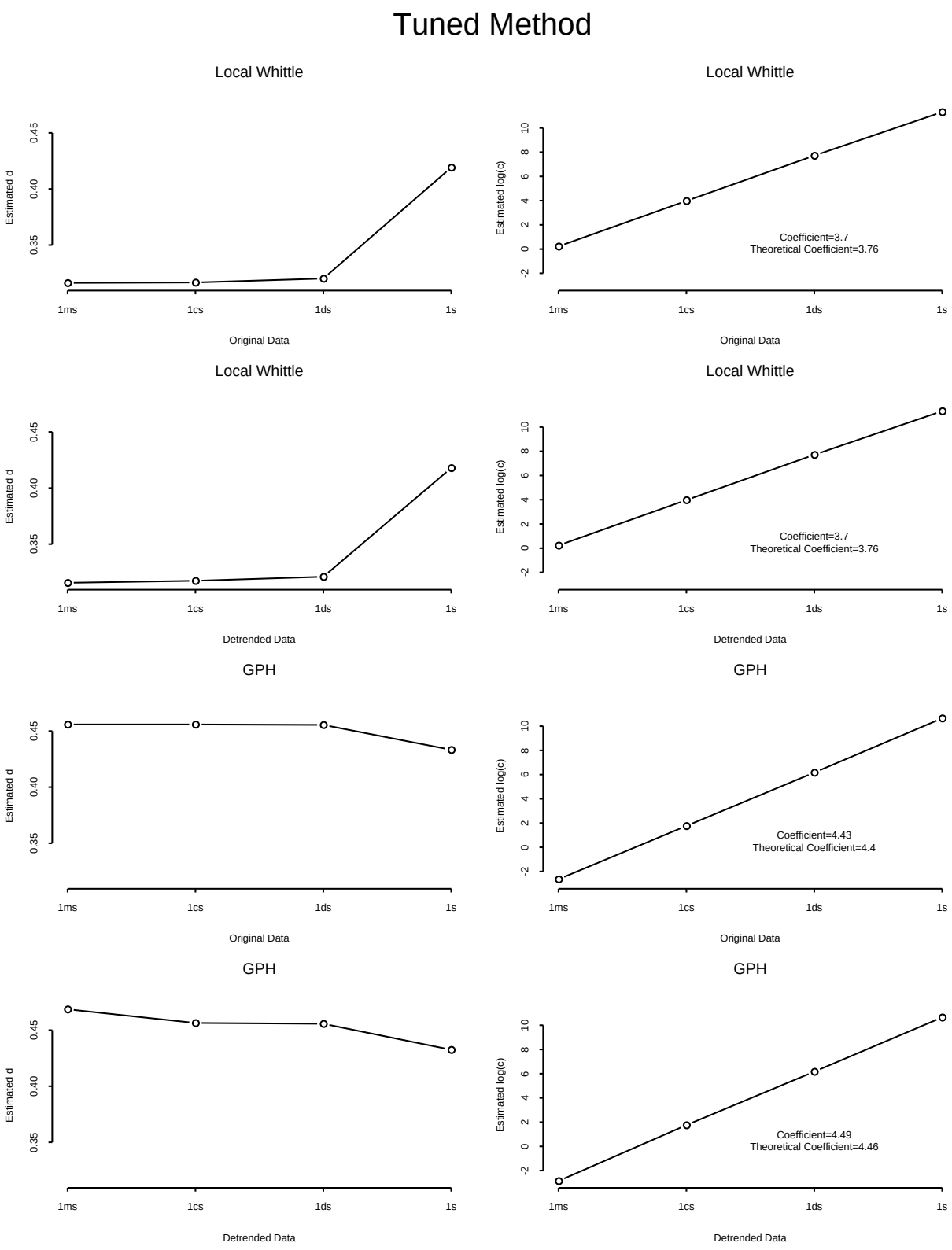

Figure 4.10: Estimates vs Aggregation, Tuned Method, Apr 10 21:30. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

4.2.2 Simulated Data and Lab Data

The results that we have observed with the UNC data repeat without significant difference for the simulated and the lab data. The only difference worth noticing is that the estimates, as function of m, are more stable when compared to the real data due to the fact the the data are "better behaved". This fact allowed us to choose $m = 50$ for the automatic method with the FARIMA data so that we could use the same value across all aggregation levels.

It is worth noticing that, in the case of the simulated data, both estimators came very close to the actual value of the parameter d. Unfortunately, the time needed to produce one simulated series did not allow us to perform a study of the variability of the estimates. This can be a possible further development.

For brevity, here we present only the plots for one of the FGN simulations and one of FARIMA series. Detrending has virtually no effect, except for a mild effect for `fgn_pipiras`. For this reason, detailed plots of the estimates for the detrended data are omitted.

We maintained the rule $k = 10^j$ and the notation $1ms, 1cs, 1ds, 1s$ to indicate $j = 0, 1, 2, 3$ for consistency with the UNC data. For the data set `fgn_zhu` that implied that we could only estimate the parameters for the first three aggregation levels since it contains only 524288 observations.

4.2.3 Conclusions

The analysis showed consistent results for both estimators across all data sets, whether the data was real or simulated.

Unfortunately, the estimation of the MSE of $\hat{d}$ did not unveil much about the variation of m_{opt} across aggregation levels. However, we uncovered interesting facts about the estimators of c and d. We can summarize them as follows.

- For a given m, the estimates of long-range dependence parameter d do not vary significantly across aggregation levels. This finding is consistent with our theoretical results about the GPH estimator in Chapter 2. One can conjecture that a similar result holds for the other semiparametric periodogram-based estimator. It can be intuitively explained by considering that aggregation is a form of "compression" of data that "looses" the high-frequency characteristics but preserves the long-term features. This intuition is reflected in the fact that the only the right tail of the periodogram is lost to aggregation, while the left tail, that is the relevant fraction to both estimators, remains essentially unchanged in its most important feature as far as long-range dependence is concerned, that is the slope of the its logarithm as a function of the log frequencies.

- For any fixed value of m, the estimates of the scale parameter c follow almost exactly the linear relation (4.2). On one side this may leads to think that, as with $\hat{d}$, one can retrieve an equally valid estimate of the scale parameter c of the original process from the estimate of c_k, the scale parameter of the aggregated process. On the other hand, one may conjecture that, since c, in most cases, is dependent on the variance of the original process, better estimates may be obtained when the high-frequency characteristics of the data are not lost to aggregation. This intuition suggests that the optimal approach to estimating the scale parameter may differ from the best approach to estimate d. If an accurate estimation of the variance of the process is as relevant as the the estimation of the long-range dependence effect, this may lead to two different estimation strategies.

The above facts and the findings in Chapter 2 suggest some possible developments of our research. First we can attempt and compute the asymptotic MSE of the parameter c in the same fashion that in Chapter 2 is done for the GPH estimator of d, and study its variation with

aggregation. Second, similar results can be developed for the local Whittle estimator. Third, simulation studies can be performed to measure the MSE of c in finite samples for some selected models. For this purpose, however, faster and less memory-demanding algorithms will have to be employed.

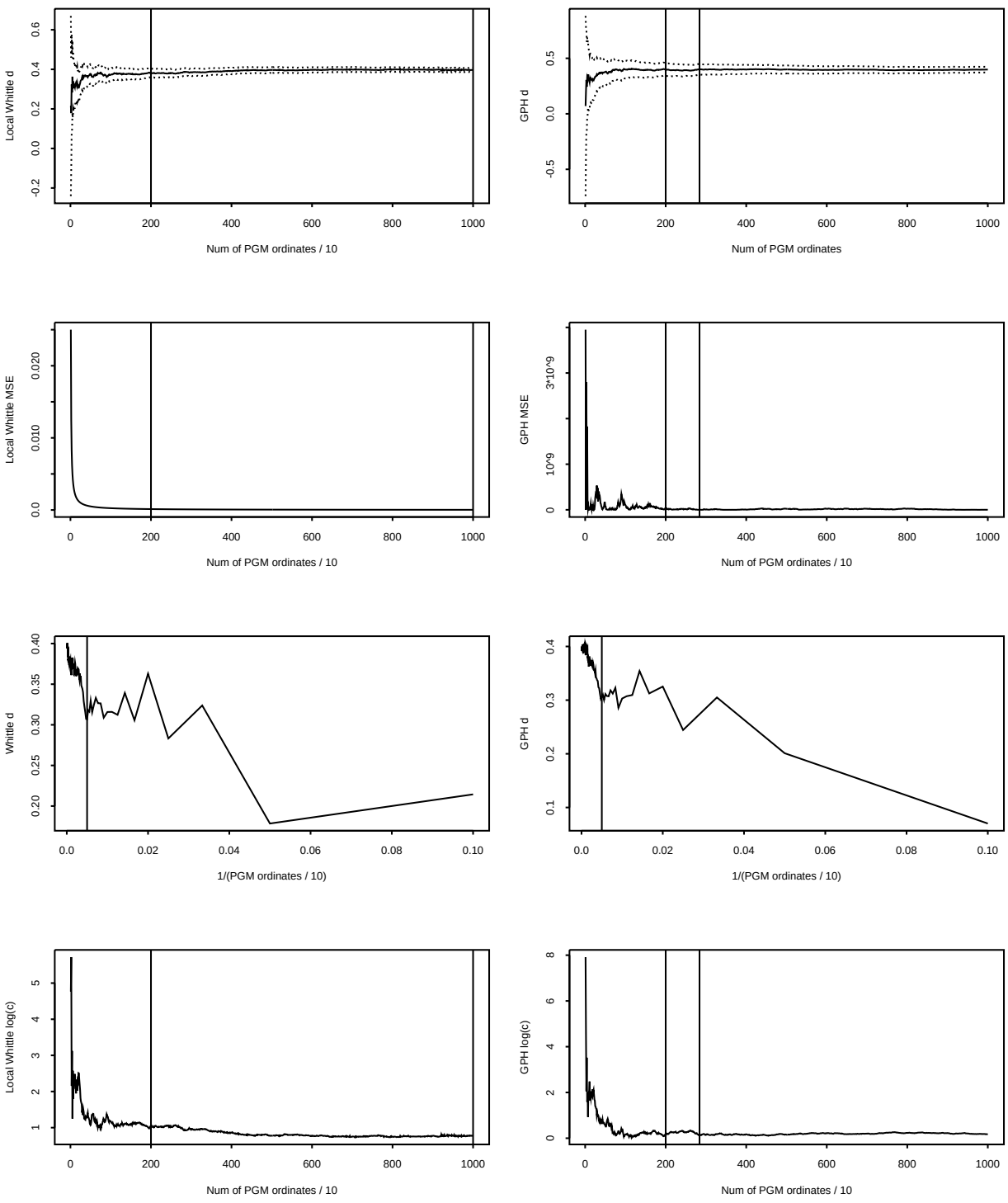

Figure 4.11: FGN Pipiras, 1ms. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

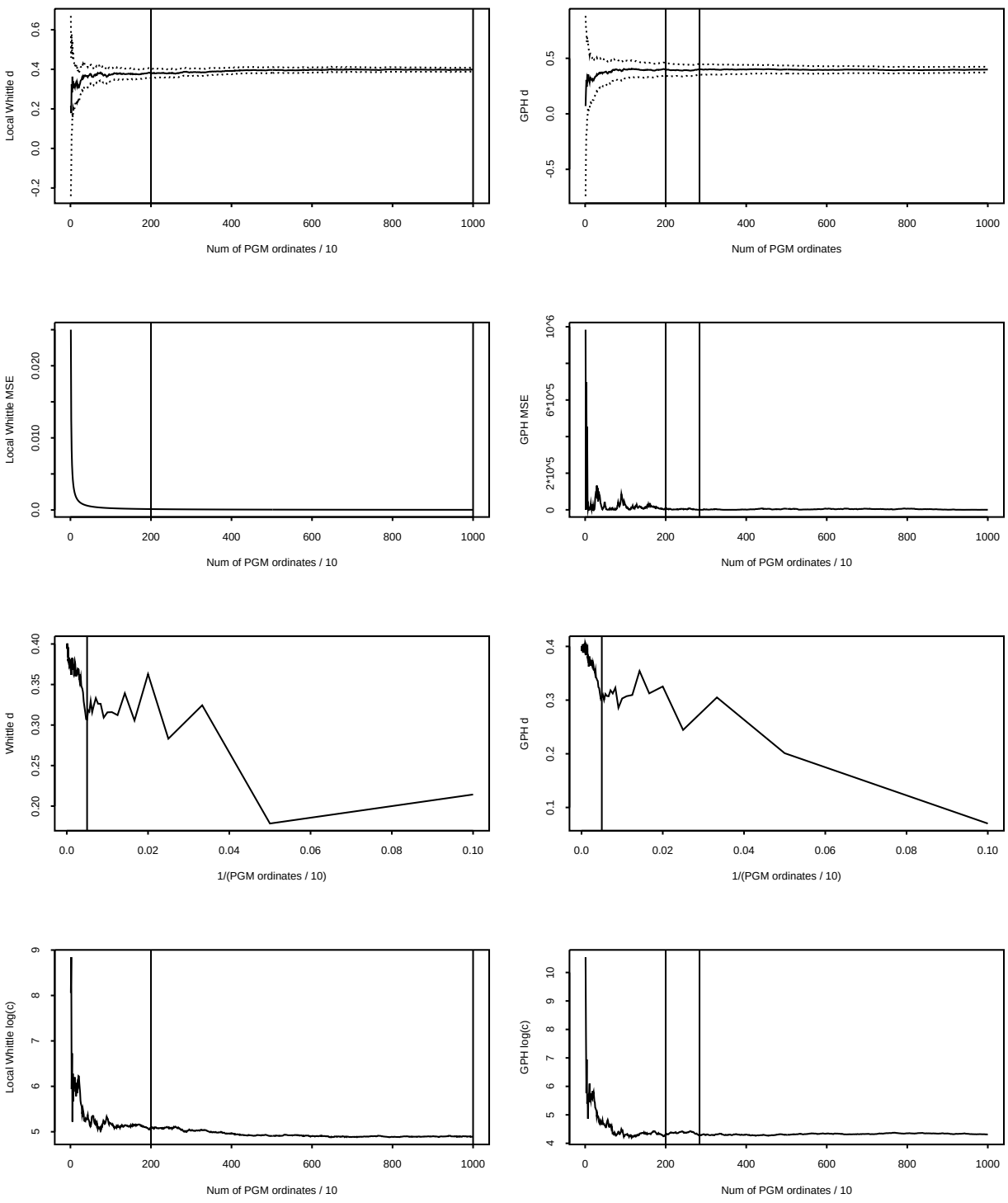

Figure 4.12: FGN Pipiras, 1/100-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

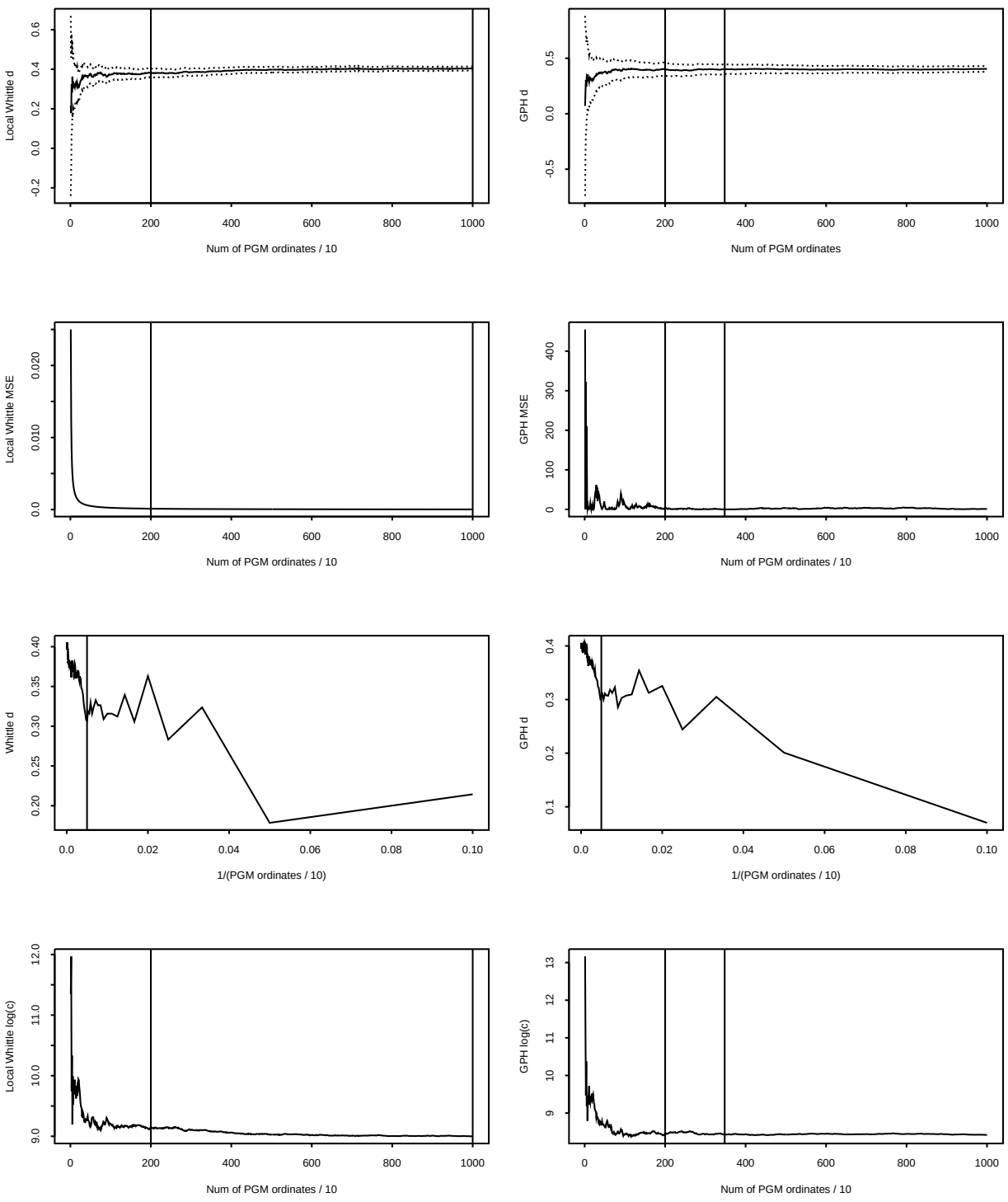

Figure 4.13: FGN Pipiras, 1/10-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

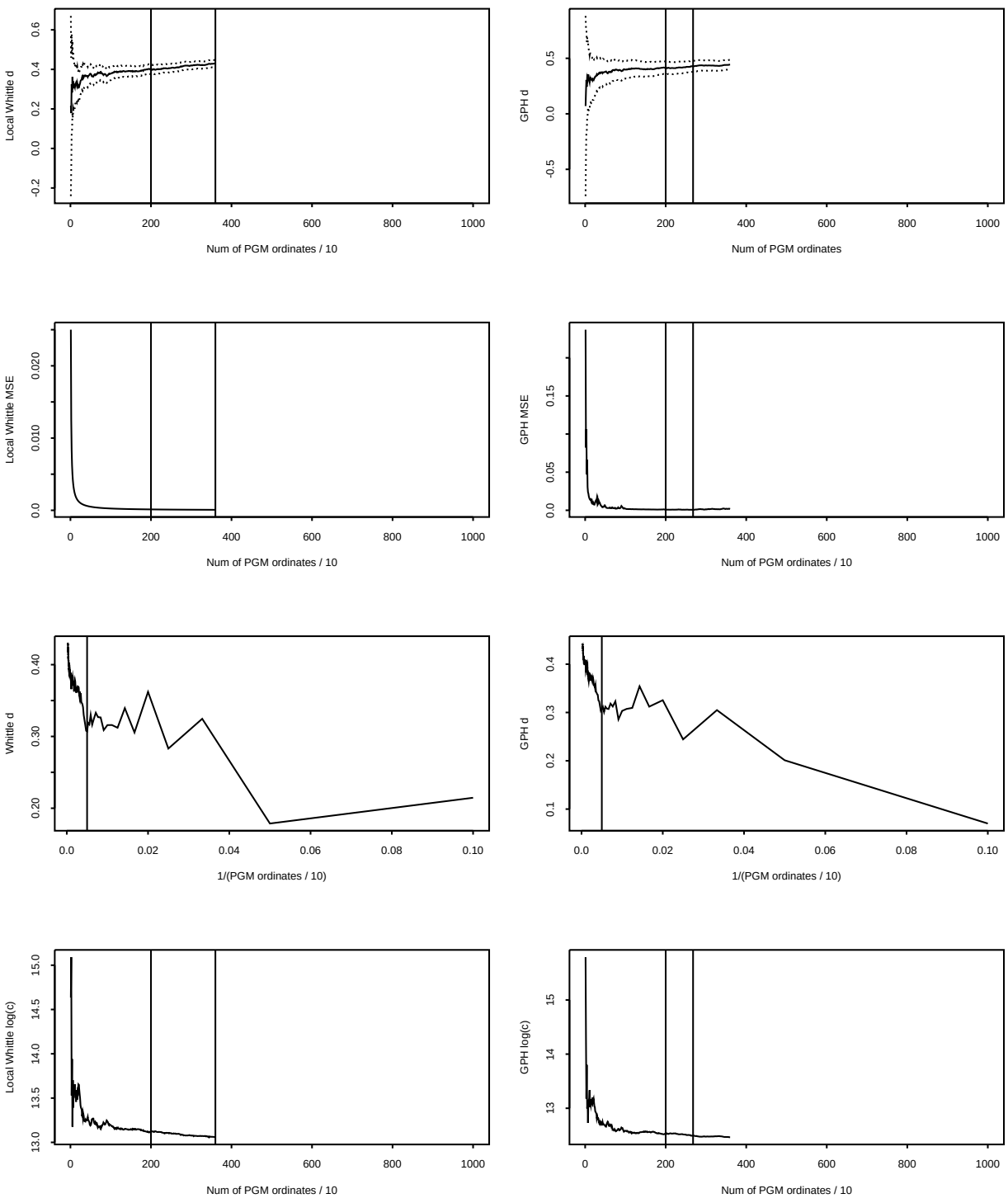

Figure 4.14: FGN Pipiras, 1 Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

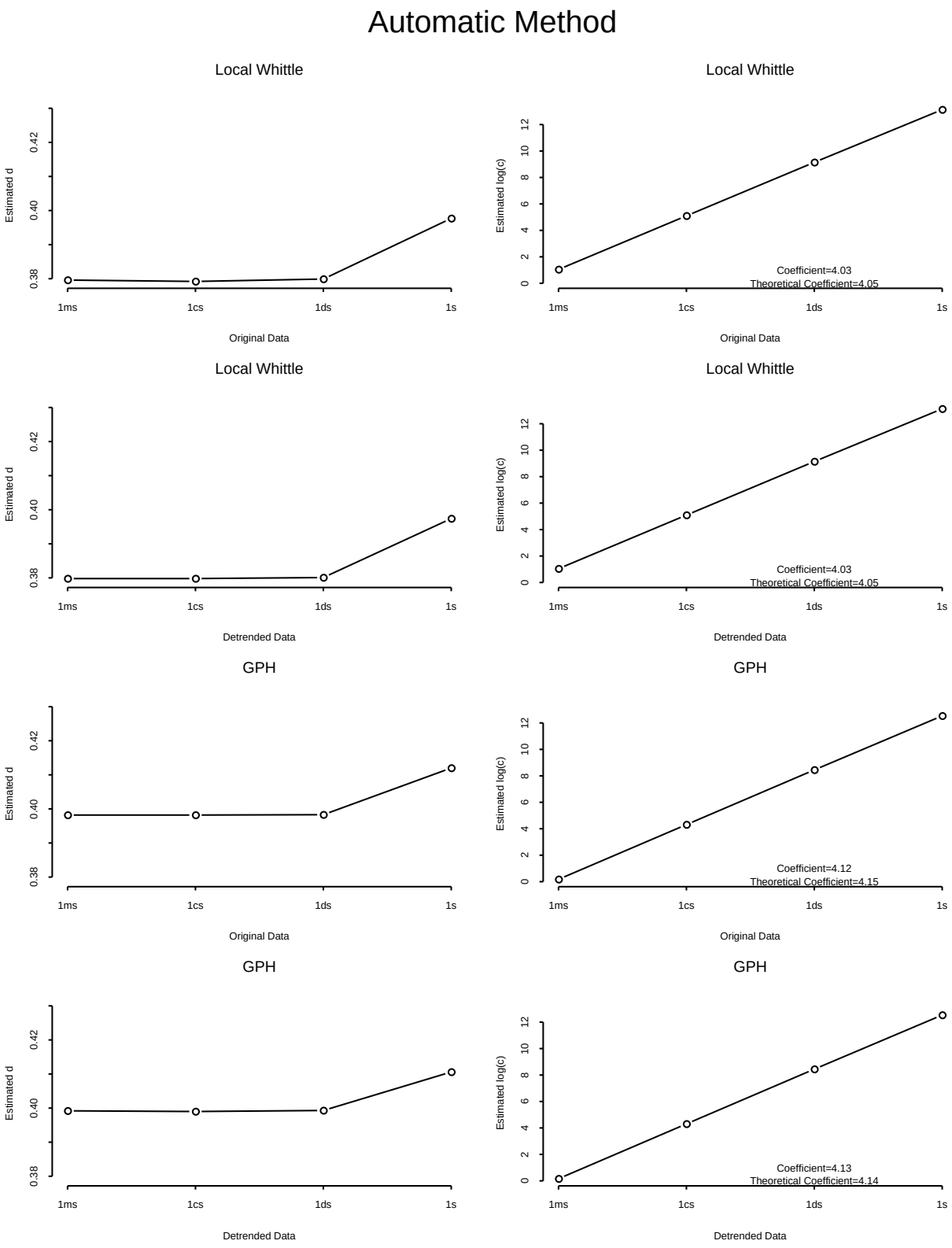

Figure 4.15: Estimates vs Aggregation, Automatic Method, FGN Pipiras. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

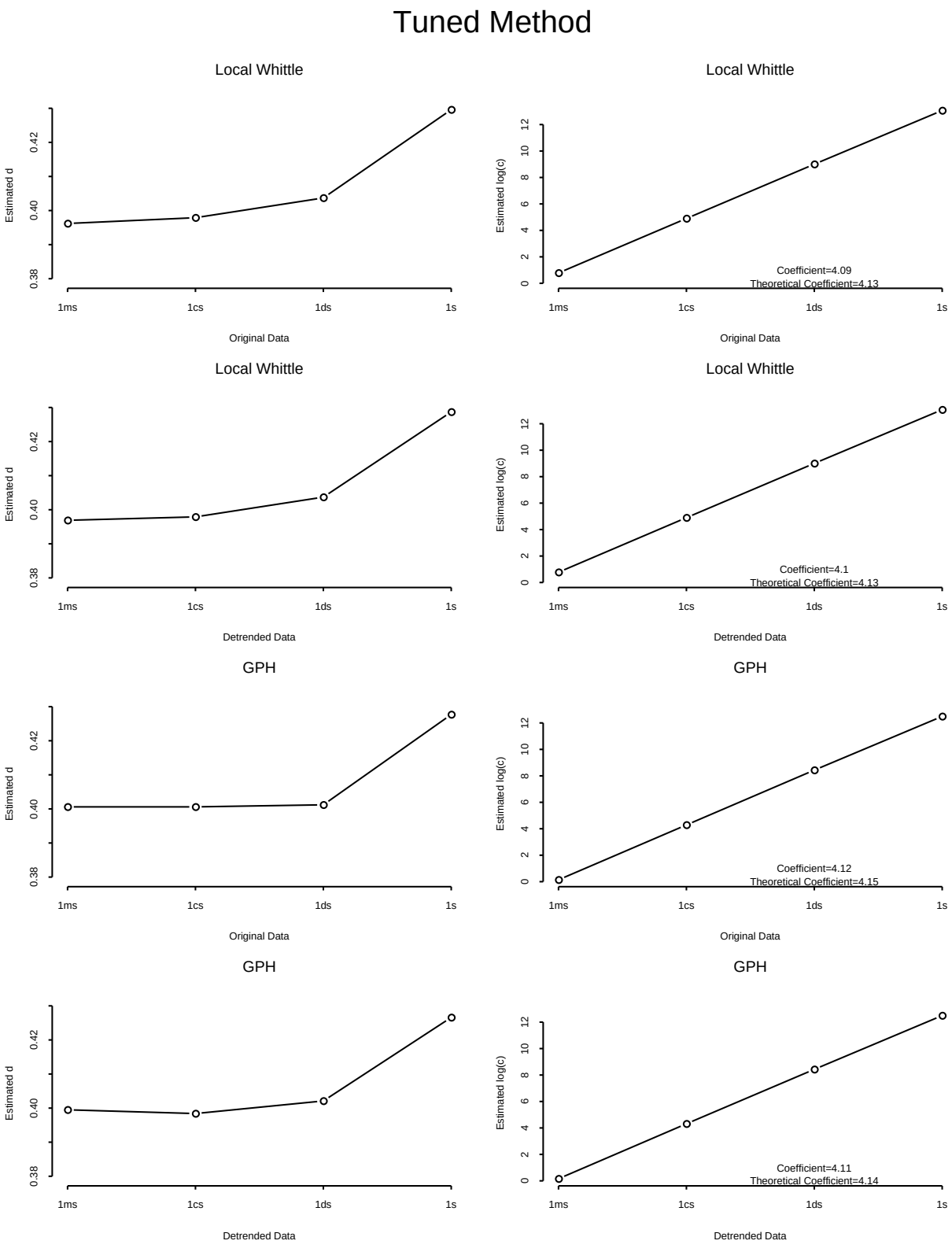

Figure 4.16: Estimates vs Aggregation, Tuned Method, FGN Pipiras. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

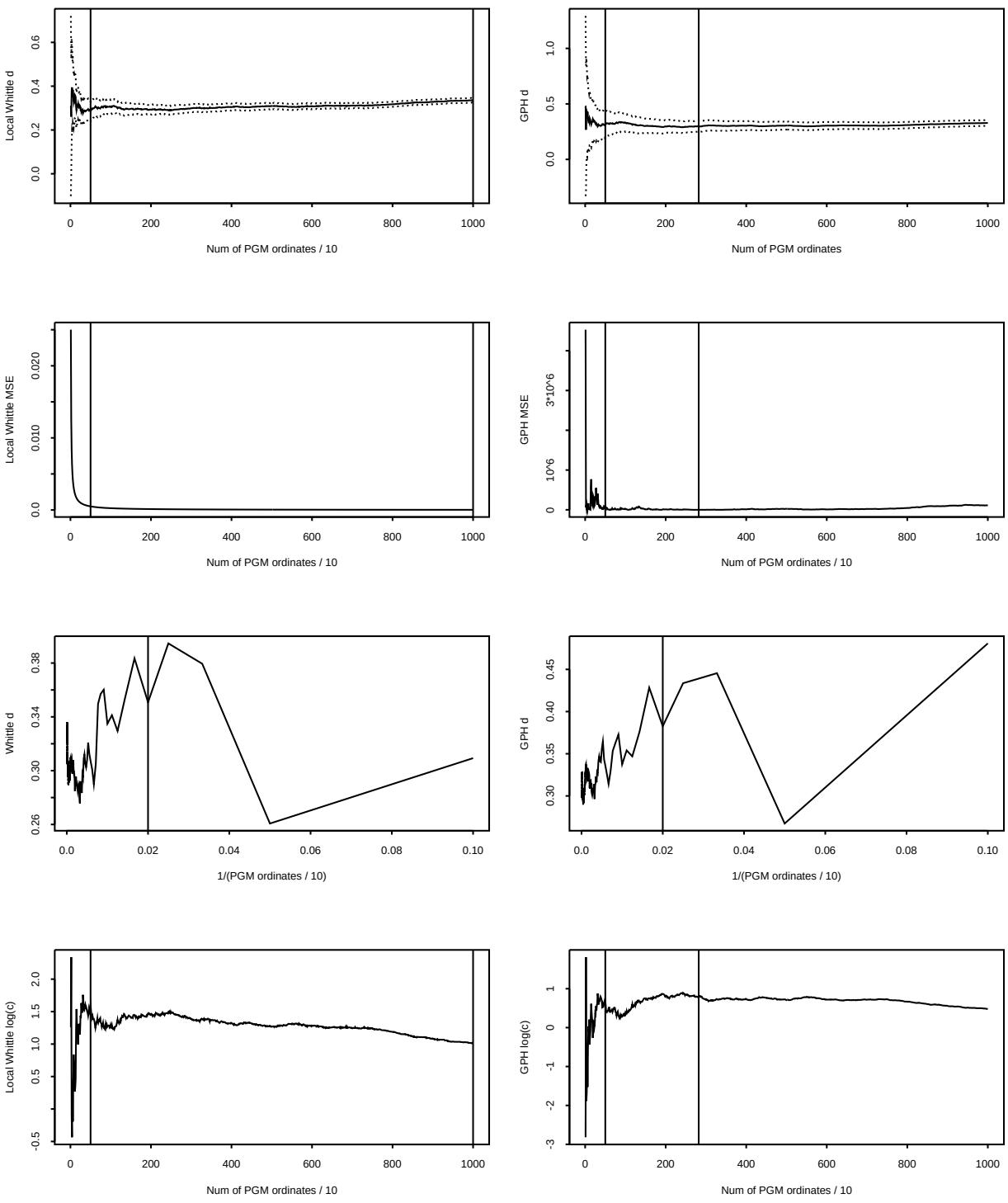

Figure 4.17: FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5, 1ms. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

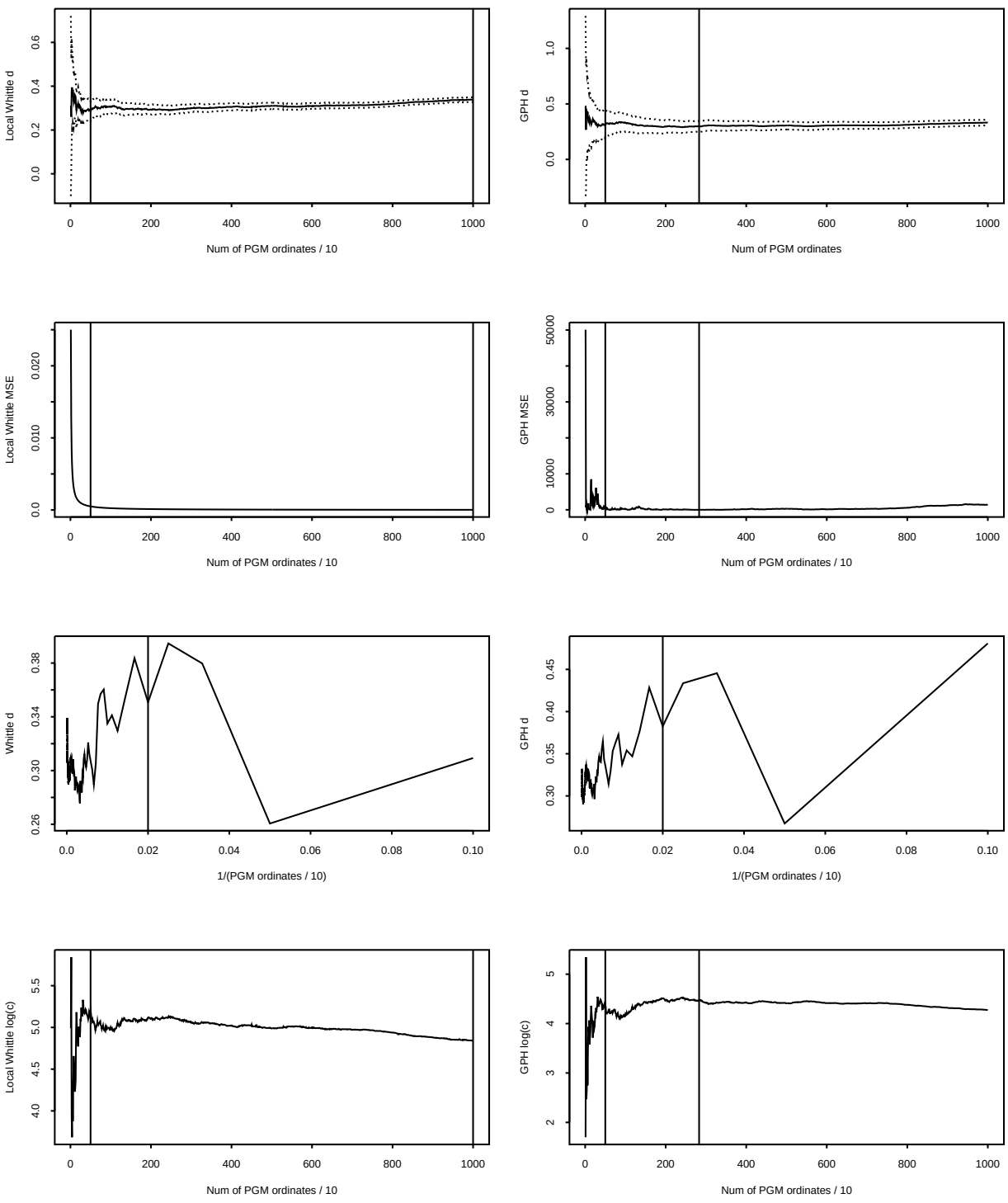

Figure 4.18: FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5, 1/100-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

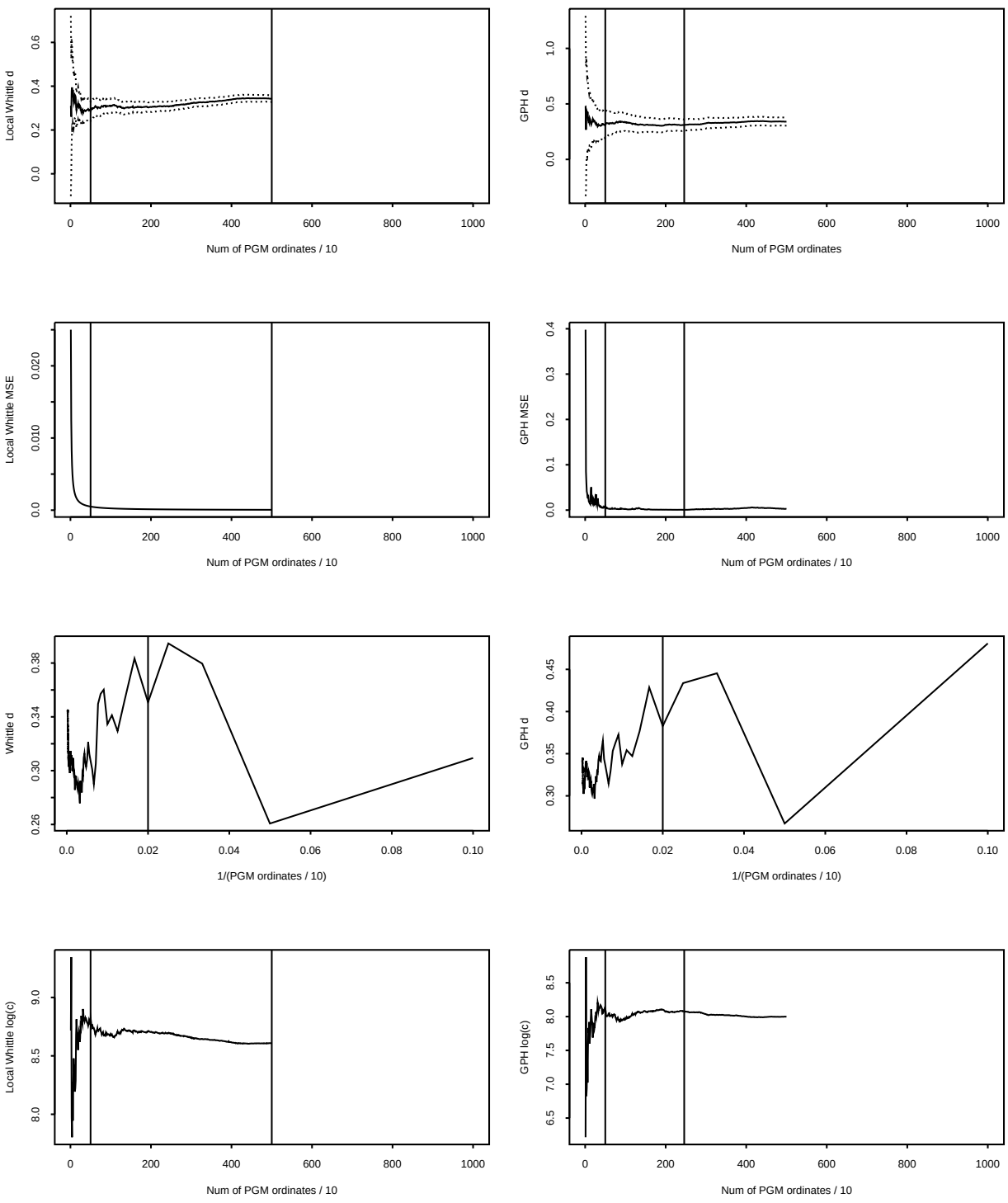

Figure 4.19: FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5, 1/10-th Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

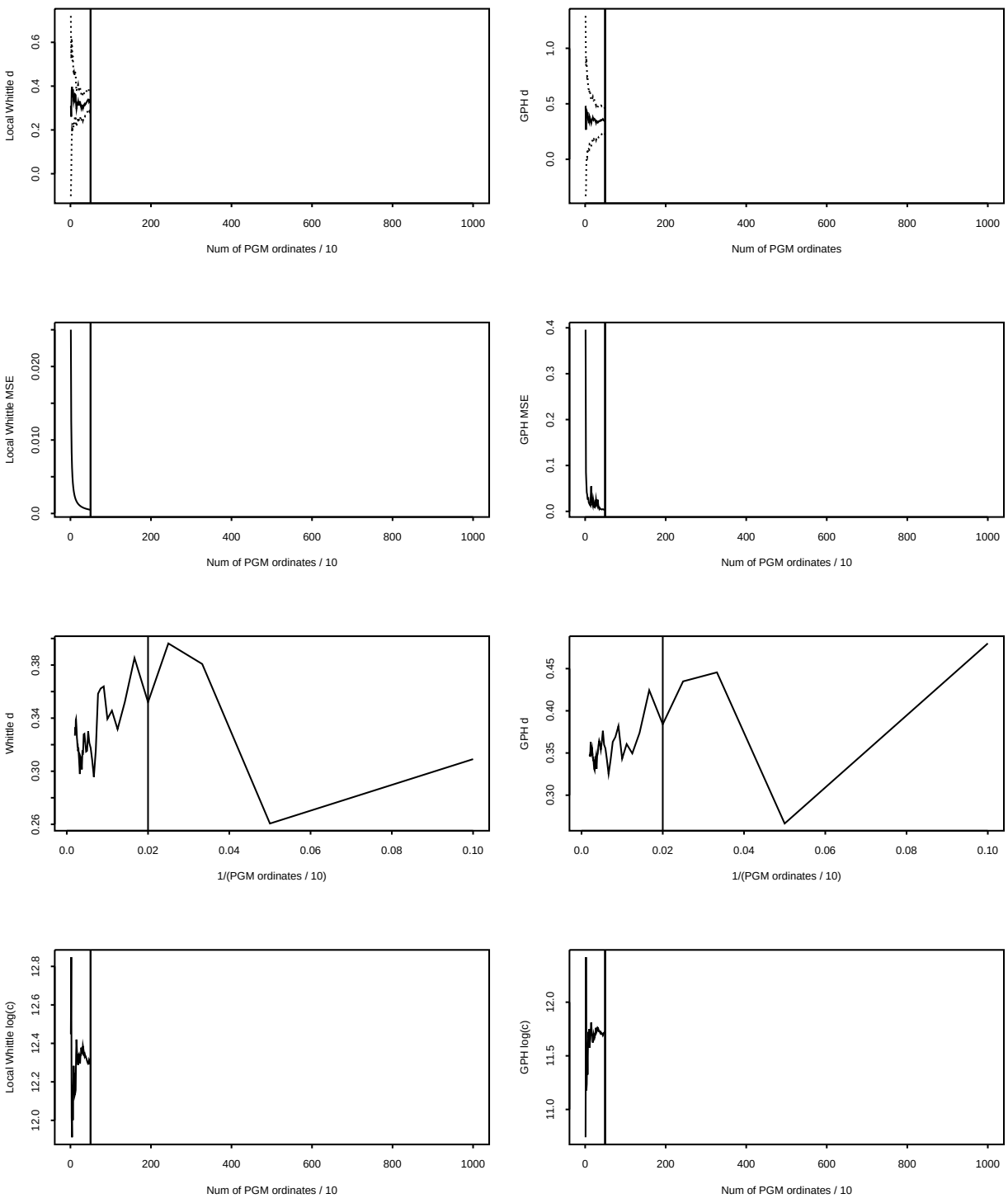

Figure 4.20: FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5, 1 Sec. The left column illustrates the LW estimates, while the right column the GPH estimates. In the first row the estimates are plotted against the $m/10$. The second row shows the estimates of the MSE against $m/10$. The third row shows the estimated d against $10/m$. The last row shows the estimates of $\log c$ against $m/10$

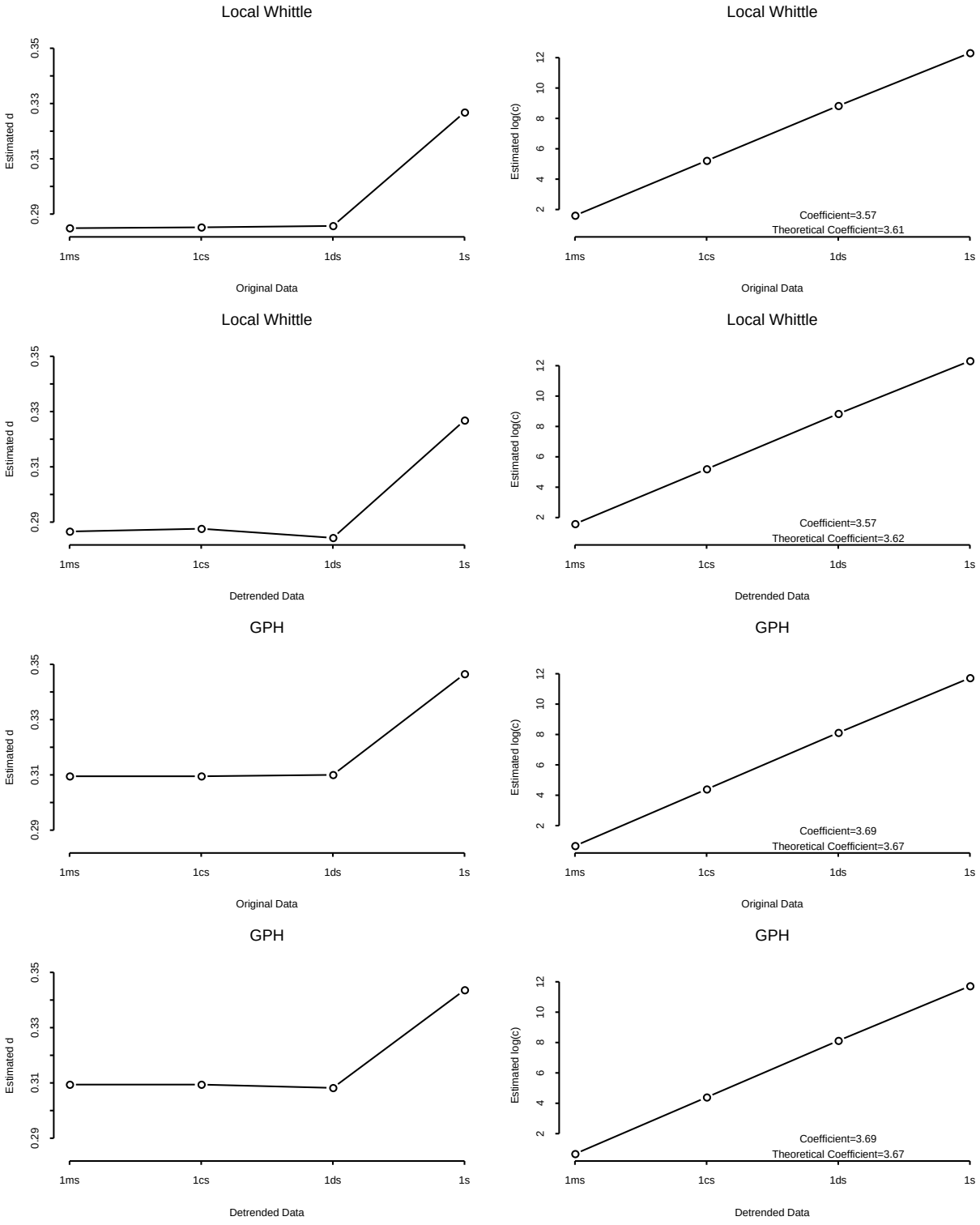

Figure 4.21: Estimates vs Aggregation, Automatic Method, FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

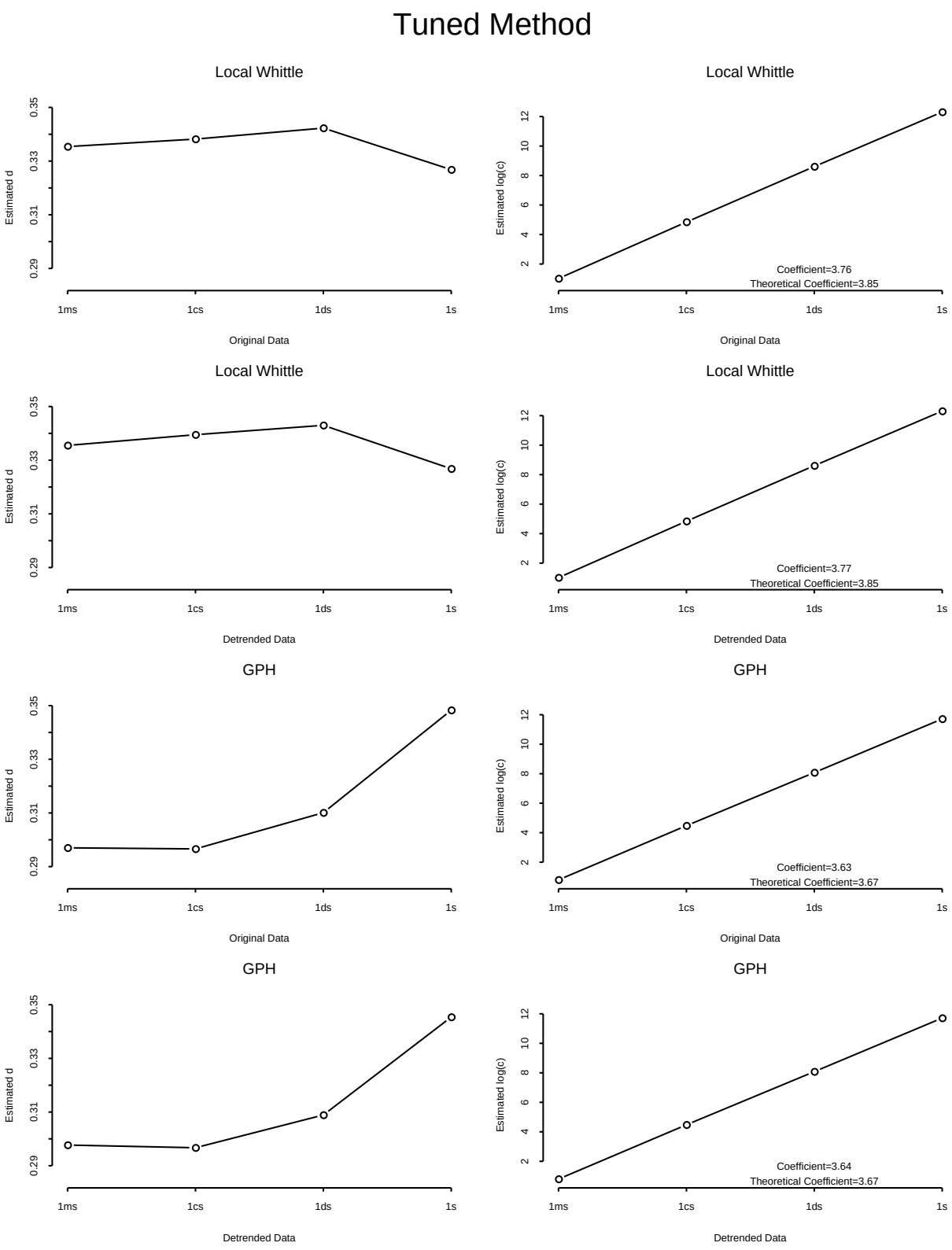

Figure 4.22: Estimates vs Aggregation, Tuned Method, FARIMA(1,.3,1), ar(1)=.9, ma(1)=.5. The left column illustrates the estimates of d versus the aggregation level. The right columns shows instead the estimates of $\log c$. The first two rows show the LW estimates, while the last two the GPH estimates. The first one of each employs the original data, while, in the second row, the data has been detrended prior to estimation

Chapter 5

Conclusions

We analyzed the effect that temporal aggregation has on long-range dependence. We first considered the effect that temporal aggregation has on LRD processes. Making only local assumptions on the spectral density in a neighborhood of the origin we showed in Proposition 2.1.1 that the fractional order of a LRD processes is invariant under aggregation. Long-range dependence is a phenomenon that characterizes the process at low frequencies. Temporal aggregation is a form of data compression that looses the high-frequency characteristics of the data, but preserves the low-frequency behavior. We confirmed this result with our empirical analysis of the UNC network data. Provided sufficient data is available at all aggregations levels, the estimates of the LRD parameter d is invariant under aggregation, regardless whether the bandwidth parameter m used in the estimation is fixed across aggregation levels (automatic method), or it is chosen to minimize the asymptotic MSE (tuned method). Furthermore, we established that the logarithm of the scale parameter varies linearly with aggregation. Again, this fact is confirmed in our empirical analysis of the UNC network data in Chapter 4.

We then considered the effect of aggregation on the finite sample properties of periodogram-based estimators of long-range dependence. In particular, we looked at the efficiency of the estimators as measured by the MSE, and how the MSE-optimal choice of the bandwidth parameter m

varies with aggregation. Based on asymptotic expression for the MSE of the estimators, Proposition 2.2.1 and Proposition 2.2.2 show that, for $b < 2$, the MSE and MSE-optimal choice of m for the GPH and LW estimators, respectively, are invariant under aggregation.

We considered also the effect of aggregation when the LRD process is not observed directly. Instead, a process that is the sum of an LRD process and a white noise process is observed. For such a case, we considered the LW estimator and the ELW estimator of Hurvich and Ray (2001). For the ELW estimator, it was not possible to derive any expression for the MSE. Therefore, the analysis was conducted numerically. We derived an expression for the asymptotic MSE of the LW estimator that can be used to evaluate numerically the effect of aggregation. The bias of the LW estimator in Equation (2.52) is composed by a component due to short-range dependence, plus a component that depends on the variance of the noise. The two components may have antagonist effects on the bias. The component due to the noise generates a positive bias for the LRD stationary range $0 < d < 0.5$. The SRD component can generate either a positive or negative bias depending on the sign of the parameter γ in Equation (2.52). For a FARIMA(1,d,0) process, γ can be approximated by $-\theta/(1-\theta)^2 + d/12$ (Delgado and Robinson, 1996), where θ is the autoregressive parameter. This implies that, for positive θ, γ is negative except for small values of θ and relatively large values of d. The presence of noise, therefore, can effectively reduce the MSE of the LW estimator when there is positive SRD correlation. This is the case, for example, of Figure 5.1, where, for $d = 0.1$, the MSE of the LW estimator decreases for values of ns greater than 0.5.

The numerical analysis in Chapter 3 shows that the asymptotic expressions (2.33) of the MSE of the LW estimator derived by Henry and Robinson (1996) grossly overestimate the bias for large m. In particular, the LW bias does not seem to follow a power 4 relationship with respect to m, as the theoretical results would predict. The LW estimator performs better than expected at large

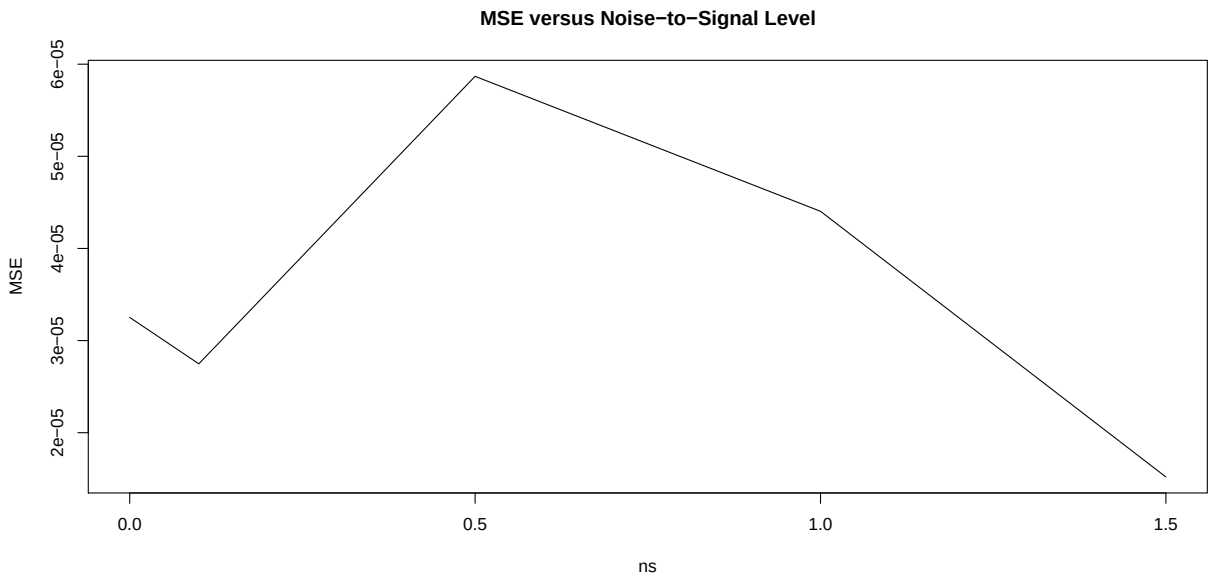

Figure 5.1: Optimal MSE of the LW estimator vs. noise-to-signal, $d = 0.1$, $N = 2^{19}$

frequencies, and it seems to be affected only moderately by the bias induced by SRD components. The minimal value of the MSE for FARIMA(0,d,0) processes is attained almost always at the last available frequency, which indicates that the variance component of the MSE dominates the bias component. The asymptotic expression of the MSE overestimates the bias also when a white noise and an AR(1) component are added, although the minimal MSE is not attained any longer at the maximum m, a sign that the bias/variance trade-off becomes relevant. The same can be said regarding the MSE expression (A.16) which accounts for the presence of the noise component. For the practitioner, this seems to suggest to choose a value of m larger than the one indicated by minimizing the estimated MSE according to the procedure suggested by Henry and Robinson (1996). It has to be noted that while equation (2.33) does not predict accurately the MSE of the LW estimator when the true values of the parameters are used, when it is used for finding the minimal MSE on real data with estimated parameters, as it is the case of the "tuned method" employed for the UNC network data, it often leads to select the highest frequency available as the

value that minimizes the MSE. This behavior is consistent with the behavior observed by the MSE in the numerical analysis.

The LW estimator performs at least as well as the ELW estimator for all values of the noise-to-signal ratio, and for all levels of aggregation. The better asymptotic properties of the bias component of the ELW estimator do not translate to the finite sample, where the variance component of the MSE is dominant. This is particularly true for small values of d, being the asymptotic variance of the ELW estimator equal to $\frac{1+2d}{16d^2}$ (Hurvich et al., 2005), which diverges as d converges to zero. Unless the estimation of the volatility of the noise component is required, as it is the case of LMSV models, it seems hard to recommend the ELW in place of the LW estimator for the estimation of the LRD parameter, even when the presence of an additive noise component is suspected.

When no bias term is present, that is, when the process is pure LRD, the MSE of the LW estimator depends uniquely on the variance, which decreases exponentially with m. Therefore, the minimal MSE is attained at the highest value of m available. This implies that, for pure LRD processes, aggregation is deleterious because it reduces the number of available frequencies. If a bias component is present, because of either a noise component or a SRD components, temporal aggregation can have beneficial effects. The numerical analysis in Chapter 3 shows that the MSE of the LW estimator is essentially unaffected by temporal aggregation, provided that the number of frequencies is not reduced to the point where the variance effect becomes dominant. When the values of d and ns are large, corresponding to a larger noise component of the bias, higher levels of aggregation can be used before the variance effect become dominant.

The variance of the ELW estimator dominates the MSE of the estimator to a greater extent than the variance of the LW estimator does. It is not surprising, therefore, that aggregation affects

the ELW estimator negatively at lower levels of aggregation than it does for the LW estimator, and more so for small values of d. As for the LW estimator, with large values of d and ns, corresponding to a larger noise component of the bias, higher levels of aggregation can be used before the variance effect becomes dominant. Because the ELW estimator is less robust with respect to aggregation than the LW estimator, we would again recommend against using the ELW estimator to estimate d if there is no need to estimate the volatility of the noise.

To determine how much aggregation is acceptable, a rule of thumb can be to estimate d using the LW estimator according to the procedure suggested by Henry and Robinson (1996) for a range of aggregation levels. The performance of the estimator is invariant under aggregation, provided enough frequencies are available to reach the optimal value of m. If the estimated value $\hat{d}$ is constant over all aggregation levels, one can reasonably infer that the variance effect has not become dominant over the whole range of frequencies at any level of aggregation. Therefore, an estimate that attain the minimal MSE can be obtained at the highest level of aggregation considered.

Appendix A

The overall procedure to compute the MSE LW-type estimators is inspired by the proofs in Smith (1989), Henry and Robinson (1996) and Hurvich et al. (1998) and can be summarized as follows. The Local Whittle estimator minimizes the contrast

$$h = \sum_{i=1}^{m} \left(\log f_i + \frac{y_i}{f_i} \right) \tag{A.1}$$

where $m < N$ is a bandwidth parameter chosen by the user in a way that balances the trade-off between bias and variance.

Let's write $\hat{\theta}$ for the value that minimizes (A.1), and θ_0 for the true value of θ. Indicate by ∇h the gradient of h and by H its Hessian matrix.

If $\hat{\theta}$ is in a neighborhood of θ_0, one can use the following Taylor approximation

$$\nabla h(\hat{\theta}) = 0 = \nabla h(\theta_0) + H(\theta_0)(\hat{\theta} - \theta_0) + r_N \tag{A.2}$$

where $H(\theta)$ is the Hessian matrix computed at θ. If the remainder r_N is negligible, the asymptotic properties of $\hat{\theta} - \theta_0$ are equal to the properties of

$$\hat{\theta} - \theta_0 \approx -H^{-1}(\theta_0)\nabla h(\theta_0) \tag{A.3}$$

In practice, we will replace $\nabla h(\theta_0)$ by an approximation and H by its expectation H^*. Hence, we

will have

$$E(\hat{\theta} - \theta_0) \approx -H^{*-1}(\theta_0) E\nabla h(\theta_0) \tag{A.4}$$

and

$$Var(\hat{\theta} - \theta_0) \approx H^{*-1}(\theta_0) \tag{A.5}$$

A.1 The Bias of the LW Estimator with Added Noise

In this section we assume that the underlying model is the composed by a LRD process plus a white noise, that is, it has the form of Equation (2.41). However, the local Whittle estimator employs the familiar form of the spectral density $f(\lambda) = c\lambda^{-2d}$ in the expression of the contrast. In other words, we are assuming that the model has been misspecified in the estimation step.

The Whittle contrast is the usual

$$h = \sum_{i=1}^{m} \left\{ \log c - 2d \log \lambda_i + \frac{I(\lambda_i)}{c\lambda_i^{-2d}} \right\}$$

Let $x_i := \log \lambda_i$ and $I_i = I(\lambda_i)$.

For the purpose of evaluating the bias and variance of the estimator, assume that the expected value of the periodogram ordinates is

$$E\{I_i\} = \exp(\beta_0 + \beta_1 x_i)\{1 + \gamma\lambda_i^{\delta} + \alpha\lambda_i^{-\beta_1} + o(\lambda_i^{min(-\beta_1,\delta)})\} \tag{A.6}$$

where $\alpha = \frac{\sigma_\varepsilon^2}{2\pi c}$, $\beta_0 = \log c$, and $\beta_1 = -2d$. The parameters β_0 and β_1 are the targets of the estimation.

Write l_j for $\frac{\partial L}{\partial \beta_j}$, l_{jk} for $\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}$, $j, k = 0, 1$. Then

$$\begin{aligned} h_0 &= \sum_i \{1 - I_i \exp(-\beta_0 - \beta_1 x_i)\}, \\ h_1 &= \sum_i x_i \{1 - I_i \exp(-\beta_0 - \beta_1 x_i)\}, \\ h_{00} &= \sum_i I_i \exp(-\beta_0 - \beta_1 x_i), \\ h_{01} &= \sum_i x_i I_i \exp(-\beta_0 - \beta_1 x_i), \\ h_{11} &= \sum_i x_i^2 I_i \exp(-\beta_0 - \beta_1 x_i). \end{aligned}$$

We replace each of these by its expected value evaluated by (A.6) and ignore the $O(\lambda_i^{min(-\beta_1, \delta)})$ terms in (A.6). Then

$$\begin{aligned} h_{00}^* &= m, \\ h_{01}^* &= \sum x_i, \\ h_{11}^* &= \sum_i x_i^2. \end{aligned}$$

and the expected value of the Hessian matrix is

$$\begin{aligned} H &= \begin{pmatrix} m & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}, \\ H^{-1} &= \frac{1}{m \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & m \end{pmatrix}. \end{aligned} \tag{A.7}$$

Evaluating the sums,

$$\begin{aligned}
\sum_{i=1}^{m} x_i &= \sum_{i=1}^{m} \left(\log i + \log \frac{2\pi}{N} \right) \\
&\approx \int_0^m \left(\log x + \log \frac{2\pi}{N} \right) dx \\
&= m \log m - m + m \log \frac{2\pi}{N}, \qquad \text{(A.8)} \\
\sum_{i=1}^{m} x_i^2 &= \sum_{i=1}^{m} \left(\log i + \log \frac{2\pi}{N} \right)^2 \\
&= \sum_{i=1}^{m} \log^2 i + 2 \log \frac{2\pi}{N} \sum_{i=1}^{m} \log i + m \log^2 \frac{2\pi}{N} \\
&\approx \int_0^m \log^2 x dx + 2 \log \frac{2\pi}{N} \int_0^m \log x dx + m \log^2 \frac{2\pi}{N} \\
&= m \log^2 m - 2m \log m + 2m + 2 \log \frac{2\pi}{N} (m \log m - m) + m \log^2 \frac{2\pi}{N}
\end{aligned}$$

from which follows that

$$m \sum x_i^2 - (\sum x_i)^2 \sim m^2 \qquad \text{(A.9)}$$

From the discussion at the beginning of this section, we have that

$$\begin{aligned}
\hat{\theta} - \theta_0 &\approx -H^{-1} \nabla h(\theta_0), \\
E\left\{ \hat{\theta} - \theta_0 \right\} &\approx -H^{-1} E\left\{ \nabla h(\theta_0) \right\}, \qquad \text{(A.10)} \\
Var\{\hat{\theta} - \theta_0\} &\approx H^{-1}. \qquad \text{(A.11)}
\end{aligned}$$

And so

$$Var\{\hat{\beta}_1\} \approx \frac{1}{m}. \qquad \text{(A.12)}$$

To complete the calculation we must evaluate the asymptotic expectations of h_0 and h_1.

Using (A.6) we have

$$\begin{aligned} E\{h_0\} &\approx \sum_{i=1}^{m}\{1-(1+\gamma\lambda_i^{\delta}-\alpha\lambda_i^{-\beta_1})\} \\ &= -\gamma\left(\frac{2\pi}{N}\right)^{\delta}\sum_{i=1}^{m} i^{\delta}+\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\sum_{i=1}^{m} i^{-\beta_1} \\ &\approx -\gamma\left(\frac{2\pi}{N}\right)^{\delta}\int_0^m x^{\delta}dx-\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\int_0^m x^{-\beta_1}dx \\ &= -\gamma\left(\frac{2\pi}{N}\right)^{\delta}\frac{m^{\delta+1}}{\delta+1}-\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\frac{m^{1-\beta_1}}{1-\beta_1} \end{aligned} \tag{A.13}$$

and

$$\begin{aligned} E\{h_1\} &\approx \sum_{i=1}^{m} x_i\{1-(1+\gamma\lambda_i^{\delta}-\alpha\lambda_i^{-\beta_1})\} \\ &= -\gamma\left(\frac{2\pi}{N}\right)^{\delta}\sum_{i=1}^{m} i^{\delta}\left(\log i+\log\frac{2\pi}{N}\right) \\ &\quad +\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\sum_{i=1}^{m} i^{-\beta_1}\left(\log i+\log\frac{2\pi}{N}\right) \\ &\approx -\gamma\left(\frac{2\pi}{N}\right)^{\delta}\int_0^m x^{\delta}\left(\log x+\log\frac{2\pi}{N}\right)dx \\ &\quad +\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\int_0^m x^{-\beta_1}\left(\log x+\log\frac{2\pi}{N}\right)dx \end{aligned} \tag{A.14}$$

Then the bias of $\hat{\beta}_1$ is approximately

$$\begin{aligned}
& \frac{1}{m^2}(\sum_{i=1}^{m} x_i h_0 - m h_1) \\
= \; & \frac{1}{m^2}\left[\left(m\log m - m + m\log\frac{2\pi}{N}\right)\left\{-\gamma\left(\frac{2\pi}{N}\right)^{\delta}\frac{m^{\delta+1}}{\delta+1}\right\}\right. \\
& \left(m\log m - m + m\log\frac{2\pi}{N}\right)\left\{-\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\frac{m^{1-\beta_1}}{1-\beta_1}\right\} \\
& \left.+m\gamma\left(\frac{2\pi}{N}\right)^{\delta}\left\{\frac{m^{\delta+1}\log m}{\delta+1} - \frac{m^{\delta+1}}{(\delta+1)^2} + \frac{m^{\delta+1}}{\delta+1}\log\frac{2\pi}{N}\right\}\right] \\
& \left.-m\alpha\left(\frac{2\pi}{N}\right)^{-\beta_1}\left\{\frac{m^{1-\beta_1}\log m}{1-\beta_1} - \frac{m^{1-\beta_1}}{(1-\beta_1)^2} + \frac{m^{1-\beta_1}}{1-\beta_1}\log\frac{2\pi}{N}\right\}\right] \\
= \; & \frac{\gamma\delta}{(\delta+1)^2}\left(\frac{2\pi m}{N}\right)^{\delta} - \frac{\alpha\beta_1}{(1-\beta_1)^2}\left(\frac{2\pi m}{N}\right)^{-\beta_1}
\end{aligned} \tag{A.15}$$

Combining (A.12) and (A.15), the approximate MSE of $\hat{\beta}_1$ is given by

$$\frac{1}{m} + \frac{\gamma^2\delta^2}{(\delta+1)^4}\left(\frac{2\pi m}{N}\right)^{2\delta} + \frac{\alpha^2{\beta_1}^2}{(1-\beta_1)^4}\left(\frac{2\pi m}{N}\right)^{-2\beta_1} - 2\frac{\gamma\delta\alpha\beta_1}{(\delta+1)^2(1-\beta_1)^2}\left(\frac{2\pi m}{N}\right)^{\delta-\beta_1} \tag{A.16}$$

The optimal value of m can only be found numerically, and it will depend on the value of d.

A.2 The Extended Local Whittle

Suppose the model is characterized by a spectral density $f(\lambda,\theta)$, for some vector of parameters θ. Let $f_i := f(\lambda_i)$ be the value of the density corresponding to the Fourier frequency $\lambda_i = \frac{2\pi i}{N}$, $i = 1,\ldots,[N/2]$. Similarly, let $y_i := I(\lambda_i)$ be the ith periodogram ordinate.

For the model with added noise, we will assume that

$$f_i = \alpha\left\{1 + \beta\left(\frac{i}{m}\right)^{-\gamma}\right\} \tag{A.17}$$

where $\theta = (\alpha, \beta, \gamma)$ is the set of parameters. The long range dependence effect is driven by the parameter $\gamma = 2d$.

For a long-memory process, the periodogram is a biased estimator of the spectral density in a neighborhood of the origin (Hurvich and Beltrao, 1994). Therefore, we assume that

$$E(y_i) = \alpha \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\} \left\{ 1 + \delta \left(\frac{i}{m} \right)^{\tau} + o\left[\left(\frac{i}{m} \right)^{\tau} \right] \right\} \tag{A.18}$$

for some $\delta, \tau > 0$.

We will now proceed to compute a heuristic expression for H^* for the ELW estimator.

The Whittle contrast (A.1) is

$$h = \sum_{i=1}^{m} \left[\log \alpha + \log \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\} + y_i \alpha^{-1} \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-1} \right] \tag{A.19}$$

Denote the derivatives of h by $h_{\theta_i} := \frac{\partial h}{\partial \theta_i}$ and $h_{\theta_i, \theta_j} := \frac{\partial h}{\partial \theta \partial \theta_i}$. Then, we have, for the gradient

$$\begin{aligned}
h_\alpha &= \sum_{i=1}^{m} \left[\alpha^{-1} y_i \alpha^{-2} \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-1} \right] \\
h_\beta &= \sum_{i=1}^{m} \left(\frac{i}{m} \right)^{-\gamma} \left[\left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-1} - y_i \alpha^{-1} \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-2} \right] \\
h_\gamma &= -\beta \sum_{i=1}^{m} \left(\frac{i}{m} \right)^{-\gamma} \log \left(\frac{i}{m} \right) \left[\left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-1} - y_i \alpha^{-1} \left\{ 1 + \beta \left(\frac{i}{m} \right)^{-\gamma} \right\}^{-2} \right]
\end{aligned}$$

And for the Hessian matrix,

$$
\begin{aligned}
h_{\alpha\alpha} &= \sum_{i=1}^{m}\left[-\alpha^{-2}+2y_i\alpha^{-3}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-1}\right] \\
h_{\beta\beta} &= \sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-2\gamma}\left[-\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}+2y_i\alpha^{-1}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-3}\right] \\
h_{\gamma\gamma} &= \beta\sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-\gamma}\log^2\left(\frac{i}{m}\right)\left[\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-1}-y_i\alpha^{-1}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}\right] \\
&\quad +\beta^2\sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-2\gamma}\log^2\left(\frac{i}{m}\right)\left[-\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}+2y_i\alpha^{-1}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-3}\right] \\
h_{\alpha\beta} &= \sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-\gamma}\left[y_i\alpha^{-2}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}\right] \\
h_{\alpha\gamma} &= -\beta\sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-\gamma}\log\left(\frac{i}{m}\right)\left[y_i\alpha^{-2}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}\right] \\
h_{\beta\gamma} &= -\sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-\gamma}\log\left(\frac{i}{m}\right)\left[\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-1}-y_i\alpha^{-1}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}\right] \\
&\quad -\beta\sum_{i=1}^{m}\left(\frac{i}{m}\right)^{-2\gamma}\log\left(\frac{i}{m}\right)\left[-\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}+2y_i\alpha^{-1}\left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-3}\right]
\end{aligned}
$$

Replacing f_i for y_i in the equations above, we find the expected values

$$\begin{aligned}
h^*_{\alpha\alpha} &= m\alpha^{-2} \\
h^*_{\beta\beta} &= \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{-2\gamma} \left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2} \\
h^*_{\gamma\gamma} &= \beta^2 \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{-2\gamma} \log^2\left(\frac{i}{m}\right) \left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2} \\
h^*_{\alpha\beta} &= \alpha^{-1} \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{-\gamma} \left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-1} \\
h^*_{\alpha\gamma} &= -\beta\alpha^{-1} \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{-\gamma} \log\left(\frac{i}{m}\right) \left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-1} \\
h^*_{\beta\gamma} &= -\beta \sum_{i=1}^{m} \left(\frac{i}{m}\right)^{-2\gamma} \log\left(\frac{i}{m}\right) - \left\{1+\beta\left(\frac{i}{m}\right)^{-\gamma}\right\}^{-2}
\end{aligned}$$

Replacing sums with integrals, we can write

$$\begin{aligned}
h^*_{\beta\beta} &\approx \int_0^m \left(\frac{y}{m}\right)^{-2\gamma} \left\{1+\beta\left(\frac{y}{m}\right)^{-\gamma}\right\}^{-2} dy \\
&= m\int_0^1 x^{-2\gamma}(1+\beta x^{-\gamma})^{-2} dx \qquad \text{(A.20)}
\end{aligned}$$

In (A.20), operate the substitution

$$x = \beta^{1/\gamma}\left(\frac{u}{1-u}\right)^{-1/\gamma} \qquad \text{(A.21)}$$

Then $h^*_{\beta\beta}$ becomes

$$\begin{aligned} h^*_{\beta\beta} &\approx m\beta^{1/\gamma-2}\gamma^{-1}\int_{\beta/(\beta+1)}^{1} u^{1-1/\gamma}(1-u)^{1/\gamma-1}du \\ &= m\beta^{1/\gamma-2}\gamma^{-1}B\left(2-\frac{1}{\gamma},\frac{1}{\gamma}\right)\left\{1-I_{\beta/(1-\beta)}\left(2-\frac{1}{\gamma},\frac{1}{\gamma}\right)\right\} \end{aligned} \tag{A.22}$$

where

$$I_x(a,b) = \frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}dt \tag{A.23}$$

is an incomplete beta function, and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Similarly, for the other elements of H^*, we have

$$\begin{aligned} h^*_{\gamma\gamma} &\approx m\gamma^{-1}\frac{\partial}{\partial\eta}\frac{\partial}{\partial\xi}\left|\beta^{1/\gamma+\eta+\xi}B\left(2-\frac{1}{\gamma}\{1+\eta+\xi\};\frac{1}{\gamma}\{1+\eta+\xi\}\right)\right. \\ &\quad \left.\left\{1-I_{\beta/(1-\beta)}\left(2-\frac{1}{\gamma}\{1+\eta+\xi\};\frac{1}{\gamma}\{1+\eta+\xi\}\right)\right\}\right|_{\eta=0,\xi=0} \end{aligned} \tag{A.24}$$

$$\tag{A.25}$$

$$\begin{aligned} h^*_{\alpha,\beta} &\approx m\alpha^{-1}\beta^{1/\gamma-1}\gamma^{-1}B\left(1-\frac{1}{\gamma},\frac{1}{\gamma}\right) \\ &\quad \left\{1-I_{\beta/(1-\beta)}\left(1-\frac{1}{\gamma},\frac{1}{\gamma}\right)\right\} \end{aligned} \tag{A.26}$$

$$\tag{A.27}$$

$$\begin{aligned} h^*_{\alpha\gamma} &\approx -m\frac{\partial}{\partial\eta}\left|\beta^{(1+\eta)/\gamma}\gamma^{-1}B\left(1-\frac{1+\eta}{\gamma};\frac{1+\eta}{\gamma}\right)\right. \\ &\quad \left.\left\{1-I_{\beta/(1-\beta)}\left(1-\frac{1+\eta}{\gamma};\frac{1+\eta}{\gamma}\right)\right\}\right|_{\eta=0} \end{aligned} \tag{A.28}$$

$$\tag{A.29}$$

$$h^*_{\beta\gamma} \approx -m\frac{\partial}{\partial\eta}\left|\beta^{(1+\eta)/\gamma-1}\gamma^{-1}B\left(2+\frac{1+\eta}{\gamma};-\frac{1+\eta}{\gamma}\right)\right.$$
$$\left.\left\{1-I_{\beta/(1-\beta)}\left(2+\frac{1+\eta}{\gamma};-\frac{1+\eta}{\gamma}\right)\right\}\right|_{\eta=0} \tag{A.30}$$

A.3 The Mean Squared Error of the Geweke-Porter Hudak Estimator

In this section we extend the result of Hurvich et al. (1998) regarding the MSE of the GPH estimator to a model similar to the one employed by Robinson (1995a). We will assume a Gaussian series y_t whose spectral density near the origin has the form

$$f(\lambda) = \alpha\lambda^{-2d}\{1+\beta\lambda^{\delta}+o(\lambda^{\delta})\} \tag{A.31}$$

and that $m, n \to \infty$ with m going to infinity at a rate faster than Kn^{γ}, for $\gamma > 2/3$. We will show that, under these conditions, the MSE is approximately given by

$$\frac{\pi^2}{24m} + \frac{\beta^2\delta^2}{4(\delta+1)^4}\left(\frac{2\pi m}{N}\right)^{2\delta} \tag{A.32}$$

Note that the bias component of the GPH estimator does not depend on the scale parameter α.

Taking the logarithm of the spectral density, and using a Taylor expansion, we can rewrite

$$\begin{aligned}\log f(\lambda) &= \log\alpha - 2d\log\lambda + \log\{1+\beta\lambda^{\delta}+o(\lambda^{\delta})\}\\ &\approx \log\alpha - 2d\log\lambda + \beta\lambda^{\delta}+o(\lambda^{\delta})\end{aligned}$$

or

$$\log \frac{f(\lambda)}{I(\lambda)} I(\lambda) = \log \alpha - 2d \log \lambda + \beta \lambda^{\delta} + o(\lambda^{\delta})$$

which implies

$$\log I_j = (\log \alpha - C) - 2dx_j + \beta \lambda_j^{\delta} + \varepsilon_j$$

where $I_j = I(\lambda_j)$, $\lambda_j = \frac{2\pi j}{N}$, $x_i = \log \lambda_j$, $\varepsilon_j = \frac{I_j}{f_j} + C$, and $C = 0.5777$ is Euler's constant.

Let $a_j = (x_i - \bar{x})$ and $S_{xx} = \sum_1^m (x_i - \bar{x})^2$. Then the GPH estimator of d is

$$\begin{aligned} \hat{d} &= -\frac{1}{2} \frac{\sum_1^m a_j \log I_j}{S_x x} \\ &= d - \frac{1}{2} \frac{a_j \left[\beta \lambda_j^{\delta} + o(\lambda^{\delta}) \right]}{S_{xx}} - \frac{1}{2} \frac{\sum_1^m a_j \varepsilon_j}{S_{xx}} \end{aligned}$$

which implies that the bias is

$$\hat{d} - d = -\frac{1}{2S_{xx}} a_j \left[\beta \lambda_j^{\delta} + o(\lambda^{\delta}) \right] - \frac{1}{2S_{xx}} \sum_1^m a_j \varepsilon_j \tag{A.33}$$

with expected value

$$E(\hat{d} - d) = -\frac{1}{2S_{xx}} \sum_i^m a_j \left[\beta \lambda_j^{\delta} + o(\lambda^{\delta}) \right] - \frac{1}{2S_{xx}} \sum_1^m a_j E(\varepsilon_j) \tag{A.34}$$

The variance of $\hat{d}$ is

$$\begin{aligned} Var(\hat{d}) &= \frac{1}{4S_{xx}^2} Var(\sum_1^m a_j \epsilon_j) \\ &= \frac{1}{4S_{xx}^2} \sum_1^m a_j^2 Var(\varepsilon_j) + \frac{1}{2S_{xx}^2} \sum_{j=1}^m \sum_{k=j}^m a_j a_k Cov(\varepsilon_j, \varepsilon_k) \end{aligned} \tag{A.35}$$

Now, we have (Hurvich and Beltrao, 1994) (or also from the above discussion),

$$S_{xx} = m + o(m) \tag{A.36}$$

In order to evaluate the asymptotic expression of the bias, we first need to find asymptotic for the first term in Equation (A.34).

$$\sum a_j \left[\beta\lambda_j^\delta + o(\lambda_j^\delta)\right] = \beta \sum a_j \lambda_j^\delta + \sum a_j o(\lambda_j^\delta)$$

We have:

$$\begin{aligned}
\sum a_j \lambda_j^\delta &= \sum (x_i - \bar{x})\lambda_j^\delta \\
&= \sum (\log \lambda_j - 1/m \sum \log \lambda_j)\lambda_j^\delta \\
&= \sum \lambda_j^\delta \log \lambda_j - \frac{1}{m} \sum \lambda_j^\delta \sum \log \lambda_j
\end{aligned}$$

and

$$\begin{aligned}
\sum \lambda_j^\delta &= \sum \left(\frac{q\pi j}{N}\right)^\delta \\
&\approx \left(\frac{q\pi}{N}\right)^\delta \int_0^m x^\delta dx \\
&= \left(\frac{q\pi}{N}\right)^\delta \frac{m^{\delta+1}}{\delta+1}
\end{aligned}$$

$$\begin{aligned}
\sum \lambda_j^\delta \log \lambda_j &= \sum \left(\frac{2\pi j}{N}\right)^\delta \log j + \sum \left(\frac{2\pi j}{N}\right)^\delta \log \frac{2\pi}{N} \\
&= \left(\frac{2\pi}{N}\right)^\delta \sum j^\delta \log j + \left(\frac{2\pi}{N}\right)^\delta \log \frac{2\pi}{N} \sum j^\delta
\end{aligned}$$

$$\sum j^{\delta} \approx \frac{m^{\delta+1}}{\delta+1}$$

integrating by parts,

$$\begin{aligned}\sum j^{\delta} \log j &\approx \int_0^m x^{\delta} \log x dx \\ &= \frac{m^{\delta+1}}{\delta+1} \log m - \frac{m^{\delta+1}}{(\delta+1)^2}\end{aligned}$$

Hence

$$\sum \lambda^{\gamma} \log \lambda_j \approx \left(\frac{2pi}{N}\right)^{\delta} \left\{\frac{m^{\delta+1}}{\delta+1} \log m - \frac{m^{\delta+1}}{(\delta+1)^2}\right\} + \left(\frac{2pi}{N}\right)^{\delta} \log \frac{2\pi}{N} \frac{m^{\delta+1}}{\delta+1}$$

and

$$\begin{aligned}\sum a_j \lambda_j^{\delta} &\approx \left(\frac{2pi}{N}\right)^{\delta} \left\{\frac{m^{\delta+1}}{\delta+1} \log m - \frac{m^{\delta+1}}{(\delta+1)^2} + \log \frac{2\pi}{N} \frac{m^{\delta+1}}{\delta+1}\right. \\ &\quad \left. -\frac{m^{\delta+1}}{\delta+1}\left(\log m - 1 + \log \frac{2\pi}{N}\right)\right\} \\ &= \left(\frac{2\pi}{N}\right)^{\delta} \left\{\frac{\delta}{(\delta+1)^2} m^{\delta+1}\right\}\end{aligned}$$

which implies

$$-\frac{\beta}{2S_{xx}} \sum a_j \lambda_j^{\delta} \approx -\frac{\beta\delta}{2(\delta+1)^2} \left(\frac{2\pi m}{N}\right)^{\delta} \tag{A.37}$$

This results is equivalent to Lemma 1 in Hurvich et al. (1998), and to the bias term in MSE computation in Smith (1989). However, to fully evaluate the expected value of the bias, we still need to show that the remainder terms are negligible.

We now need the expected value and variance-covariances of $\varepsilon_j = \log \frac{I_j}{f_j} + C$. Following Hurvich et al. (1998), we compute first $Cov(\varepsilon_j, \varepsilon_k), j > k$.

Rewrite

$$\frac{I_j}{f_j} = \left(\frac{A_j}{f_j^{1/2}}\right)^2 + \left(\frac{B_j}{f_j^{1/2}}\right)^2$$

where

$$A_j = \frac{1}{2\pi N}^{1/2} \sum_{t=0}^{N-1} y_t \cos(\lambda_j t)$$

$$B_j = \frac{1}{2\pi N}^{1/2} \sum_{t=0}^{N-1} y_t \sin(\lambda_j t)$$

For the discrete orthogonality of the Fourier coefficients, it follows that $E(A_j) = E(B_j) = 0$.

Since y_t is Gaussian, and since $\varepsilon_j = \log \frac{I_j}{f_j} + C$, the joint distribution of $(\varepsilon_j, \varepsilon_k)$ is determined by the covariance matrix Σ of

$$\nu := \left(\frac{A_j}{f_j^{1/2}}, \frac{B_j}{f_j^{1/2}}, \frac{A_k}{f_k^{1/2}}, \frac{B_k}{f_k^{1/2}}\right) = (\nu_1, \nu_2, \nu_3, \nu_4)' \tag{A.38}$$

The distribution of ν is multivariate normal with mean zero and variance-covariance matrix Σ.

Let

$$\begin{aligned} \alpha_{j,k} \quad := \quad & \max \Bigg\{ |Cov(\frac{A_j}{f_j^{1/2}}, \frac{A_k}{f_k^{1/2}})|, |Cov(\frac{A_j}{f_j^{1/2}}, \frac{B_k}{f_k^{1/2}})|, \\ & |Cov(\frac{B_j}{f_j^{1/2}}, \frac{A_k}{f_k^{1/2}})|, |Cov(\frac{B_j}{f_j^{1/2}}, \frac{B_k}{f_k^{1/2}})| \Bigg\} \end{aligned} \tag{A.39}$$

Lemmas 2 and 3 of Hurvich and Deo (1999) or Theorem 2 Robinson (1995a) apply, so that $Cov(\varepsilon_j, \varepsilon_j) = O(\alpha_{j,k}^2)$ uniformly for $\log^2 m \leq j < j \leq m$, and $\alpha_{j,k} = O(\log j/k)$ uniformly for $1 \leq k < j \leq m$.

The final result follows from a straightforward application of Lemmas 4-8 of Hurvich et al. (1998).

Appendix B

In this Appendix we report the result by Cao (2002) that links the spectral density of the basic process to the spectral density of the aggregate process.

Let X_t, $t \in \mathbb{Z}$, be a stationary process with covariance function $\gamma_x(k)$, $k \in \mathbb{Z}$, and let

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_x(k) e^{-i\lambda k} \qquad \lambda \in [0, 2\pi) \tag{B.1}$$

be its spectral density.

Denote by $\tilde{X}_t$ the spectral representation of X_t, that is $\tilde{X}_t$ is an orthogonal process with mean zero such that

$$X_t = \int_0^{2\pi} e^{i\lambda t} d\tilde{X}(\lambda) \tag{B.2}$$

$\tilde{X}_t$ has the following properties:

$$E[d\tilde{X}(\lambda)\overline{d\tilde{X}(\lambda')}] = 0 \qquad \text{if } \lambda \neq \lambda' \tag{B.3}$$

$$E[|d\tilde{X}(\lambda)|^2] = f_X(\lambda) d\lambda \tag{B.4}$$

Lemma B.0.1 (Jin Cao's result)**.** *Let*

$$Y_t^{(k)} := X_{(t-1)k+1} + X_{(t-1)k+2} + \cdots + X_{tk} \tag{B.5}$$

Then the spectrum of $Y_t^{(k)}$ *is*

$$f_k(\lambda) = \frac{1}{k} \sum_{j=0}^{k-1} \frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda+2\pi j}{2k}} f_X \left(\frac{\lambda + 2\pi j}{k} \right) \tag{B.6}$$

Proof. Let

$$g(\lambda) = 1 + e^{i\lambda} + \cdots + e^{i(k-1)\lambda} \tag{B.7}$$

Using the spectral representation for X_t, and noting that $e^{i\lambda x}$ has period 2π, we have

$$\begin{aligned}
Y_t^{(k)} &= \sum_{j=0}^{k-1} \int_0^{2\pi} e^{i\lambda(mk-j)} d\tilde{X}(\lambda) \\
&= \sum_{j=0}^{k-1} \int_0^{2\pi} e^{i\lambda mk} g(\lambda) d\tilde{X}(\lambda) \\
&= \sum_{j=0}^{k-1} \int_0^{2k\pi} e^{i\lambda m} g\left(\frac{\lambda}{k}\right) d\tilde{X}\left(\frac{\lambda}{k}\right) \\
&= \sum_{j=0}^{k-1} \int_{2\pi j}^{2\pi(j+1)} e^{i\lambda m} g\left(\frac{\lambda}{k}\right) d\tilde{X}\left(\frac{\lambda}{k}\right) \\
&= \int_0^{2\pi} e^{i\lambda m} \sum_{j=0}^{k-1} g\left(\frac{\lambda + 2\pi j}{k}\right) d\tilde{X}\left(\frac{\lambda + 2\pi j}{k}\right) \\
&= \int_0^{2\pi} e^{i\lambda m} d\tilde{Y}(\lambda)
\end{aligned}$$

where

$$d\tilde{Y}(\lambda) = \sum_{j=0}^{k-1} g\left(\frac{\lambda + 2\pi j}{k}\right) d\tilde{X}\left(\frac{\lambda + 2\pi j}{k}\right) \tag{B.8}$$

is an orthogonal process and, hence, is the spectral representation of $Y_t^{(k)}$.

Applying Equation (B.4), and noting that $|g(\lambda)|^2 = \frac{\sin^2 \frac{k\lambda}{2}}{\sin^2 \frac{\lambda}{2}}$,

$$\begin{aligned} E[|d\tilde{Y}(\lambda)|^2] &= f_k(\lambda) d(\lambda) \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \left| g\left(\frac{\lambda + 2\pi j}{k}\right) \right|^2 f_X\left(\frac{\lambda + 2\pi j}{k}\right) d\lambda \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \frac{\sin^2 \frac{\lambda}{2}}{\sin^2 \frac{\lambda + 2\pi j}{2k}} f_X\left(\frac{\lambda + 2\pi j}{k}\right) \end{aligned}$$

□

Bibliography

Abry, P., Flandrin, P., Taqqu, M. S., and Veich, D. (2003), "Self-similarity and long-range dependence through the weavelet lens," in Doukhan et al. (2003), pp. 527–556.

Abry, P., Gonçalves, P., and Flandrin, P. (1995), "Wavelet Spectrum estimation and $1/f$ processes," in *Wavelets and Statistics*, eds. Antoniadis, A. and Oppenheim, G., New York: Springer, no. 115 in Lectures Notes in Statistics, pp. 15–30.

Abry, P. and Sellan, F. (1996), "The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation," *Applied and Computational Harmonic Analysis*, 3, 377–383.

Abry, P., Taqqu, M. S., Flandrin, P., and Veitch, D. (2000), "Wavelets for the analysis, estimation and synthesis of scaling data," New York: Wiley, Wiley Interscience, pp. 39–88, nociteparwil00.

Abry, P., Veitch, D., and Flandrin, P. (1998), "Long range-dependence: revisiting aggregation with wavelets," *J. of Time Series Analysis*, 19, 253–266.

Adenstat, R. (1974), "On large sample estimation of the mean of a stationary random sequence," *Annals of Statistics*, 2, 1095–1107.

Adler, R. J., Feldman, R. E., and Taqqu, M. S. (eds.) (1998), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Boston: Birkhäuser.

Andrews, D. W. K. and Guggenberger, K. (2000), "Bias-reduced log-periodogram estimator for the long-range parameter," Preprint.

Andrews, D. W. K. and Sun, Y. (2004), "Adaptive local polynomial Whittle estimation of long-range dependence," *Econometrica*, 72, 569–614.

Bardet, J.-M., Lang, G., Moulines, E., and Soulieres, P. (2000), "Wavelet estimator of long-range dependence," *Statistical Inference for Stochastic Processes*, 3, 85–99.

Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., Stoev, S., and Taqqu, M. S. (2003), "Semi-parametric estimation for the long-range dependence parameter: a survey," in Doukhan et al. (2003), pp. 557–578.

Barnett, V. and Turkman, F. K. (eds.) (1993), *Statistics for the environment*, Chichester: Wiley and Sons.

Beran, J. (1993), "Fitting long-memor models by generalized linear regression," *Biometrika*, 80, 817–822.

— (1994), *Statistics for Long Memory Processes*, Monograph in Statistics and Applied Probability, 61, Boca Raton, Florida: Chapman & Hall/CRC.

Bhattacharya, R. N., Gupta, V. K., and Waymire, E. (1983), "The Hurst Effect under Trends," *Journal of Applied Probability*, 20, 649–662.

Bingham, N., Goldie, C., and Treugels, J. (1987), *Regular Variation*, vol. 27 of *Enciclopedia of Mathematics and its Applications*, Cambridge, UK: Cambridge Universisty Press.

Bloomfield, P. (1973), "An exponential model for the spectrum of a scalar time series," *Biometrika*, 60, 217–226.

Bollerslev, T. and Mikkelsen, O. H. (1996), "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73, 151–184.

Box, G. and Jenkins, G. (1971), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Brockwell, P. J. and Davis, R. A. (1991), *Time series: theory and methods*, Springer Series in Statistics, New York: Springer, 2nd ed.

Cao, J. (2002), "A note on the spectral density of aggregate processes," Unpublished note.

Cassandro, M. and Jona-Lasinio, G. (1978), "Critical behavior and probabily theory," *Adv. Physics*, 27, 913–941.

Chambers, M. J. (1998), "Long Memory and Aggregation in Macroeconomic Time Series," *International Economic Review*, 39, 1053–1072.

Cheung, Y. and Diebold, F. (1994), "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean," *Journal of Econometrics*, 62, 301–316.

Cox, D. R. (1984), "Long-range dependence: a review," in David and David (1984), pp. 55–74, .

Crato, N. and Ray, B. K. (2002), "Semi-parametric smoothing estimator for long-memory processes with added noise," Accepted for publication, *Journal of Statistical Planning and Inference*, 30 July 2001.

Dahlhaus, R. (1989), "Efficient parameter estimation of self-similar processes," *Annals of Statistics*, 17, 1749–1766.

Danzig, P., Mogul, J., Paxson, V., and Schwartz, M. (2000), "Web Site: http://ita.ee.lbl.gov/html/contrib/fft-fgn.html," .

David, H. and David, H. (eds.) (1984), *Statistics: An Appraisal. Proceeding 50th Anniversary Conference*, Iowa University Press.

Delgado, M. and Robinson, P. M. (1996), "Optimal bspectral bandwidth for long memory," *Statistical Sinica*, 97–112.

Deo, R. S. (1997), "Asymptotic Theory for certain regression models with long memory errors," *J. of Time Series Analysis*, 18, 385–394.

Deo, R. S. and Hurvich, C. M. (2001), "On the log periodogram regression estimator of the parameter in long memory stochastic volatility models," *Econometric Theory*, 17, 686–710.

— (2003), "Estimation of long memory in volatility," in Doukhan et al. (2003), pp. 313–324.

Dietrich, C. and Newsam, G. (1997), "Fast and exact simulation of stationary Gaussian process through circulant embedding of the covariance matrix," *SIAM J. Sci. Comput.*, 18, 1088–1107.

Doukhan, P., Oppenheim, G., and Taqqu, M. S. (eds.) (2003), *Theory and Applications of Long-Range Dependence*, Basel: Birkhäuser.

Embrechts, P. and Maejima, M. (2002), *Self similar Processes*, Princeton Series in Applied Mathematics, Princeton, NJ: Princeton University Press.

Fairfield Smith, H. (1938), "An empirical law describing heterogeneity in the yields of agricultural crops," *J. Agric. Sci.*, 28, 1–23.

Flandrin, P. (1989), "On the spectrum of fractional Brownian motion," *IEEE Transactions on Information Theory*, IT-35, 197–199.

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

Geweke, J. and Porter-Hudak, S. (1983), "The estimation and application of long-memory time series model," *Journal of Time Series Analysis*, 4, 221–238.

Giraitis, L. and Robinson, P. M. (2003), "Edgeworth expansions for semiparametric Whittle estimation of long memory," *Annals of Statistics*, 41, 1325–1375.

Giraitis, L. and Surgailis, D. (1990), "A central limit theorem for quadratic forms in strongly dependent random variables and its application to asymptotical normality of Whittle's estimate," *Probability Theory and Related Fields*, 86, 87–104.

Giraitis, L. and Taqqu, M. S. (1999), "Whittle Estimator for finite-variance non-Gaussian time series with long memory," *Annals of Statistics*, 27, 178–203.

Graf, H. P. (1983), "Long-range correlations and estimation of the self-similarity parameter," Ph.D. thesis, ETH, Zürich.

Granger, C. W. J. (1966), "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34, 150–161.

— (1980a), "Long memory relationship and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

Granger, Clive W. J. and. Joyeux, R. (1980b), "An introduction to long-range time series models and fractional differencing," *J. of Time Series Analysis*, 1, 15–30.

Hannan, E. J. (1973), "The asymptotic theory of linear time series models," *Journal of Applied Probability*, 10, 130–145.

Hannig, J., Marron, J., and RH Riedi, R. (2001), "Zooming statistics: Inference across scales," *Journal of the Korean Statistical Society*, 30, 327–345.

Harvey, A. C. (1998), "Long memory in stochastic volatility," in *Forecasting Volatility in Financial Markets*, eds. Knight, J. and Satchell, S., London: Butterworth-Heinmann.

Haslett, J. and Raftery, A. E. (1989), "Space-time modeling with long-memory dependence: assessing Ireland's wind power resource (with discussion)," *Applied Statistics*, 38, 1–50.

Heeke, H. (1991), "Statistical multiplex gain for variable bit rate codecs in ATM networks," *Int. J. Digit. Analog. Commun. Syst.*, 4, 182–189.

Henry, M. and Robinson, P. M. (1996), "Bandwidth choice in Gaussian semiparametric estimation of long-range dependence," in Robinson and Rosenblatt (1996), pp. 220–232.

Heyde, C. and Gay, G. (1993), "Smoothed Periodogram asymptotics and estimation for processes and fields with long-range dependence," *Stochastic Processes and Their Application*, 45, 169–187.

Heyman, D., Tabatabai, A., and Lakshman, T. (1991), "Statistical analysis and simulation of video teleconferencing in ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, 2, 49–59.

Higuchi, T. (1988), "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, 31, 277–283.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrica*, 68, 165–176.

Hosoya, Y. (1997), "Limit theory with long-range dependence and statistical inference of related models," *Annals of Statistics*, 25, 105–137.

Hurst, H. (1951), "Long term storage capacity of reservoirs," *Trans. Am. Soc. Civil Engineers*, 161, 770–799.

Hurvich, C. M. and Beltrao, K. I. (1993), "Asymtoptics for the low-frequency ordinates of the periodogram of a long-memory time series," *J. of Time Series Analysis*, 14, 455–472.

— (1994), "Automatic semiparametric estimation of the memory parameter of a long memory time series," *J. of Time Series Analysis*, 15, 285–382.

Hurvich, C. M. and Brodsky, J. (2001), "Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models," *J. of Time Series Analysis*, 22, 221–249.

Hurvich, C. M. and Deo, R. (1999), "Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series," *J. of Time Series Analysis*, 20, 331–341.

Hurvich, C. M., Deo, R., and Brodsky, J. (1998), "The mean squared error of Geweke and Porter-Hudak's estimator of a long-memory time-series," *J. of Time Series Analysis*, 19, 19–46.

Hurvich, C. M., Moulines, E., and Soulier, P. (2002), "The FEXP estimator for non-Gaussian, potentially non stationary processes," *Stochastic Processes and their Application*, 97, 307–340.

— (2005), "Estimating Long Memory in Volatility," *Econometrica*, 73, 1283–1328.

Hurvich, C. M. and Ray, B. K. (2001), "The local Whittle Estimator of long memory stochastic volatility," *Journal of Financial Econometrics*, 1, 445–470.

Janacek, G. J. (1987), "Determining the degree of differencing for time series via log spectrum," *J. of Time Series Analysis*, 3, 177–183.

Jones, P. and Briffa, K. (1992), "Global surface air temperature variations during the twentieth century, part 1, Spatial, temporal and seasonal detail," *Holocene*, 2, 165–179.

Kelly, F. P., Zachary, S., and Ziedins, I. (eds.) (1996), *Stochastic Networks*, Oxford: Oxford University Press.

Kokoszka, P. S. and Bhansali, R. J. (1999), "Estimation of the long memory parameter by fitting fractional differenced autoregressive models," Preprint.

Kolmogorov, A. N. (1941), "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," *Dokl. Akad. Nauk SSSR 30. Translated by V. Levin. Reprinted in Proc. R. Soc. Lond*, 434, 9–13.

Künsch, H. R. (1987), "Statistical aspects of self-similar processes," in Prohorov and Sazonov (1987), pp. 67–74, .

Lahiri, S. N. (2003), "A necessary and sufficient condition for aymptotic independence of discrete Fourier trasform under short- and long-range dependence," *Annals of Statistics*, 31, 613–641.

Lamperti, J. W. (1962), "Semi-stable stochastic processes," *Transaction of the American Mathematical Society*, 1962, 62–78.

Le, L. A., Jeffay, K. J., and Smith, F. D. (2003), "The effects of Active Queue Managment on Web performance," in *Proc. of ACM SIGCOMM 2003*, pp. 265–276.

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

Leland, W. E. and Wilson, D. V. (1991), "High time-resolution measurement and analysis of LAN traffic: Implication for LAN interconnection," in *Proceedings of the IEEE INFOCOM '91, Bal Harbor, FL*, pp. 1360–1366.

Lo, A. W. (1991), "Long memory in stock market prices," *Econometrica*, 59, 1279–1313.

Mandelbrot, B. B. (1975), "Limit theorems of the self-normalized range for weakly and strongly dependent processes," *Z. Whar. verw. Geb.*, 31, 271–285.

— (1977), *Fractals: Form, Chance and Dimension*, San Francisco: Freeman.

— (1983), *The Fractal Geometry of Nature*, New York: W.H. Freeman and Company.

Mandelbrot, B. B. and Ness, J. W. V. (1968), "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, 10, 422–437.

Mandelbrot, B. B. and Wallis, J. (1968), "Noah, Joseph, and Operational Hydrology," *Water Resources Research*, 4, 909–918.

— (1969a), "Computer Experiments with Fractional Gaussian Noises," *Water Resources Research*, 5, 228–267.

— (1969b), "Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence," *Water Resources Research*, 5, 967–988.

— (1969c), "Some Long-Run Properties of Geophysical Records," *Water Resources Research*, 5, 321–340.

Moulines, E. and Soulier, P. (1999), "Broadband log-periodogram regression of time series with long-range dependence," *Annals of Statistics*, 27, 1415–1439.

— (2003), "Semiparametric Spectral Estimation for Fractional Processes," in Doukhan et al. (2003), pp. 251–302.

Park, C., Hernandez-Campos, F., Le, L., Marron, J. S., Park, J., Pipiras, V., Smith, F. D., Smith,

R. L., Trovero, M., and Zhu, Z. (2005), "Long Range Dependence Analysis of Internet Traffic," Submitted to Statistical Science.

Paxson, V. (1995), "Fast Approximation of Self-Similar Network Traffic," *IEEE/ACM Transaction on Networking*, 2, 316–336.

Pipiras, V. (2005), "Wavelet-based simulation of fractional Brownian motion revisited," *Applied and Computational Harmonic Analysis*, 19, 49–60.

Prohorov, Y. and Sazonov, V. V. (eds.) (1987), *Proceeding of the First World Congress of the Bernoulli Society*, vol. 1, Utrecht, VNU Science Press.

Robinson, P. M. (1995a), "Log Periodogram regression of time series with long range dependence," *The Annals of Statistics*, 23, 1048–1072.

— (1995b), "Gaussian semiparametric estimation of long range dependence," *The Annals of Statistics*, 23, 1630–1661.

Robinson, P. M. and Henry, M. (2000), "Higher Order semiparametric M-estimation of long-memory," *Journal of Econometrics*, forthcoming.

Robinson, P. M. and Rosenblatt, M. (eds.) (1996), *Athens Conference on Applied Probability and Time Series, Volume II: Time series analysis in memory of E.J. Hannan*, no. 115 in Lecture Notes in Statistics, New York: Springer.

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

Smith, R. L. (1989), "Estimating Long Range Dependence from the Lower Tail of the Periodogram," Unpublished.

— (1993), "Long-range dependence and global warming," in Barnett and Turkman (1993), pp. 141–161, .

Taqqu, M. (2003), "Fractional Brownial Motion and Long-Range Dependence," in Doukhan et al. (2003), pp. 5–38.

Taqqu, M. S. and Teverovsky, V. (1998), "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series," in Adler et al. (1998), pp. 177–217, .

Taqqu, M. S., Teverovsky, V., and Willinger, W. (1997), "Is network traffic self-similar or multifractal?" *Fractals*, 5, 63–73.

Taqqu, Murad, S. and Teverovsky, V. (1996), "Semi-parametric graphical estimation techniques for long-memory data," in Robinson and Rosenblatt (1996), pp. 420–432.

Teles, P., Wi, W. W., and Crato, N. (1999), "The use of Aggregate Time Series in Testing for Long Memory," *Bulletin of the International Statistical Institute*, 52nd Session, 341–342.

Tousson, O. (1925), *Mémoire sur l'Histoire du Nil*, Mémoire de l'Institute d'Egypte.

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

Velasco, V. (2000), "Non-Gaussian log-periodogram regression," *Econometric Theory*, 16, 44–79.

Vervaat, W. (1987), "Properties of general self similar processes," *Bulletin International Statistical Institute*, 52, 199–216.

Wei, W. W. (1990), *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley.

Whittle, P. (1953), "Estimation and Information in stationary time series," *Ark. Mat.*, 2, 423–434.

— (1956), "On the variation of yield variance with plot size," *Biometrika*, 49, 337–343.

— (1962), "Topographic correlation, power law covariance function, and diffusion," *Biometrika*, 49, 340–314.

Willinger, W., Paxson, V., Riedi, R. H., and Taqqu, M. S. (2003), "Long-range dependence and data network traffic," in Doukhan et al. (2003), pp. 373–408.

Willinger, W., Taqqu, M. S., and Erramilli, A. (1996), "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks," in Kelly et al. (1996), pp. 339–366, .

Wood, A. T. A. and Chan, G. (1996), "Simulation of stationary Gaussian processes in $[0,1]^d$," *Journal of Computational and Graphical Statistics*, 3, 409–432.

Yajima, Y. (1989), "A central limit theorem for Fourier transforms of strongly dependent stationary processes," *J. of Time Series Analysis*, 10, 375–383.