

ESTIMATION OF LONG-RANGE DEPENDENCE

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2010

By
Oskar Vivero
School of Electrical and Electronic Engineering

Contents

Abstract	11
Declaration	12
Copyright	13
Acknowledgements	15
1 Introduction	17
1.1 Thesis Overview	20
1.2 Thesis Contributions	21
2 Long-Range Dependence	23
2.1 Introduction	23
2.2 Definitions of Long Range Dependence	24
2.3 Models with LRD	27
2.3.1 Fractional Brownian Motion	27
2.3.2 Fractional Gaussian Noise	29
2.3.3 Fractionally Differenced Noise	31
2.3.4 ARFIMA Processes	33
2.4 Synthesis of Gaussian LRD Processes	39
2.4.1 Circulant Embedding Method	39
Appendices	44
2.A Proofs of Theorems 2.3.2, 2.3.5 and 2.3.8	44
3 Estimation of LRD Models	48
3.1 Introduction	48
3.2 Semi-parametric Methods	49

3.2.1	Rescaled Adjusted Range	50
3.2.2	De-trended Fluctuation Analysis	52
3.2.3	Log-Periodogram Regression	56
3.3	Likelihood Based Methods	57
3.3.1	Maximum Likelihood Estimate	58
3.3.2	Whittle’s Approximation to the MLE	62
3.3.3	Circulant Embedding Estimator	65
3.4	Chapter Conclusions	72
Appendices		72
3.A	Proof of Theorem 3.3.6	72
3.B	Useful lemmas	76
4	Regularised Estimators	80
4.1	Introduction	80
4.2	Problem Statement	81
4.3	Regularisation	84
4.4	Maximum Likelihood for Error Corrupted Observations	86
4.5	Conclusions	90
5	Numerical Results	91
5.1	Error-Free Experiments	91
5.1.1	Description of the Experiments	91
5.1.2	Results and Discussion	92
5.2	Error Corrupted Experiments	95
5.2.1	Description of the Experiments	95
5.2.2	Results and Discussion	95
5.3	Real World Examples	98
5.3.1	Ethernet Data	98
5.3.2	Ocular Accommodation	102
Appendices		105
5.A	Monte Carlo Studies for the Error-Free Case	105
5.B	Monte Carlo Studies for the Error-Corrupted Case	119
6	Conclusions & Future Work	126
6.1	General Conclusions	126

6.2 Future Work	127
Bibliography	129
Final Word Count = 20,019	

List of Tables

- 5.1 Ethernet Data, Byte Series 99
- 5.2 Ethernet Data, Packet Series 99
- 5.3 Ocular Accommodation Estimates 103

List of Figures

1.1	Resistor-Capacitor Transmission Line	19
1.2	Linear system approximation to LRD	19
2.1	Correlation and Spectral Density of FDN	35
2.2	Correlation of an ARFIMA process	36
2.3	Spectral Density of an ARFIMA process	37
3.1	R/S: Estimate	52
3.2	DFA: Estimates	55
3.3	DFA: De-trending procedure	55
3.4	PSD: Estimate	57
3.5	MLE: Estimate	61
4.1	Spectral fit of a of the Whittle estimate	81
4.2	Condition number of the covariance matrix of fGN	83
4.3	Optimisation surface for Regularised Estimate	89
5.1	Monte Carlo Simulations FDN, N=1024	94
5.2	Monte Carlo Simulations Error-Corrupted FDN, 15 dB SNR	97
5.3	Dataset corresponding to the Ethernet byte and packet time-series	100
5.4	Ethernet dataset estimates	101
5.5	Dataset corresponding to subject ED, trial number 4	103
5.6	Ocular Accommodation, Spectral Density Estimates	104
5.7	Monte Carlo Simulations fGN, N=64	105
5.8	Monte Carlo Simulations fGN, N=128	106
5.9	Monte Carlo Simulations fGN, N=512	107
5.10	Monte Carlo Simulations fGN, N=1024	108
5.11	Monte Carlo Simulations fGN, N=2048	109
5.12	Monte Carlo Simulations fGN, N=16384	110

5.13	Monte Carlo Simulations FDN, N=64	111
5.14	Monte Carlo Simulations FDN, N=128	112
5.15	Monte Carlo Simulations FDN, N=512	113
5.16	Monte Carlo Simulations FDN, N=2048	114
5.17	Monte Carlo Simulations FDN, N=16384	115
5.18	Monte Carlo Simulations ARFIMA, Bias for N=1024	116
5.19	Monte Carlo Simulations ARFIMA, Bias for N=2048	117
5.20	Monte Carlo Simulations ARFIMA, Bias for N=16384	118
5.21	Monte Carlo Simulations Error-Corrupted FDN, 30 dB SNR . . .	119
5.22	Monte Carlo Simulations Error-Corrupted FDN, 10 dB SNR . . .	120
5.23	Monte Carlo Simulations Error-Corrupted FDN, 5 dB SNR	121
5.24	Monte Carlo Simulations Error-Corrupted fGN, 30 dB SNR . . .	122
5.25	Monte Carlo Simulations Error-Corrupted fGN, 15 dB SNR . . .	123
5.26	Monte Carlo Simulations Error-Corrupted fGN, 10 dB SNR . . .	124
5.27	Monte Carlo Simulations Error-Corrupted fGN, 5 dB SNR	125

List of Symbols

$[\cdot]^*$	Vector or matrix conjugate transpose
$[\cdot]^T$	Vector or matrix transpose
$\stackrel{D}{=}$	Equality of the finite-dimensional distributions
$\mathbf{E}[\cdot]$	Expectation operator
\mathbb{J}	Finite set of non-negative integers, i.e., $\mathbb{J} = \{0, 1, \dots, N - 1\}$
$\hat{\theta}_{\text{CE}}$	CE Estimates of the model parameters
$\hat{\theta}_{\text{DFA}}$	DFA estimate of the LRD coefficient
$\hat{\theta}_{\text{MLE}}$	MLE of the model parameters
$\hat{\Theta}_{\text{OLS}}$	Ordinary Least Squares solution to the linear regression problem
$\hat{\theta}_{\text{PSD}}$	Log-periodogram regression estimate of the LRD coefficient
$\hat{\theta}_{\text{R/S}}$	R/S estimate of the LRD coefficient
$\hat{\theta}_{\text{WE}}$	Whittle's Estimates of the model parameters
$\ \cdot\ _{\text{HS}}$	Hilbert-Schmidt norm
\mathbb{Z}	Set of all integers, i.e., $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$
∇^d	Fractional differencing operator
Π	Forward-shift permutation matrix
\mathbb{Z}^+	Set of all non-negative integers, i.e., $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$
\mathbb{R}	Set of real numbers
\triangleq	Equal by definition
$P \sim Q$	Asymptotic equivalent matrices
q^{-1}	Standard backward shift operator
$u = \exp\left(\frac{i2\pi}{N}\right)$	Root of unity

List of Acronyms

ARFIMA Autoregressive Fractionally Integrated Moving Average

ARMA Autoregressive Moving Average

BIBO Bounded Input Bounded Output

CE Circulant Embedding

CEM Circulant Embedding Method

DFA De-trended Fluctuation Analysis

DFT Discrete Fourier Transform

FARIMA Fractional Autoregressive Integrated Moving Average

FFT Fast Fourier Transform

fBM fractional Brownian Motion

fGN fractional Gaussian Noise

FD Fractionally Differenced

FDN Fractionally Differenced Noise

iid independent identically distributed

LTI Linear Time-Invariant

LRD Long-range dependence

MLE Maximum Likelihood Estimate

MSE Mean Squared Error

OLS Ordinary Least Squares

PEM Prediction Error Methods

R/S Rescaled Adjusted Range

H-ss self-similar

SRD Short-range dependence

SNR signal-to-noise-ratio

STD Standard Deviation

si stationary increments

WE Whittle's Estimate

Abstract

A set of observations from a random process which exhibit correlations that decay slower than an exponential rate is regarded as long-range dependent. This phenomenon has stimulated great interest in the scientific community as it appears in a wide range of areas of knowledge. For example, this property has been observed in data pertaining to electronics, econometrics, hydrology and biomedical signals.

There exist several estimation methods for finding model parameters that help explain the set of observations exhibiting long-range dependence. Among these methods, maximum likelihood is attractive, given its desirable statistical properties such as asymptotic consistency and efficiency. However, its computational complexity makes the implementation of maximum likelihood prohibitive.

This thesis presents a group of computationally efficient estimators based on the maximum likelihood framework. The thesis consists of two main parts. The first part is devoted to developing a computationally efficient alternative to the maximum likelihood estimate. This alternative is based on the circulant embedding concept and it is shown to maintain the desirable statistical properties of maximum likelihood.

Interesting results are obtained by analysing the circulant embedding estimate. In particular, this thesis shows that the maximum likelihood based methods are ill-conditioned; the estimators' performance will deteriorate significantly when the set of observations is corrupted by errors. The second part of this thesis focuses on developing computationally efficient estimators with improved performance under the presence of errors in the observations.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the Copyright) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the Intellectual Property) and any reproductions of copyright works in the thesis, for example graphs and tables (Reproductions), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf>), in any relevant Thesis restriction declarations deposited in the University Library, The University Librarys regulations (see <http://www.manchester.ac.uk/library/aboutus/regulations>) and in The Universitys policy on presentation of Theses

Publications

Sections of this thesis have been reported in:

- (1) O. Vivero and W. Heath, “On MLE methods for dynamical systems with fractionally differenced noise spectra”. Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 28th Chinese Control Conference. Shanghai, China 2009.

General Chair’s Recognition Award for Interactive Papers

- (2) O. Vivero and W. Heath, “Regularised Estimators for fractionally differenced processes”. Accepted for the 4th IFAC Workshop on Fractional Differentiation, Spain, 2010.
- (3) O. Vivero and W. Heath, “Regularised Estimators for fractional Gaussian Noise”. Accepted for the 49th IEEE Conference on Decision and Control, Atlanta, USA, 2010.
- (4) O. Vivero and W. Heath, “A Regularised Estimator for Long-Range Dependent Processes”. Provisionally accepted as a regular paper in Automatica.

Acknowledgements

There are a lot of people that in one way or another helped me in completing this thesis. There is not enough space to thank you all, however, I would like to highlight the following people.

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. William P. Heath. I have learnt a lot from Will during the past few years. His guidance, encouragement and generosity with his time have been invaluable. His sincere and enthusiastic approach to research have shaped my perception of what the whole ordeal is about.

I would like to thank my colleagues at the Control Systems Centre. Thank you for the useful discussions and the healthy environment to work in.

My deepest thanks to Conor Leahy and Luis-Diaz Santana for sharing the data on ocular accommodation.

I would also like to thank my parents, Oscar and Margarita, my brother, Edgar and my grandma Elda. Without your continuous support, encouragement, love and terrific sense of humour I wouldn't be where I am today.

I am grateful with my friends Caleb, Carmen, Cinthya, Francisco and Martin. Thank you for sharing the great moments and for helping me through the hard ones. No matter where destiny takes us in the future, know that our friendship is one of the things I treasure the most.

*To my mum,
my dad
and my brother.*

Chapter 1

Introduction

Several physiological mechanisms have a common purpose: to maintain the interior of a biological entity stable in the face of stress. This concept was later extended into what is termed today as *homeostasis*, that is, the tendency towards a relatively stable equilibrium between interdependent elements [15]. This definition of homeostasis is highly related to what is known in the control systems theory literature as Bounded Input Bounded Output (BIBO) stability [54], that is, in some sense a small variation in the system will produce a variation in a steady fashion.

However, as pointed out by Buchman [16] a healthy biological system exhibits a high level of variability. This behaviour is inherently different to the concept of steady state for control systems theory. The irregular behaviour of normal (healthy) physiology might be explained by the set of interactions that are present in such systems [15]. It is due to such levels of networked interactions that these systems are regarded as highly complex. This high level of complexity might be represented by nonlinear dynamical systems [16, 31, 107].

An alternative way to represent such high variability events is through stochastic processes which exhibit long-range dependence [16]. In an informal fashion, a long-range dependent stochastic process might be described as one whose observations are highly correlated regardless of the distance between them [9]. That is to say that the correlation function of the stochastic process decays at a very slow rate. This behaviour is unusual for most stochastic processes [90]. For example, common linear processes such as Autoregressive Moving Average (ARMA) processes and finite state Markov chains lead to exponentially fast decaying correlation functions.

There is empirical evidence of the existence of stochastic processes with slowly decaying correlations in various fields of study, such as, economics [81, 98], communications [63, 61, 82], electronics [57], physiology [38, 92], hydrology [59, 73] and image texture recognition [67].

The “unusual” behaviour of long-range dependent processes has many statistical implications. For instance, a set of observations of a stationary process whose correlation structure decays slower than exponentially fast is reminiscent of a non-stationary process [70, 90].

Another important implication is that the correlation structure of long-range dependence (LRD) implies that the correlations stop being absolutely summable. This has a direct effect on the rate of convergence of the strong law of large numbers for long-range dependent processes. For instance, processes whose correlations decay exponentially fast are guaranteed to have a convergence rate of order \sqrt{N} , whereas LRD processes are not. This point is further illustrated in example 2.2.6.

In terms of modelling, it is not a simple task to construct a process whose correlation function decays to zero at a slower than exponentially fast rate. Keshner [57] proposed employing high order lumped linear systems in order to approximate long-range dependence (see example 1.0.1). A parsimonious approach on representing long-range dependence is to construct a model from the frequency domain. Briefly, a finite variance, stationary process whose spectral density function diverges to infinity at a certain rate has a correlation function that decays at an appropriate slow rate [90]. A formal and precise description of such model construction is presented in chapter 2.

Example 1.0.1. Consider the simple linear system of a finite resistor capacitor (RC) transmission line in figure 1.1. Keshner [57] showed that as the length of the transmission line grows large, its spectral density function takes the form $f(\omega) = 1/\omega^\theta$ for some finite nonzero θ . This system might be represented by a high order linear model with a finite amount of state variables. Each state variable contributes a pole and a zero to the overall frequency response of the model (see figure 1.2). In order to properly summarise the spectral density function (and equivalently the correlation function), the lumped linear system would require hundreds of state variables [57]. The task of estimating hundreds of parameters to represent such behaviour is daunting. A more parsimonious approach is attained by the use of LRD models capable of representing such behaviour.

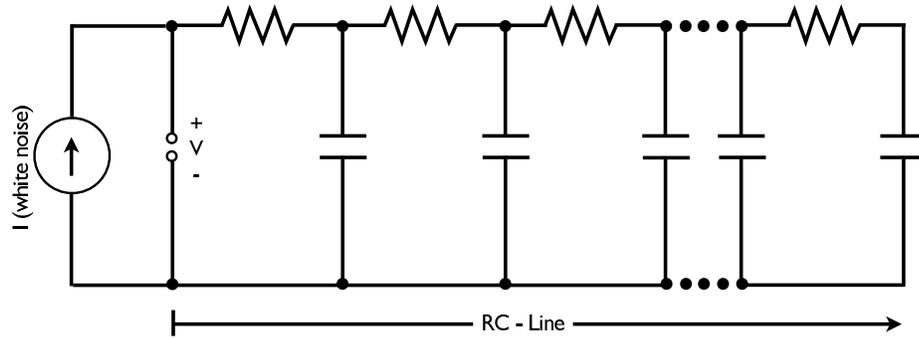


Figure 1.1: Representation of a continuous resistor-capacitor transmission line excited by a white noise current source.

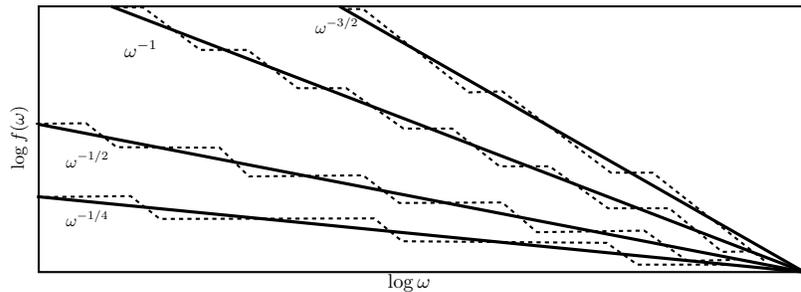


Figure 1.2: Finite dimensional linear system approximation to the spectral density function $\omega^{-\theta}$. The segmented lines represent the spectral density functions of the finite dimensional linear systems. The continuous line represents the spectral densities $\omega^{-\theta}$.

The estimation of the parameters of stochastic processes exhibiting long-range dependence is not a trivial task. This is due to the complexity of the statistics of the models portraying long-range dependence and due to the computational difficulties involved in the implementation of traditional estimation methods. There are several alternatives available in the literature. However, they are either under performing in terms of the convergence of the estimates or difficult to implement.

Maximum likelihood is a method with desirable statistical properties even under long-range dependence. Nevertheless, it has been reported as a methodology that is difficult to implement as it is computationally expensive [see for example 79]. Furthermore, under long-range dependence, the performance of maximum likelihood under measurement errors deteriorates.

This thesis is concerned with the estimation of stochastic processes which exhibit long-range dependence. In particular, we are concerned in finding the set of parameter values that makes the occurrences of a set of observations exhibiting

LRD more likely. The goal of this thesis is to address the computational implementation problem of maximum likelihood while achieving an estimator that produces accurate results under error-free and error-corrupted measurements scenarios.

1.1 Thesis Overview

Chapter 2: Long-Range Dependence

This chapter surveys the concept of long-range dependence, the most renowned models available and a method for generating Gaussian realisations with long-range dependence. Section 2.2 provides a formal mathematical definition of long-range dependence and its properties. Section 2.3 surveys the fractional Brownian Motion, fractional Gaussian Noise and fractionally differenced models. The similarities and differences between these models along with their properties are discussed. Section 2.4 surveys a computationally efficient method for generating exact realisations of Gaussian long-range dependent processes.

Chapter 3: Estimation of Long-Range Dependent Models

This chapter surveys several semi-parametric and parametric methods for estimating the parameters of the long-range dependent models presented in chapter 2. The advantages and disadvantages of the semi-parametric and parametric approaches are discussed. The properties of the parametric methods are presented along with a discussion of their implementation.

Chapter 4: Regularised Estimators

This chapter introduces a group of parametric estimators geared towards fitting the long-range dependent models of chapter 2 under the presence of errors in the measurements.

Chapter 5: Numerical Results

This chapter evaluates some of the available methods for estimating the parameters of the long-range dependent models presented in chapter 2. These methods are evaluated for several sample lengths and a wide range of parameter values of the long-range dependent coefficient. An evaluation between the parametric

methods of chapter 3 and 4 is provided for errors at different levels. Furthermore, the performance of the estimators is illustrated briefly on data from two real-world examples.

Chapter 6: Conclusions and Future Work

This chapter presents the general concluding remarks of the thesis and provides comments on future work.

1.2 Thesis Contributions

Circulant Embedding Estimator

The Circulant Embedding estimator presented in chapter 3, section 3.3.3 is novel. We prove its asymptotic properties and discuss its advantages and disadvantages in comparison to other estimation methods.

Ill-conditioning of the likelihood based methods

In chapter 3, section 3.3.3, we prove that the likelihood based estimators are ill-conditioned. This fact has been discussed briefly in the literature. However, there is no prior formal proof that the estimators are ill-conditioned to the best of the thesis author's knowledge.

Regularised Estimators

In the light of the ill-conditioning of the likelihood based methods a set of novel regularised estimators are proposed in chapter 4. These estimators are designed to correct the bias produced by errors in the data observations. The chapter also provides a novel approach for choosing the regularisation parameter. Under further restrictions it is shown that the regularised estimates are optimal in the sense of maximum likelihood. Finally, an estimate of the signal-to-noise ratio of the data is provided.

Along with the three main contributions, this thesis provides in chapter 2, section 2.4, conjectures on the non-negative definiteness of the Circulant Embedding Matrix for ARFIMA covariances. If proven, this result would ensure that a method for producing fast and exact synthetic realisations of Gaussian processes

can be extended to ARFIMA models. Furthermore, it would provide the user with an indication of the applicability of the Circulant Embedding estimator for ARFIMA processes with relatively small datasets.

Chapter 2

Long-Range Dependence

2.1 Introduction

Scientists in diverse fields have often observed that far apart correlations from a set of observations decay to zero at a much slower rate than one would expect from independent data or data that follows the traditional Markov-type or ARMA models. This phenomenon is important as it appears in areas such as geophysics, hydrology, turbulence, economics, electronics and communications among others. The processes that exhibit this correlation behaviour are referred to as long-range dependent (LRD) or long-memory processes. These processes manifest as the presence of cycles of any order in the observations. This is widely known as the *Joseph effect*, a term coined by Mandelbrot and Wallis [70] based on:

“Seven years are coming, during which there will be a great surplus of food all over Egypt. These will be followed by seven years of famine, when all the surplus in Egypt will be forgotten. The famine will ravage the land.” Genesis 41, 29-30.

Such behaviour became apparent from Hurst’s studies in 1951 of the Nile river discharge time-series, where he observed that there is a tendency of wet years to cluster into wet periods and dry years to cluster into drought periods.

Processes with correlations that decay at such slow rates often exhibit a divergence to infinity on their spectral density function [90]. This behaviour is akin to what is often referred to as “inverse frequency” or flicker noise [57, 68, 96]: terms used to describe a stochastic process whose sample spectral density or periodogram is of the form $f(\omega) = |\omega|^{-\theta}$ for some finite nonzero θ .

This chapter surveys the concept of long-range dependence. The contents are as follows: section 2.2 provides a formal definition of long-range dependence; section 2.3 surveys the most renowned models designed to specifically portray long-range dependence properties; section 2.4 presents an exact and efficient method to compute properly a synthetic realisation of a long-range dependent process.

Hereafter, $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ denotes the set of all integers, \mathbb{Z}^+ denotes the set of all non-negative integers, $\mathbb{J} = \{0, 1, 2, \dots, N-1\}$ denotes a finite set of non-negative integers and \mathbb{R} denotes the set of all reals.

2.2 Definitions of Long Range Dependence

The work throughout this thesis is focused in the context of weakly stationary processes. Formally, they are defined as follows:

Definition 2.2.1. Let $X(t)$, $t \in \mathfrak{J}$ (where \mathfrak{J} denotes an index set, i.e. $\mathfrak{J} = \mathbb{R}$ if $X(t)$ is continuous or $\mathfrak{J} = \mathbb{Z}$ if $X(t)$ is discrete) be a stochastic process with finite first and second moments

$$\begin{aligned}\mu(t) &= \mathbf{E}[X(t)] \\ \gamma(t, s) &= \mathbf{E}[X(t) - \mathbf{E}X(t)][X(s) - \mathbf{E}X(s)]^*\end{aligned}$$

where $\mathbf{E}[\cdot]$ denotes the expectation operator and $[\cdot]^*$ is the conjugate transpose. If the first and second moments of $X(t)$ are invariant to time-shifts, that is,

$$\begin{aligned}\mu(t) &= \mu(t + \tau) \\ \gamma(t, s) &= \gamma(t + \tau, s + \tau)\end{aligned}$$

for any τ such that $t + \tau, s + \tau \in \mathfrak{J}$, then $X(t)$ is said to be a *weakly stationary* process.

As mentioned in the previous chapter, there is empirical evidence in various fields of study where a set of observations exhibits a slow decay on its correlations. Hence, it seems natural to formally define long-range dependence in terms of the behaviour of the covariance function of a stochastic process. However, it is important to note that there are several, not always equivalent, definitions of long-range dependence available in the literature. A comprehensive review of the available definitions of LRD can be found in [9, 80, 90, 100]. For the remainder of the thesis, definition 2.2.2 is sufficient.

Definition 2.2.2. The weakly stationary process $X = X_k$, $k \in \mathbb{Z}$, is said to be *long-range dependent* if its covariance function,

$$\gamma(\tau) = \mathbf{E}[X_k X_{k+\tau}] , \quad (2.1)$$

is such that

$$\lim_{\tau \rightarrow \infty} \frac{\gamma(\tau)}{\mathcal{A} \tau^{-\theta_{\text{LRD}}}} = 1 \quad (2.2)$$

for some finite, non-zero constant \mathcal{A} and $\theta_{\text{LRD}} \in (0, 1)$.

Remark 2.2.3. It is important to note that definition 2.2.2 is more restrictive than the available definitions in the literature. In general, the literature employs slowly varying functions at infinity instead of the finite, non-zero constant \mathcal{A} . In such cases, the spectral density function of the process is not always guaranteed to exist [46].

Of particular interest is to analyse the implications of definition 2.2.2. It follows that under this definition, the covariances, and thus the correlations, of X are not absolutely summable. Furthermore, the spectral density function of X has a pole in the origin. This is formalised in theorem 2.2.4.

Theorem 2.2.4. *Let $X = X_k$, $k \in \mathbb{Z}$, be a weakly stationary process exhibiting LRD with parameter $\theta_{\text{LRD}} \in (0, 1)$. Then:*

(a) *The covariance function of X is not absolutely summable, that is*

$$\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| = \infty . \quad (2.3)$$

(b) *The spectral density function of X ,*

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i\tau\omega) , \quad (2.4)$$

is such that

$$\lim_{\omega \rightarrow 0} \frac{f(\omega)}{\mathcal{B} |\omega|^{\theta_{\text{LRD}}-1}} = 1 \quad (2.5)$$

for some finite, non-zero constant \mathcal{B} .

Proof. The proofs are standard results, see [80, 90, 100] and the references within. \square

Remark 2.2.5. Point (b) in theorem 2.2.4 states that at low frequencies the spectral density function of a LRD process behaves as $|\omega|^{1-\theta_{\text{LRD}}}$. This implies that at the origin, the spectral density diverges as a power law, and thus a log-log chart of the spectral density function will appear as a straight line with slope $\theta_{\text{LRD}} - 1$. A large part of the available literature defines LRD based on this property, and commonly refers to them as “one over frequency” or “inverse frequency” processes.

Example 2.2.6. One of the reasons to concentrate on the lack of summability of correlations is the interest in understanding the order of magnitude of the convergence of the strong law of large numbers for processes exhibiting long-range dependence. To illustrate, let X_k , $k \in \mathbb{J}$, be a set of observations of a zero-mean process with finite variance, covariance function $\gamma(\tau)$ and partial sums

$$S_x = \frac{1}{N} \sum_{k=1}^N X_k .$$

Suppose X_k is such that

$$\sum_{t=0}^{N-1} \sum_{s=0}^{N-1} \gamma(|t-s|) \leq \mathcal{C}N^\Omega$$

for some $\mathcal{C} < 0$ and $\Omega \in (0, 1)$. Then the result of Ninness [78] shows that for any $\alpha > \Omega/2$

$$\frac{1}{N^\alpha} \sum_{k=0}^{N-1} X_k \xrightarrow{\text{a.s.}} 0 \tag{2.6}$$

as $N \rightarrow \infty$, where a.s. means almost surely. To illustrate, we look at the following two examples:

(a) Suppose

$$\lim_{\tau \rightarrow \infty} \frac{\gamma(\tau)}{\exp(-\lambda\tau)} = 1 .$$

with $\lambda > 0$. Then we have

$$\begin{aligned} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} \gamma(|t-s|) &\leq \mathcal{C}N \sum_{\tau=0}^{N-1} \gamma(\tau) \\ &= \mathcal{C}N \int_0^N \exp(-\lambda\tau) d\tau \\ &\leq \frac{1}{\lambda} \bar{\mathcal{C}}N . \end{aligned}$$

Thus, for the case of exponentially fast decaying covariance functions, equation (2.6) implies that the rate of convergence of S_x is of at least \sqrt{N} .

(b) Suppose X_k is LRD such that $\gamma(|t-s|)$ is as in equation (2.2). Then

$$\begin{aligned} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} \gamma(|t-s|) &\leq \mathcal{C}N \sum_{\tau=0}^{N-1} \gamma(\tau) \\ &\leq \bar{\mathcal{C}}N \int_0^N \tau^{-\theta_{\text{LRD}}} d\tau \\ &= \frac{1}{1-\theta_{\text{LRD}}} \bar{\mathcal{C}}N^{2-\theta_{\text{LRD}}} . \end{aligned}$$

Thus, equation (2.6) implies that for LRD processes, the convergence of S_x is not longer guaranteed to be of \sqrt{N} .

2.3 Models with LRD

This section presents the most renowned finite variance models that exhibit long-range dependence along with most of their properties. All of the weakly stationary models presented in this section satisfy equation (2.2) and thus are compliant with definition 2.2.2.

2.3.1 Fractional Brownian Motion

In the 1960's, Mandelbrot suggested using models in order to represent processes exhibiting a slowly decaying covariance function. In [69], he introduced the now widely known fractional Brownian Motion (fBM) model. The discovery of fBM was by Kolmogorov in his work on the spiral of Wiener and phenomenological theory of turbulence for large Reynolds numbers [75]. However, it was Mandelbrot

who pointed out its relevance to more general applications and who developed many of its properties along with Van Ness in [69] and many other subsequent publications.

Definition 2.3.1. A Gaussian process $B_H(t)$, with coefficient $H \in (0, 1)$ and $t, \tau \in \mathbb{R}$, is called a fractional Brownian Motion if

$$B_H(\Omega t) \stackrel{D}{=} \Omega^H B_H(t) \quad (2.7)$$

for any $\Omega > 0$ and

$$B_H(t + \tau) - B_H(\tau) \stackrel{D}{=} B_H(t) - B_H(0) \quad (2.8)$$

where $\stackrel{D}{=}$ denotes equality of the finite-dimensional distributions [100].

Equation (2.7) and (2.8) denote that fBM is a self-similar (H-ss) process with coefficient H that possesses stationary increments (si), that is fBM is H-sssi. Fractional Brownian Motion is a process that has been widely studied in the literature. Comprehensive reviews can be found in [9, 100]. Theorem 2.3.2 summarises most of the properties of fBM.

Theorem 2.3.2. *Let a process $B_H = B_H(t)$, $s, t \in \mathbb{R}$, be fBM as in definition 2.3.1, and let $\beta = [H, \nu^2]$, then*

- (a) $B_H(0) = 0$ almost surely.
- (b) $\mathbf{E}[B_H] = 0$ for all $t \in \mathbb{R}$.
- (c) $B_H(-t) = -B_H(t)$.
- (d) The variance of fBM is given by

$$\mathbf{E}[B_H^2] = |t|^{2H} \nu^2 . \quad (2.9)$$

- (e) The covariance function of fBM, $\gamma_H(s, t)$, is given by

$$\begin{aligned} \gamma_\beta(s, t) &= \mathbf{E}[B_H(t)B_H(s)] \\ &= \frac{\nu^2}{2} [|t|^{2H} + |s|^{2H} - |t - s|^{2H}] . \end{aligned} \quad (2.10)$$

Proof. See appendix 2.A. □

Remark 2.3.3. Given the properties of fBM in theorem 2.3.2 we have

- (a) fBM is a zero mean non-stationary Gaussian process.
- (b) When $H = 1/2$, fBM reduces to Brownian Motion.
- (c) If the process is H-ss, then its finite dimensional distributions on the positive real line are completely determined by those on any interval of finite length [100].

2.3.2 Fractional Gaussian Noise

While fBM is a non-stationary process (see theorem 2.3.2), its increments are stationary. This makes them very useful in terms of parameter estimation. We refer to these stationary increments as fractional Gaussian Noise (fGN) and define them as follows.

Definition 2.3.4. The process X_k , $k \in \mathbb{Z}$, is said to be fGN if there exists $B_H(\cdot)$ such that

$$X_k \triangleq B_H(k+1) - B_H(k) \quad (2.11)$$

where $B_H(\cdot)$ is given by definition 2.3.1.

Along with fBM, fractional Gaussian Noise has been widely studied in the literature. Comprehensive reviews can be found in [9, 80, 100]. Theorem 2.3.5 summarises the properties of fGN.

Theorem 2.3.5. *Let a process X_k , $k, \tau \in \mathbb{Z}$, be fGN as in definition 2.3.4 and let $\beta = [H, \nu_2]$, then*

- (a) $\mathbf{E}[X_k] = 0$
- (b) $\mathbf{E}[X_k^2] = \mathbf{E}[B_H^2(1)] = \nu^2$
- (c) *The covariance function of X_k is given by*

$$\begin{aligned} \gamma_\beta(k) &= \mathbf{E}[X_\tau X_{\tau+k}] \\ &= \frac{\nu^2}{2} (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) . \end{aligned} \quad (2.12)$$

- (d) *Let $k \neq 0$. Then $\gamma_\beta(k) = 0$ if $H = 1/2$, $\gamma_\beta(k) < 0$ if $H \in (0, 1/2)$, and $\gamma_\beta > 0$ if $H \in (1/2, 1)$.*

(e) The spectral density function of fGN is given by

$$f_{\beta}(\omega) = \frac{2\nu^2(1 - \cos \omega)}{\pi[H\Gamma(2H) \sin H\pi]^{-1}} \sum_{k=-\infty}^{\infty} |2\pi k + \omega|^{-2H-1} . \quad (2.13)$$

Proof. See appendix 2.A. □

Remark 2.3.6. Given the properties of fGN in theorem 2.3.5 we have:

(a) If $H \neq 1/2$ then [see p. 11 in 100] the covariance function of fGN behaves asymptotically as

$$\lim_{k \rightarrow \infty} \frac{\gamma_{\beta}(k)}{\nu^2 H(2H - 1)|k|^{2H-2}} = 1 . \quad (2.14)$$

Note that this behaviour is as defined in 2.2.2.

(b) If $\omega \rightarrow 0$ the spectral density function is given by [See corollary 2.1, p. 53 in 9]

$$f_{\beta}(\omega) = \nu^2 U_H |\omega|^{1-2H} + O(|\omega|^{\min(3-2H, 2)}) \quad (2.15)$$

with

$$U_H = (2\pi)^{-1} \sin(H\pi)\Gamma(2H + 1) \quad (2.16)$$

where $\Gamma(\cdot)$ denotes the gamma function. Note that the behaviour of the spectral density function of fGN at small frequencies is as defined in equation (2.5).

- (c) If $H \in (0, 1)$, fGN is a weakly-stationary, invertible (i.e., the errors between the outputs of the process and a set of forecasts tends to zero [43, 56]) Gaussian process [80, 100]. If $H \in (1/2, 1)$ the process exhibits LRD. If $H \in (0, 1/2)$ the process is anti-persistent (i.e., all $\tau \neq 0$ correlations are negative).
- (d) There is no known closed form solution for the spectral density of fGN. The vexing infinite sum in (2.13) makes the spectral density function difficult to compute. An approach is to directly truncate the terms in the sum, such as in [9], where only 200 elements are added. A different approximation is provided by Paxson [82]. This approximation consists in truncating the midpoint infinite sum between the upper and lower bounds of the spectral

density. The approximation will improve arbitrarily provided the amount of terms included within the truncation is increased. In particular, [82] uses a level of truncation such that the spectral density may be approximated by

$$\begin{aligned} \tilde{f}_\beta(\omega) &= \mathcal{U}(\omega, H) [|\omega|^{-2H-1} + \mathcal{V}(\omega, H)] \\ &\approx f_\beta(\omega) \end{aligned}$$

with

$$\begin{aligned} \mathcal{U}(\omega, H) &= 2 \sin(\pi H) \Gamma(2H + 1) (1 - \cos \omega) \\ \mathcal{V}(\omega, H) &= g_1^{-2H-1} + p_1^{-2H-1} + g_2^{-2H-1} + p_2^{-2H-1} + g_3^{-2H-1} + p_3^{-2H-1} \\ &\quad + \frac{g_3^{-2H} + p_3^{-2H} + g_4^{-2H} + p_4^{-2H}}{8H\pi} \end{aligned}$$

and

$$\begin{aligned} g_k &= 2k\pi + \omega \\ p_k &= 2k\pi - \omega . \end{aligned}$$

[82] reports a performance that diverges slightly to the method of [9], however there is an improved computational performance.

2.3.3 Fractionally Differenced Noise

Granger and Joyeux [44] and Hosking [50] developed the concept of Fractionally Differenced Noise (FDN) independently in the 1980's. The properties of this model draw several parallels with those of fGN, particularly, the behaviour of the covariance and spectral density functions at large lags and low frequencies respectively. In terms of parameter estimation, the Fractionally Differenced (FD) family of models is favoured over fGN due to its flexibility and by the simplicity of the computation of the spectral density function [63].

Definition 2.3.7. The process X_k , $k \in \mathbb{Z}$, is said to be FDN with fractional dimension d if

$$e_k \triangleq \nabla^d X_k \tag{2.17}$$

is a sequence of independent identically distributed (iid) random variables with

variance $\nu_e^2 < \infty$ and ∇^d is a linear operator given by

$$\begin{aligned}\nabla^d &= (1 - q^{-1})^d \\ &= \sum_{j=0}^{\infty} \binom{d}{j} (-q^{-1})^j \\ &= 1 - dq^{-1} - \frac{1}{2}d(1-d)q^{-2} - \frac{1}{6}d(1-d)(2-d)q^{-3} - \dots\end{aligned}\tag{2.18}$$

with q^{-1} being the standard backward shift operator such that

$$X_{k-\tau} = q^{-\tau} X_k ,\tag{2.19}$$

and $\tau \in \mathbb{Z}^+$.

Fractionally differenced noise has been widely studied in the literature. Theorem 2.3.8 is due to [50] and summarises the properties of FDN.

Theorem 2.3.8. *Let a process X_k , $k, \tau \in \mathbb{Z}$, be FDN with index $d \in (-1/2, 1/2)$ as in definition 2.3.7 and let $\beta = [d, \nu_e^2]$ then*

(a) *The spectral density of X_k is given by*

$$f_\beta(\omega) = \frac{\nu_e^2}{2\pi[2\sin(\omega/2)]^{2d}} .\tag{2.20}$$

with $\omega \in [-\pi, \pi]$ frequencies.

(b) *The covariance function of X_k is given by*

$$\begin{aligned}\gamma_\beta(k) &= \mathbf{E} [x_\tau x_{\tau+k}] \\ &= \frac{\nu_e^2}{\sqrt{\pi}} \left[\frac{\Gamma(1/2 - d)\Gamma(k + d)}{4^d \Gamma(d)\Gamma(1 + k - d)} \right] .\end{aligned}\tag{2.21}$$

Here, $\Gamma(\cdot)$ denotes the gamma function.

Proof. See appendix 2.A. □

Remark 2.3.9. Given the properties of FDN in theorem 2.3.8 we have

(a) If $d \neq 0$, we have by the limit of the ratio of gamma functions [see eq. 6.1.46, p. 257 in 1]

$$\lim_{k \rightarrow \infty} \frac{\gamma_\beta(k)}{\nu_e^2 U_d k^{2d-1}} = 1\tag{2.22}$$

and

$$U_d = \frac{1}{\sqrt{\pi}} \left[\frac{\Gamma(1/2 - d)}{4^d \Gamma(d)} \right]. \quad (2.23)$$

The behaviour of the covariance function of FDN is as defined in 2.2.2.

(b) If $d \neq 0$, then we have [50]

$$\lim_{\omega \rightarrow 0} \frac{f_\beta(\omega)}{|\omega|^{-2d}} = 1. \quad (2.24)$$

The behaviour of the spectral density function of FDN is as defined in equation (2.5).

- (c) From equations (2.14) and (2.22) we have that as the lags of the covariance functions of fGN and FDN grow large, the LRD coefficients, d and H are related by $H = d + 1/2$.
- (d) FDN is asymptotically self-similar [50].
- (e) If $d \in (-1/2, 1/2)$, FDN is a weakly-stationary, invertible, Gaussian process [50]. Particularly, if $d \in (0, 1/2)$ then FDN exhibits LRD since its covariance is not absolutely summable. If $d \in (-1/2, 0)$ the process is anti-persistent.
- (f) It is not necessary that the innovations e_k be Gaussian. The only requirement for FDN to be well defined is that the mean and variance are finite [100].

2.3.4 ARFIMA Processes

The FDN and fGN processes excel at modelling long memory behaviour of a process, i.e., the LRD coefficient θ_{LRD} can be represented by either d or H respectively. Nevertheless, there is a disadvantage inherited from the rigidity of their covariance, that is, the subtleties of the short-term correlations, or short-range dependence (SRD), cannot be addressed by means of these models [73]. Hosking [50] addressed this issue by extending the framework of FDN and ARMA processes into what is known as Autoregressive Fractionally Integrated Moving Average (ARFIMA), or Fractional Autoregressive Integrated Moving Average (FARIMA) processes. The ARFIMA model provides more flexibility as it retains the eventual hyperbolic decay of the correlations while encompassing a wider

range of short term behaviour. Figures 2.1 and 2.2 illustrate the differences between a pure LRD model and the ARFIMA model which incorporates both LRD and SRD.

Definition 2.3.10. Let a process X_k , $k \in \mathbb{Z}$, be FDN with parameter d as in definition 2.3.7. Then the process Y_k is said to be ARFIMA(n_a, d, n_b) if

$$A(q)Y_k \triangleq B(q)X_k, \quad (2.25)$$

where

$$\begin{aligned} A &= A(q) = 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{n_a}q^{-n_a} \\ B &= B(q) = 1 + b_1q^{-1} + b_2q^{-2} + \dots + b_{n_b}q^{-n_b} \end{aligned} \quad (2.26)$$

are finite dimensional polynomials with no common roots.

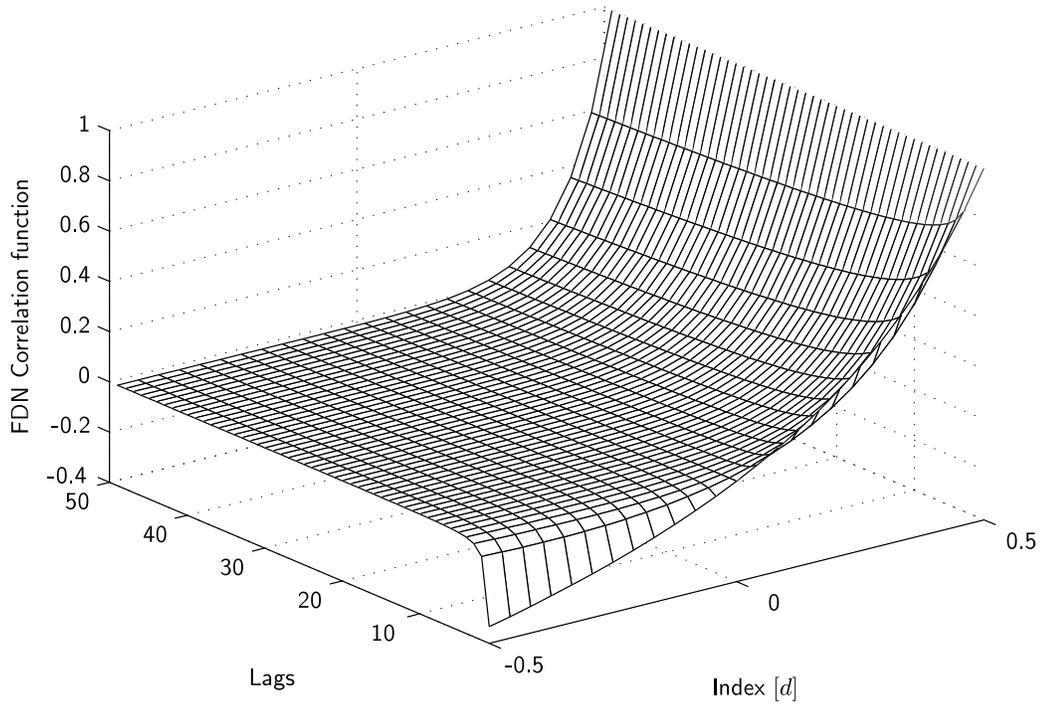
The ARFIMA process can be understood as an ARMA process with FDN innovations, i.e., a sequence of FDN that has been filtered by a rational transfer function $B(q)/A(q)$. Thus, if $A(q) = B(q)$ then we have an ARFIMA(0, d , 0) process which is equal to FDN.

Remark 2.3.11. We have that

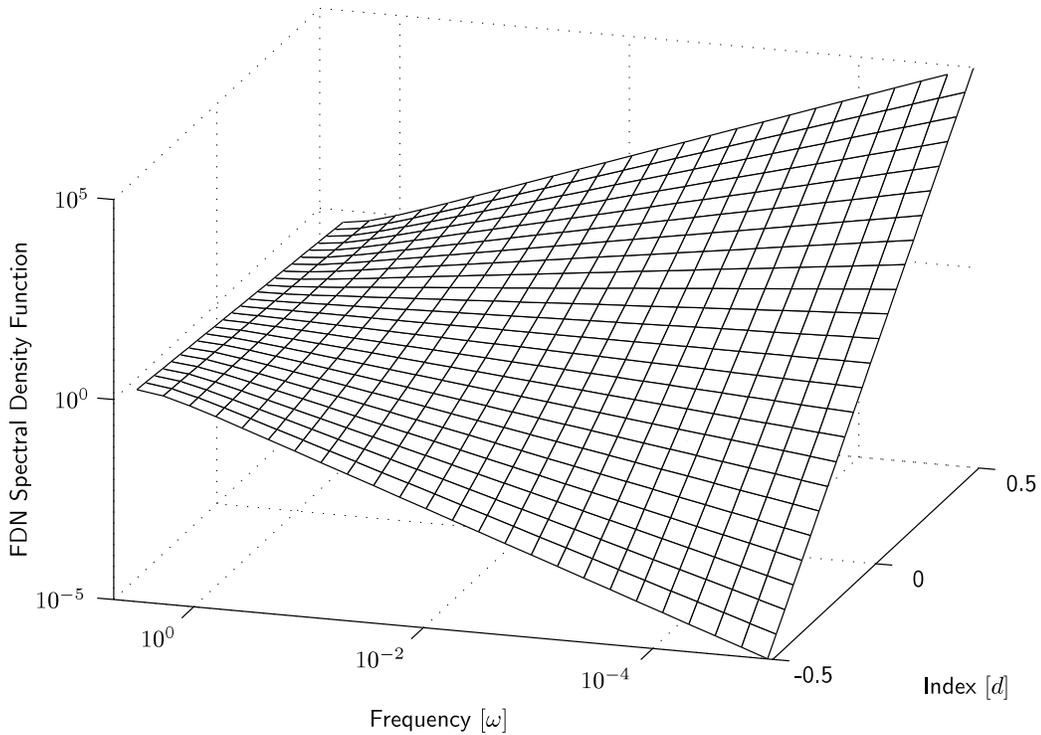
- (a) if $B(q)/A(q)$ is analytic outside the unit circle and $d \in (-1/2, 1/2)$ then the ARFIMA(n_a, d, n_b) process is invertible, causal and stationary [see Theorem 2, p. 170 in 50].
- (b) An explicit expression for the covariance of an ARFIMA process is difficult to find [9]. Nevertheless, it is possible to decompose its covariance function into its ARMA and FDN contributions by means of a convolution, i.e.,

$$\begin{aligned} \gamma_\beta(\tau) &= \gamma_{\text{ARMA}}(\tau) * \gamma_{\text{FDN}}(\tau) \\ &= \sum_{k=-\infty}^{\infty} \gamma_{\text{ARMA}}(k) \gamma_{\text{FDN}}(k - \tau) \end{aligned} \quad (2.27)$$

with $\beta = [d, A, B, \nu_e^2]$, $\gamma_{\text{ARMA}}(\cdot)$ the covariance function of an ARMA process with unity innovations variance and polynomials $A(q)$ and $B(q)$. Furthermore, $\gamma_{\text{FDN}}(\cdot)$ is the covariance function of FDN given by (2.21). The covariance of the ARMA process can be found, for example, by solving the

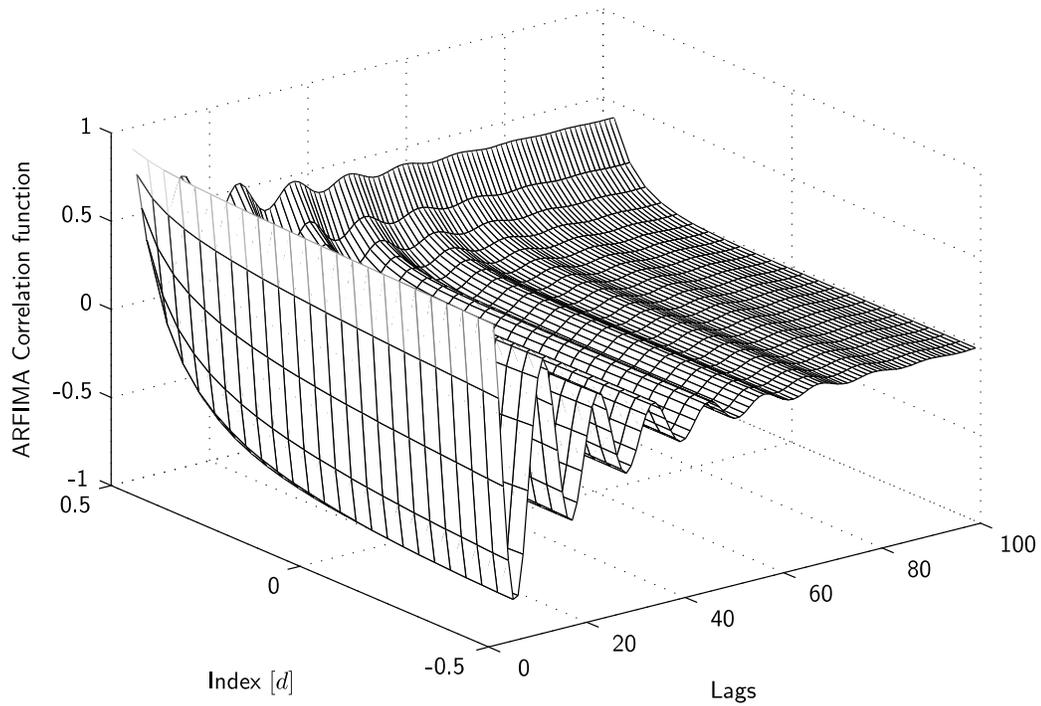


(a) Correlation Function of FDN

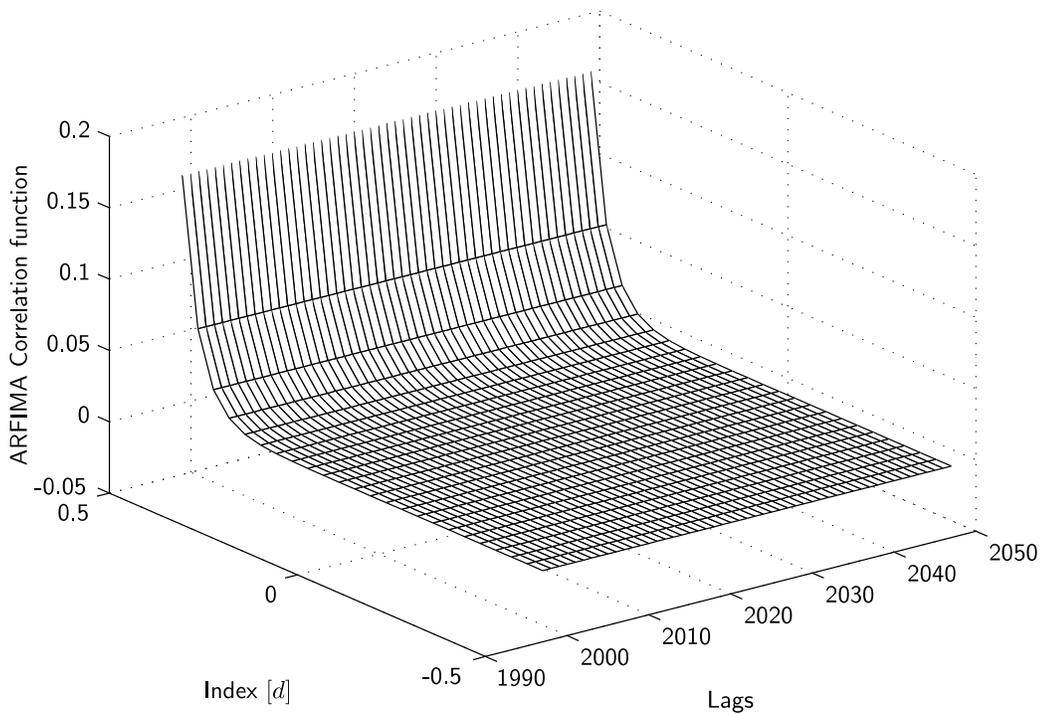


(b) Spectral Density Function of FDN

Figure 2.1: The figure illustrates the behaviour of FDN for $d \in (-1/2, 1/2)$. In figure 2.1a the correlation function of FDN is shown. When $d > 0$, the long term correlations are not negligible. Figure 2.1b shows the spectral density function of FDN. It can be observed that at small frequencies, the spectral density function of FDN exhibits a power law relationship of the form w^{-2d} , which is referred to as one over frequency behaviour.

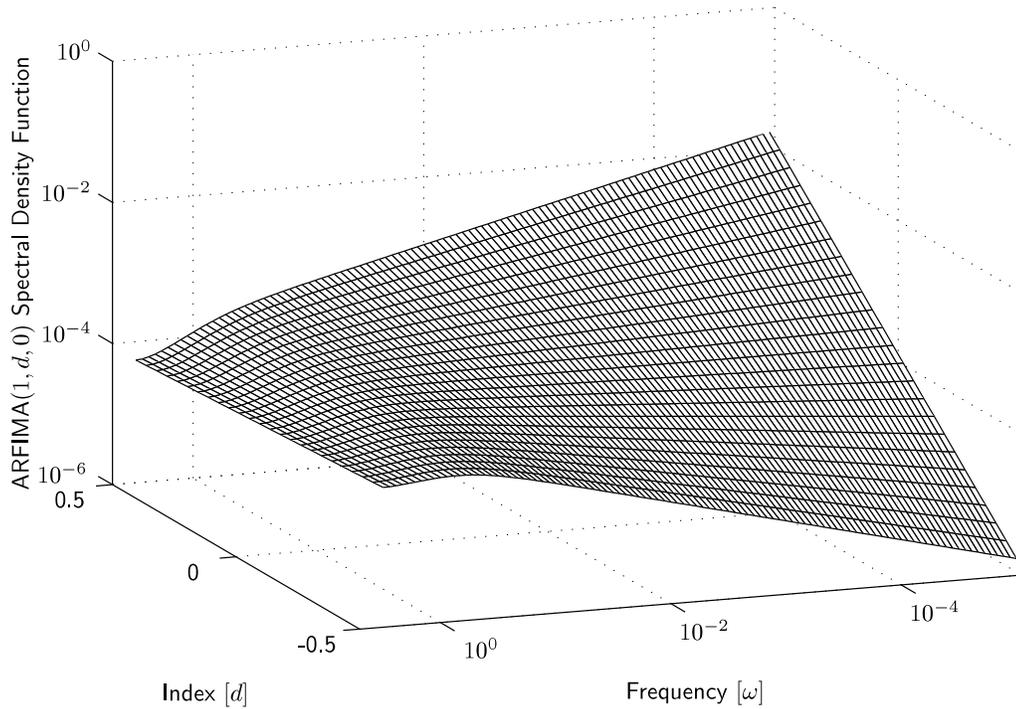


(a) Correlation Function of an ARFIMA process at small lags

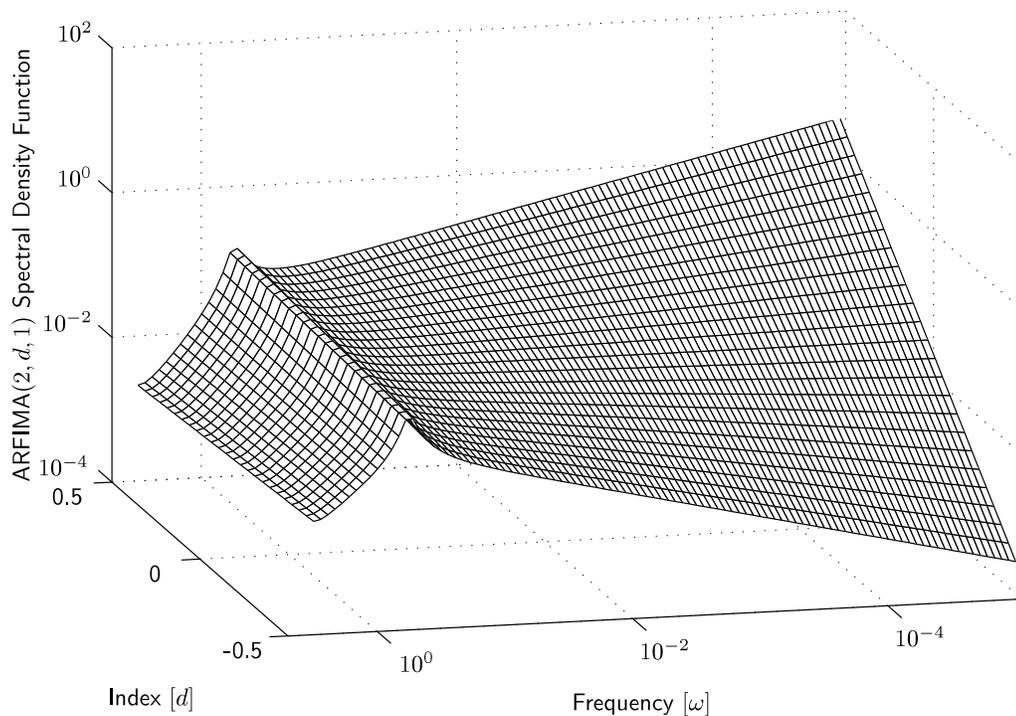


(b) Correlation Function of an ARFIMA process at large lags

Figure 2.2: The figure illustrates the behaviour of the correlations of an ARFIMA process. In figure 2.2a it is possible to observe that the short term behaviour is driven by the ARMA process. Figure 2.2b portrays the behaviour of the same ARFIMA process at large lags. The behaviour resembles that of FDN at large lags.



(a) Spectral Density Function of an ARFIMA(1, d , 0) Process



(b) Spectral Density Function of an ARFIMA(2, d , 1) Process

Figure 2.3: The figure illustrates the shape of two different ARFIMA processes. It is possible to observe the flexibility provided at higher frequencies due to the ARMA contributions to the process. Note that in contrast with typical spectral density function charts, the frequency decreases from left to right rather than right to left.

Yule-Walker equations [see for example 94]. Algorithms to numerically evaluate the covariance function of an ARMA process are given by [14, 72, 95]. Once the covariances of the ARMA and FDN processes have been found we can truncate the infinite sum in (2.27) to obtain an approximation. The convolution can be implemented via the Fast Fourier Transform (FFT) algorithm. Bertelli and Caporin [10] established truncation bounds such that the errors in the approximation of the convolution are reduced. They suggest employing a bound m such that $\gamma_{\text{ARMA}}(\tau) < 1 \times 10^{-j}$ for all $\tau : |\tau| > m$, with $j \geq 6$. This rule leads to a computation of the covariance function of an ARFIMA(n_a, d, n_b) process with a precision up to $1 \times 10^{-j-1}$.

A different approach is given by Sowell [97] which involves the computation of hypergeometric functions. This method is much more computationally expensive.

Note that the efficient and accurate computation of the covariance function of an ARFIMA process is of great importance when computing some of the estimation methods in chapter 3 and 4.

- (c) We have from [14, p. 525] that the covariance function of the ARFIMA process behaves asymptotically as

$$\lim_{k \rightarrow \infty} \frac{\gamma_{\beta}(k)}{\zeta_d k^{2d-1}} = 1 \quad (2.28)$$

with $\zeta_d \neq 0$ a function independent of k .

- (d) The spectral density function of an ARFIMA process is given by the spectral factorisation theorem [see for example 3, 65], i.e.,

$$f_{\beta}(\omega) = f_{\text{ARMA}}(\omega) f_{\text{FDN}}(\omega) \quad (2.29)$$

with $f_{\text{FDN}}(\cdot)$ given by equation (2.20) and $f_{\text{ARMA}}(\cdot)$ the spectral density function of an ARMA process with unity innovations variance and polynomials $A(q)$ and $B(q)$.

2.4 Synthesis of Gaussian LRD Processes

Being capable of efficiently and accurately producing artificial realisations of LRD stochastic processes is of great interest. The synthetic data can be used to provide some insight on the small sample statistics of such processes, in Monte Carlo studies for evaluating the performance of different parameter estimators [see for example 27, 37, 40, 49, 86, 102] or for generating persistent excitation for experiment design [89].

There exist several methods capable of generating a realisation of a Gaussian LRD process. However, there is a trade-off between their accuracy and computational efficiency. The exact (up to the precision of the computations) methods are, in general, computationally expensive and thus impractical from an implementation point of view. They are based on matrix factorisation algorithms such as the Choleski decomposition and the Levinson-Durbin algorithm. The Choleski decomposition method has a complexity of order $\mathcal{O}(M^3)$ for a sequence of length M . The Levinson-Durbin algorithm has a computational complexity of $\mathcal{O}(M^2)$.

Approximate methods are computationally more efficient, e.g., the direct spectral method has a computational complexity of $\mathcal{O}(M \log M)$. However, it suffers from periodicity, i.e., the end of the simulated sequence is highly correlated with the beginning, and aliasing [6]. A comprehensive review of several methods for synthesising LRD processes is given by Bardet et al. [6].

Davies and Harte [1987] introduced a fast and exact algorithm for producing simulated realisations of fGN. Dietrich and Newsam [1997] showed that this method can be applied to synthesising sequences of stationary Gaussian processes. Its computational complexity is of order $\mathcal{O}(N \log N)$, with $N \geq 2(M-1)$. This algorithm relies on embedding the $M \times M$ covariance matrix of the process, Σ , into a larger $N \times N$ circulant matrix, C . If C is non-negative definite (written $C \geq 0$), then the method is guaranteed to properly generate a Gaussian stationary sequence. If $C \geq 0$, then the Circulant Embedding Method (CEM) or Davies-Harte algorithm, is exact in the sense that the simulated path has the correct covariances up to a multiplicative factor of N [30].

2.4.1 Circulant Embedding Method

It is known that the covariance matrix of a weakly stationary complex stochastic process is positive definite [35] and hermitian [66]. If the weakly stationary process

is real then its covariance matrix is positive definite and symmetric. Furthermore, we have that the covariance matrix has a Toeplitz structure [94].

Definition 2.4.1. A square matrix \mathcal{T} of dimension M is said to be Toeplitz if

$$\mathcal{T}_{j,k} = t_{j-k} \quad (2.30)$$

and $j, k = 0, 1, \dots, M - 1$.

Definition 2.4.1 implies that the matrix \mathcal{T} is such that its elements are constant along all diagonals, i.e.,

$$\mathcal{T} = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(M-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \vdots \\ \vdots & & & \ddots & \\ t_{M-1} & & \cdots & & t_0 \end{bmatrix}.$$

Circulant matrices are a common special case of Toeplitz matrices. These occur when every row of the matrix is a right cyclic shift of the row above it.

Definition 2.4.2. A square matrix C of dimension N is said to be a circulant if

$$C = \begin{bmatrix} c_0 & c_{-1} & c_{-2} & \cdots & c_{-(N-1)} \\ c_{-(N-1)} & c_0 & c_{-1} & & \\ c_{-(N-2)} & c_{-(N-1)} & c_0 & & \vdots \\ \vdots & & & \ddots & \\ c_{-1} & c_{-2} & \cdots & & c_0 \end{bmatrix}. \quad (2.31)$$

The eigenstructure of a circulant matrix is considered to be simple as it is possible to explicitly find the eigenvectors and eigenvalues of such matrices [36]. We have that the eigenvectors of a circulant matrix are the columns of the Fourier matrix [26]

$$F^* = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & u & u^2 & \cdots & u^{N-1} \\ 1 & u^2 & u^4 & \cdots & u^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u^{N-1} & u^{2(N-1)} & \cdots & u^{(N-1)(N-1)} \end{bmatrix} \quad (2.32)$$

where

$$u = \exp\left(\frac{i2\pi}{N}\right) \quad (2.33)$$

is the root of unity. Thus, we have

$$C = F^* \Lambda F \quad (2.34)$$

with

$$\Lambda = \text{diag}\{\lambda_k\}, \quad k \in \mathbb{J} \quad (2.35)$$

and

$$\lambda_k = \sum_{\tau=0}^{N-1} c_\tau u^{-k\tau}. \quad (2.36)$$

Equation (2.36) implies that the eigenvalues of a circulant matrix C can be obtained by computing the Discrete Fourier Transform (DFT) of the sequence c_τ . Therefore, diagonalising a circulant matrix is computationally inexpensive due to the FFT algorithm.

Another of their attractive properties is that any circulant of composite order $N = mn$ is automatically a block circulant matrix in which each block matrix is Toeplitz. The blocks are of order $n \times n$ and the arrangement of the blocks is $m \times m$, e.g.,

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & c_4 & c_5 \\ c_5 & c_0 & c_1 & c_2 & c_3 & c_4 \\ c_4 & c_5 & c_0 & c_1 & c_2 & c_3 \\ c_3 & c_4 & c_5 & c_0 & c_1 & c_2 \\ c_2 & c_3 & c_4 & c_5 & c_0 & c_1 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_0 \end{bmatrix}_{N \times N}$$

or

$$C = \begin{bmatrix} C_0 & C_1 & C_2 \\ C_2 & C_0 & C_1 \\ C_1 & C_2 & C_0 \end{bmatrix}_{m \times m}$$

with

$$C_0 = \begin{bmatrix} c_0 & c_1 \\ c_5 & c_0 \end{bmatrix}_{n \times n}, \quad C_1 = \begin{bmatrix} c_2 & c_3 \\ c_1 & c_2 \end{bmatrix}_{n \times n}, \quad C_2 = \begin{bmatrix} c_4 & c_5 \\ c_3 & c_4 \end{bmatrix}_{n \times n},$$

$m = 3$ and $n = 2$.

Using this property, we can embed a Toeplitz matrix into a larger circulant matrix. This idea is not new as circulant embedding matrices have been used to speed up algebraic computations of Toeplitz matrices [26], for covariance matrix estimation [28], for synthesis of Gaussian sequences [23, 25, 30, 85], for providing bounds of the eigenvalues of the Toeplitz matrix [36] and for estimating the parameters of LRD processes (see chapter 3).

Let $\Sigma = \Sigma(\beta)$ denote the $M \times M$ covariance matrix of a weakly stationary Gaussian process, Y , with parameters β . Its elements are given by

$$\Sigma_{j,k} = \Sigma_{j,k}(\beta) = \gamma_{\beta}(|j - k|) \quad (2.37)$$

with $\gamma_{\beta}(\cdot)$ the covariance function of the process, $j, k = 0, 1, \dots, M - 1$. Equation (2.37) implies that $\Sigma = \Sigma^T$ with $[\cdot]^T$ the matrix transpose, that is, Σ is symmetric.

Davies and Harte [25] proposed a particular circulant embedding of Σ for generating synthetic realisations of the Gaussian stationary process Y . Let c_{β} be a vector of length $N \geq 2(M - 1)$ with elements

$$\begin{aligned} c_{\beta}(k) &= \gamma_{\beta}(k) , & k &= 0, 1, \dots, M - 1 \\ c_{\beta}(N - k) &= \gamma_{\beta}(k) , & k &= 1, 2, \dots, M - 2 . \end{aligned} \quad (2.38)$$

If $N > 2(M - 1)$ then the entries $c(M), \dots, c(N - M + 1)$ are arbitrary, or conveniently chosen. The optimal case $N = 2(M - 1)$ is said to be a minimal embedding.

Let C be an $N \times N$ symmetric circulant matrix with first row or column elements c_{β} . Then, provided that C is non-negative definite, the Davies-Harte algorithm (see algorithm 1) will yield an exact artificial realisation of length M of the Gaussian process with covariance function $\gamma_{\beta}(\cdot)$.

Perrin et al. [85] and Craigmile [23] showed respectively that artificial sequences of fGN and FDN can be produced via the CEM. Nevertheless, there is no general result that ensures the nonnegative definiteness of the embedding of an ARFIMA process [6]. Dietrich and Newsam [30] showed that if the covariance function of a process has finite support and is convex decreasing then the CEM is guaranteed to be non-negative definite. The covariance function of an ARFIMA process is not guaranteed to be convex decreasing. Furthermore, it does not have finite support, thus the result of [30] is not applicable.

From experience, we have noted that the circulant embedding is non-negative

Algorithm 1 Davies-Harte Algorithm for generating synthetic realisations of a stationary Gaussian Process

- 1: Generate an M -vector of covariances γ_β .
 - 2: Form the vector c of length $N \geq 2(M - 1)$ as in equation (2.38).
 - 3: **if** $N > 2(M - 1)$ **then**
 - 4: Select appropriate entries for $c(M), \dots, c(N - M + 1)$.
 - 5: **end if**
 - 6: Compute $\lambda = Fs$ by using the Fast Fourier Transform (FFT) algorithm.
 - 7: **if** $\lambda \geq 0$ **then**
 - 8: Generate the complex vector of size N , $x = x_1 + ix_2$ with x_1 and x_2 independent Gaussian simulations with zero mean and unit variance.
 - 9: Form the vector $\lambda^{1/2}x$.
 - 10: Compute $[y_1, y_2] = y = F\lambda^{1/2}x$ using the FFT algorithm. This step yields two independent vectors with Gaussian distribution and covariance C .
 - 11: **else if** $N > 2(M - 1)$ **and** $\lambda < 0$ **then**
 - 12: Choose different entries for $c(M), \dots, c(N - M + 1)$.
 - 13: Return to step 6.
 - 14: **end if**
 - 15: Select any consecutive values of y_k , $k = 1, 2$. These vectors will be Gaussian distributed with zero mean and covariance Σ .
 - 16: To generate two additional vectors go to step 8.
-

definite provided its dimension is large such that most of the SRD contributions of the ARMA covariance function are encapsulated in C . This hypothesis is written formally in conjecture 2.4.3 and 2.4.4.

Conjecture 2.4.3. *Let C be the $N \times N$, $N = 2(M - 1)$, circulant embedding of the covariance matrix of an ARFIMA(0, d , n_b) stationary Gaussian process. If $M > n_b + 1$, then the eigenvalues of C are such that condition 2.4.5 is satisfied. Thus, by lemma 2.4.6 we have $C \geq 0$.*

Conjecture 2.4.4. *Let C be the $N \times N$, $N = 2(M - 1)$, circulant embedding of the covariance matrix of an ARFIMA(n_a , d , n_b) stationary Gaussian process. If M is large such that $\gamma_{ARMA}(k)/\gamma_{ARMA}(0) < 1 \times 10^{-6}$ for all $k : |k| > M$, then the eigenvalues of C satisfy condition 2.4.5. Thus, by lemma 2.4.6 we have $C \geq 0$.*

Condition 2.4.5. *Let $f_\beta(\omega_k) \geq 0$ be the spectral density function of the Gaussian process evaluated at frequencies $\omega_k = i2\pi k/N$ and $k \in \mathbb{J}$. Then*

$$\max_k \left[\frac{|\lambda_k - 2\pi f_\beta(\omega_k)|}{2\pi f_\beta(\omega_k)} \right] \leq 1 . \quad (2.39)$$

Lemma 2.4.6. *The circulant matrix with eigenvalues λ_k is non-negative definite if condition 2.4.5 holds.*

Proof. We have $\lambda_k - 2\pi f_\beta(\omega_k) \leq 2\pi f_\beta(\omega_k)$ and $2\pi f_\beta(\omega_k) - \lambda_k \leq 2\pi f_\beta(\omega_k)$, that is $0 \leq \lambda_k \leq 4\pi f_\beta(\omega_k)$. \square

2.A Proofs of Theorems 2.3.2, 2.3.5 and 2.3.8

Proof of Theorem 2.3.2. Most of these properties are a direct consequence of fBM being an H-sssi process.

- (a) The proof follows immediately from the self-similarity condition in (2.7).
- (b) From the self-similarity property we have

$$\mathbf{E}[B_H(\Omega t)] = \Omega^H \mathbf{E}[B_H(t)] .$$

From stationary increments condition we have

$$\begin{aligned} \mathbf{E}[B_H(\Omega t)] &= \mathbf{E}[B_H(\Omega t) - B_H(t)] + \mathbf{E}[B_H(t)] \\ &= \mathbf{E}[B_H(t) - B_H(0)] + \mathbf{E}[B_H(t)] \\ &= \Omega \mathbf{E}[B_H(t)] . \end{aligned}$$

Therefore, in order for $\Omega^H \mathbf{E} B_H = \Omega \mathbf{E} B_H$ to hold for $H \in (0, 1)$, $\mathbf{E} B_H = 0$.

- (c) From the stationary increments condition we have

$$B_H(-t) - B_H(0) \stackrel{D}{=} B_H(-t+h) - B_H(h)$$

which holds only for $h = t$. Hence

$$B_H(-t) - B_H(0) \stackrel{D}{=} B_H(0) - B_H(t) .$$

- (d) From the self-similarity condition we have

$$\begin{aligned} \mathbf{E}[B_H^2(t)] &= \mathbf{E}[B_H(|t|\text{sign } t)B_H(|t|\text{sign } t)] \\ &= |t|^{2H} \mathbf{E}[B_H(\text{sign } t)B_H(\text{sign } t)] . \end{aligned}$$

From property (c) we get

$$\mathbb{E} [B_H^2(t)] = |t|^{2H} \mathbb{E} [B_H^2(1)] ,$$

with $\mathbb{E} [B_H^2(1)] = \nu^2$.

(e) First, we have

$$\mathbb{E}[B_H(t) - B_H(s)]^2 = \mathbb{E} [B_H^2(t)] + \mathbb{E} [B_H^2(s)] - 2\mathbb{E}[B_H(t)B_H(s)] .$$

Let $\tau = t - s$, then we have

$$B_H(s + \tau) - B_H(s) \stackrel{D}{=} B_H(\tau) - B_H(0) ,$$

thus, by the property of self-similarity we get

$$\begin{aligned} 2\mathbb{E}[B_H(t)B_H(s)] &= \mathbb{E} [B_H^2(t)] + \mathbb{E} [B_H^2(s)] - \mathbb{E} [B_H(\tau) - B_H(0)]^2 \\ &= |t|^{2H} \mathbb{E} [B_H^2(1)] + |s|^{2H} \mathbb{E} [B_H^2(1)] \\ &\quad - |t - s|^{2H} \mathbb{E} [B_H^2(1)] \\ &= [|t|^{2H} + |s|^{2H} - |t - s|^{2H}] \nu^2 \end{aligned}$$

with $\mathbb{E} [B_H^2(1)] = \nu^2$.

□

Proof of Theorem 2.3.5. The proof from this properties follow from the properties of fBM processes and the definition of fGN.

(a) From the stationary increments condition we have

$$\mathbb{E} [X_k] = \mathbb{E}[B_H(k + 1) - B_H(k)] .$$

From theorem 2.3.2 we know that $\mathbb{E} B_H(t) = 0$ for all $t \in \mathbb{R}$ and thus the result.

(b) From the stationary increments condition we have

$$\begin{aligned}\mathbb{E}[X_k^2] &= \mathbb{E}[B_H(1) - B_H(0)]^2 \\ &= \mathbb{E}[B_H^2(1)] \\ &= \nu^2 .\end{aligned}$$

(c) We have

$$\begin{aligned}\gamma_\beta(k) &= \mathbb{E}[X_\tau X_{\tau+k}] \\ &= \mathbb{E}[\{B_H(\tau+1) - B_H(\tau)\} \{B_H(\tau+k+1) - B_H(\tau+k)\}] \\ &= \mathbb{E}[B_H(\tau+1)B_H(\tau+k+1)] - \mathbb{E}[B_H(\tau+1)B_H(\tau+k)] \\ &\quad - \mathbb{E}[B_H(\tau)B_H(\tau+k+1)] + \mathbb{E}[B_H(\tau)B_H(\tau+k)] .\end{aligned}$$

The result follows immediately from the covariance function of fBM in equation (2.10).

(d) Let $g(k) = k^{2H}$ and $k \geq 1$. The function $g(k)$ is strictly convex for $H \in (1/2, 1)$, hence

$$g(k+1) - 2g(k) + g(k-1) > 0 .$$

This implies that $\gamma_\beta(k) > 0$ when $H \in (1/2, 1)$. If $H \in (0, 1/2)$ then $g(k)$ is strictly concave, hence

$$g(k+1) - 2g(k) + g(k-1) < 0 .$$

which implies that $\gamma_\beta(k) < 0$ when $H \in (0, 1/2)$. For $H = 1/2$ the proof is trivial.

(e) See [100].

□

Proof of Theorem 2.3.8. (a) Recall that the spectral density of a sequence of iid random variables is given by

$$f_e(\omega) = \frac{\nu_e^2}{2\pi}$$

Then from the spectral factorisation theorem [see for example 3, 65] we get

$$\begin{aligned} f_\beta(\omega) &= f_e \left[\frac{1}{1 - e^{-i\omega}} \right]^d \left[\frac{1}{1 - e^{i\omega}} \right]^d \\ &= \frac{f_e}{[2 - 2 \cos(\omega)]^d} \\ &= \frac{\nu_e^2}{2\pi [2 \sin(\omega/2)]^{2d}} . \end{aligned}$$

(b) From [44, 50] we get that for $d < 1/2$

$$\begin{aligned} \gamma_\beta(k) &= \int_{-\pi}^{\pi} e^{-i\omega k} \phi_x(\omega, d) d\omega \\ &= \frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!} . \end{aligned}$$

We have from [eq 6.1.5, p. 255, in 1]

$$\Gamma(z+1) = z\Gamma(z) = z! .$$

Then we may write

$$\frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!} = \frac{\cos(\pi k) \Gamma(1-2d)}{\Gamma(k+1-d) \Gamma(1-k-d)}$$

The result follows from the reflection and duplication properties of the gamma function [eq. 6.1.17 and 6.1.18, p. 256, in 1], i.e.,

$$\begin{aligned} \Gamma(z)\Gamma(1-z) &= \pi \csc(\pi z) \\ \Gamma(2z) &= (2\pi)^{-1/2} 2^{2z-1/2} \Gamma(z)\Gamma(z+0.5) . \end{aligned}$$

□

Chapter 3

Estimation of Long-range Dependent Models

3.1 Introduction

The goal of estimation is to find the parameter values that will enable a mathematical model to explain, in some sense, the behaviour of a set of observations along with the underlying process producing them. Much of the traditional parameter estimation has been developed around a flexible set of discrete-time models which arise from the foundations of Linear Time-Invariant (LTI) system theory [see for example 3, 12, 65]. There are several parameter estimation methods for such family of models, e.g., Box-Jenkins “back-forecasting” algorithm, Prediction Error Methods (PEM) or Kalman filtering implementations.

For the case of the LRD family of models, the implementation of such methods is no longer straight forward. However, there exist several alternatives for performing parameter estimation for such models. Their frames of analysis range from the parametric to the semi-parametric. Some of them are based in the time domain, while others work in the frequency domain. The finite dimensional properties of most of these estimators have been evaluated in Monte-Carlo studies [see for example 27, 37, 40, 49, 86, 102].

The chapter is organised as follows: section 3.2 of this chapter surveys some of the semi-parametric estimation methods available; section 3.3 discusses in detail likelihood based estimation methods under a parametric frame of analysis; here, we present the maximum likelihood estimate, the Whittle estimate and a novel Circulant Embedding estimate, and show their asymptotic properties; section 3.4

presents the concluding remarks of the chapter.

3.2 Semi-parametric Methods

The semi-parametric approach may seem an advantageous strategy for estimating the LRD coefficient of a stochastic process provided that its SRD behaviour is not of central interest. The asymptotic behaviour of the covariance and spectral density functions of the LRD processes presented in chapter 2 indicate that the short term behaviour of the model is almost irrelevant at very low frequencies and very long lags. Thus, estimates of the LRD coefficient may be obtained by taking advantage of the aforementioned properties. If the SRD behaviour is not of central interest, the semi-parametric estimates become fairly attractive as they avoid model order selection and other global assumptions of the spectral density and covariance function of the process. However, this comes at the price of a slower rate of convergence of such methods in comparison to parametric estimates based on models with a correctly specified structure [88]. Nevertheless, for a sufficiently large set of observations the precision of the semi-parametric estimates can be expected to be adequate. This, along with their computational simplicity makes them attractive as simple diagnostic tools under a wide range of scenarios.

Among these methods we have: Higuchi's method (see [102] and the references within), Geweke and Porter-Hudak method (see [102] and the references within), the rescaled adjusted range [71], the de-trended fluctuation analysis [84] and a log-periodogram regression [9]. In this section we review the rescaled adjusted range, the de-trended fluctuation analysis and the log-periodogram regression. All three of these methods perform an Ordinary Least Squares (OLS) regression at some point in their computation, hence, we briefly describe this procedure.

Ordinary Least Squares Regression

The method of ordinary least squares (OLS) can be traced back to Carl Friedrich Gauss in 1794 [4]. Let $(x_{k,m}, y_k)$, $k, m \in \mathbb{J}$ with $m \leq k$ be a finite set of observations given by a linear combination, i.e.

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} x_{0,m}^T \\ x_{1,m}^T \\ \vdots \\ x_{N-1,m}^T \end{bmatrix} \Theta + \begin{bmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \eta_{N-1} \end{bmatrix}$$

or in vector form

$$Y = X\Theta + \eta \quad (3.1)$$

where Θ is a matrix containing N vectors of unknown parameters θ_m and η is a vector of deviations or noise. The residual sum of squares in vector form is given by

$$V_{\text{OLS}} = [Y - X\Theta]^T [Y - X\Theta] . \quad (3.2)$$

Given V , we have the standard result

$$\begin{aligned} \hat{\Theta}_{\text{OLS}} &= \arg \min_{\Theta} V_{\text{OLS}} \\ &= (X^T X)^{-1} X^T Y . \end{aligned} \quad (3.3)$$

We refer to $\hat{\Theta}_{\text{OLS}}$ as the ordinary least squares solution to the linear regression problem in equation (3.1). Under certain conditions, the OLS solution can be shown to have optimality properties for various statistical models [see 4, 18, 93].

3.2.1 Rescaled Adjusted Range

First introduced by Hurst in the 1950's, the Rescaled Adjusted Range (R/S) method is one of the most well known semi-parametric estimators for the LRD coefficient. It is discussed in detail in [71] and it has been subject to Monte-Carlo studies in [49, 102] and more recently for small sample lengths in [27].

Let X_k , $k \in \mathbb{J}$, be a series of sequential observations of a weakly stationary

LRD process with partial sums

$$Y_k = \sum_{\tau=0}^k X_\tau \quad (3.4)$$

and sample variance

$$S_k^2 = \frac{1}{k} \sum_{\tau=0}^k X_\tau^2 - \frac{1}{k^2} Y_k^2 . \quad (3.5)$$

The R/S statistic is given by [71]

$$\frac{R_k}{S_k} = \frac{1}{S_k} \left[\max_{0 \leq t \leq k} \left(Y_t - \frac{t}{k} Y_k \right) - \min_{0 \leq t \leq k} \left(Y_t - \frac{t}{k} Y_k \right) \right] . \quad (3.6)$$

Asymptotically, we have [102]

$$\lim_{k \rightarrow \infty} \frac{\mathbb{E} [R_k/S_k]}{\Omega_\theta k^\theta} = 1 \quad (3.7)$$

where $\Omega_\theta < \infty$ is a positive function independent of k and θ is the LRD coefficient of X_k . Equation (3.7) implies that for large values of k , $\log R_k/S_k$ should be randomly scattered around a straight line with slope $\theta = d - 1/2$ for FD models and $\theta = H$ for fGN [9].

Algorithm 2 briefly describes the procedure for estimating the LRD coefficient via the R/S statistic. Figure 3.1 depicts an estimate from a fGN sequence obtained by employing the R/S statistic.

Algorithm 2 R/S Estimation Method for LRD observations

- 1: Subdivide the observations X_k into M segments of length $m = N/M$.
- 2: Compute the R_k/S_k statistic for all possible lags, k , per segment such that $k + m \leq N$.
- 3: The estimate of the LRD coefficient is given by finding the solution

$$\hat{\Theta} = [\hat{\theta}_{R/S} \quad \hat{\theta}_0]$$

to the OLS regression for the ordered pair $(k, R_k/S_k)$.

- 4: Compute the LRD coefficient given by $\hat{\theta}_{R/S} = \hat{d}_{R/S} - 1/2$ for the FD family of models and $\hat{\theta}_{R/S} = \hat{H}_{R/S}$ for fGN.
-

There exist several modified versions of the R/S statistic. They are based on replacements to the sample variance and some perform better than others. A

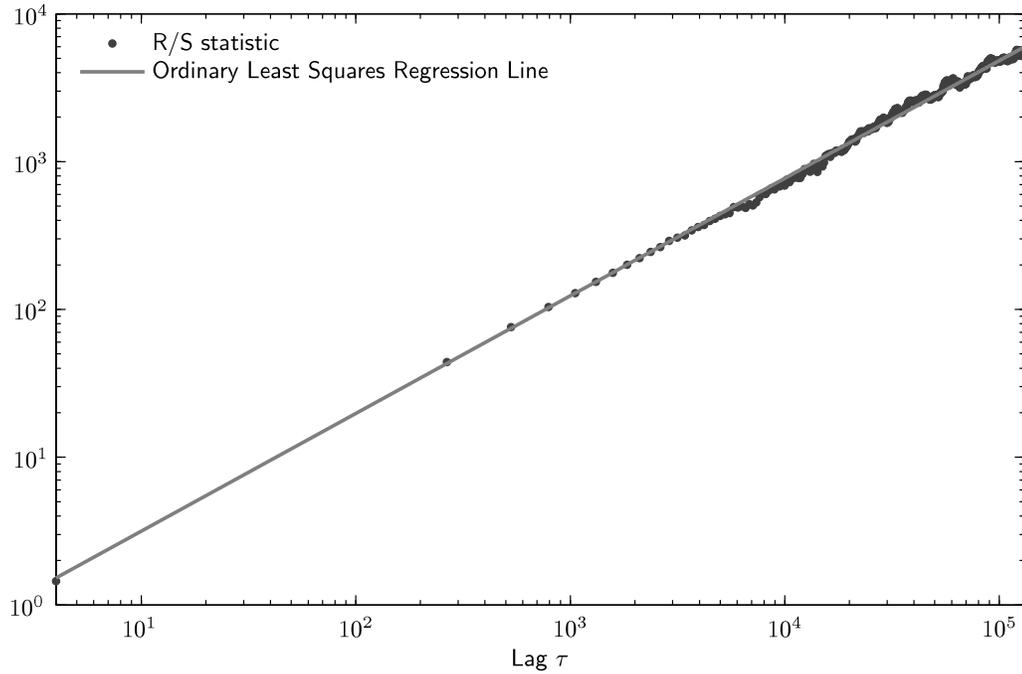


Figure 3.1: The figure depicts the R/S statistic of a sequence of fGN observations X_k with nominal LRD coefficient $H = 0.8$, and the estimate of the LRD coefficient obtained by fitting an ordinary least squares regression line to the R/S ordinates. The estimate obtained via this method was of $\hat{H} = 0.7959$.

comprehensive review of the R/S method and its modified versions is given by [9, 49] and the references within.

3.2.2 De-trended Fluctuation Analysis

De-trended Fluctuation Analysis (DFA) is a method based on classical random walk studies for estimating the LRD coefficient of a process in a semi-parametric frame of analysis. First introduced by Peng et al. [84], DFA is frequently used in the area of physiological data processing, particularly for estimating the LRD coefficient in heartbeat signals of healthy and sick individuals [see references 41, 52, 53, 83, 87, 92] and more recently for modelling bimanual coordination [103]. This method has been consistently reported as one of the most accurate estimators for the LRD coefficient in various Monte-Carlo studies [27, 40, 86, 102].

Bardet and Kammoun [5] showed that DFA possesses a reasonable (but not optimal) convergence rate when the sequence of observations are not subject to a trend. Chen et al. [22], Hu et al. [51] and Kantelhardt et al. [55], performed

heuristic studies of the DFA method under the presence of non-linearities associated with examples of trends (linear, sinusoidal and power-law). They deduced that the estimate obtained through DFA is affected by a trade-off between the effects of such non-linearities and the LRD coefficient of a process. The effects of non-stationarities over the estimates produced by DFA was studied in Chen et al. [21].

Let X_k , $k \in \mathbb{J}$, be a sequence of observations from a weakly stationary LRD process. The k^{th} sequence of partial sums of X_k is given by Y_k as in equation (3.4). Let

$$Y_\tau^m = \left[Y_{m\tau} \quad Y_{m\tau+1} \quad \cdots \quad Y_{m\tau+m-1} \right]^T \quad (3.8)$$

denote the $\tau \in \{0, 1, \dots, M\}$ segment of length $m = N/M$ of the partial sum sequence Y_k and let

$$Z_\tau^m = \begin{bmatrix} Z_{m\tau} \\ Z_{m\tau+1} \\ \vdots \\ Z_{m\tau+m-1} \end{bmatrix} = Y_\tau^m - \begin{bmatrix} m\tau & 1 \\ m\tau+1 & 1 \\ \vdots & \vdots \\ m\tau+m-1 & 1 \end{bmatrix} \bar{\Theta} \quad (3.9)$$

denote the τ^{th} de-trended segment of the partial sum sequence, where

$$\bar{\Theta} = \begin{bmatrix} \bar{\theta}_1 \\ \bar{\theta}_0 \end{bmatrix}$$

is the solution to the OLS regression for the ordered pair (K, Y_τ^m) with

$$K = \left\{ m\tau \quad m\tau+1 \quad \cdots \quad m\tau+m-1 \right\} . \quad (3.10)$$

The standard deviation of the de-trended segment Z_τ^m is given by the de-trended fluctuation function

$$\mathcal{F}_\tau^2(m) = \frac{1}{m} \sum_{k=0}^{m-1} [Z_{m\tau+k}]^2 \quad (3.11)$$

The de-trended fluctuation function of the entire set of segments is given by

$$\mathcal{F}^2(m) = \frac{1}{M} \sum_{\tau=0}^M \mathcal{F}_\tau^2(m) . \quad (3.12)$$

Taqqu et al. [102] showed that the de-trended fluctuation function behaves

asymptotically as

$$\lim_{m \rightarrow \infty} \frac{\mathbb{E} \mathcal{F}(m)}{\Omega_\theta m^\theta} = 1 \quad (3.13)$$

where Ω_θ is a positive function independent of m . Equation (3.13) suggests that for large values of m , $\log F(m)$ should be randomly scattered around a straight line with slope $\theta = d - 1/2$ for FD models and $\theta = H$ for fGN.

Algorithm 3 outlines the DFA procedure for obtaining an estimate of the LRD coefficient from a sequence of observations X_k . Algorithm 4 outlines the procedure for computing the de-trended fluctuation function for a segment of length m .

Figure 3.2 depicts a series of estimates obtained by employing the DFA method on a set of sequences of fGN with different LRD coefficient H . Figure 3.3 depicts the de-trending procedure given by Z_τ^m in equation (3.9) for a particular segment length m .

Algorithm 3 DFA Estimation Method for LRD observations

- 1: Select a set of segment lengths \mathbb{M} .
- 2: Compute the de-trended fluctuation function in equation (3.12) for each segment length $m \in \mathbb{M}$ (see algorithm 4).
- 3: The estimate of the LRD coefficient is given by finding the solution

$$\hat{\Theta} = [\hat{\theta}_{\text{DFA}} \quad \hat{\theta}_0]$$

to the OLS regression for the ordered pair $(m, \mathcal{F}(m))$.

- 4: Compute the LRD coefficient by $\hat{\theta}_{\text{DFA}} = \hat{d}_{\text{DFA}} - 1/2$ for the FD family of models and $\hat{\theta}_{\text{DFA}} = \hat{H}_{\text{DFA}}$ for fGN.
-

Algorithm 4 De-trended Fluctuation Function for a Segment of Length m

- 1: Compute the N partial sums Y_k in equation (3.4) of the sample X_k , $k \in \mathbb{J}$.
- 2: Subdivide the sequence Y_k into M segments, Y_τ^m , of length $m = N/M$ as in equation (3.8).
- 3: Find the solution

$$\bar{\Theta} = [\bar{\theta}_1 \quad \bar{\theta}_0]$$

to the OLS regression for the ordered pair (T, Y_τ^m) with T given by equation (3.10).

- 4: De-trend each partial sum segment Y_τ^m as in equation (3.9)
 - 5: Compute the de-trended fluctuation function $\mathcal{F}(m)$ from equation (3.12).
-

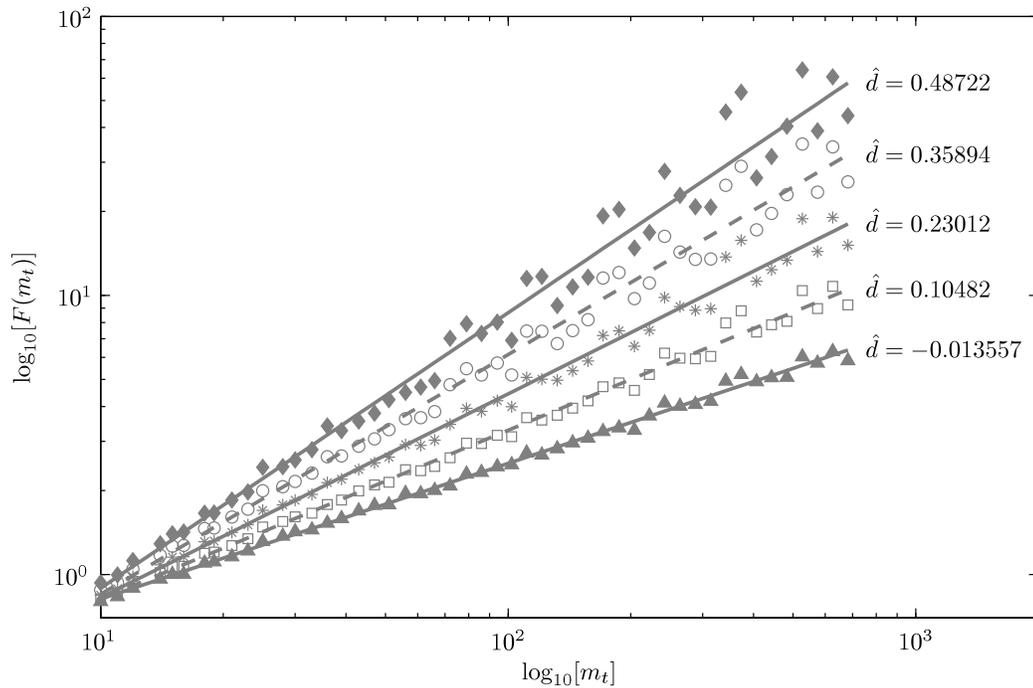


Figure 3.2: Estimates obtained by the DFA method. The DFA was applied to sequences of FDN with coefficient $d = \{0, 0.11, 0.22, 0.34, 0.5\}$ and of length $N = 4096$.

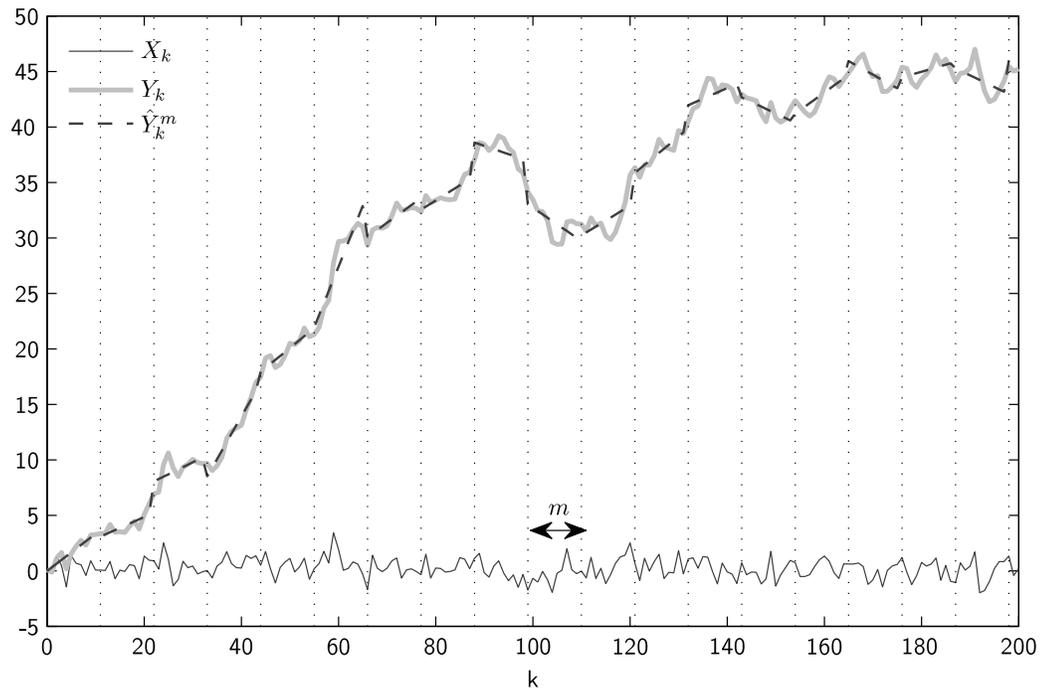


Figure 3.3: The figure depicts a sequence of observations X_k , its partial sums Y_k and the least squares linear interpolation \hat{Y}_k^m for segments of length $m = 47$ (see steps 1 through 3 in algorithm 4).

3.2.3 Log-Periodogram Regression

We have from point (b) in theorem 2.2.4, that the spectral density function of a weakly stationary LRD process takes the form

$$\lim_{\omega \rightarrow 0} \frac{f(\omega)}{\Omega |\omega|^{-\theta}} = 1$$

for some finite, non-zero constant, Ω . Here $f(\omega)$ denotes the spectral density function of the process and the LRD coefficient is given by $\theta = 2d$ for the FD family of models and $\theta = 2H - 1$ for fGN.

Given the behaviour of f as $\omega \rightarrow 0$ and the fact that the periodogram of a sequence of observations is an estimate of f , then a natural estimator for the LRD coefficient is given by fitting a line to the log-periodogram ordinates at small frequencies.

Let X_k , $k \in \mathbb{J}$, be a series of sequential observations of a weakly stationary LRD process. The periodogram of X_k is given by

$$I(\omega) = \frac{1}{2\pi N} \left| \sum_{k=0}^{N-1} X_k \exp(ik\omega) \right|^2 \quad (3.14)$$

with frequencies $\omega \in [-\pi, \pi]$. Let

$$\hat{\Theta} = \begin{bmatrix} \hat{\theta}_{\text{PSD}} & \hat{\theta}_0 \end{bmatrix} \quad (3.15)$$

denote the solution to the OLS regression for the ordered pair $(\omega, I(\omega))$ then $\hat{\theta}_{\text{PSD}} = 2\hat{d}_{\text{PSD}}$ for the FD family of models and $\hat{\theta}_{\text{PSD}} = 2\hat{H}_{\text{PSD}} - 1$ for fGN. Figure 3.4 depicts an estimate from a fGN sequence obtained by employing the log-periodogram regression method.

As noted by [86, 102], the range of frequencies included into the OLS regression must be carefully selected as the asymptotic behaviour of the spectral density function of the LRD process only holds at small frequencies. This means that the occurrences at larger frequencies can significantly bias the estimate of the LRD coefficient. Taqqu et al. [102] suggest a heuristic range of small frequencies to be included in the OLS regression that only accounts for 10% of the total frequencies in the range $\omega \in (0, \pi]$.

There are as many modifications to this method as there are ways to compute a periodogram. These alternatives are justified as the periodogram is not

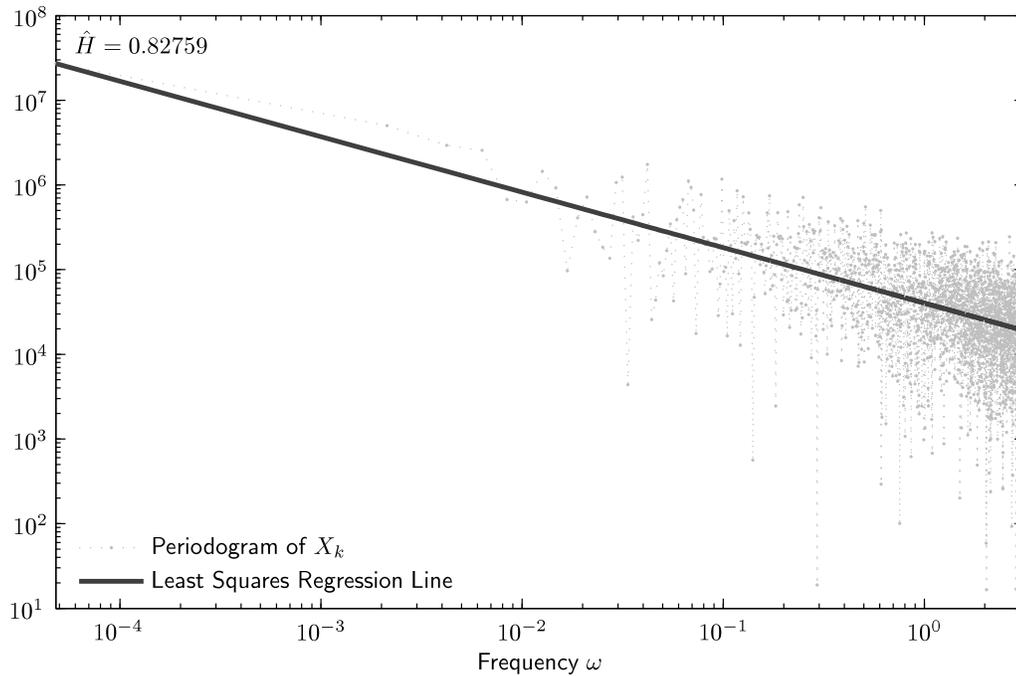


Figure 3.4: The periodogram of a sequence of fGN observations X_k with nominal LRD coefficient $H = 0.8$, and the estimate of the LRD coefficient obtained by fitting a least squares line to the periodogram ordinates are illustrated. The estimate obtained via this method was of $\hat{H} = 0.827$

a minimum variance estimate of the spectral density function [13]. For example, computing a smoothed periodogram, i.e., subdividing the record into segments, computing the periodogram ordinates per segment and averaging the results will yield variance reduction properties. A related approach is given by Welch [106] which consists on subdividing the record into segments, using a function to window the data segments, computing the periodogram of each windowed segment and averaging. For the case of unevenly spaced data records Scargle [91] provides a modification to the periodogram. The Lomb-Scargle periodogram was recently used by [61] to report LRD behaviour in records of ocular accommodation in which there are large segments of missing data.

3.3 Likelihood Based Methods

In the previous section we discussed heuristic methods for estimating the LRD coefficient. They are useful as a first informal diagnostic tool for assessing if a set of observations exhibits LRD. These methods only provide a useful approximation

of the asymptotic behaviour of the process under study. However, they do not provide information of the short-term properties of the process. For such cases, more refined methods which model the entire correlation (or spectral density) structure are available. One possible approach is to use parametric models and estimate their parameters by likelihood based methods. These methods allow the characterisation of both SRD and LRD characteristics of a set of observations and are more efficient [9]. However, they inherit a set of difficulties such as:

- (a) Model structure selection: this is a problem that arises in parametric methods and it is beyond the scope of this thesis. However, as noted by [9] model order selection can be assessed via the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC).
- (b) Computational implementation.

In this section we will present the likelihood based methods along with their properties and assess their computational implementation.

3.3.1 Maximum Likelihood Estimate

R.A. Fisher introduced the method of maximum likelihood in the early 1900's [2]. This method has become standard in several applications, mainly due to its desirable statistical properties under reasonable conditions, e.g., consistency and efficiency. In Fisher's words

“The likelihood that any parameter (or set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed.”

Problem Statement. Maximum likelihood aims to find the set of parameter values that makes the observation occurrences more likely. That is, given a set of parameters β , it attempts to find for a set of observations X

$$\hat{\beta}_{\text{MLE}} = \arg \max_{\beta} L(\beta|X) \quad (3.16)$$

where $L(\beta|X)$ is the likelihood function.

Let $X = X_k$, $k \in \mathbb{J}$, denote a series of equidistant observations. Suppose that these observations constitute a realisation of a real, discrete, weakly stationary, invertible Gaussian process. Let $\beta = [\theta, \sigma^2]$ be a vector containing the model parameters θ and the innovations variance σ^2 . The probability density function (PDF) of such process is given by

$$p(X|\beta) = \left[\frac{1}{(2\pi)^N \det \Sigma} \right]^{1/2} \exp \left[-\frac{1}{2} X^T \Sigma^{-1} X \right] \quad (3.17)$$

where $\Sigma = \Sigma(\beta) > 0$ is an $N \times N$ symmetric Toeplitz matrix with elements $\gamma_\beta(\tau)$ generated by the covariance function of the process [34]. Note that for the long-range dependent processes presented in the previous chapter, the covariance function of the process is given by equation (2.12) for the case of fGN and equation (2.21) for the case of FDN.

The Maximum Likelihood Estimate (MLE), $\hat{\beta}_{\text{MLE}} = [\hat{\theta}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}^2]$, is given by equation (3.16) with

$$L(\beta|x) = p(X|\beta) . \quad (3.18)$$

Since the logarithm is a monotonic function, we can obtain the MLE, $\hat{\beta}_{\text{MLE}}$, from

$$\hat{\beta}_{\text{MLE}} = \arg \min_{\beta} Q(\beta|X) \quad (3.19)$$

where

$$\begin{aligned} Q(\beta|X) &= -\log L(\beta|X) \\ &= \frac{1}{2} [X^T \Sigma^{-1} X + N \log(2\pi) + \log \det \Sigma] \end{aligned} \quad (3.20)$$

is the negative Gaussian log-likelihood function of the set of observations. It is standard to perform the optimisation in equation (3.19) only through the parameters θ rather than $\beta = [\theta, \sigma^2]$. To do this we have the following lemma.

Lemma 3.3.1. *Let a factorisation of the covariance matrix, Σ , be given by*

$$\sigma^2 R = \kappa \Sigma \quad (3.21)$$

where $R = R(\theta) > 0$ is a symmetric Toeplitz matrix with first row elements

$$\sigma^2 r_\theta(\tau) = \kappa \gamma_\beta(\tau) \quad (3.22)$$

with $\tau \in \mathbb{J}$ and $\kappa > 0$ a scaling constant. Then the value that minimises the cost function $Q(\beta|X)$ in terms of the innovations variance, σ^2 , is given by

$$\hat{\sigma}_{MLE}^2 = \frac{\kappa}{N} X^T R^{-1} X . \quad (3.23)$$

Proof. The proof is a standard result, see for example [67, 105]. \square

From lemma 3.3.1 we have that the MLE, $\hat{\theta}_{MLE}$, is given by

$$\hat{\theta}_{MLE} = \arg \min_{\theta} V(\theta|X) \quad (3.24)$$

where

$$\begin{aligned} 2V(\theta|X) = & N [\log(2\pi) + 1 - \log(N)] \\ & + \log(\det R) + N \log [X^T R^{-1} X] . \end{aligned} \quad (3.25)$$

Note that for the case of LRD models we have

$$\theta = \begin{cases} H & \text{if the model is fGN} \\ d & \text{if the models is FDN} \\ [d, a_1, \dots, a_{n_a}, b_1, \dots, b_{n_b}] & \text{if the model is ARFIMA}(n_a, d, n_b) \end{cases} , \quad (3.26)$$

$$\sigma^2 = \begin{cases} \nu^2 & \text{if the model is fGN} \\ \nu_e^2 & \text{if the models is FDN or ARFIMA}(n_a, d, n_b) \end{cases} , \quad (3.27)$$

and

$$\kappa = \begin{cases} 1 & \text{if the model is fGN} \\ \sqrt{\pi} & \text{if the models is FDN or ARFIMA}(n_a, d, n_b) \end{cases} . \quad (3.28)$$

Furthermore, the MLE estimate σ_{MLE}^2 can be computed after obtaining θ_{MLE} .

Dahlhaus [24] and Yajima [110] have shown that the MLE, $\hat{\theta}_{MLE}$, given by equation (3.24) is consistent, asymptotically normal and efficient for fGN, Gaussian FDN and Gaussian ARFIMA. This means that asymptotically the MLE, $\hat{\theta}_{MLE}$ becomes unbiased and that its asymptotic variance achieves the Cramér-Rao Lower Bound. These are desirable statistical properties in an estimator.

Remark 3.3.2. The minimum of $V(\theta|X)$ in equation (3.25) is difficult to find analytically. As an alternative, optimisation algorithms such as line search type methods can be employed for finding the MLE, $\hat{\theta}_{MLE}$. Nevertheless, such methods

require multiple evaluations of $V(\theta|X)$, which implies the calculation of several matrix inverses and determinants. This renders the implementation of the MLE as impractical [32, 37, 40, 79, 86, 104].

The direct inversion of R has a computational complexity of order $\mathcal{O}(N^3)$; hence, several algorithms have been developed in order to reduce the computational burden of the aforementioned operations. The Cholesky decomposition algorithm used in [73] for estimation fGN has a computational complexity of order $\mathcal{O}(N^3/6)$. The Levinson-Durbin algorithm used in [67] for the estimation of fGN has an algorithmic complexity of $\mathcal{O}(2N^2)$. The expectation maximisation algorithm presented in [29] for estimating FDN has a computational complexity of order $\mathcal{O}(N^2/2)$. The recursive Kalman filter of [19] has a computational complexity of $\mathcal{O}(N \times m^2)$ for obtaining an estimate from an m -truncated state space representation of a LRD process.

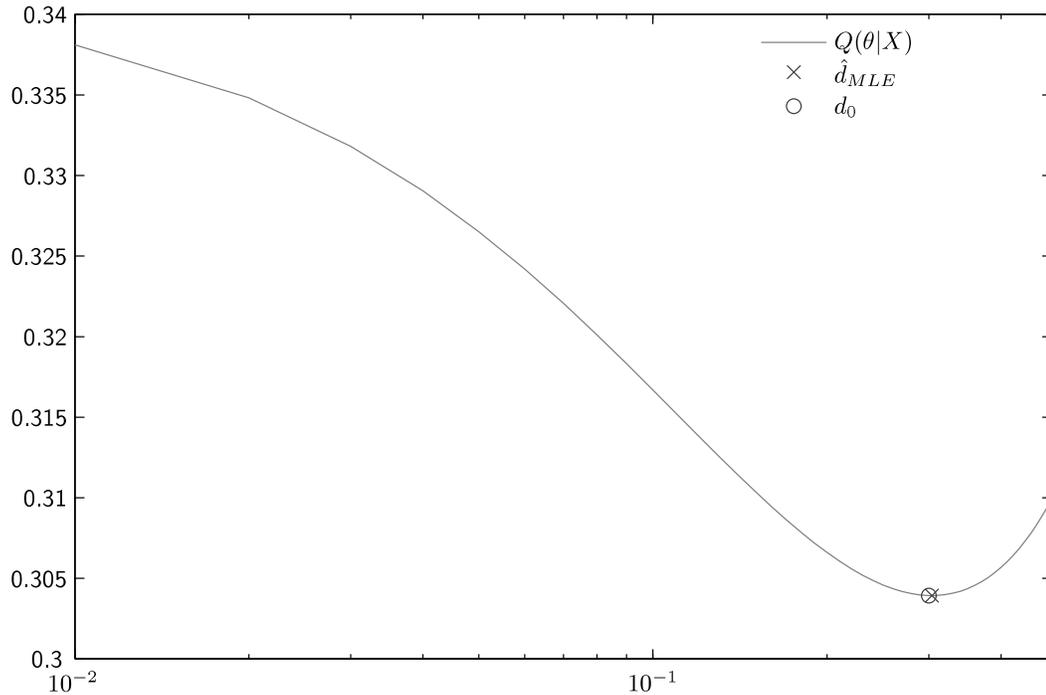


Figure 3.5: The figure depicts the cost function $V(\theta|X)$ for a sequence of observations X_k with nominal LRD coefficient $d = 0.3$ and the MLE estimate $\hat{d}_{MLE} = 0.3038$.

3.3.2 Whittle's Approximation to the MLE

As mentioned in the previous section, the MLE has very attractive statistical properties. However, its implementation is difficult due to the high computational complexity involved. This has been a recurring problem even for short range dependent processes. Whittle [108] proposed to approximate the quadratic form in $Q(\beta|X)$ such that the explicit computations of the inverse and the determinant of the covariance matrix are avoided. In general terms, this approximation is given by the ratio of the periodogram of the time-series and the spectral density function of the process. The order of operations required to compute Whittle's approximation is of $\mathcal{O}(N \log N)$ due to the fast Fourier Transform (FFT) algorithm.

Whittle's approach yields a computationally efficient approximation to the Gaussian likelihood function, $Q(\beta|X)$. Nevertheless, it does not appear to have been used much for SRD observations, arguably, due to the introduction of the Box-Jenkins "back-forecasting" algorithm and Kalman filtering methods [17, 48]. For the case of LRD, the implementation of the Box-Jenkins and Kalman filtering approaches is no-longer straight forward (see remark 3.3.2). Thus, Whittle's approximation has gathered importance as the estimation of the model parameters can be performed without the computational burden of the MLE [17].

Let $X = X_k$, $k \in \mathbb{J}$, be a set of equidistant observations. Suppose that these observations constitute a realisation of a real, discrete, weakly stationary Gaussian process whose spectral density function, $f_\beta(\omega)$, is integrable everywhere in $\omega \in [-\pi, \pi]$. Then Whittle's approximation to the Gaussian likelihood function is given (up to an additive constant) by [see for example 48, 101, 108]

$$\tilde{Q}(\beta|X)_{\text{WE}} = \int_{-\pi}^{\pi} \log f_\beta(\omega) d\omega + \int_{-\pi}^{\pi} \frac{I(\omega)}{f_\beta(\omega)} d\omega \quad (3.29)$$

where $I(\cdot)$ denotes the periodogram of the set of observations given by equation (3.14).

These approximations follow from the limiting behaviour of the Toeplitz determinant described by Szegö [see equation 1.11 in 45], i.e.,

$$\lim_{N \rightarrow \infty} \log \det \Sigma = \int_{-\pi}^{\pi} \log f_\beta(\omega) d\omega$$

and that discrete Fourier Transform (DFT) approximately diagonalises the covariance matrix, Σ , the diagonal elements being $2\pi f_\beta(\omega_k)$ [see p. 224 in 48, and the references within]. Then, by approximating the integrals in (3.29) with Riemann sums over the set of frequencies $\omega_k = 2\pi k/N$ we have

$$Q(\beta|X)_{\text{WE}} = \sum_{k=0}^{N-1} \frac{I(\omega_k)}{f_\beta(\omega_k)} + \sum_{k=0}^{N-1} \log f_\beta(\omega_k) .$$

It is important to note that Whittle's assumption on the spectral density function is no longer valid for the LRD case since $f_\beta(\omega)$ possesses a singularity at $\omega = 0$. However, Fox and Taqqu [39] showed that the ratio $I(\omega)/f_\beta(\omega)$ has a limiting distributional behaviour at the singularity, i.e., there is mutual compensation as both $I(\omega)$ and $f_\beta(\omega)$ "blow up" at $\omega = 0$. This fact justifies using Whittle's approximation of the Gaussian Likelihood function for estimating the parameters of a LRD process.

In this section we present a development of the Whittle Estimate with a matrix analysis focus as in [17]. Let $G = G(\beta) \geq 0$ be a circulant matrix

$$G = F^* \Phi F , \tag{3.30}$$

where F is the Fourier matrix given by equation (2.32) and

$$\Phi = \Phi(\beta) = \text{diag}\{f_\beta(\omega_k)\} \tag{3.31}$$

with $f_\beta(\omega_k)$ the spectral density function of the process evaluated at frequencies $\omega_k = i2\pi k/N$. Suppose that G is asymptotically equivalent to the covariance matrix Σ , i.e., their difference tends to zero as N tends to infinity (see definition 3.3.5), then we may write

$$\begin{aligned} 2Q_{\text{WE}}(\beta|X) &= N \log(2\pi) + X^T G^{-1} X + \log(\det G) \\ &= N \log(2\pi) + (FX)^* \Phi^{-1} (FX) + \sum_{k=0}^{N-1} \log f_\beta(\omega_k) \\ &= N \log(2\pi) + \sum_{k=0}^{N-1} \frac{I(\omega_k)}{f_\beta(\omega_k)} + \sum_{k=0}^{N-1} \log f_\beta(\omega_k) \\ &\approx 2Q(\beta|X) \end{aligned} \tag{3.32}$$

where $I(\omega_k)$ is the periodogram of the time-series evaluated at frequencies ω_k . Thus, Whittle's Estimate (WE) is given by

$$\hat{\beta}_{\text{WE}} = \arg \min_{\beta} Q_{\text{WE}}(\beta|X) . \quad (3.33)$$

Since $f_{\beta}(\omega)$ is an even function with period of 2π [see pp. 116 in 13] then we may perform the sums in (3.32) up to $(N-1)/2$. Furthermore, we may omit frequency $\omega = 0$ to obtain a mean-correction of X_k [88] and so that $I(\omega)$ is asymptotically unbiased [see corollary 5.2.1 in 13, pp. 123]. Then, from lemma 3.3.1 we have

$$\hat{\theta}_{\text{WE}} = \arg \min_{\theta} V_{\text{WE}}(\theta|X) \quad (3.34)$$

with

$$\begin{aligned} 2V_{\text{WE}}(\theta|X) = & N [\log(2\pi) + 1 - \log(N)] \\ & + N \log \sum_{k=1}^{(N-1)/2} \frac{I(\omega_k)}{f_{\theta}(\omega_k)} + \sum_{k=1}^{(N-1)/2} \log f_{\theta}(\omega_k) \end{aligned} \quad (3.35)$$

and $f_{\theta}(\cdot) = (\kappa/\sigma^2)f_{\beta}(\cdot)$. For the case of SRD processes, Hannan [47] showed that Whittle's estimate is asymptotically consistent, efficient and normal. For the LRD case, Fox and Taqqu [39] showed that the WE is asymptotically consistent and normal, while Dahlhaus [24] established its asymptotic efficiency.

Algorithm 5 presents the pseudocode for computing the cost function $V_{\text{WE}}(\theta|X)$.

Algorithm 5 Computing the Cost Function of the CE Estimate

- 1: Compute the spectral density function of the chosen model for the frequencies $\omega_k = i2\pi k/N$, $k = 1, 2, \dots, N-1$. See chapter 2 for details on the individual models.
 - 2: Compute the periodogram of the time series as per equation (3.14).
 - 3: Truncate the vectors containing the spectral density function and periodogram values up to $N-1/2$.
 - 4: Compute $V_{\text{WE}}(\theta|X)$ as per equation (3.35).
-

3.3.3 Circulant Embedding Estimator

In order to cope with the computational burden of the MLE, this section proposes a method which consists in embedding the covariance matrix of the process into an asymptotically equivalent circulant matrix. The embedding provides an algorithm with a computational complexity of order $\mathcal{O}(N \log N)$ given the simplicity of the eigenstructure of the circulant matrix. This method is referred to as the Circulant Embedding (CE) estimator.

The development of the CE estimator will be presented and its asymptotic properties will be shown in this section. The properties of the circulant embedding matrix are used to show the relationship between the MLE, the WE and the CE Estimator.

Let $C = C(\theta)$ be an $N \times N$ symmetric Toeplitz matrix with first row elements

$$c_\theta(k) = r_\theta(k) + s_\theta(k) \quad (3.36)$$

with $k \in \mathbb{J}$. Here, $r_\theta(\cdot)$ is given by equation (3.22) and denotes the first row elements of the symmetric Toeplitz matrix R defined in (3.21). The term $s_\theta(\cdot)$ is given by

$$s_\theta^T = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ r_\theta(m_1) - r_\theta(m_1 + 1) \\ r_\theta(m_1 - 1) - r_\theta(m_1 + 2) \\ r_\theta(m_1 - 2) - r_\theta(m_1 + 3) \\ \vdots \\ r_\theta(2) - r_\theta(N - 2) \\ r_\theta(1) - r_\theta(N - 1) \end{bmatrix} \quad (3.37)$$

with $m_1 = (N - 1)/2$ and N odd, or

$$s_\theta^T = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ r_\theta(m_2 - 1) - r_\theta(m_2 + 1) \\ r_\theta(m_2 - 2) - r_\theta(m_2 + 2) \\ r_\theta(m_2 - 3) - r_\theta(m_2 + 3) \\ \vdots \\ r_\theta(2) - r_\theta(N - 2) \\ r_\theta(1) - r_\theta(N - 1) \end{bmatrix} \quad (3.38)$$

with $m_2 = N/2$ and N even. Note that s_θ is the first row of a symmetric Toeplitz such that

$$C = R + S . \quad (3.39)$$

First, we will show that the symmetric Toeplitz matrix C with first row elements given by the perturbed vector c_θ is a circulant matrix.

Lemma 3.3.3. *The symmetric Toeplitz matrix C with first row elements c_θ given by equation (3.36) is a circulant matrix.*

Proof. Suppose N is even. Let R and S be symmetric Toeplitz matrices with first row elements r_θ and s_θ respectively. Let r_θ be given by equation (3.22) and s_θ be given by equation (3.38), such that

$$C = R + S .$$

Let the forward-shift permutation matrix be given by

$$\Pi = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} .$$

Then, it suffices to show that

$$\Pi R - R \Pi + \Pi S - S \Pi = 0 \quad (3.40)$$

holds [see theorem 3.1.1 in 26, pp. 67]. By performing the algebra on the left hand side of equation (3.40) we have

$$\Pi R - R\Pi + \Pi S - S\Pi = \begin{bmatrix} \varphi(1) & 0 & 0 & \cdots & 0 \\ \varphi(2) & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \varphi(N-1) & 0 & 0 & \cdots & 0 \\ 0 & \varphi(1) & \varphi(2) & \cdots & \varphi(N-1) \end{bmatrix}$$

where

$$\varphi(\tau) = r(\tau) - r(N - \tau) + s(\tau) - s(N - \tau)$$

for $\tau = \{1, 2, \dots, N - 1\}$. By substituting equation (3.38) we have that $\varphi(\tau) = 0$ for all τ and hence the result. The proof is similar for N odd. \square

Corollary 3.3.4. *Let $C = R + S$ be a matrix with first row elements, c_θ , given by equation (3.36). If*

$$r_\theta(k) = \mathcal{A}\delta_k$$

with δ_k the Kroenecker delta given by

$$\delta_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}. \quad (3.41)$$

and some finite constant \mathcal{A} then $C = R$.

Proof. The result follows immediately from the residual elements s_θ in equations (3.38) for N even and (3.37) for N odd. \square

The $N \times N$ matrix C may be seen as an embedding of a symmetric Toeplitz matrix \bar{R} of dimension m_j , $j = 1, 2$ and $m_1 = (N - 1)/2$ if N is odd or $m_2 = N/2$ if N is even. That is, the matrix \bar{R} is encapsulated within the larger matrix C . Note that C is the optimal case of the embedding proposed by Davies and Harte [25] for generating artificial realisations of length m_j for fGN (see chapter 2, section 2.4.1).

The simplicity of the eigenstructure of the circulant matrix C allows for efficient computational manipulations. Thus, provided that C behaves similarly to the covariance matrix R , we can develop an estimator based on the circulant embedding. That is, if we can establish the asymptotic equivalency between C and

R , then we can justify an estimator based on C as an asymptotically equivalent approximation to the MLE.

Definition 3.3.5. We say that a pair of $N \times N$ non-random matrices P and Q are asymptotically equivalent, written $P \sim Q$, if

$$\lim_{N \rightarrow \infty} \|P - Q\|_{\text{HS}} = 0 \tag{3.42}$$

where $\|\cdot\|_{\text{HS}}$ is the Hilbert-Schmidt (HS) norm of a matrix and it is given by

$$\|P - Q\|_{\text{HS}}^2 \triangleq \frac{1}{N} \text{trace} [(P - Q)^T (P - Q)] \ . \tag{3.43}$$

In theorem 3.3.6 we show that C is an approximation to the MLE in the sense that the HS-norm of the matrix of residuals S , $\|S\|_{\text{HS}}$, decays to zero as N grows large.

Theorem 3.3.6. *Let C be a symmetric circulant matrix with first row elements, c_θ , given by equation (3.36). Then, $C \sim R$ for all $d \in (-1/2, 1/4)$ for the case of the FD family of models or $H \in (0, 3/4)$ for the case of fGN.*

Proof. See appendix 3.A. □

Theorem 3.3.6 shows that the circulant matrix C is asymptotically equivalent to the matrix R . This implies that asymptotically, the eigenstructure of both matrices is equivalent. Therefore, an estimator based on the CE is justified at least for the validity regions $d \in (-1/2, 1/4)$ and $H \in (0, 3/4)$ for the models of chapter 2.

$$\hat{\theta}_{\text{CE}} = \arg \min_{\theta} V_{\text{CE}}(\theta|X) \tag{3.44}$$

where

$$\begin{aligned} 2V_{\text{CE}}(\theta|X) = & N [\log(2\pi) + 1 - \log(N)] \\ & + \sum_{k=0}^{N-1} \log \lambda_k + N \log (X^T C^{-1} X) \end{aligned} \tag{3.45}$$

is the CE approximation to the Gaussian likelihood and λ_k , is the k^{th} eigenvalue of C . The pseudocode for computing $V_{\text{CE}}(\theta|X)$ is presented in algorithm 6.

Algorithm 6 Computing the Cost Function of the CE Estimate

- 1: Generate the sequence of covariance values r_θ of length N as per equation (3.22).
 - 2: Generate the sequence of residuals s_θ as per equation (3.38) or (3.37).
 - 3: Obtain the sequence c_θ as per equation (3.36).
 - 4: Compute the eigenvalues λ_k by computing the FFT of c_θ .
 - 5: Compute the inverse of the matrix by use of the FFT. In MatLab code this would be `ic = real(fft(1./eC/N))` where `eC` is a row vector containing the set of eigenvalues computed in the previous step.
 - 6: Use the eigenvalues to compute equation (3.45). Note that the term $X^T C^{-1} X$ can be computed by repeated use of the FFT. In MatLab code this would be `x'*ifft(fft(ic')*.fft(x))` were `x` is a column vector containing the set of observations and `ic` is a row vector containing the inverse of the matrix C .
-

We now proceed to justify the CE estimator for the entire intervals $d \in (-1/2, 1/2)$ and $H \in (0, 1)$ for the models of chapter 2. Furthermore, we show that the CE estimator is asymptotically normal, consistent and efficient.

Theorem 3.3.7. *For all $d \in (-1/2, 1/2)$ and $H \in (0, 1)$, the CE Estimate, $\hat{\theta}_{CE}$, in equation (3.44) is asymptotically equivalent to the WE, $\hat{\theta}_{WE}$, in equation (3.34).*

Proof. We have that the matrices C and G , given by equations (3.39) and (3.30) respectively, are circulant matrices. Then it suffices to show that their eigenvalues are asymptotically equivalent.

In the limit, the k^{th} eigenvalue of C is given by [26]

$$\lim_{N \rightarrow \infty} \lambda_k = \lim_{N \rightarrow \infty} \sum_{\tau=0}^{N-1} c_\theta(\tau) u^{-k\tau}$$

with u the root of unity given by equation (2.33). Since $r_\theta(\cdot)$ is an even function, then by equation (3.36) we can write

$$\begin{aligned} \lim_{N \rightarrow \infty} \lambda_k &= r_\theta(0) + 2 \lim_{N \rightarrow \infty} \sum_{\tau=1}^{N-1} r_\theta(\tau) u^{-k\tau} \\ &= \sum_{\tau=-\infty}^{\infty} r_\theta(\tau) u^{-k\tau} . \end{aligned}$$

By the definition of spectral density function [see for example 3, 74] we have

$$\lim_{N \rightarrow \infty} \lambda_k = 2\pi f_\theta(\omega_k) \quad (3.46)$$

with $\omega_k = i2\pi k/N$. Since $f_\theta(\omega_k)$ are the eigenvalues of G then the result follows. \square

Corollary 3.3.8. *The CE estimator, $\hat{\theta}_{CE}$ is asymptotically normal, consistent and efficient.*

Proof. The result follows immediately from theorem 3.3.7. \square

Theorem 3.3.7 establishes that the CE estimator is asymptotically equivalent to the WE. Hence, by theorem 3.3.6, we have addressed for the intervals $d \in (-1/2, 1/4)$ and $H \in (0, 3/4)$ the subtle yet important point on the quality of the approximation of the WE to the MLE, a point left unaddressed by Whittle as pointed out in [17]. Corollary 3.3.8 justifies the CE estimator along the entire ranges $d \in (-1/2, 1/2)$ and $H \in (0, 1)$.

Further analysis into equation 3.46 shows that the matrices C , R and G are ill-conditioned.

Theorem 3.3.9. *The circulant matrix C with first row elements c_θ given by equation (3.36) is ill-conditioned, that is,*

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{cond}(C) &= \lim_{N \rightarrow \infty} \frac{\max \lambda_k}{\min \lambda_k} \\ &= \infty \end{aligned} \quad (3.47)$$

for all $d \in (-1/2, 1/2)$ and $d \neq 0$ for the case of the FD family of models and $H \in (0, 1)$ and $H \neq (1/2)$ for the case of fGN.

Proof. We have from equation (3.46)

$$\begin{aligned} \lim_{N \rightarrow \infty} \max_k \lambda_k &= \frac{2\pi\kappa}{\sigma^2} \max f_\beta(\omega) \\ \lim_{N \rightarrow \infty} \min_k \lambda_k &= \frac{2\pi\kappa}{\sigma^2} \min f_\beta(\omega) . \end{aligned}$$

Let $d < 0$ for the FD family of models or $H < 1/2$ for fGN, then

$$\begin{aligned} \min f_\beta(\omega) &= \lim_{\omega \rightarrow 0} f_\beta(\omega) \\ &= 0 \end{aligned}$$

and $0 < \max f_\beta(\omega) < \infty$. If $d > 0$ for the FD family of models or $H > 1/2$ for fGN, then

$$\begin{aligned} \max f_\beta(\omega) &= \lim_{\omega \rightarrow 0} f_\beta(\omega) \\ &= \infty \end{aligned}$$

and $0 < \min f_\beta(\omega) < \infty$. Thus the result. \square

Theorem 3.3.10. *The circulant matrix G , given by equation (3.30), is asymptotically ill-conditioned for all $d \in (-1/2, 1/2)$ and $d \neq 0$ for the case of the FD family of models and $H \in (0, 1)$ and $H \neq (1/2)$ for the case of fGN.*

Proof. The result follows immediately from theorems 3.3.7 and 3.3.9. \square

Theorem 3.3.11. *The symmetric Toeplitz matrix R , given by equation (3.21), is asymptotically ill-conditioned for all $d \in (-1/2, 1/4)$ and $d \neq 0$ for the case of the FD family of models and $H \in (0, 3/4)$ and $H \neq (1/2)$ for the case of fGN.*

Proof. The result follows immediately from theorems 3.3.6 and 3.3.9. \square

Theorem 3.3.11 only guarantees the ill-conditioning of the matrix R for the intervals $d \in (-1/2, 1/4)$, $d \neq 0$ and $H \in (0, 3/4)$, $H \neq 1/2$ for the models of chapter 2. Nevertheless, numerical evidence suggests that this behaviour is still present for $d > 1/4$ and $H > 3/4$. This is illustrated in figure 4.2.

Under the ill-conditioning circumstances, we have that in the presence of measurement errors, the computation of $C^{-1}X$, $G^{-1}X$ and $R^{-1}X$ in $V_{\text{CE}}(\theta|X)$, $V_{\text{WE}}(\theta|X)$ and $V(\theta|X)$ respectively may be poor [42] and thus the estimators may present a significant bias.

The ill-conditioning issue was mentioned hinted at in [37, 67]. The simulation studies in [67] showed that the MLE has a degraded performance when the set of observations has been contaminated with errors that account to a signal-to-noise-ratio (SNR) of 30 decibels (dB). Chen et al. [20] mentions, without a formal proof, that the MLE is ill-conditioned and provides a solution by means of a preconditioned conjugate gradient (PCG) algorithm. The PCG has a computational complexity of order $\mathcal{O}(N[\log N]^{5/2})$ to invert the matrix R . Nevertheless, the determinant of the preconditioned covariance matrix must be approximated (which can be computationally expensive). In the simulation studies presented in [20], it is shown that the WE slightly outperforms the PCG when the observations are not corrupted with noise.

3.4 Chapter Conclusions

Parameter estimation of LRD processes is important as they appear to be ubiquitous in several scientific areas. There exists several semi-parametric estimation methods available in the literature. These can be used as a starting point for the analysis of the time-series at hand. However, as noted by [9, 88] and the Monte Carlo studies of [102], parametric estimation methods based on a parsimonious model will be a preferred approach given their statistical properties. Among these, the likelihood based methods excel in accuracy; however, their computational implementation is not straightforward.

In order to simplify the implementation of the maximum likelihood estimate, one might use an approximate approach provided by either the Whittle Estimate or the Circulant Embedding Estimate. The CE estimate may be seen as a time-domain approach for performing the Whittle approximation. This is in the vein of Blackman-Tukey's approach to construct a power spectrum versus Welch's method. There are clear advantages to using the CE estimate and the WE depending on the model selection. For example, if the model selected is fGN then we have from point (d) in remark 2.3.6 that the spectral density function must be approximated and thus it makes sense to compute the covariances and perform the embedding. The converse is also true, that is, when the covariance function is difficult to evaluate, one might prefer to use the WE over the CE estimate. Furthermore, from the Monte Carlo studies presented in chapter 5 we can observe that the CE estimate is more accurate than the WE when the length of the observations is small, a reason why the WE is criticised [48].

Furthermore, we have shown that all three likelihood based methods presented in the chapter are sensitive to error corruptions due to the ill-conditioning of the covariance matrices involved in computing the estimates.

3.A Proof of Theorem 3.3.6

The proof for this theorem will be divided into three sections: 1) For the case $d = 0$ and $H = 1/2$, 2) for antipersistent processes, that is, $d < 0$ and $H < 1/2$ and 3) for LRD processes in the ranges $d \in (0, 1/4)$ and $H \in (1/2, 3/4)$.

- (a) Let $d = 0$ for the FD family of models or $H = 1/2$ for the case of fGN. From point (a) in lemma 3.B.1 and corollary 3.3.4 we have $S = 0$ and thus

the result.

- (b) Suppose N is even and let $d < 0$ for the FD family of models and $H < 1/2$. Then from the definition of the Hilbert-Schmidt norm and the construction of the matrix S we can write

$$\|S\|_{\text{HS}}^2 = \frac{2}{N} \sum_{k=1}^{m_2-1} (m_2 - k) |s_\theta(m_2 + k)|^2 \quad (3.48)$$

with $m_2 = N/2$ and s_θ given by equation (3.38).

- (i) Let $r_\theta(\cdot)$ denote the scaled covariance matrix of a fGN process or FDN process. From point (d) in lemma 3.B.1 we have

$$\begin{aligned} \|S\|_{\text{HS}}^2 &< \frac{2}{N} \sum_{k=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 - k)|^2 \\ &= \frac{2}{N} \sum_{k=1}^{m_2-1} k |r_\theta(k)|^2 \\ &= \Upsilon \frac{z_1}{N} \end{aligned}$$

with

$$z_1 = \begin{cases} \sum_{k=1}^{m_2-1} k [|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}]^2 & \text{for fGN} \\ \sum_{k=1}^{m_2-1} k \left[\frac{\Gamma(k+d)}{\Gamma(k+1-d)} \right]^2 & \text{for FDN} \end{cases} \quad (3.49)$$

and

$$\Upsilon = \begin{cases} \frac{1}{2} & \text{for fGN} \\ 2 \left| \frac{\Gamma(1/2-d)}{4^d \Gamma(d)} \right|^2 & \text{for FDN} \end{cases} . \quad (3.50)$$

Thus we have that

$$\lim_{N \rightarrow \infty} \|S\|_{\text{HS}}^2 < \Upsilon \lim_{N \rightarrow \infty} \frac{z_1}{N} .$$

From lemmas 3.B.2 and 3.B.3 we have that

$$\lim_{N \rightarrow \infty} z_1 < \infty .$$

Therefore

$$\lim_{N \rightarrow \infty} \|S\|_{\text{HS}}^2 = 0$$

holds for all $H \in (0, 1/2)$ and $d \in (-1/2, 0)$. The proof is similar for N odd.

- (ii) Let $r_\theta(\cdot)$ denote the scaled covariance matrix of an ARFIMA process. By the triangle inequality [see for example equation 3.2.5, p. 11 in 1] we have

$$\|S\|_{\text{HS}}^2 < \frac{2}{N} z_3$$

with

$$z_3 = \sum_{k=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 - k)|^2 + \sum_{k=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 + k)|^2 . \quad (3.51)$$

Thus

$$\lim_{N \rightarrow \infty} \|S\|_{\text{HS}}^2 < 2 \lim_{N \rightarrow \infty} \frac{z_3}{N} .$$

From lemma 3.B.4 we have that

$$\lim_{N \rightarrow \infty} z_3 < \infty .$$

Therefore

$$\lim_{N \rightarrow \infty} \|S\|_{\text{HS}}^2 = 0$$

holds for all $d \in (-1/2, 0)$. The proof is similar for N odd.

- (c) Suppose N be even and let $m_2 = N/2$. Furthermore, let $d \in (0, 1/4)$ for the FD family of models and let $H \in (1/2, 3/4)$ for fGN.

- (i) Let r_θ denote the covariance function of either FDN or fGN. Since r_θ is non-negative for all $d > 0$ and $H > 1/2$ for these two models we can write

$$\|S\|_{\text{HS}}^2 < \frac{2}{N} \sum_{k=1}^{m_2-1} k [r_\theta(k)]^2 .$$

It follows from equations (2.14) and (2.22) that for a sufficiently large

$$\mathcal{M} < \infty$$

$$r_\theta \leq \mathcal{M}k^{\theta_{\text{LRD}}}$$

holds for all $k \in \mathbb{Z}^+, \mathbb{Z}^+\{1, 2, \dots\}$, where

$$\theta_{\text{LRD}} = \begin{cases} 2d - 1 & \text{for FDN} \\ 2H - 2 & \text{for fGN} \end{cases} .$$

Thus, it follows that

$$\frac{N}{2} \|S\|_{\text{HS}}^2 < \mathcal{M} \sum_{k=1}^{m_2-1} k [g(k)]^2 .$$

Approximating the sum on the right hand side with the Euler-Maclaurin formula yields

$$\frac{N}{2} \|S\|_{\text{HS}}^2 < \mathcal{M} \int_1^{m_2-1} t [g(t)]^2 dt + \mathcal{K}$$

for some finite constant \mathcal{K} . Hence, for the case of FDN we have

$$m_2 \|S\|_{\text{HS}}^2 < \frac{\mathcal{M}}{4d} \left[\frac{(m_2 - 1)^{4d}}{m_2} - \frac{1}{m_2} \right] + \frac{\mathcal{K}}{m_2}$$

and for the case of fGN we have

$$m_2 \|S\|_{\text{HS}}^2 < \frac{\mathcal{M}}{4H - 2} \left[\frac{(m_2 - 1)^{4H-2}}{m_2} - \frac{1}{m_2} \right] + \frac{\mathcal{K}}{m_2}$$

Therefore

$$\lim_{m_2 \rightarrow \infty} \|S\|_{\text{HS}}^2 = 0$$

holds for all $d \in (0, 1/4)$ and $H \in (1/2, 3/4)$. The proof is similar for N odd.

- (ii) Let r_θ be the covariance function of and ARFIMA process. The result follows from equation (3.64), (2.28) and the development of the previous point.

□

3.B Useful lemmas

Lemma 3.B.1. *Suppose $X = X_k$, $k \in \mathbb{J}$ is a stationary sequence of either fGN or FDN with covariance function $\gamma_\beta(k)$. Let $r_\theta(k)$ be given by equation (3.22). Then*

(a) *If $H = 1/2$ for fGN or $d = 0$ for FDN we have*

$$r_\theta(k) = \frac{\sigma^2}{\kappa} \delta_k \quad (3.52)$$

with δ_k the Kroenecker delta in equation (3.41).

(b) *If $H \neq 1/2$ for fGN, $d \neq 0$ for FDN then*

$$\left| \frac{r_\theta(k+1)}{r_\theta(k)} \right| < 1 . \quad (3.53)$$

(c) *Let N be even and $\tau = 0, 1, \dots, m_2 - 1$ with $m_2 = N/2$. If $H \neq 1/2$ for fGN and $d \neq 0$ for FDN then*

$$|r_\theta(m_2 - \tau) - r_\theta(m_2 + \tau)| \leq |r_\theta(m_2 - \tau)| \quad (3.54)$$

with equality if and only if $\tau = 0$.

(d) *Let N be odd and $\tau = 0, 1, \dots, m_1 - 1$, $m_1 = (N - 1)/2$. If $H \neq 1/2$ for fGN and $d \neq 0$ for FDN then*

$$|r_\theta(m_1 - \tau) - r_\theta(m_1 + \tau)| \leq |r_\theta(m_1 - \tau)| \quad (3.55)$$

with equality if and only if $\tau = 0$.

Proof. The proof is as follows.

(a) Follows immediately from the definitions of each process, that is, for such conditions, X is i.i.d.

(b) The proof for fGN is given in [100]. Let X be FDN. We have that [see equation 6.1.15 in 1]

$$\Gamma(z + 1) = z\Gamma(z) . \quad (3.56)$$

Then, we have

$$\left| \frac{r_\theta(k+1)}{r_\theta(k)} \right| = \left| \frac{k+d}{k+1-d} \right|.$$

Thus, it suffices to show that

$$|k+d| < |k+1-d| \quad (3.57)$$

holds for all $k \in \mathbb{J}$. Let $k = 0$, then we have

$$|d| < |1-d|$$

which holds for $d < 1/2$. Let $k \geq 1$, then we can write equation (3.57) as

$$k+d < 1+k-d$$

which also holds for all $d < 1/2$. Since X is stationary, then the result follows.

- (c) Let N be even. if $d > 0$ for FDN or $H > 1/2$ for fGN we have from equation (3.53)

$$r_\theta(m_2 - \tau) \geq r_\theta(m_2 + \tau) \geq 0$$

with equality if and only if $\tau = 0$. If $d < 0$ for FDN or $H < 1/2$ for fGN, then we have

$$r_\theta(m_2 - \tau) \leq r_\theta(m_2 + \tau) \leq 0$$

with equality if and only if $\tau = 0$.

- (d) The proof is similar to point (c), except $t = 0, 1, \dots, m_1 - 1$.

□

Lemma 3.B.2. *Let z_1 be given by*

$$z_1 = \sum_{k=1}^N g_d(k) \quad (3.58)$$

with

$$g_d(k) = k \left[\frac{\Gamma(k+d)}{\Gamma(k+1-d)} \right]^2. \quad (3.59)$$

If $d \in (-1/2, 0)$, then

$$\lim_{N \rightarrow \infty} z_1 < \infty$$

Proof. From [eq. 6.1.46, p. 257 in 1] we have that

$$\lim_{n \rightarrow \infty} n^{s-z} \frac{\Gamma(n+z)}{\Gamma(n+s)} = 1 . \quad (3.60)$$

Let $z = d$, $s = 1 - d$, $n = k$ and square both sides of equation (3.60) then we have

$$\begin{aligned} \lim_{k \rightarrow \infty} k^{2-4d} \left[\frac{\Gamma(k+d)}{\Gamma(k+1-d)} \right]^2 &= \lim_{k \rightarrow \infty} k^{1-4d} g_d(k) \\ &= 1 . \end{aligned}$$

Since

$$\sum_{k=1}^{\infty} \frac{1}{k^{1-4d}} < \infty$$

for $d \in (-1/2, 0)$, then the results follows by the limit comparison test, [see for example theorem 2, p. 468 in 99]. \square

Lemma 3.B.3. *Let z_1 be given by*

$$z_1 = \sum_{k=1}^N g_H(k) \quad (3.61)$$

with

$$g_H(k) = k \left[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H} \right]^2 . \quad (3.62)$$

If $H \in (0, 1/2)$, then

$$\lim_{N \rightarrow \infty} z_1 < \infty$$

Proof. From [100] we have for $k \rightarrow \infty$

$$k^2 \left[(1 + 1/k)^{2H} - 2 + (1 - 1/k)^{2H} \right] \rightarrow 2H(2H - 1) .$$

Hence

$$\frac{g_H(k)}{k^{3-4H}} \rightarrow 4H^2(2H - 1)^2 \quad (3.63)$$

as $k \rightarrow \infty$. Since

$$\sum_{k=1}^{\infty} \frac{1}{k^{3-4H}} < \infty$$

for $H \in (0, 1/2)$ then the result follows by the limit comparison test [see for example theorem 2, p. 468 in 99]. \square

Lemma 3.B.4. *Let z_3 be given by equation (3.51), i.e.,*

$$z_3 = \sum_{k=1}^{m_2-1} k |r_\theta(k)|^2 + \sum_{k=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 + k)|^2$$

with $r_\theta(\cdot)$ given by (3.22) and $m_2 = N/2$. If $d \in (-1/2, 0)$ then

$$\lim_{N \rightarrow \infty} z_3 < \infty .$$

Proof. For m_2 sufficiently large

$$|r_\theta(m_2 + \tau)| < |r_\theta(m_2 - \tau)|$$

for $\tau = 1, 2, \dots, m_2$. Hence

$$\sum_{\tau=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 + \tau)|^2 < \sum_{\tau=1}^{m_2-1} (m_2 - k) |r_\theta(m_2 - \tau)|^2 . \quad (3.64)$$

Thus, it suffices to show that first sum in z_3 is finite. From [14, p. 525] we have for $k \rightarrow \infty$

$$\frac{r_\theta(k)^2}{k^{2-4d}} \rightarrow (\varsigma_d)^2$$

with $\varsigma_d \neq 0$ and independent of k . Since

$$\sum_{k=1}^{\infty} \frac{1}{k^{1-4d}} < \infty$$

for $d \in (-1/2, 0)$ then the result follows by the limit comparison test [see for example theorem 2, p. 468 in 99]. \square

Chapter 4

Regularised Estimators

4.1 Introduction

Chapter 3 surveyed three likelihood based estimators for long-range dependent processes. Under the LRD context, the statistical properties of the likelihood based methods are desirable. However, theorems 3.3.9, 3.3.10 and 3.3.11 imply that these likelihood based methods are highly sensitive to errors in the observations.

The sensitivity of the likelihood based methods stems from the matrix inversion required to compute their respective cost functions. The sensitivity of the likelihood based methods to the presence of measurement errors is manifested by a relevant amount of bias in the parameter estimates.

This chapter provides a solution that reduces the bias produced by the presence of errors in the measurements. Furthermore, the solution corrects the ill-conditioning of the likelihood methods of chapter 3. Section 4.2 formalises the problem statement. Section 4.3 surveys a regularisation method that alleviates the ill-conditioning of a matrix. Section 4.4 proposes a solution that reduced the bias produced by the presence of errors in the measurements. The chapter conclusions are presented in section 4.5.

4.2 Problem Statement

Let $Y = Y_k$, $k \in \mathbb{J}$, be a series of equidistant observations. These observations consist of a realisation of a LRD process subject to additive errors. That is,

$$Y_k = X_k + \epsilon_k \quad (4.1)$$

with $\epsilon = \epsilon_k$ a white noise sequence with variance $\nu_\epsilon^2 < \infty$ and $X = X_k$ one of the weakly stationary, invertible LRD processes presented in chapter 2. Suppose X and ϵ are independent.

If the observations, Y , are such that $\nu_\epsilon^2 = 0$ and if the computer precision is good enough, i.e., the machine epsilon is sufficiently small, the round-off errors will not affect the performance of the likelihood based estimators [42]. This is clear from the simulation results of chapter 5.

However, if $\nu_\epsilon^2 \neq 0$, the likelihood based methods will be sensitive to the effects of ϵ . This sensitivity will be apparent in a relevant bias of the estimates $\hat{\theta}$. To illustrate, both the Whittle and CE estimation methods of chapter 3 may be interpreted as finding the best fit to the periodogram ordinates computed from the data from a family of spectral density functions. When $\nu_\epsilon^2 \neq 0$, the underlying LRD process will be dominant at lower frequencies. Nevertheless, there is a cross-over frequency at which the ϵ will be dominant. This change in the periodogram ordinates is what biases the spectral density fit as the estimators try to accommodate the occurrences at higher frequencies. This is illustrated in figure 4.1.

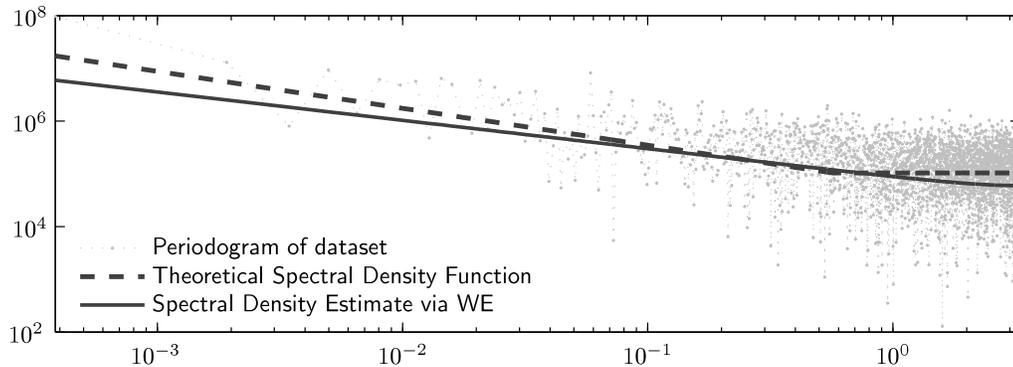
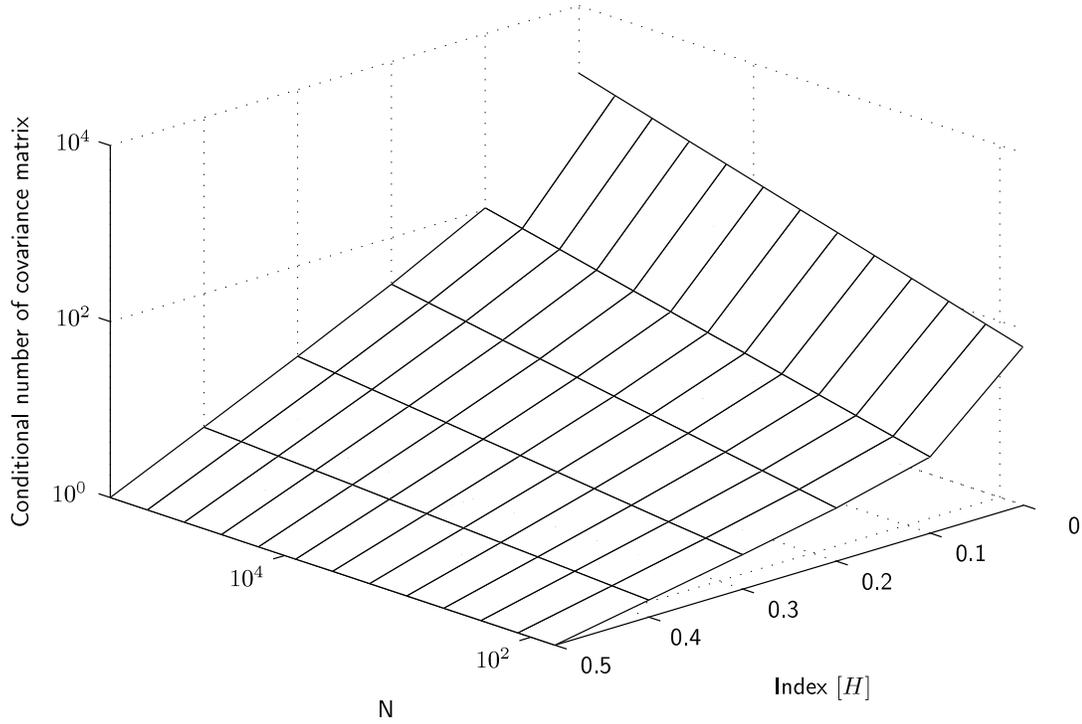


Figure 4.1: The figure portrays the periodogram ordinates computed from a set of corrupted observations y . The dashed line shows the theoretical spectral density of the process y . The continuous line shows the spectral density function of fGN with fractional index, $\hat{\theta}_{WE}$, obtained through the WE.

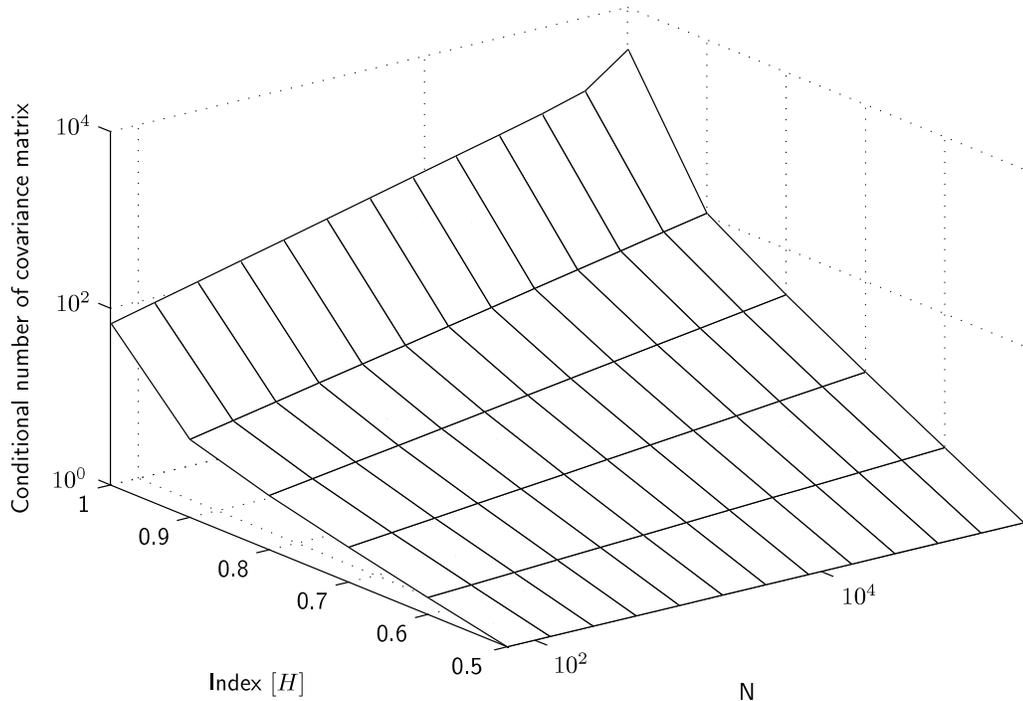
Lundahl et al. [67] noted this behaviour. He generated fGN sequences and corrupted them with additive Gaussian errors with 30db to 10 db SNR. He observed that under error corruption, the MLE performance was biased. Furthermore, he noted that the bias on the estimates was not uniform.

The bias on the parameter estimates stems from the matrix inversion required to compute their respective cost functions. Chapter 3 shows that the condition number of these matrices grows along with their dimension, that is, the larger the matrix the more ill-conditioned it becomes. Thus, under the effect of ϵ , the likelihood based estimates are difficult to compute. Furthermore, the bias of the estimates will not be reduced by collecting larger amounts of observations.

The lack of uniformity on the bias of the parameter estimates is not surprising in the light of the ill-conditioning of the covariance matrix of a LRD process. From the behaviour of the eigenvalue distribution given the LRD coefficient of the process, e.g., as $H \in (0, 1)$ or $d \in (-1/2, 1/2)$ approaches the extrema of their respective validity intervals, the condition number of the covariance matrix of the process deteriorates regardless of its dimension. Such behaviour is illustrated in figure 4.2.



(a) Condition number of Σ with $H < 1/2$



(b) Condition number of Σ with $H > 1/2$

Figure 4.2: The figure illustrates the behaviour of the conditional number of the covariance matrix of fGN. As N becomes large, the matrix condition number deteriorates, i.e., it grows large. The same phenomenon occurs as $H \rightarrow 1$ and $H \rightarrow 0$ regardless of the size of N .

4.3 Regularisation

Fitting the parameters of the underlying LRD process in Y by means of a likelihood based estimation method, such as the ones presented in chapter 3, requires the computation of

$$\begin{aligned} z &= P^{-1}Y \\ &= P^{-1}(X + \epsilon) . \end{aligned} \tag{4.2}$$

Chapter 3 shows that in the case of LRD, the matrix P is ill-conditioned. This implies that a small ϵ will introduce a large variance in the solution z . In other words, the solution z has little relationship between the cases $\nu_\epsilon^2 = 0$ and $\nu_\epsilon^2 \neq 0$.

An alternative is provided by solving a well-conditioned approximation of the ill-conditioned linear system, that is, to solve

$$\hat{z}_R = (P + \alpha I)^{-1}Y . \tag{4.3}$$

This is referred to as regularisation with $\alpha > 0$ as a regularisation parameter [77].

If P has been properly scaled, then the minimum eigenvalue of the matrix $P + \alpha I$ will be bounded below by α as P is positive definite and symmetric. It follows that α will represent a trade-off between the variance of the solution \hat{z}_R and its bias, $|\hat{z}_R - z|$.

The choice of the regularisation method, αI , guarantees that the matrix structure of P is preserved. Therefore, the computational complexity of the regularised likelihood methods is left unaffected, that is, the regularised WE and CE estimators remain with a computational complexity of order $\mathcal{O}(N \log N)$ while the regularised MLE can be solved in $\mathcal{O}(N^2)$ operations via the Levinson-Durbin algorithm. While the regularisation method choice might not be the optimal one [77], the ill-conditioning of the problem is mitigated while the WE and CE estimators remain computationally efficient.

Selection of the Regularisation Parameter

The performance of a regularisation method depends directly on the suitable choice of α . If α is selected to be small, the ill-conditioning of the matrix might not be addressed properly. However, if α is chosen to be large enough, the bias introduced by the approximation \hat{z}_R will be large.

There exist several strategies for choosing the regularisation parameter in the context of least squares and ridge regression. These strategies depend on additional information inferred from the analysed problem and its solution. Among others, the discrepancy principle [33] and the L-curve [58, 60] criterion are some of the most well known strategies for choosing the regularisation parameter [58]. However they have not been used to estimate the parameters of a LRD process. Both the discrepancy principle and the L-curve criterion are iterative strategies that require several computations of the solution \hat{z}_R . This kind of procedure would turn the parameter estimation for LRD processes into a tedious and computationally expensive task.

A different approach can be obtained through a Bayesian interpretation. Assuming z and Y are jointly distributed continuous random variables we have by Bayes rule [18]

$$p(z|Y) = \frac{p(Y|z)p(z)}{p(Y)}$$

where $p(z)$ is a *prior* density function. Suppose that $z = S\eta$ where S is a smoothing matrix that yields the regularisation method. If η and ϵ are uncorrelated random variables with zero mean and variances ν_η^2 and ν_ϵ^2 respectively, then [77]

$$\hat{\eta}_R = (S^*P^*PS + \alpha^2I)^{-1} S^*P^*Y$$

with $\alpha = \nu_\epsilon^2/\nu_\eta^2$. If $S = I$, then a particular regularisation method known as Tikhonov regularisation is recovered. To illustrate, let us look at the following example.

Example 4.3.1. Suppose η and ϵ are iid Gaussian distributed random variables with zero mean and variances ν_η^2 and ν_ϵ^2 respectively. Let $S = I$, then

$$p(z) = \frac{1}{[2\pi\nu_\eta^2]^{N/2}} \exp\left(-\frac{\|z\|^2}{2\nu_\eta^2}\right)$$

and

$$p(Y|z) = \frac{1}{[2\pi\nu_\epsilon^2]^{N/2}} \exp\left(-\frac{\|Pz - Y\|^2}{2\nu_\epsilon^2}\right)$$

with $\|\cdot\|$ the \mathcal{L}_2 norm. Hence the cost function

$$\begin{aligned} -\log p(z|Y) &= -\log p(Y|z) - \log p(z) \\ &= \|Pz - Y\|^2 + \alpha\|z\|^2 \end{aligned}$$

with $\alpha = \nu_\epsilon^2 / \nu_\eta^2$ has its minimum at

$$\hat{\eta}_R = \hat{z}_R = (P^*P + \alpha^2 I)^{-1} P^*Y .$$

For the particular case of LRD models the smoothness of S is questionable and the difficulties involved in computing the regularisation method through this approach are not investigated here. Nevertheless, in the sequel we propose a set of estimators that are related to the regularisation method $\hat{z}_R = (P + \alpha I)^{-1}Y$. The estimation of the regularisation parameter is performed simultaneously with the parameters of the LRD models.

4.4 Maximum Likelihood for Error Corrupted Observations

In the chapter introduction, we have noted that the presence of measurement errors, ϵ , will introduce a bias in the likelihood based estimates. This behaviour is accentuated by the ill-conditioning of the covariance matrix of the LRD processes. Nevertheless, this bias can be corrected by imposing statistical properties on the noise model and by taking advantage of its structure. Propositions 4.4.3 and 4.4.2 provide such solution.

The regularised estimators of propositions can also be equivalently justified from a Bayesian perspective. This follows from Bayes rule and its equivalency to maximum likelihood when a Gaussian prior is assumed. We do not pursue this approach here, however [65], points out the relationship between the Bayesian approach of Maximum A Posteriori and MLE. A much more in-depth treatment is provided in [18, 76].

Lemma 4.4.1. *Let ϵ be a realisation of an iid Gaussian process with zero mean and variance ν_ϵ^2 and let Y be given by equation (4.1). Then Y is a Gaussian process with covariance matrix*

$$\Psi = \Sigma + \nu_\epsilon^2 \tag{4.4}$$

where Σ is the covariance matrix of the LRD process X .

Proof. The proof is a standard result based on the characteristic function of a Gaussian process, see for example [74]. \square

Under lemma 4.4.1, we have from the development of the MLE in chapter 3, that the MLEs, $\hat{\beta}$ and $\hat{\nu}_\epsilon^2$ are given by

$$\left\{ \hat{\beta}, \hat{\nu}_\epsilon^2 \right\} = \arg \min_{\beta, \nu_\epsilon^2} Q(\beta, \nu_\epsilon^2 | Y) \quad (4.5)$$

where

$$2Q(\beta, \nu_\epsilon^2 | Y) = N \log(2\pi) + \log(\det \Psi) + Y^T \Psi^{-1} Y \quad (4.6)$$

is the log-likelihood and $\beta = [\theta, \sigma^2]$ is given by equations (3.26) and (3.27).

The optimisation in equation (4.5) is performed simultaneously over the parameter space $\{\theta, \sigma^2, \nu_\epsilon^2\}$. By introducing the hyper-parameters

$$\begin{aligned} \alpha &= \alpha(\sigma^2, \nu_\epsilon^2) = \frac{\kappa \nu_\epsilon^2}{\sigma^2} \\ \psi &= \psi(\sigma^2, \nu_\epsilon^2) = \frac{\sigma^2}{\kappa} \end{aligned} \quad (4.7)$$

with κ given by equation (3.28), such that

$$\Psi = \psi(R + \alpha I) \quad , \quad (4.8)$$

and by lemma 3.3.1, then the optimisation in equation (4.5) can be reduced to the parameter space $\{\theta, \alpha\}$, that is,

$$\{\hat{\theta}, \hat{\alpha}\} = \arg \min_{\theta, \alpha} V(\theta, \alpha | Y) \quad (4.9)$$

where

$$\begin{aligned} 2V_{\text{RWE}}(\theta, \alpha | Y) &= N [\log(2\pi) + 1 - \log(N)] \\ &\quad + \log(\det[R + \alpha I]) \\ &\quad + N \log(Y^T \hat{z}_{\text{RWE}}) \end{aligned} \quad (4.10)$$

and

$$\hat{z}_{\text{RWE}} = (G + \alpha I)^{-1} Y \quad . \quad (4.11)$$

In the same sense as in the previous chapter, the MLEs, $\hat{\beta}$ and $\hat{\nu}_\epsilon^2$ are expensive to compute. Hence, we introduce two computationally efficient approximations to these MLEs.

Proposition 4.4.2. *Given a series of equidistant observations Y as in equation*

(4.1) then the Regularised Whittle Estimate (RWE) is given by

$$\{\hat{\theta}_{RWE}, \hat{\alpha}_{RWE}\} = \arg \min_{\theta, \alpha} V_{RWE}(\theta, \alpha | Y) \quad (4.12)$$

with

$$\begin{aligned} 2V_{RWE}(\theta, \alpha | Y) = & N [\log(2\pi) + 1 - \log(N)] \\ & + \log(\det[G + \alpha I]) \\ & + N \log(Y^T \hat{z}_{RWE}) \end{aligned} \quad (4.13)$$

and

$$\hat{z}_{RWE} = (G + \alpha I)^{-1} Y .$$

Proposition 4.4.3. *Given a series of equidistant observations Y as in equation (4.1) then the Regularised Circulant Embedding (RCE) Estimate is given by*

$$\{\hat{\theta}_{RCE}, \hat{\alpha}_{RCE}\} = \arg \min_{\theta, \alpha} V_{RCE}(\theta, \alpha | Y) \quad (4.14)$$

with

$$\begin{aligned} 2V_{RCE}(\theta, \alpha | Y) = & N [\log(2\pi) + 1 - \log(N)] \\ & + \log(\det[C + \alpha I]) \\ & + N \log(Y^T \hat{z}_{RCE}) \end{aligned} \quad (4.15)$$

and

$$\hat{z}_{RCE} = (C + \alpha I)^{-1} Y .$$

By propositions 4.4.2 and 4.4.3 we try to find the best fit to the periodogram ordinates produced by the combination of the parameters θ and α . For the case of FDN and fGN propositions 4.4.2 and 4.4.3 extend the estimation framework from a line search to a surface search. This is illustrated in figure 4.3.

By theorem 4.4.4 shows the relationship between propositions 4.4.2 and 4.4.3 with the MLEs in equation (4.5).

Theorem 4.4.4. *The RWE and RCE estimators are asymptotically equivalent to the MLEs in equation (4.5) for $d \in (-1/2, 1/4)$ for the FD family of models and $H \in (0, 3/4)$ for fGN.*

Proof. The result follows immediately from theorems 3.3.6 and 3.3.7. \square

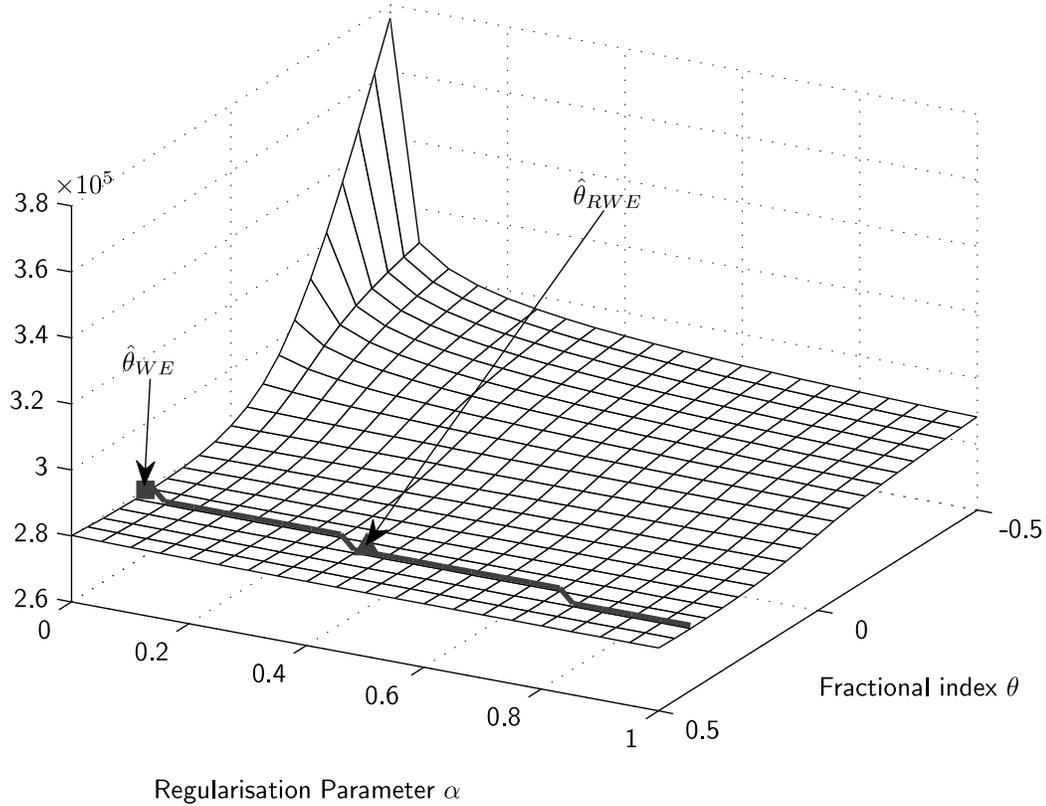


Figure 4.3: The figure illustrates the optimisation surface $Q_{\text{RCE}}(\theta, \alpha|Y)$ for a sequence of FDN. When $\alpha = 0$ the cost function $Q_{\text{CE}}(\theta|Y)$ is recovered and an estimate \hat{H}_{CE} can be found. The bold line shows the minima across the surface for each α . The estimate \hat{H}_{RCE} is the global minimum of the surface.

Finally, an estimate of the signal-to-noise ratio (SNR) can be obtained as follows.

Definition 4.4.5. Let Y be the process given by equation (4.1). Then we define the SNR of Y (in decibels) as

$$\text{SNR}_Y = 20 \log \frac{\sigma^2}{\nu_\epsilon^2} . \tag{4.16}$$

Lemma 4.4.6. Let Y be the process given by equation (4.1). Then the MLE estimate of the SNR is given by

$$\widehat{\text{SNR}}_Y = 20 \log \frac{\hat{\sigma}^2}{\nu_\epsilon^2} . \tag{4.17}$$

Proof. Proof follows immediately from the derivation of the MLEs at the beginning of this section. □

4.5 Conclusions

In this chapter we have presented a solution that corrects the bias introduced by measurement errors and that alleviates the ill-conditioning of the likelihood based methods. Under Gaussian conditions the regularised methods are optimal in the sense of maximum likelihood. Furthermore, they remain computationally efficient, that is, their order of computational complexity is $\mathcal{O}(N \log N)$.

It is important to note that while the estimate bias is reduced, the variance of the estimates is increased slightly as shown in chapter 5. While this phenomenon is expected as we are parting from an under-parameterised model to a properly parameterised one, a proof of this phenomenon is left open for further examination.

Finally, we have that the regularised likelihood methods will be slower to compute in comparison to the likelihood methods of chapter 3. This is due to the augmented parameter space for which the search must be performed, e.g., for fGN and FDN the regularised methods will optimise over a two dimensional parameter space while the standard likelihood methods will optimise over a one dimensional parameter space.

Chapter 5

Numerical Results

This chapter presents a series of numerical results that are related to the estimators presented in chapters 3 and 4. The objective is to present a fair comparison between some of the relevant estimators under the scenarios of error-free and error corrupted observations. In section 5.1 a Monte-Carlo comparison between the WE, the CE Estimator and the DFA is performed. Section 5.2 presents the comparison between the WE, CE Estimator and their regularised versions for the case of error-corrupted observations. Section 5.3 presents some real world examples in which the parameter estimation of LRD processes is useful. The appendices 5.A and 5.B contain the complete results of the Monte Carlo studies on the previous sections.

5.1 Scenario I: Sets of Error-Free Observations

5.1.1 Description of the Experiments

The numerical experiments were performed in a Monte-Carlo approach. The procedure was as follows:

- (a) The following parameter grid was employed:

$$\begin{aligned}H_0 &= \{0, 0.1, \dots, 1\} \\d_0 &= \{-0.5, -0.4, \dots, 0, 0.1, \dots, 0.5\} \\A_0(q) &= 1 + 0.5q^{-1} \\B_0(q) &= 1 - 0.3q^{-1} .\end{aligned}$$

Within this parameter grid there are 11 fGN processes, 11 FDN processes and 11 ARFIMA processes, with true parameters H_0 , d_0 , $A_0(q)$ and $B_0(q)$.

- (b) For each of the fGN, FDN and ARFIMA processes contained within the parameter grid, 100 realisations were generated. The circulant embedding algorithm described in chapter 2, section 2.4.1 was employed for generating the synthetic realisations.
- (c) For each experiment, a different realisation length was selected. The choices were

$$N = \{64, 128, 512, 1024, 2048, 16384\}$$

for fGN and FDN. For ARFIMA the choices were

$$N = \{1024, 2048, 16384\}.$$

Thus, the parameters of 3300 processes of length N were estimated. This gives a total of 16500 estimation problems solved. The objective of subdividing the estimation lengths is to analyse how the performances of the estimators improve or deteriorate with sample length.

- (d) For each one of the experiments, the parameter estimates were obtained by using the WE, the CE Estimator and the DFA. The WE and DFA methods were selected as they rated as the best performing estimators in the comparisons performed in [27, 86, 102, 101]. The PCG algorithm of [20] was not employed as it is slightly outperformed by the WE in the case where the mean of the process is unknown. Furthermore it requires various iterations due to the preconditioning. This translates into an unfair computational expense comparison. For the particular case of ARFIMA, we only compared the WE and the CE.

5.1.2 Results and Discussion

The comparison of results is performed through boxplots and the average time consumed to compute the estimates. In general terms, the DFA was underperforming in comparison to the WE and CE Estimator. For this particular method we employed a set of 200 segments which varied with lengths $m \in$

$(10, N/6)$ where possible. For the experiments with less than 200 data points we used $N - 10$ segments of the same length.

fGN Experiments

For computing the WE estimate we employed the approximation of the spectral density function of fGN by [9]. This approximation is slightly more accurate than the approximation of [82]. For a discussion on this see point (d) in remark 2.3.6.

There is no surprise in the fact that all three estimators improved when the amount of data points was increased. For small sample lengths, the CE Estimator performs better than the WE in all four measures. When the amount of observations is large, the performance improvement over the bias, STD and MSE of the CE Estimator over the WE starts to diminish. However, the time required to compute the CE Estimator in comparison to the WE is quite an improvement as noted previously in the conclusions of chapter 3. It is important to note that as $H_0 \rightarrow 1$ the performance of the WE deteriorates while the performance of the CE Estimator remains uniform. The results of each experiment are displayed in figures 5.7 through 5.12 in the appendix 5.A

FDN Experiments

Just as for fGN, all three estimators improve when the amount of data points is increased. For small sample lengths, the CE Estimator performs better than the WE in all four measures. When the amount of observations is large, the performance improvement over the bias, STD and MSE of the CE Estimator over the WE starts to diminish. The CE Estimator requires less time to be computed in comparison to the WE. Nevertheless, the difference is not as much as in the case of fGN, particularly, when N is large. The results of each experiment are displayed in figures 5.13 through 5.17 in the appendix 5.A. An example is presented in figure 5.1.

ARFIMA Experiments

As in the last two cases, the WE and the CE Estimator improve when the amount of data points is increased. For the sample lengths of $N = 1024, 2048$, the WE performs slightly better than the CE Estimator. For $N = 16384$ the estimators' performance is comparable. Nevertheless, the CE Estimator is computationally more expensive than the WE. This is due to the cost of computing the covariance of an ARFIMA process in comparison to computing its spectral density function. The results of each experiment are displayed in figures 5.18 through 5.20 in the appendix 5.A.

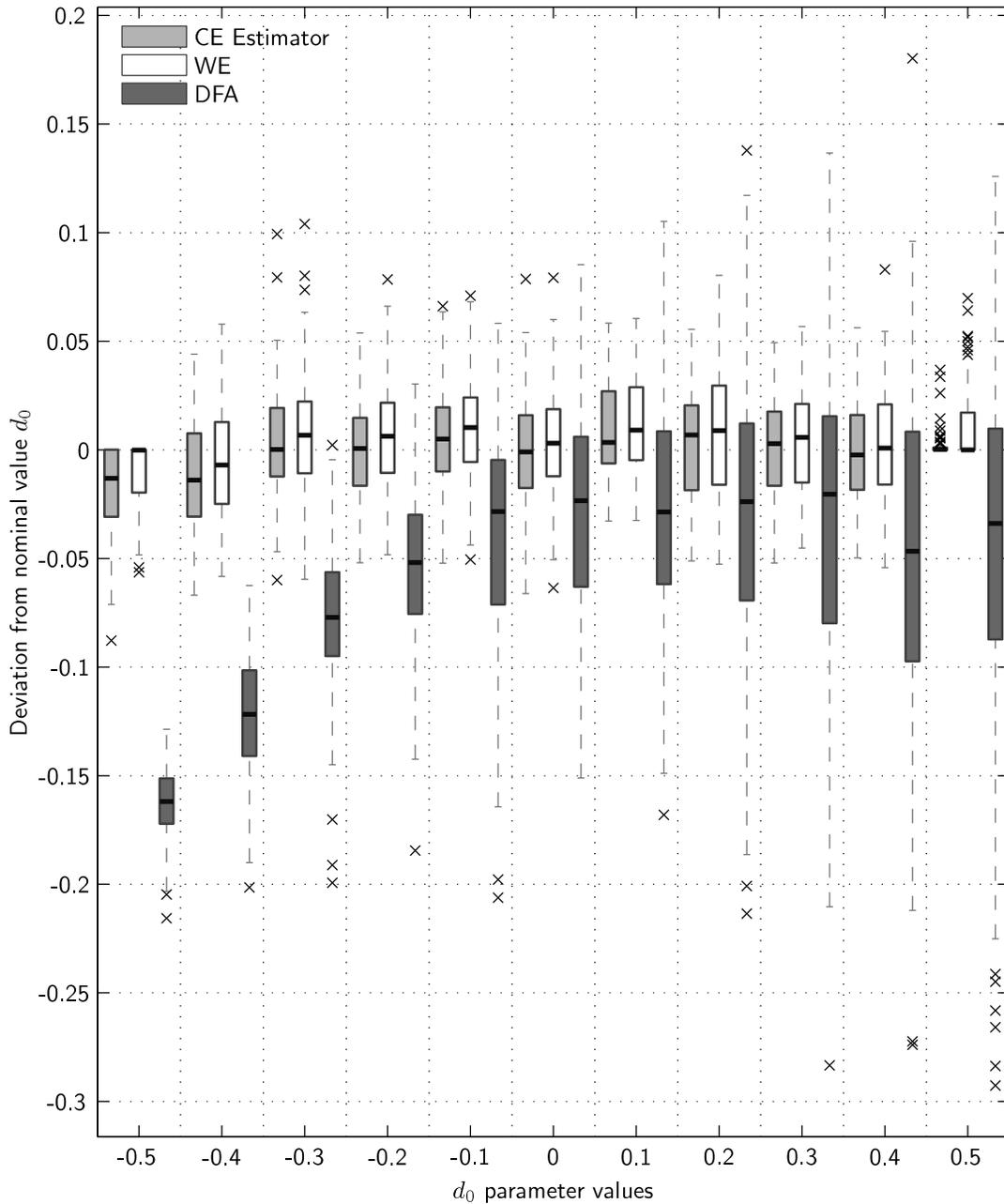


Figure 5.1: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 1024$. It can be observed that the CE Estimate outperforms the WE and the DFA for most instances of d_0 . The average computational times were of 0.0074 seconds for the CE Estimator, of 0.0125 seconds for the WE and 5.494 seconds for DFA.

5.2 Scenario II: Sets of Error Corrupted Observations

5.2.1 Description of the Experiments

The numerical experiments were performed in a Monte-Carlo approach. The procedure was as follows:

- (a) The following parameter grid was employed:

$$H_0 = \{0, 0.1, \dots, 1\}$$

$$d_0 = \{-0.5, -0.4, \dots, -0.1, 0.1, \dots, 0.5\} .$$

Within this parameter grid there are 10 fGN processes and 10 FDN processes.

- (b) For each of the fGN and FDN processes contained within the parameter grid, 100 realisations of length $N = 32768$ were generated. The circulant embedding algorithm described in chapter 2, section 2.4.1 was employed for generating the synthetic realisations.
- (c) Each of the realisations were contaminated with white Gaussian noise as a signal to noise ratio of $\text{SNR} = \{30, 15, 10, 5\}$ decibels.
- (d) For each realisation, the parameters of the models were estimated by employing the likelihood methods of chapter 3 and the regularised estimators of chapter 4. For the fGN experiments, the RWE was excluded as it is computationally much more costly than the RCE Estimator. For the FDN experiments, the WE, the CE Estimator, the RCE Estimator and the RWE are compared. For the ARFIMA experiments the CE and RCE estimators were excluded as they are computationally much more costly than their Whittle counterparts.

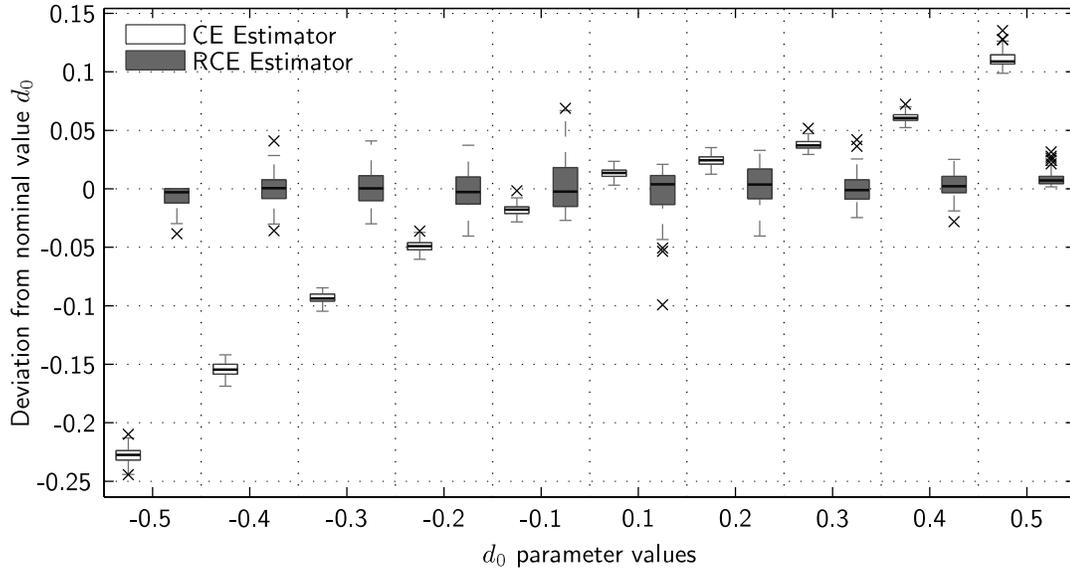
5.2.2 Results and Discussion

The result comparison is performed through the use of boxplots. As expected from the analysis in chapter 3, section 3.3.3, the CE Estimator and the WE are highly biased whenever the observations are corrupted with errors.

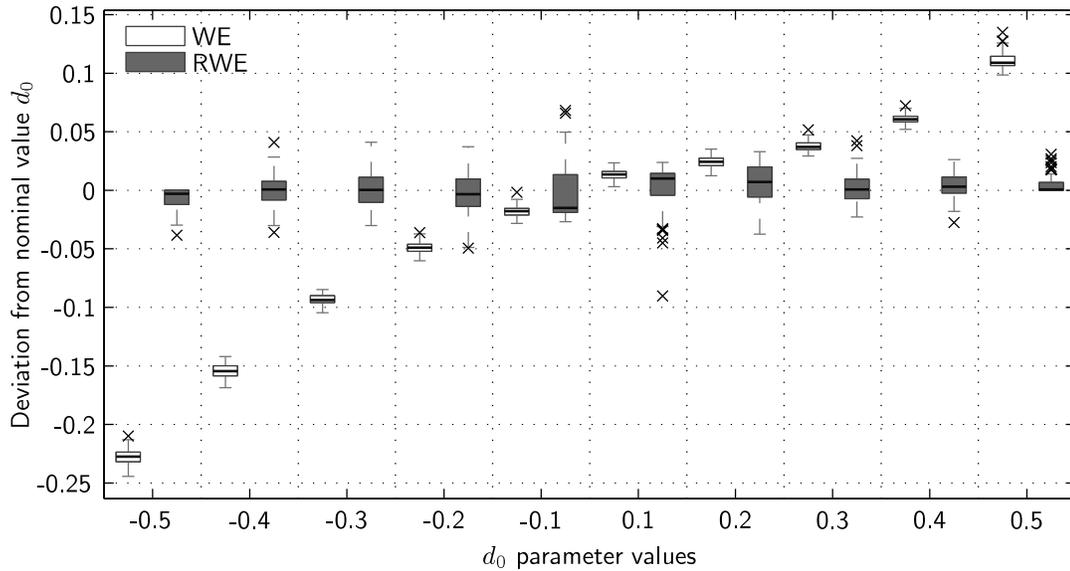
Conversely, the bias of the RCE and RWE is small for all the levels of error corruption in the experiments. It is important to note that as the SNR increases, the standard deviation of the estimates increases slightly as well. This is not a desirable characteristic of the regularised estimates. Nevertheless, it can be corrected by implementing a segmentation framework such as the one presented in [105]. Under such a framework, a smoothing of the estimates is performed such that the variance of the estimates is highly reduced.

Another important point to notice is that the regularised estimators become sensitive as the processes resembles a series of i.i.d. random variables, i.e., white noise. This behaviour is related to the choice of initial conditions.

The results of the experiments are summarised in figures 5.21 through 5.27 in the appendix 5.B. An example is given in figure 5.4.



(a) Parameter estimates for FDN via the RCE est.



(b) Parameter estimates for FDN via the RWE

Figure 5.2: Comparative results between the likelihood based estimators and their regularised versions. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed as crosses. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate's variance. The estimators were applied to 100 realisations of FDN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 15 dB.

5.3 Real World Examples

In this section we present some real world examples of time-series exhibiting long-range dependence. The difficulty when performing parameter estimation of real world data is that there is no “true” underlying model producing the observations. In the words of George Box [11]

“Essentially, all models are wrong, but some are useful.”

There is no reason to assume that the models selected for the parameter estimation of these examples are the correct ones. The topic of model order selection and model validation are outside the scope of this thesis. Instead, we compare the estimates available in the literature against the estimates obtained by applying the methods of chapters 3 and 4.

5.3.1 Ethernet Data

In this section we apply the WE and the CE Estimator to the real time series obtained from a study of Ethernet traffic performed at Bellcore Morris Research and Engineering Centre. This data was produced by [63] and it corresponds to one hour of normal traffic in August 1989. The data collected contains two time-series. One corresponds to the number of bytes per 10 milliseconds. The second time series represents the number of packets per 10 milliseconds. Leland et al. [63] and Willinger et al. [109] obtained estimates of the LRD coefficient of approximately $\hat{H} \approx 0.8$ for the byte time-series and $\hat{H} \approx 0.9$ for the packet time-series. They employed several methods such as the R/S statistic (see chapter 3, section 3.2.1), the log-periodogram method (see chapter 3, section 3.2.3) and the aggregated Whittle method [101]. Taqqu and Teverovsky [101] estimated the parameters of both the byte and packet time-series by employing the WE and the Local Whittle estimator. The results obtained in [101] are highly in accordance to those in [63] and [109].

Here, we estimate the parameters of a fGN, FDN and ARFIMA(1,d,1) models through the WE and CE Estimator. The estimates were obtained in a 2.33 GHz Intel Core 2 Duo Macbook Pro with 2 GB of RAM running MatLab R2008a. The estimates and the time to compute them are reported in tables 5.1 and 5.2. The parameters obtained by the CE Estimator are highly in accordance to the WE estimates, which in turn are equal to those reported in [101]. The notable

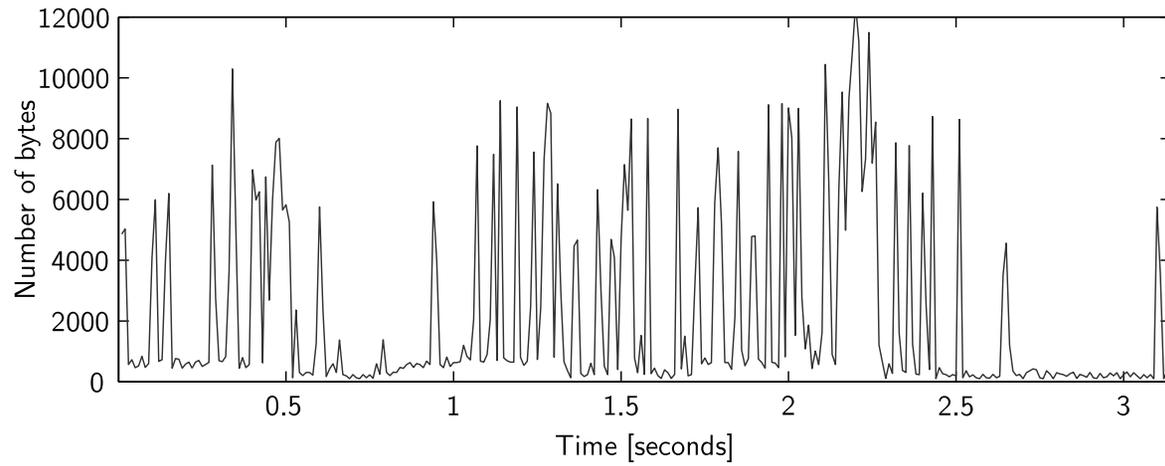
Model Structure	Whittle's Estimate		CE Estimator	
	LRD est.	Time	LRD est.	Time
fGN	0.774	238 s	0.775	25.6 s
FDN	0.318	9.6 s	0.319	15.3 s
ARFIMA(1,d,1)	0.26	80 s	0.25	358 s

Table 5.1: Comparative results between the WE and CE Estimator for the Ethernet byte time-series. The parameter estimates are highly in accordance to those reported in [63, 109, 101]. It can be observed that the CE Estimator is more efficient than the WE for computing the parameters of a fGN model. For ARFIMA models, the WE is more efficient than the CE Estimator.

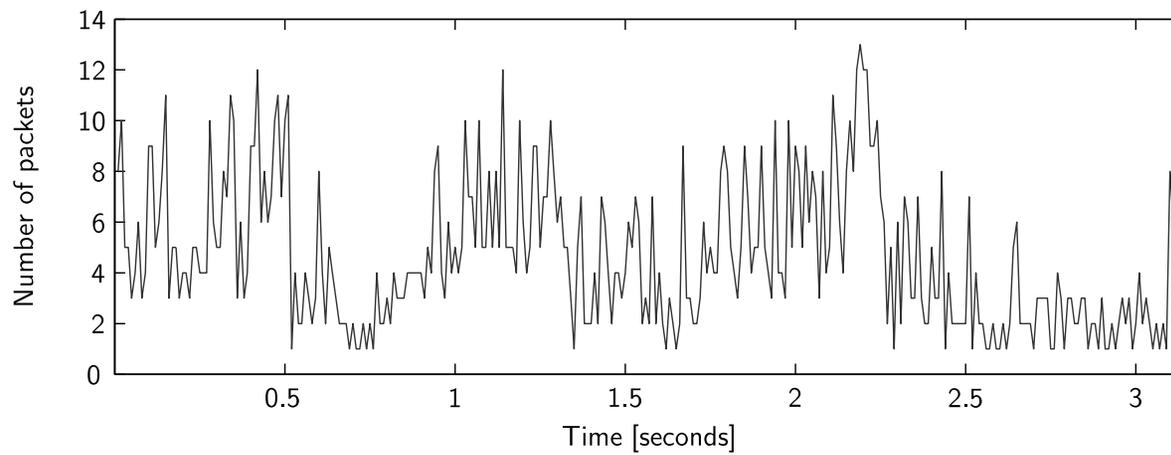
Model Structure	Whittle's Estimate		CE Estimator	
	LRD est.	Time	LRD est.	Time
fGN	0.829	260 s	0.831	15.3 s
FDN	0.389	8.8 s	0.389	14.5 s
ARFIMA(1,d,1)	0.353	93.9s	0.355	497 s

Table 5.2: Comparative results between the WE and CE Estimator for the Ethernet byte time-series. The parameter estimates are highly in accordance to those reported in [63, 109, 101]. It can be observed that the CE Estimator is more efficient than the WE for computing the parameters of a fGN model. For ARFIMA models, the WE is more efficient than the CE Estimator.

difference between the WE and the CE Estimator is the computational expense of each method. For fGN models, the CE Estimator is less expensive than the WE. The contrary is true for ARFIMA models. For the case of FDN, the CE Estimator was slightly slower than the WE. This might be due to the amount of likelihood function evaluations performed in the optimisation.

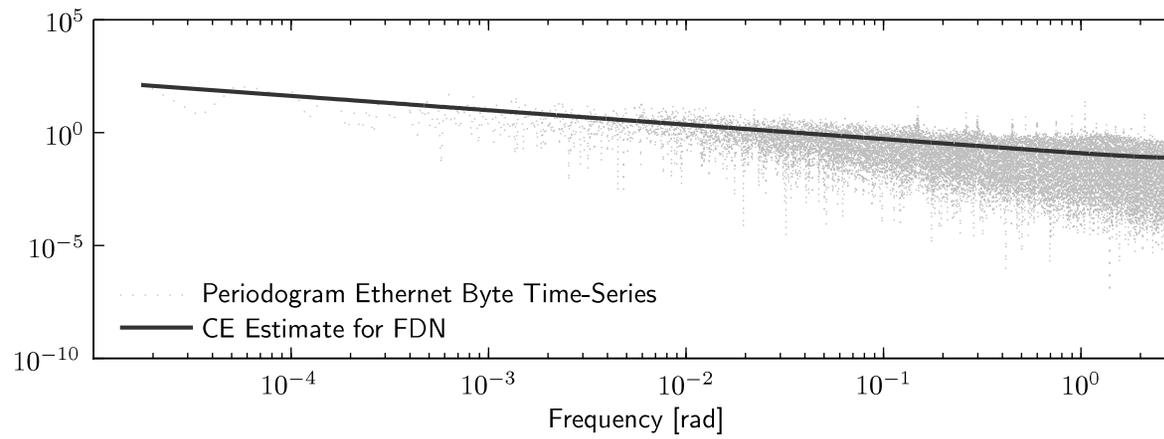


(a) Byte time-series

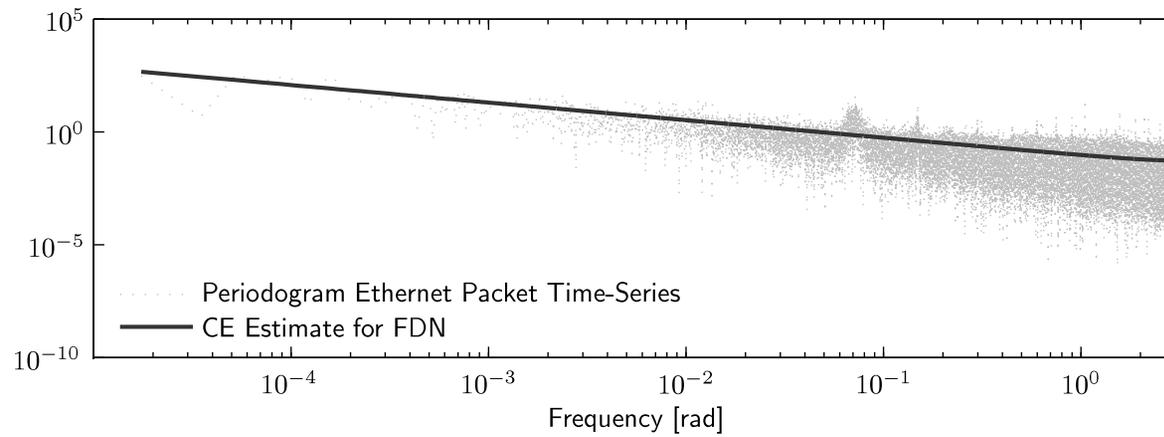


(b) Packet time-series

Figure 5.3: Time-series corresponding to the Ethernet dataset from Bellcore Morris Research and Engineering Centre



(a) Spectral Density Estimate for the Byte time-series



(b) Spectral Density Estimate for the Packet time-series

Figure 5.4: The figure depicts the periodogram of the time-series and the estimates of the spectral density functions obtained by the CE Estimator for and FDN model.

5.3.2 Ocular Accommodation

In this section, parameter estimates from a study of ocular accommodation are obtained by means of the RWE. The time-series was obtained by Leahy et al. [61] and it corresponds to subject ED, trial number 4. The data contained in the time-series represents the accommodative response of a subject sample at 173 Hz for approximately 47 seconds. This particular time-series has been reported to be stationary [61] with heavy-tailed [62] increments.

Leahy et al. [61] and Leahy [62] analysed this data and reported a power-law decay that resembles long-range dependence. The analysis performed in [61, 62] is similar to the log-periodogram analysis summarised in chapter 3, section 3.2.3. Instead of computing the standard periodogram as in equation (3.14), the Lomb-Scargle periodogram was computed. The advantage of the Lomb-Scargle periodogram to that of equation (3.14) is its robustness to unevenly sample data and missing observations.

Let X_k , $k \in \mathbb{J}$, denote a set of observations. Then Lomb-Scargle periodogram of X_k is given by

$$2I_{\text{LS}}(\omega) = \frac{\left[\sum_{k \in \mathbb{J}} X_k \cos(\omega t_k - \omega \tau) \right]^2}{\sum_{k \in \mathbb{J}} X_k \cos^2(\omega t_k - \omega \tau)} + \frac{\left[\sum_{k \in \mathbb{J}} X_k \sin(\omega t_k - \omega \tau) \right]^2}{\sum_{k \in \mathbb{J}} X_k \sin^2(\omega t_k - \omega \tau)}$$

where ω is the angular frequency, t_k corresponds to the time in which the data point X_k was sampled and τ is given by

$$\tan(2\omega\tau) = \sum_{k \in \mathbb{J}} \sin(2\omega t_k) \left[\sum_{k \in \mathbb{J}} \cos(2\omega t_k) \right]^{-1}.$$

Since the ocular accommodation time-series presents segments of missing data (see figure 5.5), [61, 62] employed the Lomb-Scargle periodogram to fit a bilinear model with slopes $\hat{\theta}_{\text{PSD1}}$ and $\hat{\theta}_{\text{PSD2}}$. The slope $\hat{\theta}_{\text{PSD1}}$ corresponds to the linear fit at low frequencies $\omega < 4\pi/173$. The $\hat{\theta}_{\text{PSD2}}$ slope corresponds to the higher frequencies $4\pi/173 < \omega < 20\pi/173$. The cross-over frequency selected by [61, 62] was of 2Hz and the estimates obtained from averaging four sets of trials was $\hat{\theta}_{\text{PSD1}} = -2\hat{d}_{\text{PSD1}} = -0.8$ and $\hat{\theta}_{\text{PSD2}} = -2\hat{d}_{\text{PSD2}} = -2.5$.

The approach presented in this section is different to that of [61, 62]. Instead of employing the entire time-series, a segment where there were no missing observations was selected for estimation. The RWE was employed to fit an ARFIMA(1, d , 0) and ARFIMA(2, d , 1) models. The results are summarised in table 5.3 and figure 5.6.

From the estimates we can observe that the RWE obtains a smaller LRD coefficient in comparison to the PSD estimate of [61, 62]. This difference can be attributed to the cut-off frequency selected by [61, 62] which does not correspond to the ones obtained by the RWE. It is important to note that neither the cut-off frequencies nor the correct values for the model coefficients can be verified without proper validation tools. Under weak stationarity, the goodness of fit test provided by [8] becomes a powerful tool. Nevertheless, under heavy-tailed distributions and non-stationarity the effectiveness of this test is an open question.

Model Parameters	ARFIMA(1, d , 0)	ARFIMA(2, d , 1)
\hat{d}	0.3298	0.3662
$\hat{A}(q)$	$1 - 0.9577q^{-1}$	$1 - 0.2015q^{-1} + 0.624q^{-2}$
$\hat{B}(q)$	1	$1 + 0.6497q^{-1}$
$\hat{\alpha}$	0.5604	0.1493

Table 5.3: Comparative results for between the ARFIMA(1, d , 0) and ARFIMA(2, d , 1) models for the Ocular Accommodation time-series corresponding to subject ED trial number 4. In comparison to the result of [61], $\hat{d}_{\text{PSD1}} = 0.4$, we have that the RWE yields a smaller LRD coefficient.

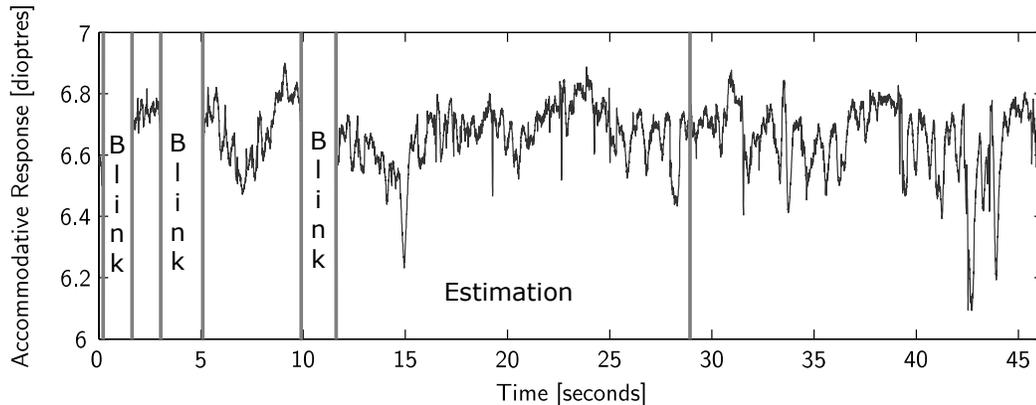


Figure 5.5: The figure depicts the dataset corresponding to subject ED, trial number 4.

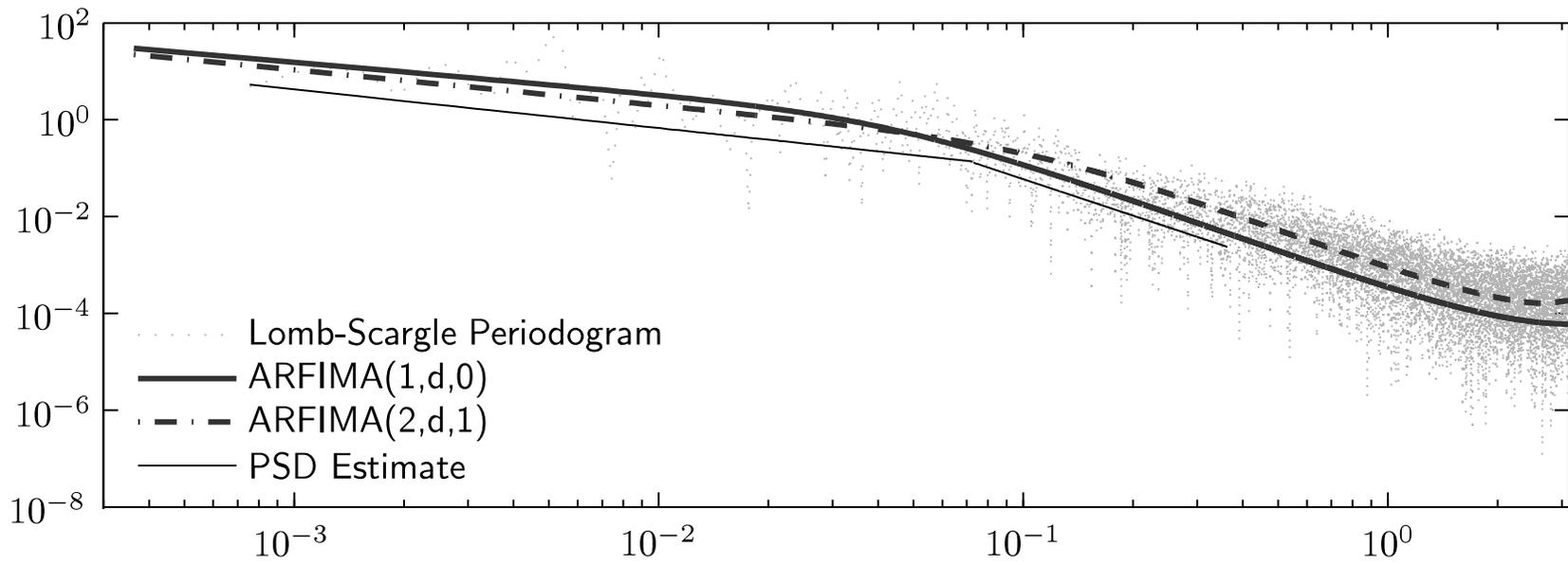


Figure 5.6: The figure depicts the Lomb-Scargle periodogram of the time-series along with the estimates of the spectral density functions obtained by the RWE. The PSD estimate of [61] is also depicted.

5.A Monte Carlo Studies for the Error-Free Case

This appendix contains the Monte Carlo studies discussed in section 5.1.

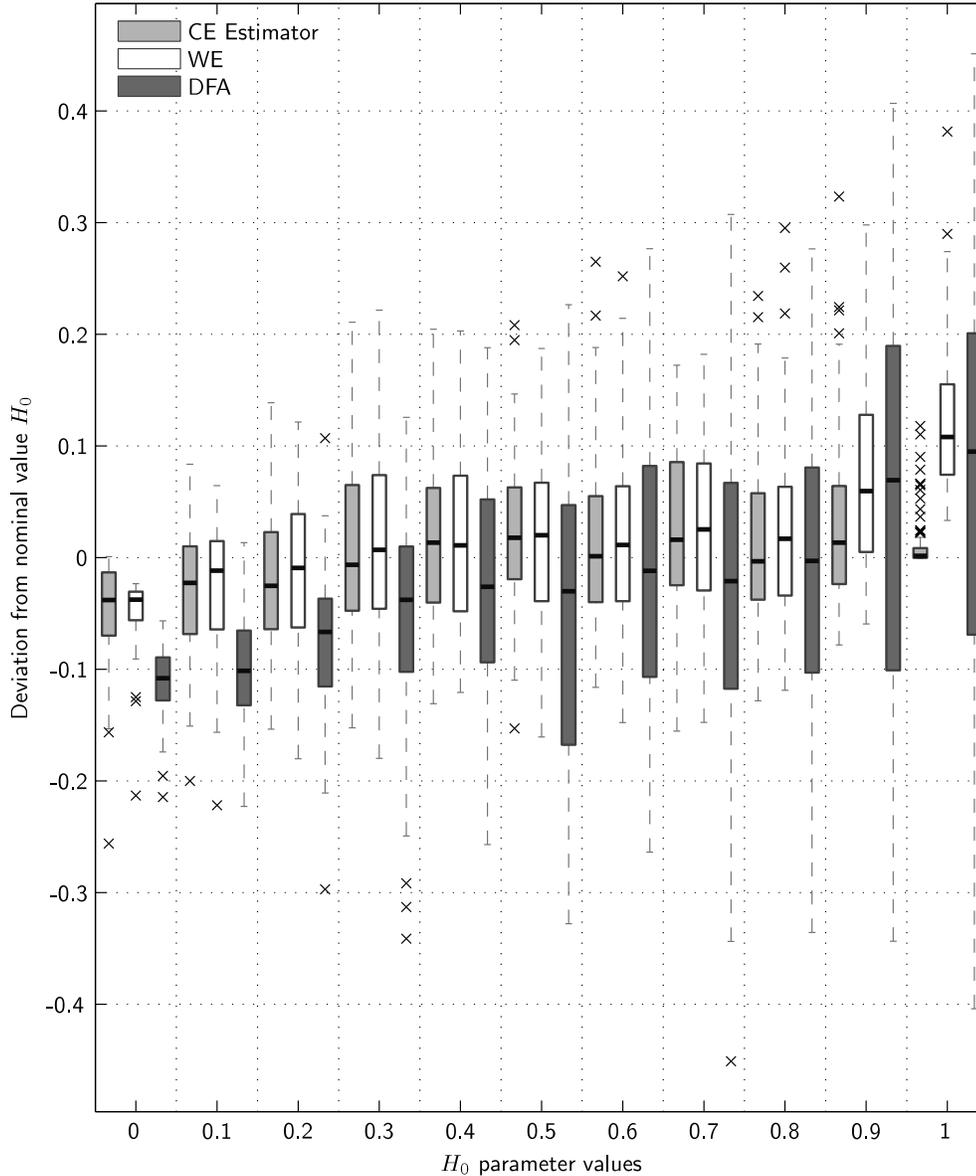


Figure 5.7: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 64$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.0047 seconds for the CE Estimator, of 0.0404 seconds for the WE and 0.3076 seconds for DFA.

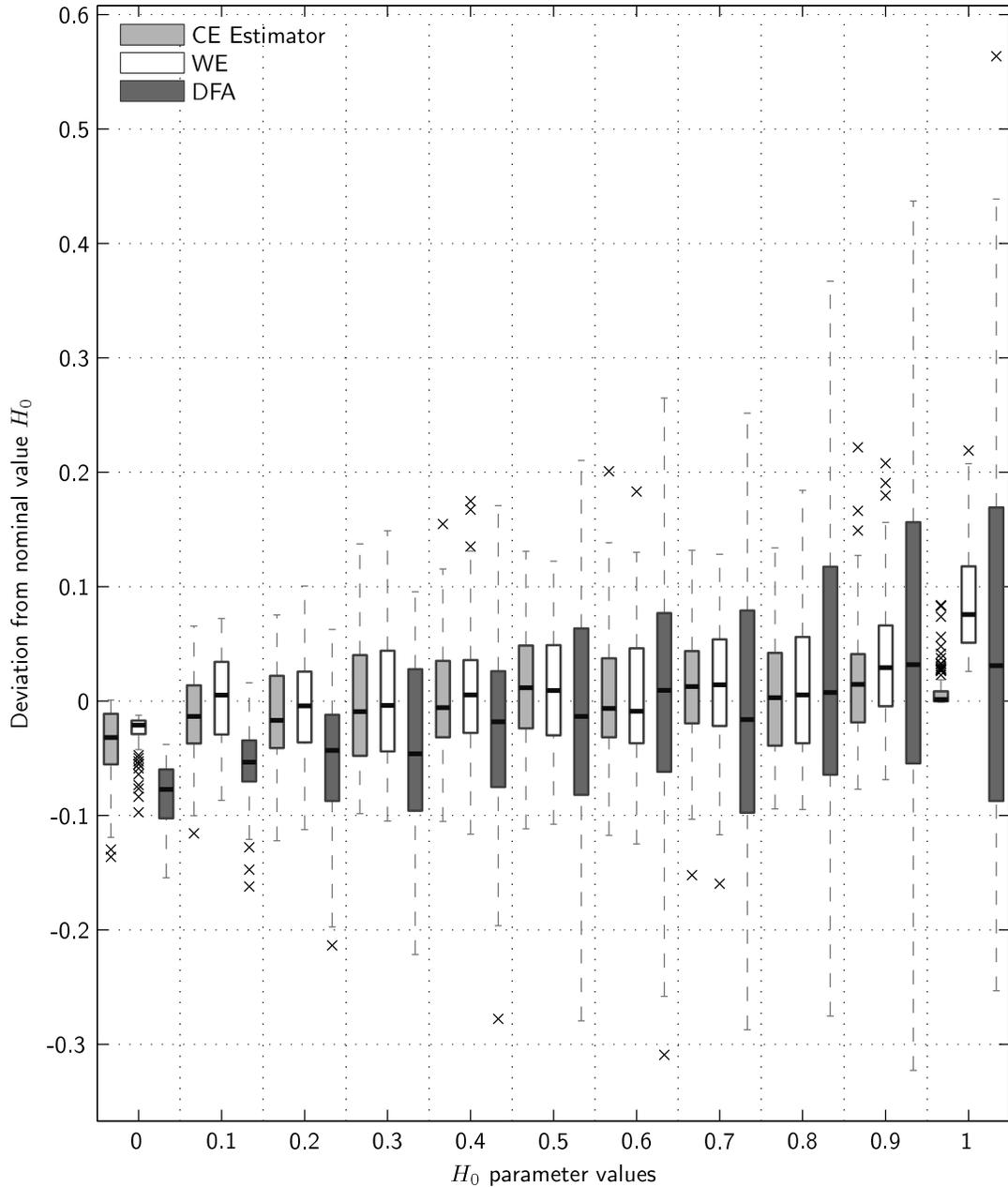


Figure 5.8: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 128$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.0051 seconds for the CE Estimator, of 0.0794 seconds for the WE and 0.6386 seconds for DFA.

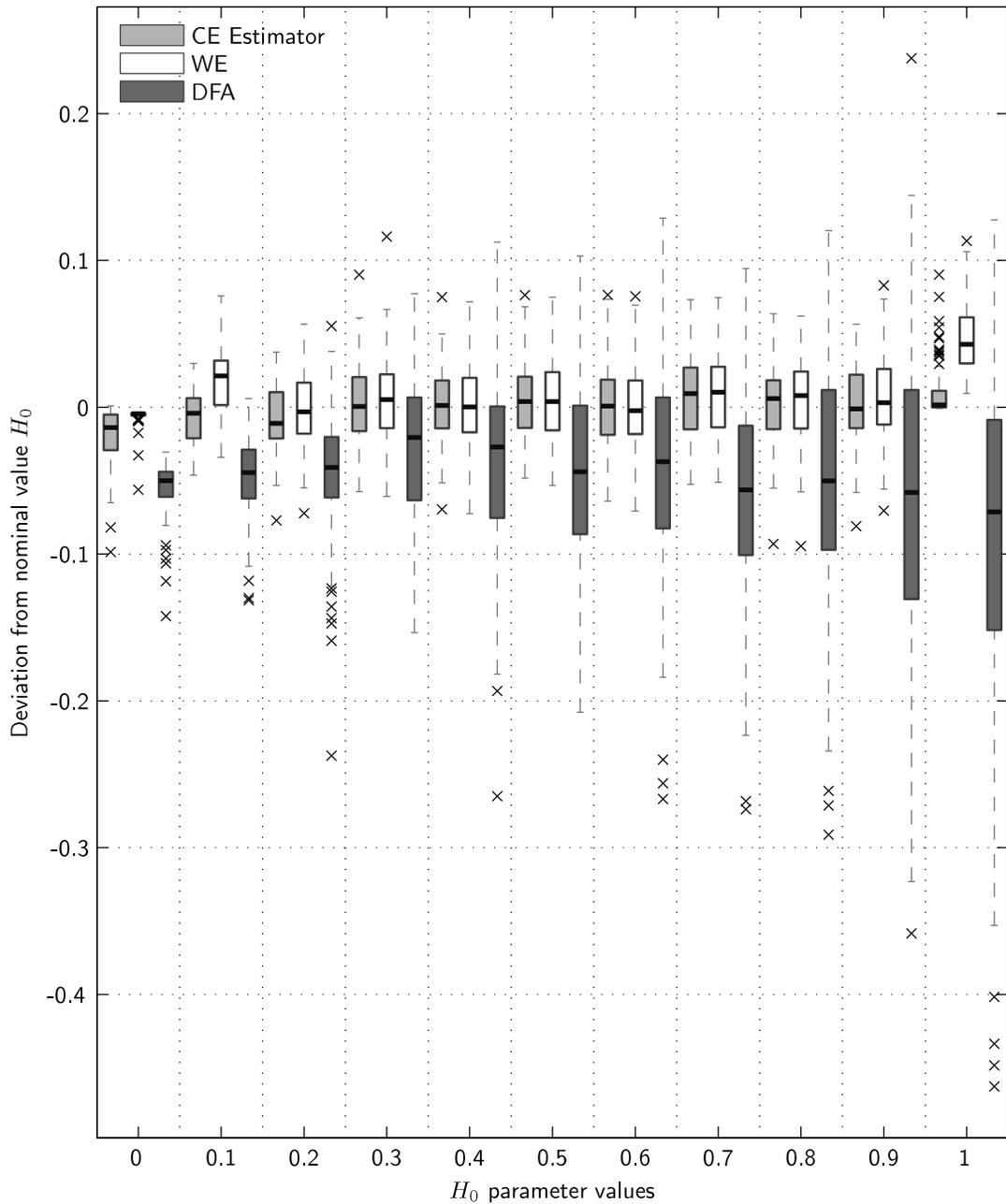


Figure 5.9: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 512$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.0103 seconds for the CE Estimator, of 0.3185 seconds for the WE and 3.4608 seconds for DFA.

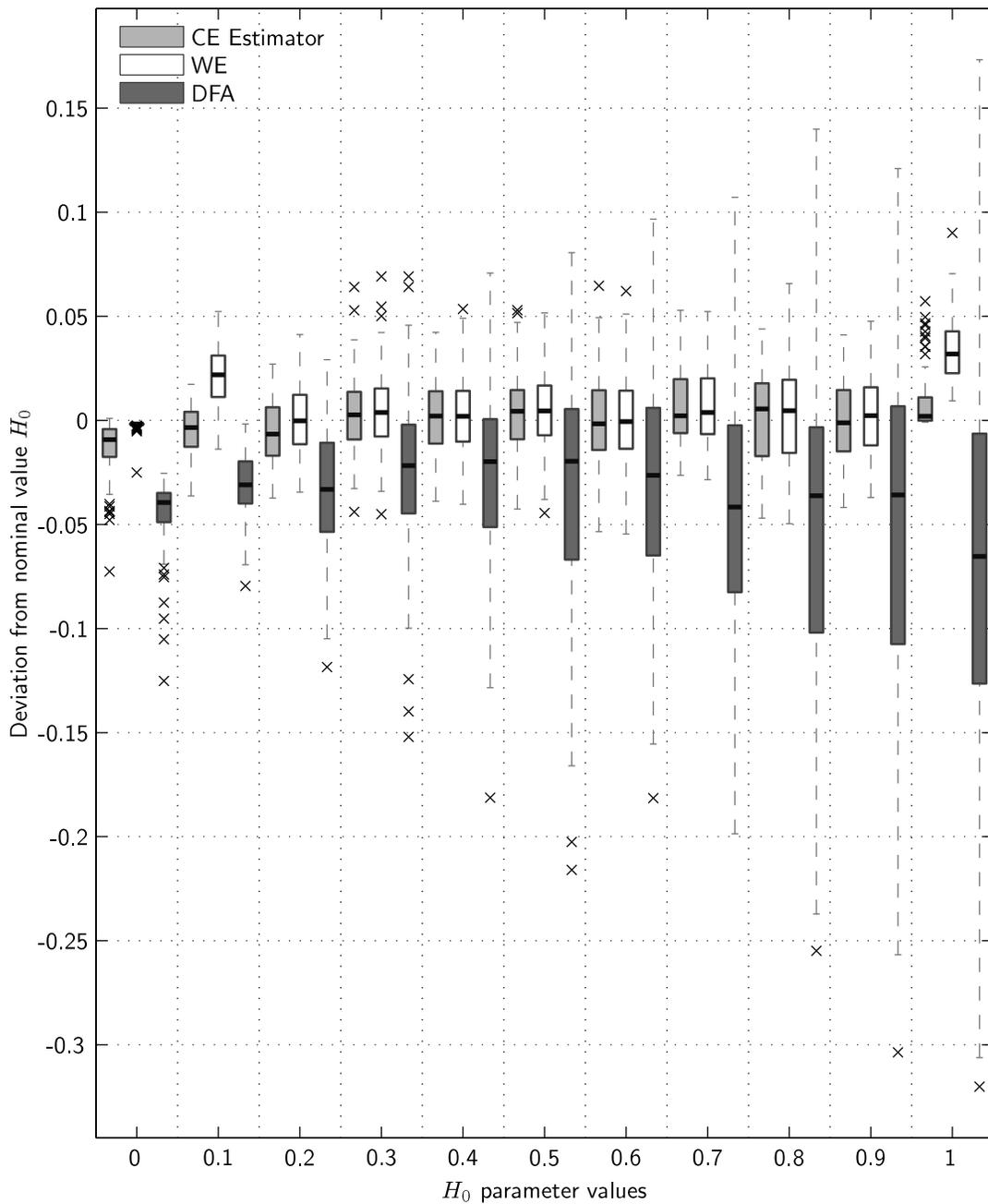


Figure 5.10: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 1024$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.0103 seconds for the CE Estimator, of 0.3185 seconds for the WE and 3.4608 seconds for DFA.

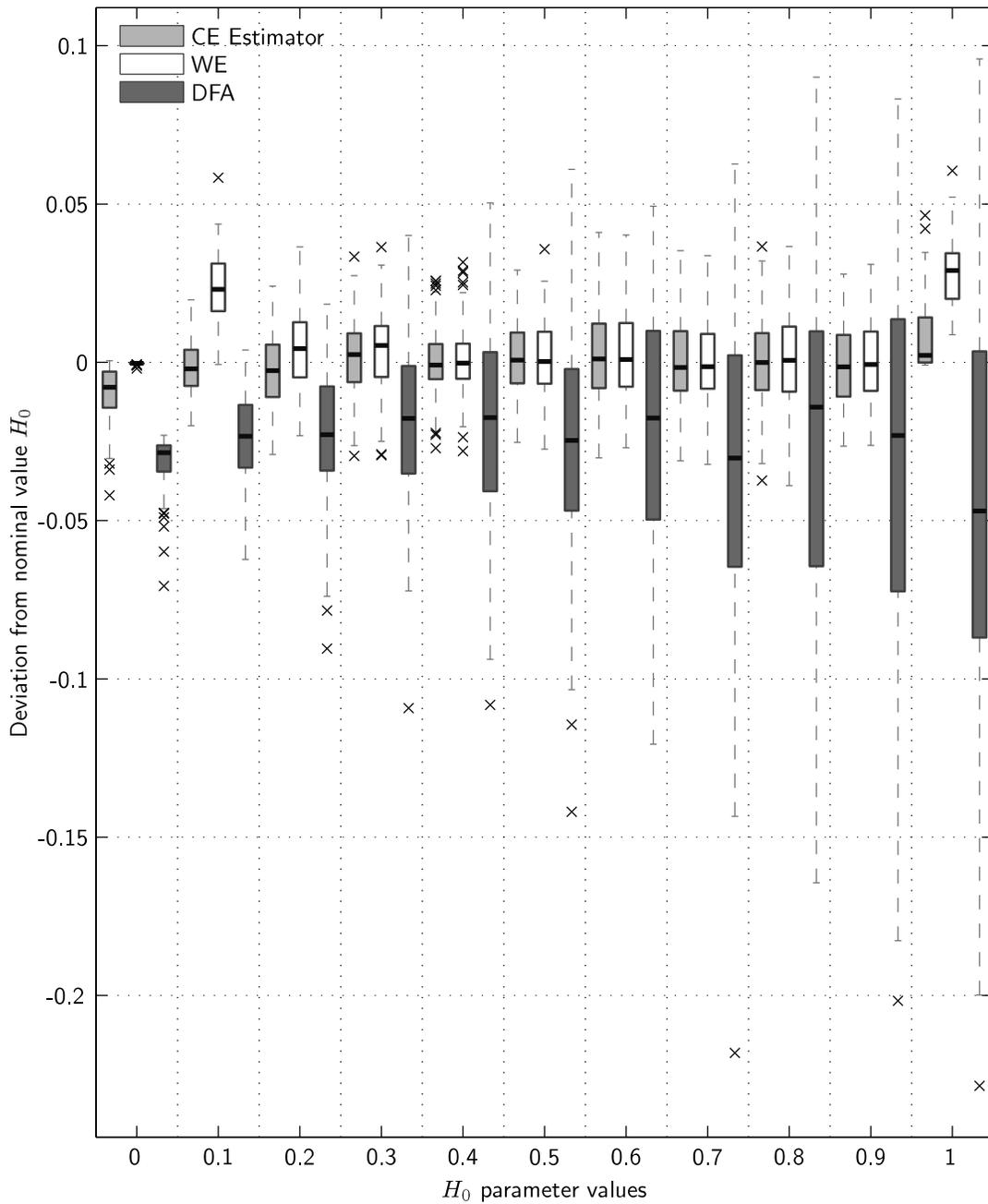


Figure 5.11: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 2048$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.034 seconds for the CE Estimator, of 1.3072 seconds for the WE and 8.8891 seconds for DFA.

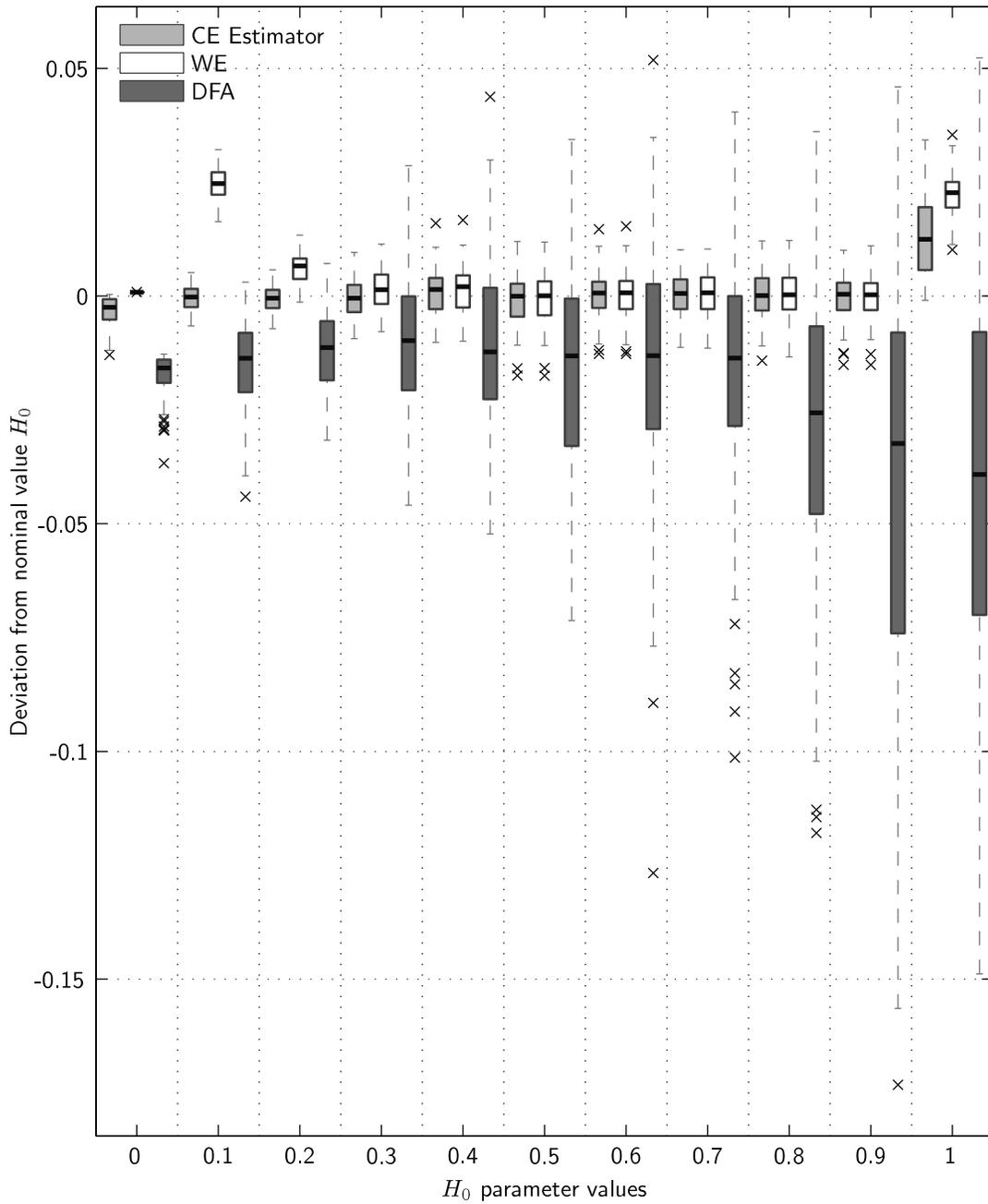


Figure 5.12: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of fGN with different nominal LRD coefficient H_0 . The realisations had a length of $N = 16384$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of H_0 . The average computational times were of 0.2766 seconds for the CE Estimator, of 10.844 seconds for the WE and 45.588 seconds for DFA.

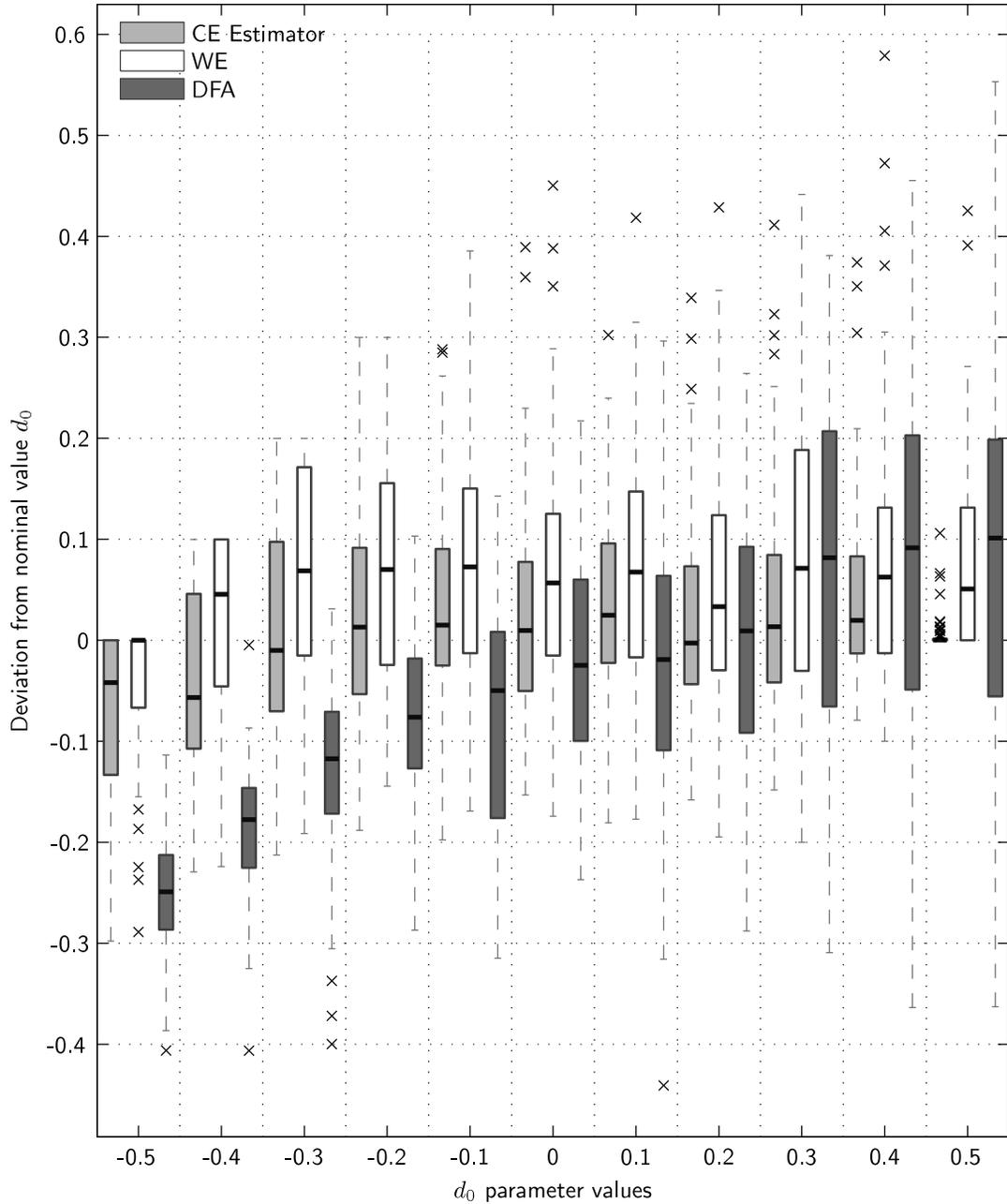


Figure 5.13: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 64$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of d_0 . The average computational times were of 0.0039 seconds for the CE Estimator, of 0.0049 seconds for the WE and 0.2915 seconds for DFA.

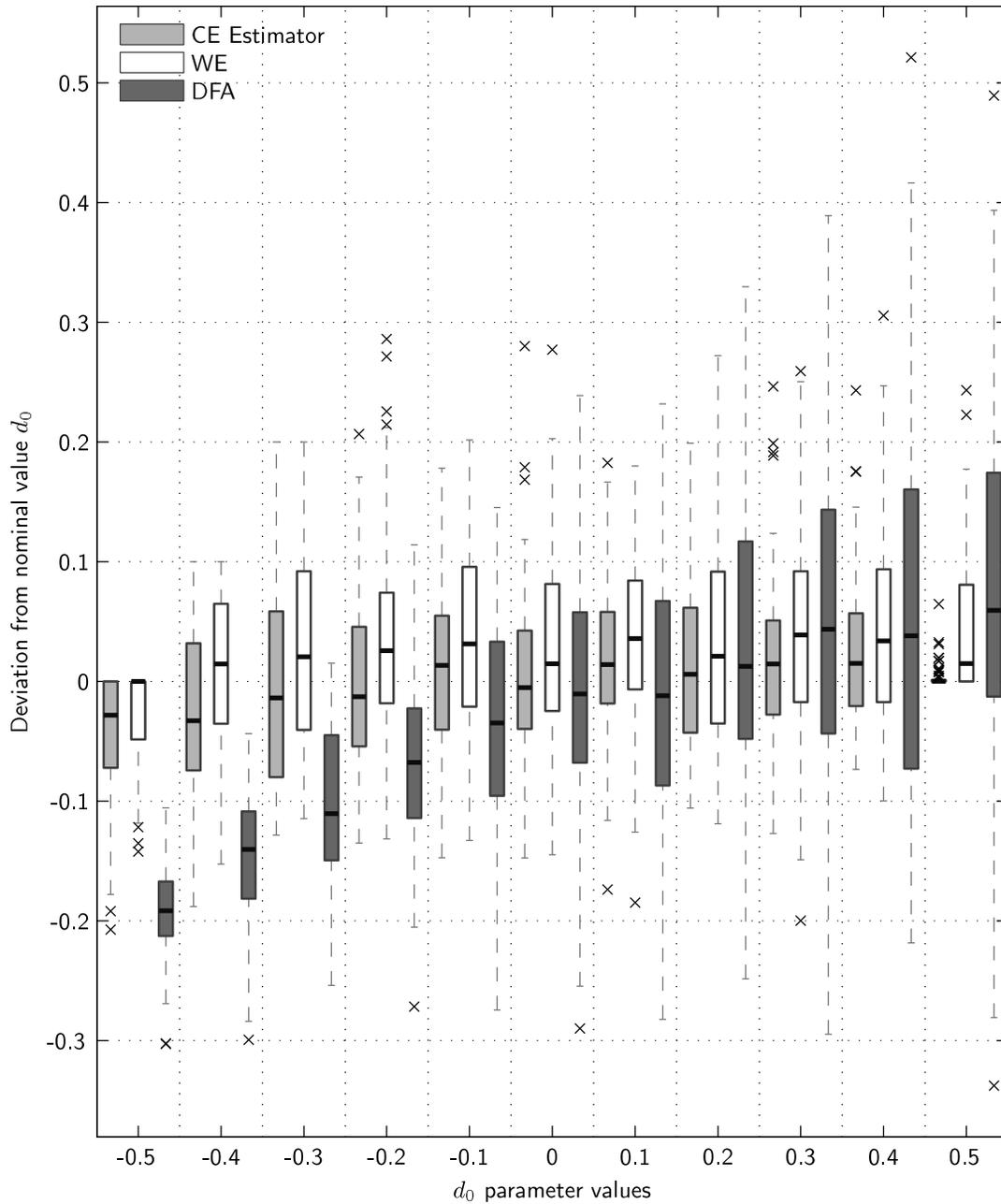


Figure 5.14: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 128$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of d_0 . The average computational times were of 0.0039 seconds for the CE Estimator, of 0.005 seconds for the WE and 0.6003 seconds for DFA.

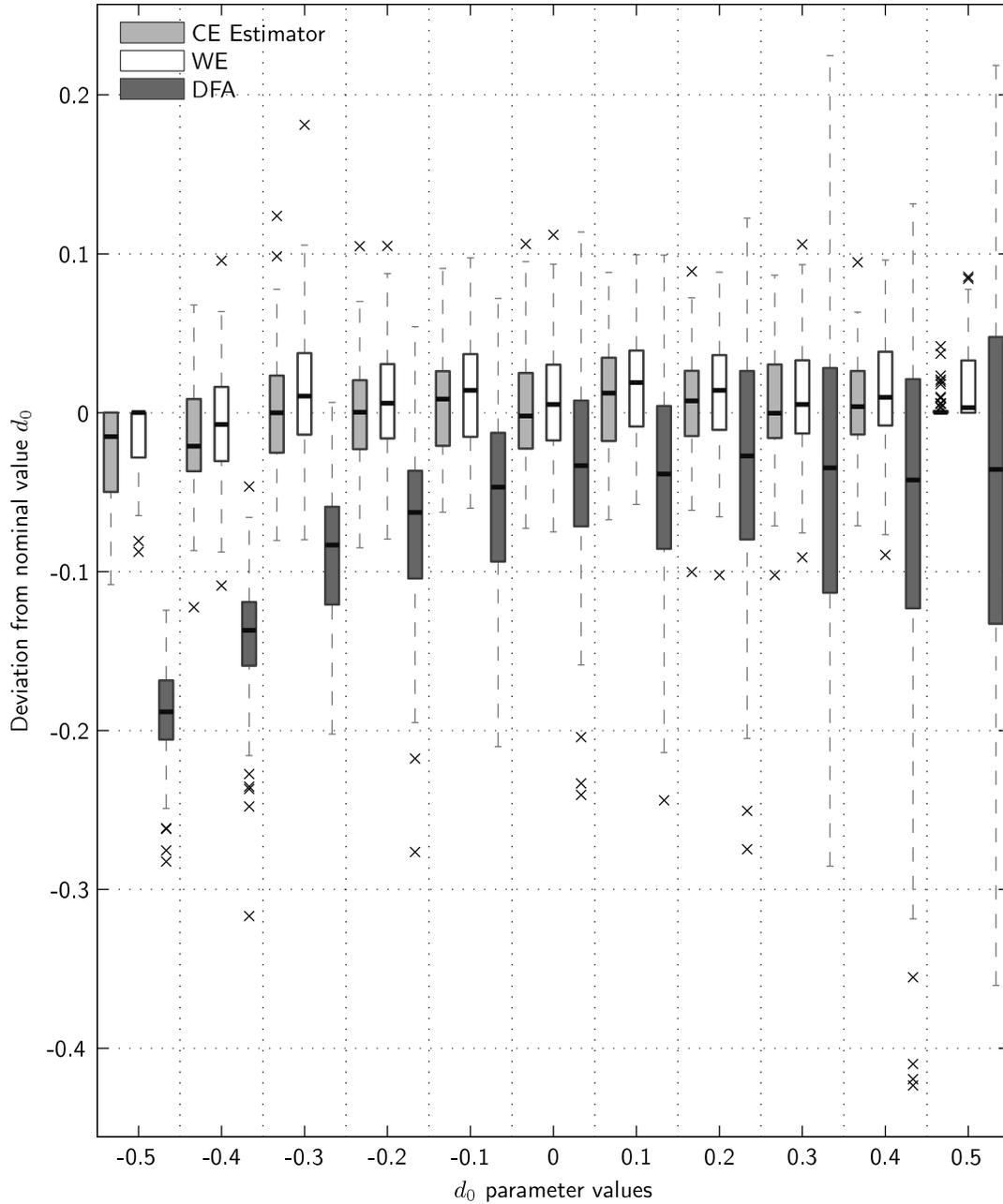


Figure 5.15: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 512$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of d_0 . The average computational times were of 0.0053 seconds for the CE Estimator, of 0.0082 seconds for the WE and 3.5108 seconds for DFA.

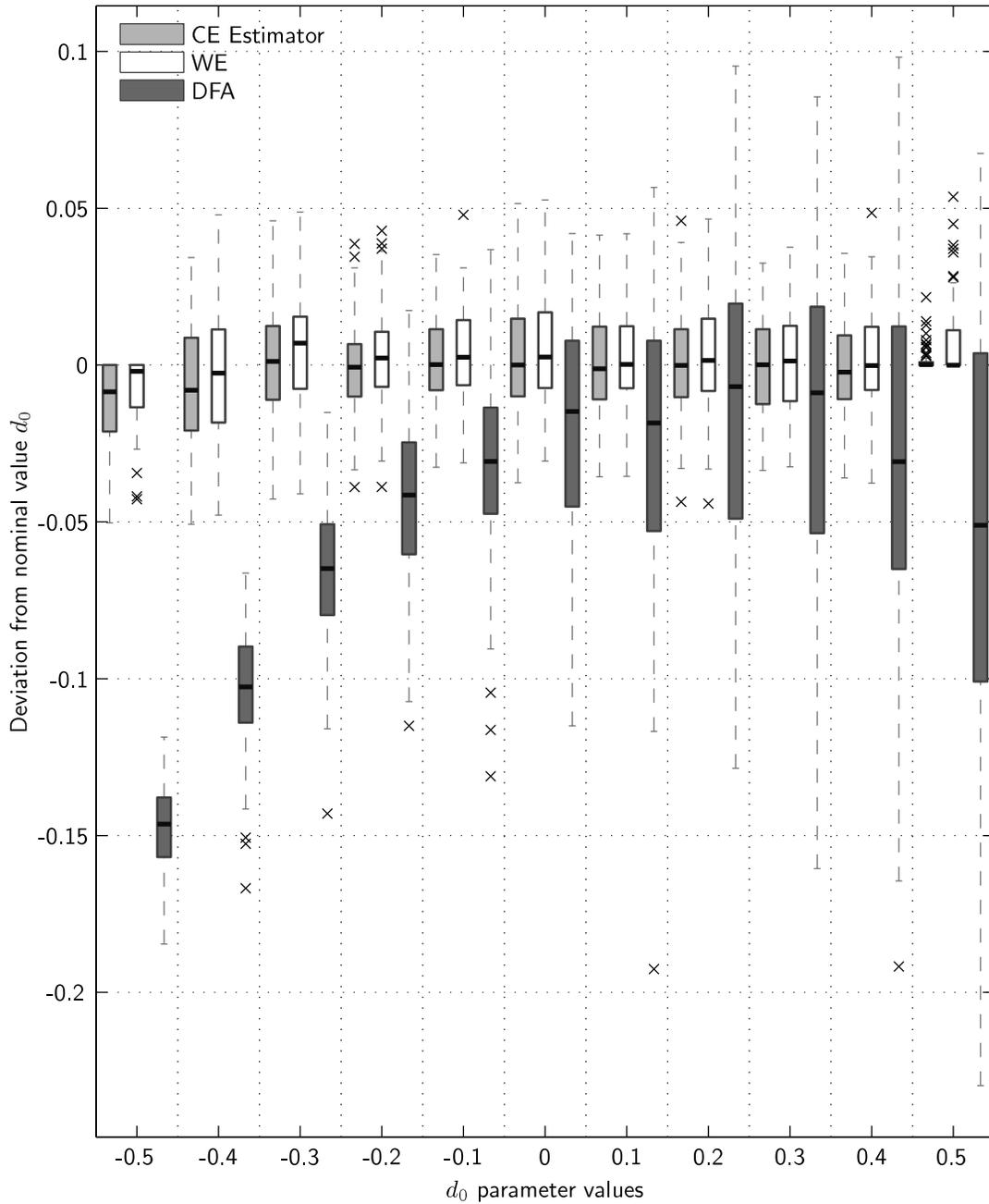


Figure 5.16: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 2048$. It can be observed that the CE Estimate outperforms the WE and DFA for most instances of d_0 . The average computational times were of 0.0131 seconds for the CE Estimator, of 0.0211 seconds for the WE and 8.961 seconds for DFA.

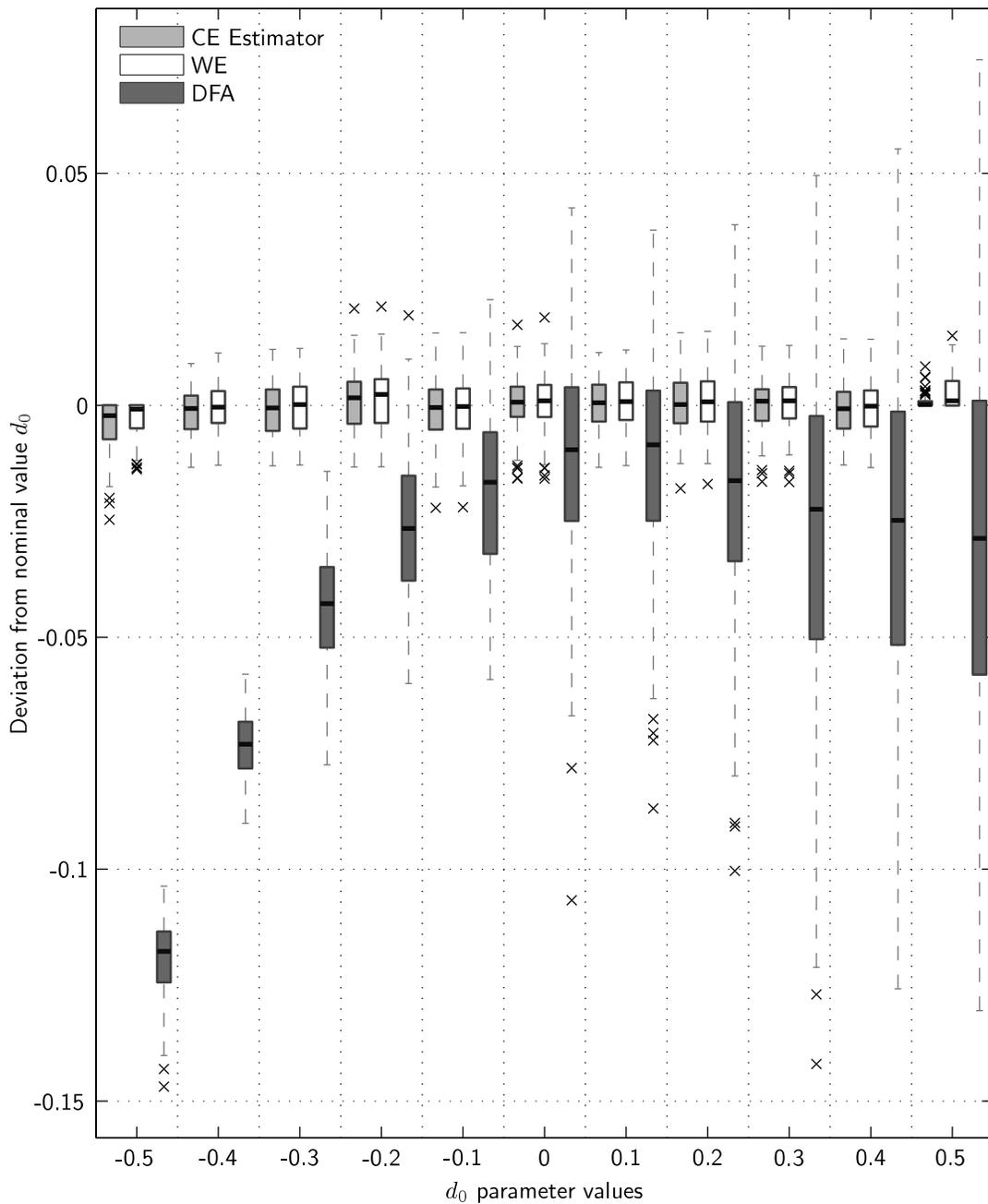


Figure 5.17: Comparative results between the CE Estimator, the WE and the DFA. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are displayed with crosses. The estimation methods were applied to 100 realisations of FDN with different nominal LRD coefficient d_0 . The realisations had a length of $N = 16384$. It can be observed that the CE Estimate and the WE perform similarly while the DFA is the estimator that performs the poorest. The average computational times were of 0.1112 seconds for the CE Estimator, 0.1578 seconds for the WE and 45.678 seconds for DFA.

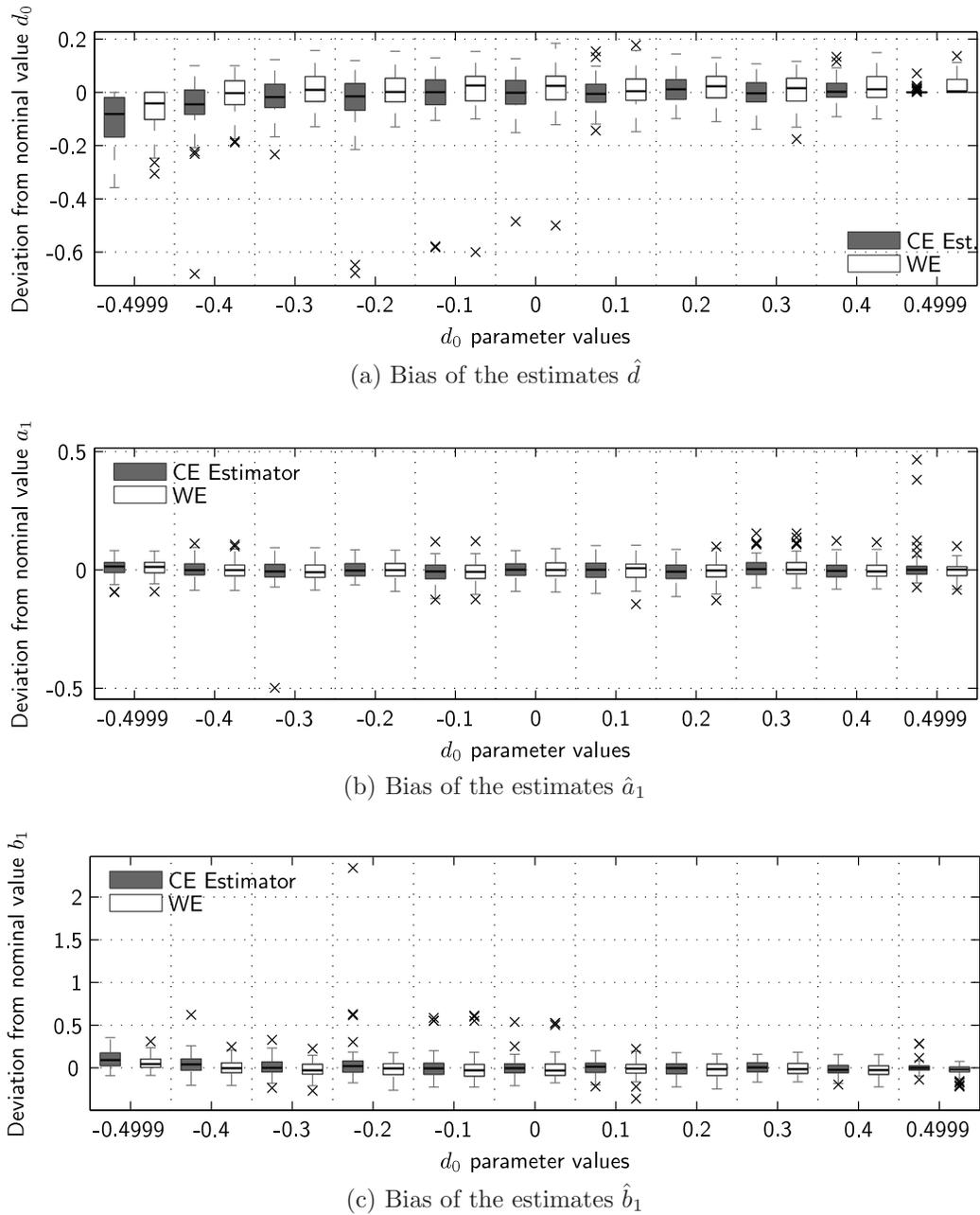
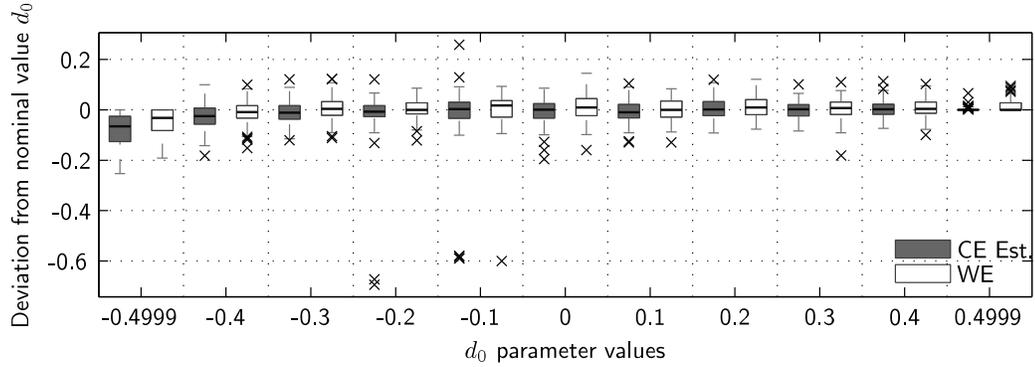
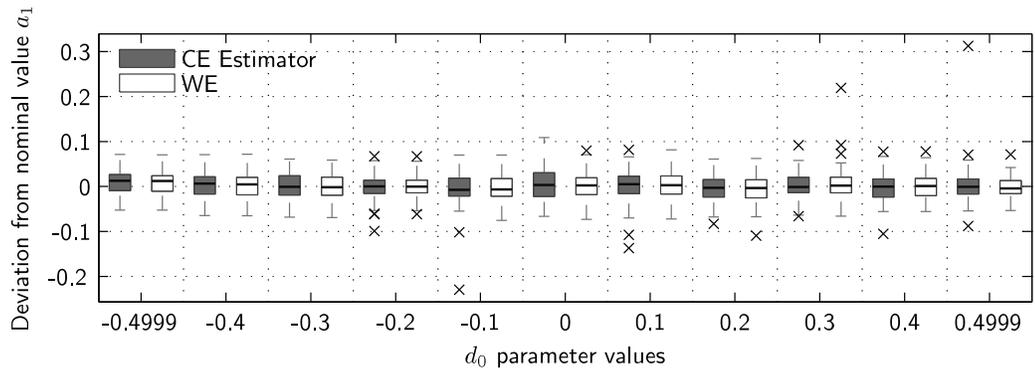


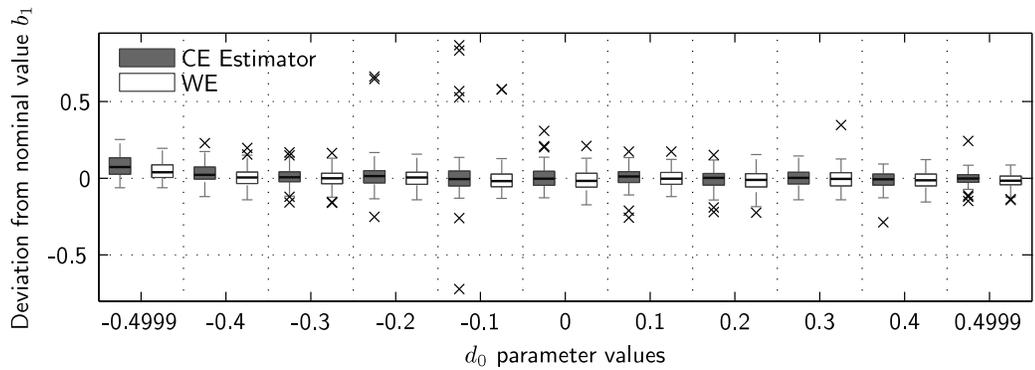
Figure 5.18: Comparative bias results between the CE Estimator and the WE. The estimation methods were applied to 100 realisations of an ARFIMA process with $A_0(q) = 1 + 0.5q^{-1}$, $B_0(q) = 1 - 0.3q^{-1}$ and different nominal LRD coefficient d_0 . These realisations had a length of $N = 1024$. It can be observed that the CE Estimator has a comparable performance to the WE in all estimates for most instances of $d_0 \geq 0$. The WE performs slightly better when $d_0 < 0$. The average computational times were of 2.6587 seconds for the CE Estimator and 0.0492 seconds for the WE.



(a) Bias of the estimates \hat{d}

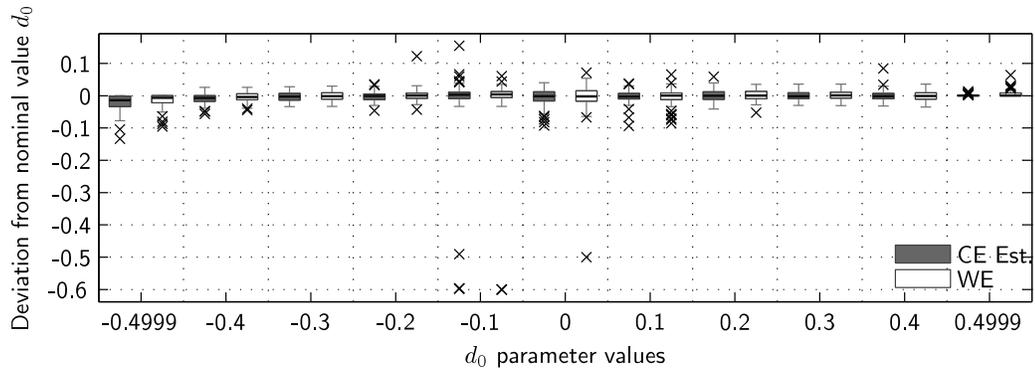


(b) Bias of the estimates \hat{a}_1

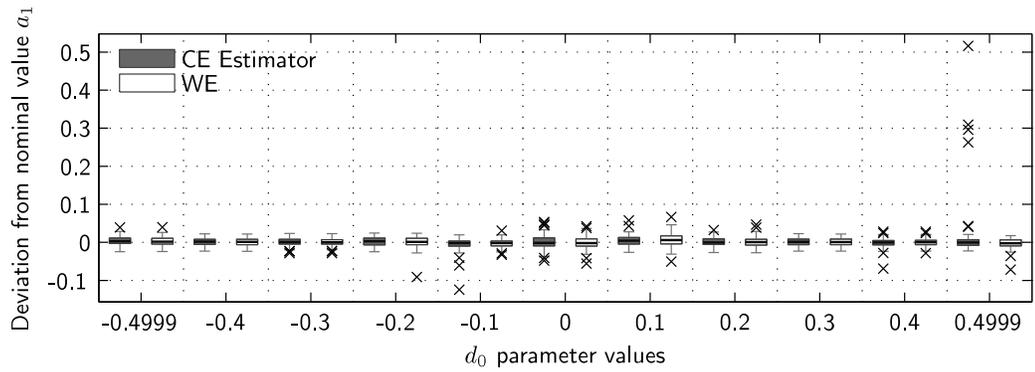


(c) Bias of the estimates \hat{b}_1

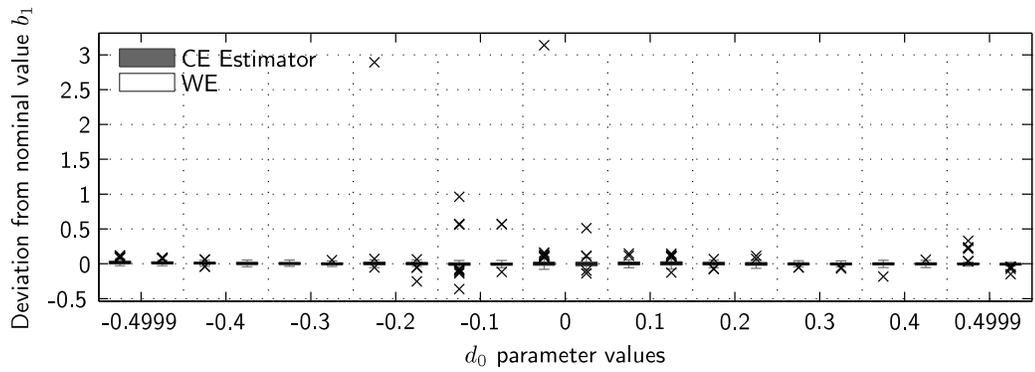
Figure 5.19: Comparative bias results between the CE Estimator and the WE. The estimation methods were applied to 100 realisations of an ARFIMA process with $A_0(q) = 1 + 0.5q^{-1}$, $B_0(q) = 1 - 0.3q^{-1}$ and different nominal LRD coefficient d_0 . These realisations had a length of $N = 2048$. It can be observed that the CE Estimator has a comparable performance to the WE for all instances of $d_0 \geq 0$. When $d_0 < 0$ we have that the WE performs slightly better. The average computational times were of 2.3882 seconds for the CE Estimator and 0.6423 seconds for the WE.



(a) Bias of the estimates \hat{d}



(b) Bias of the estimates \hat{a}_1

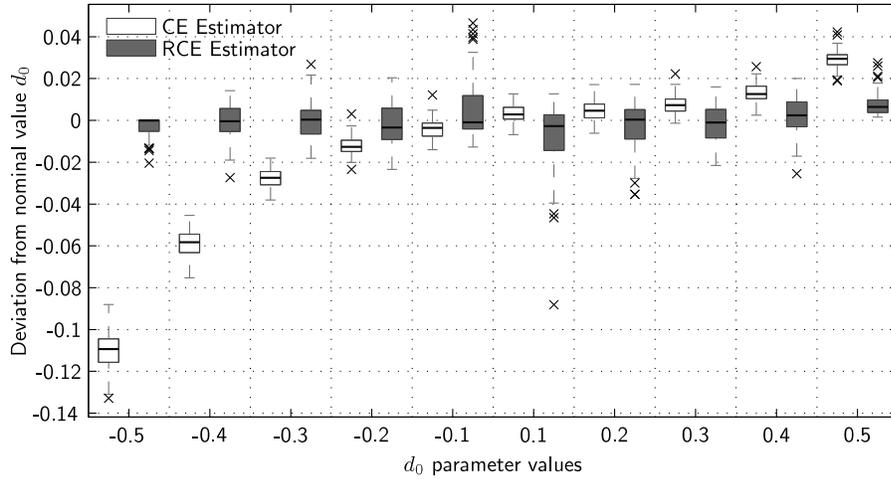


(c) Bias of the estimates \hat{b}_1

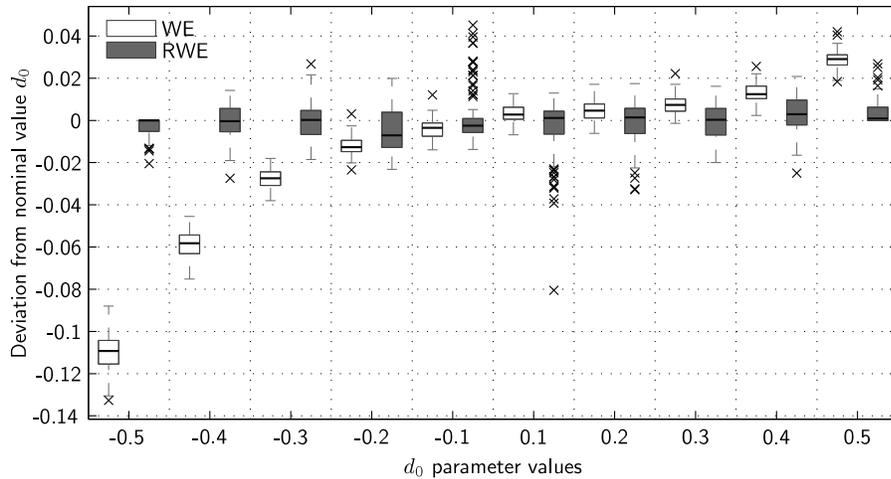
Figure 5.20: Comparative bias results between the CE Estimator and the WE. The estimation methods were applied to 100 realisations of an ARFIMA process with $A_0(q) = 1 + 0.5q^{-1}$, $B_0(q) = 1 - 0.3q^{-1}$ and different nominal LRD coefficient d_0 . These realisations had a length of $N = 16384$. It can be observed that for \hat{d} , the CE Estimator performs better for $d_0 > -0.1$. The WE performs better in most instances for the estimates \hat{A} and \hat{B} . The average computational times were of 19.3852 seconds for the CE Estimator and 4.133 for the WE.

5.B Monte Carlo Studies for the Error-Corrupted Case

This appendix contains the Monte Carlo studies discussed in section 5.2.

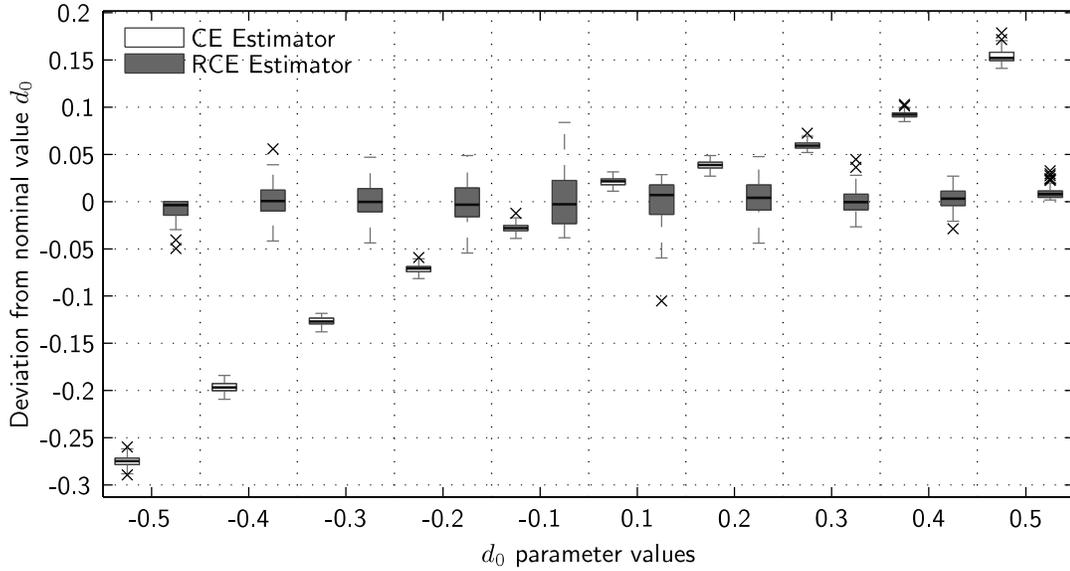


(a) Parameter estimates for FDN via the RCE est.

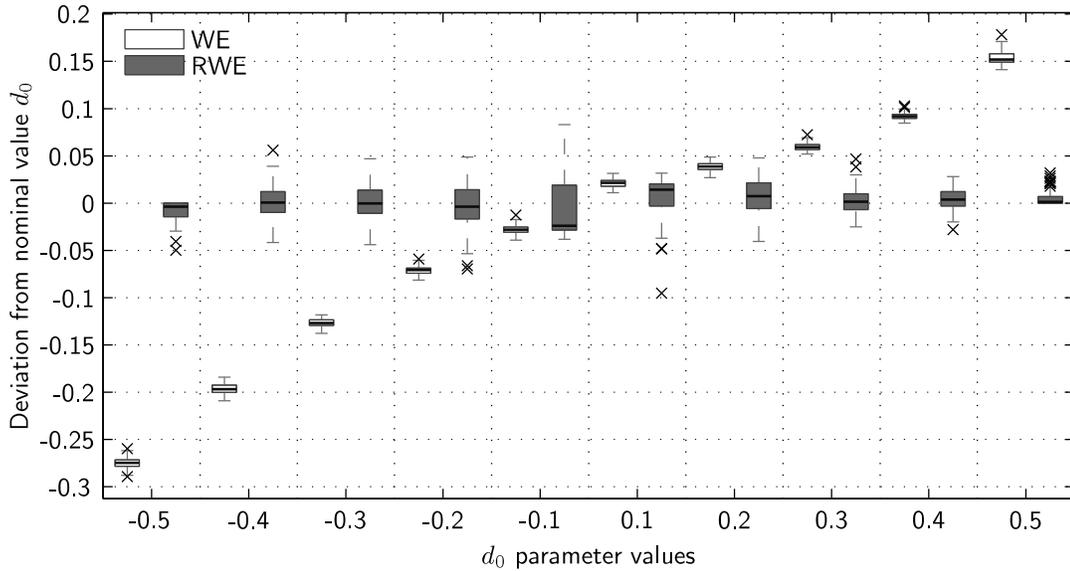


(b) Parameter estimates for FDN via the RWE

Figure 5.21: Comparative results between the likelihood based estimators and their regularised versions. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate's variance. The segmented line represents the true LRD parameter value. The estimators were applied to 100 realisations of FDN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 30 dB.

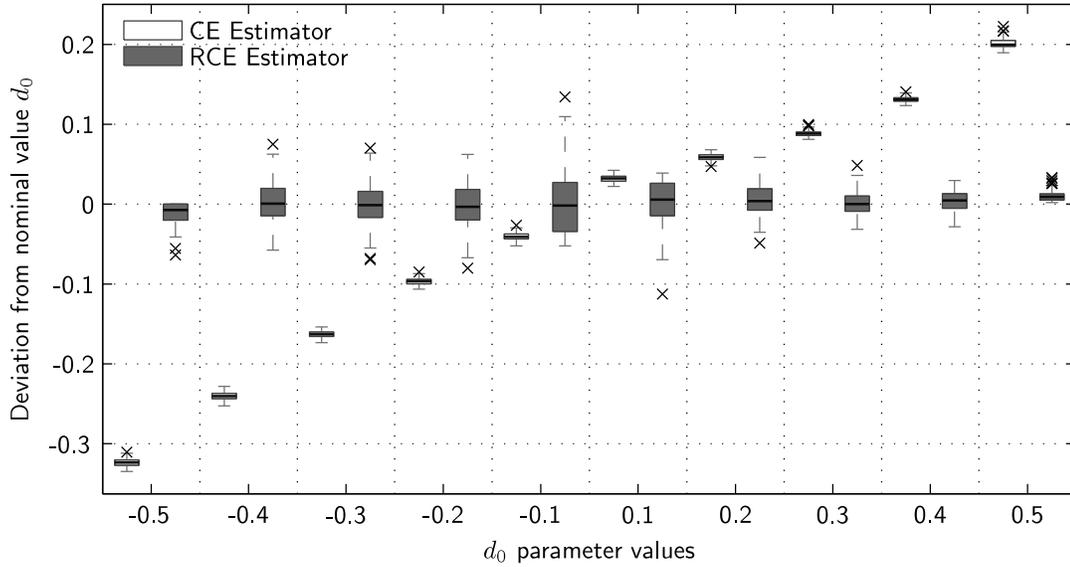


(a) Parameter estimates for FDN via the RCE est.

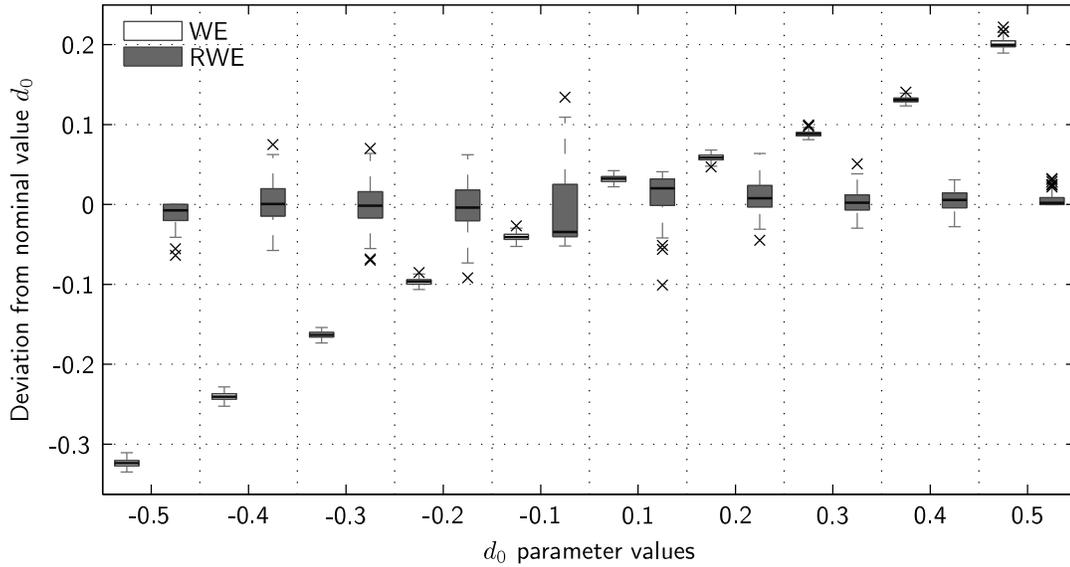


(b) Parameter estimates for FDN via the RWE

Figure 5.22: Comparative results between the likelihood based estimators and their regularised versions. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate's variance. The estimators were applied to 100 realisations of FDN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 10 dB.



(a) Parameter estimates for FDN via the RCE est.



(b) Parameter estimates for FDN via the RWE

Figure 5.23: Comparative results between the likelihood based estimators and their regularised versions. The box-plots of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate's variance. The estimators were applied to 100 realisations of FDN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 5 dB.

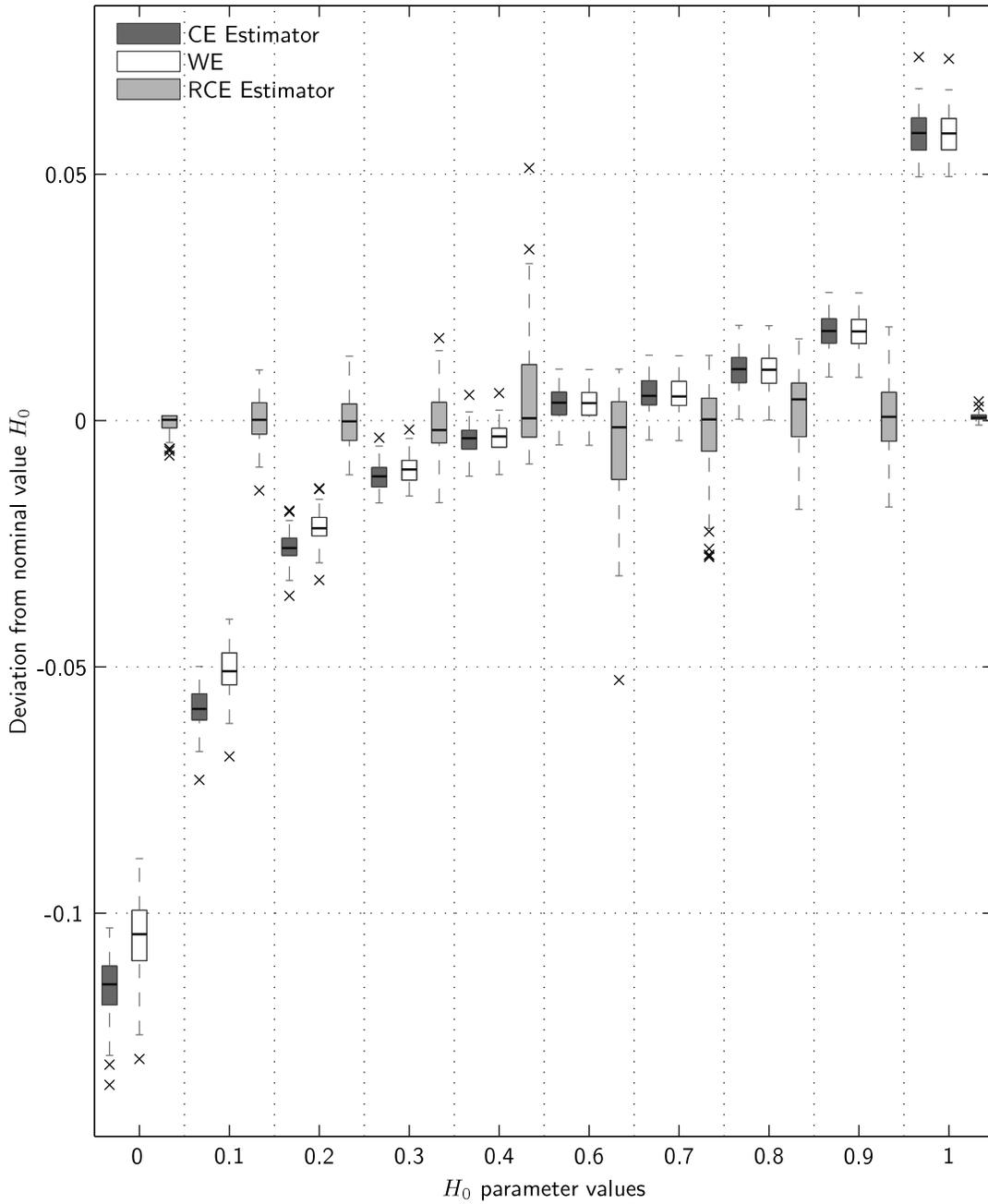


Figure 5.24: Comparative results between the likelihood based estimators and their regularised versions. The box-plot of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate's variance. The estimators were applied to 100 realisations of fGN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 30 dB.

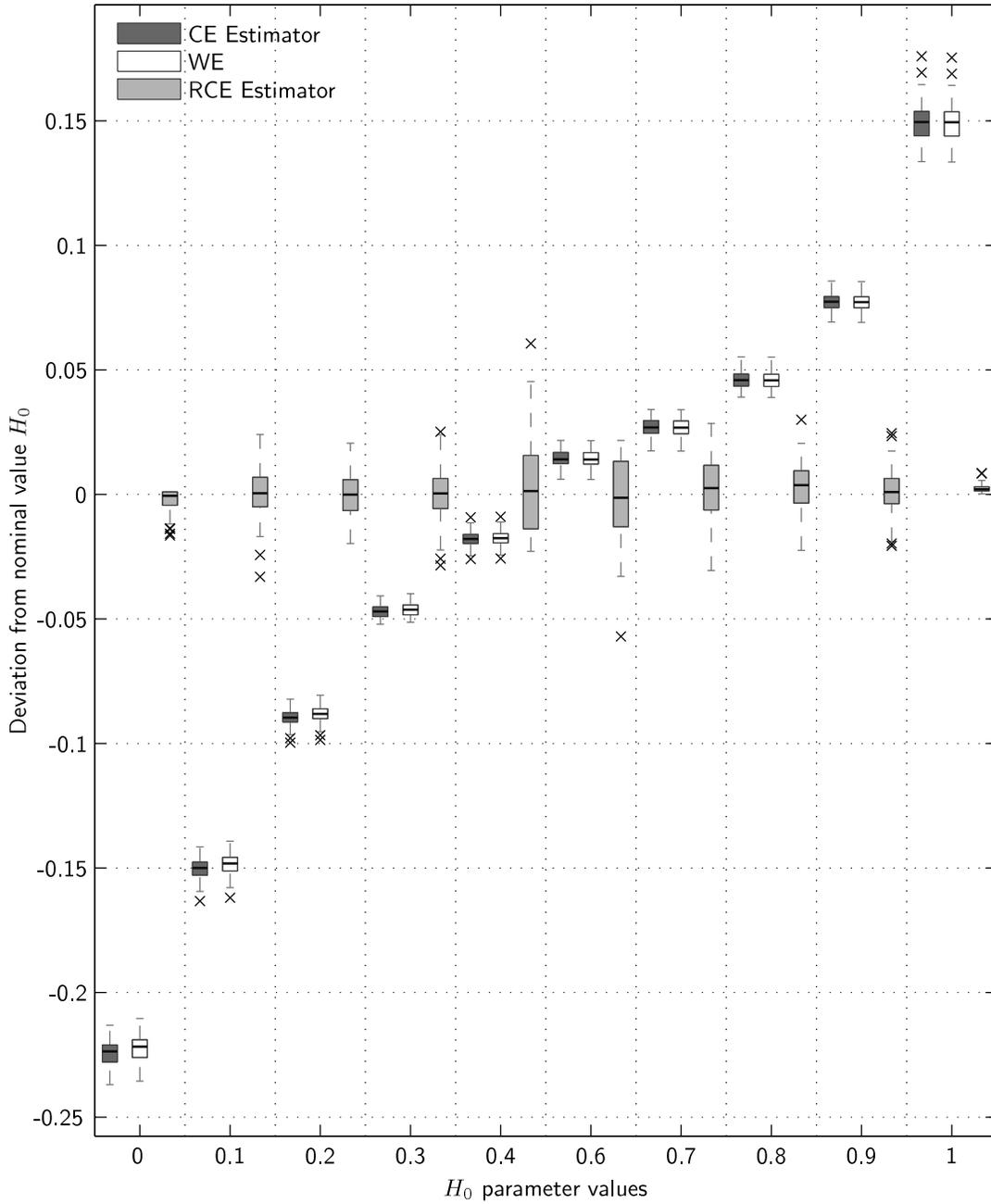


Figure 5.25: Comparative results between the likelihood based estimators and their regularised versions. The box-plot of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate’s variance. The estimators were applied to 100 realisations of fGN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 15 dB.

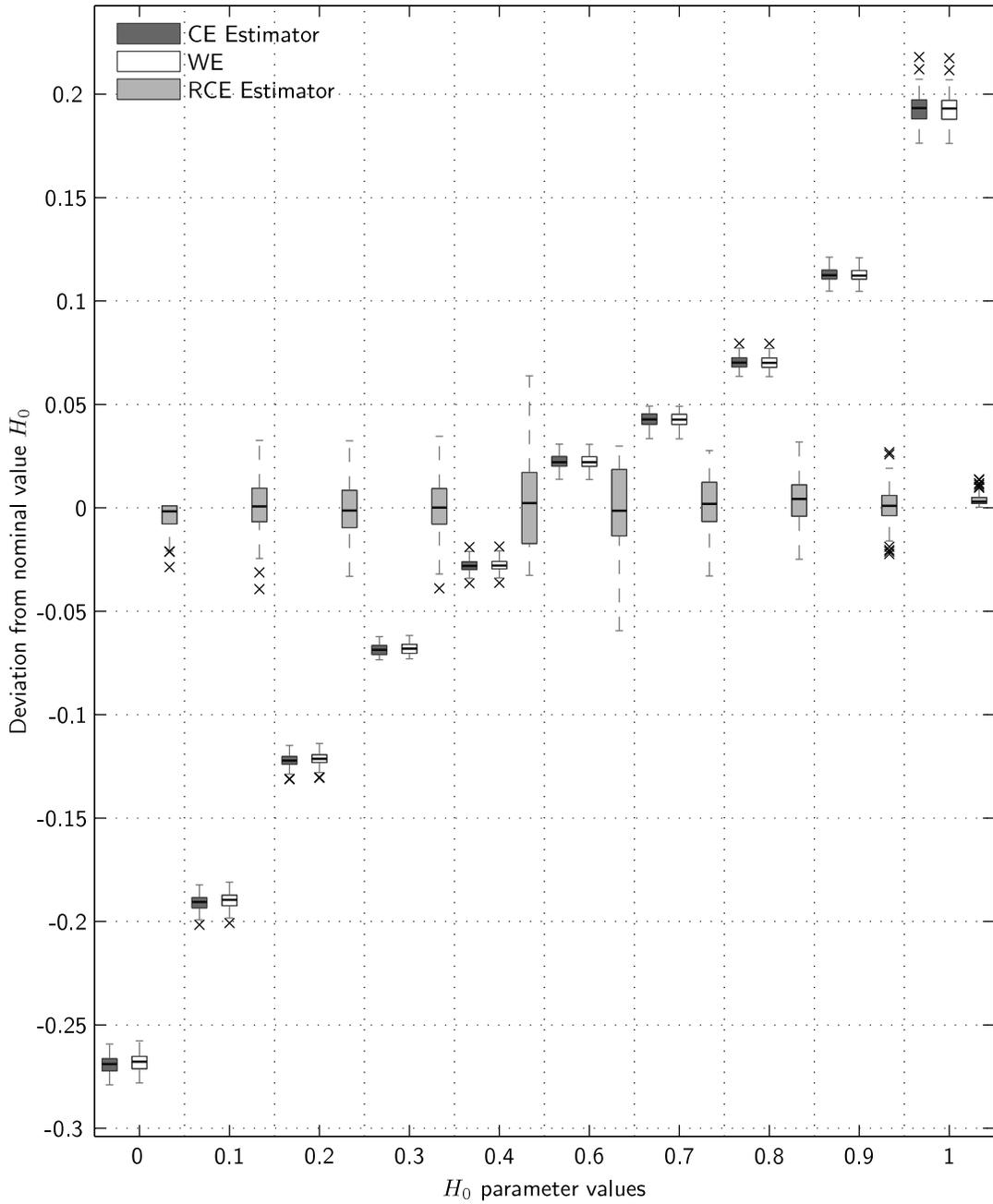


Figure 5.26: Comparative results between the likelihood based estimators and their regularised versions. The box-plot of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate’s variance. The estimators were applied to 100 realisations of fGN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 10 dB.

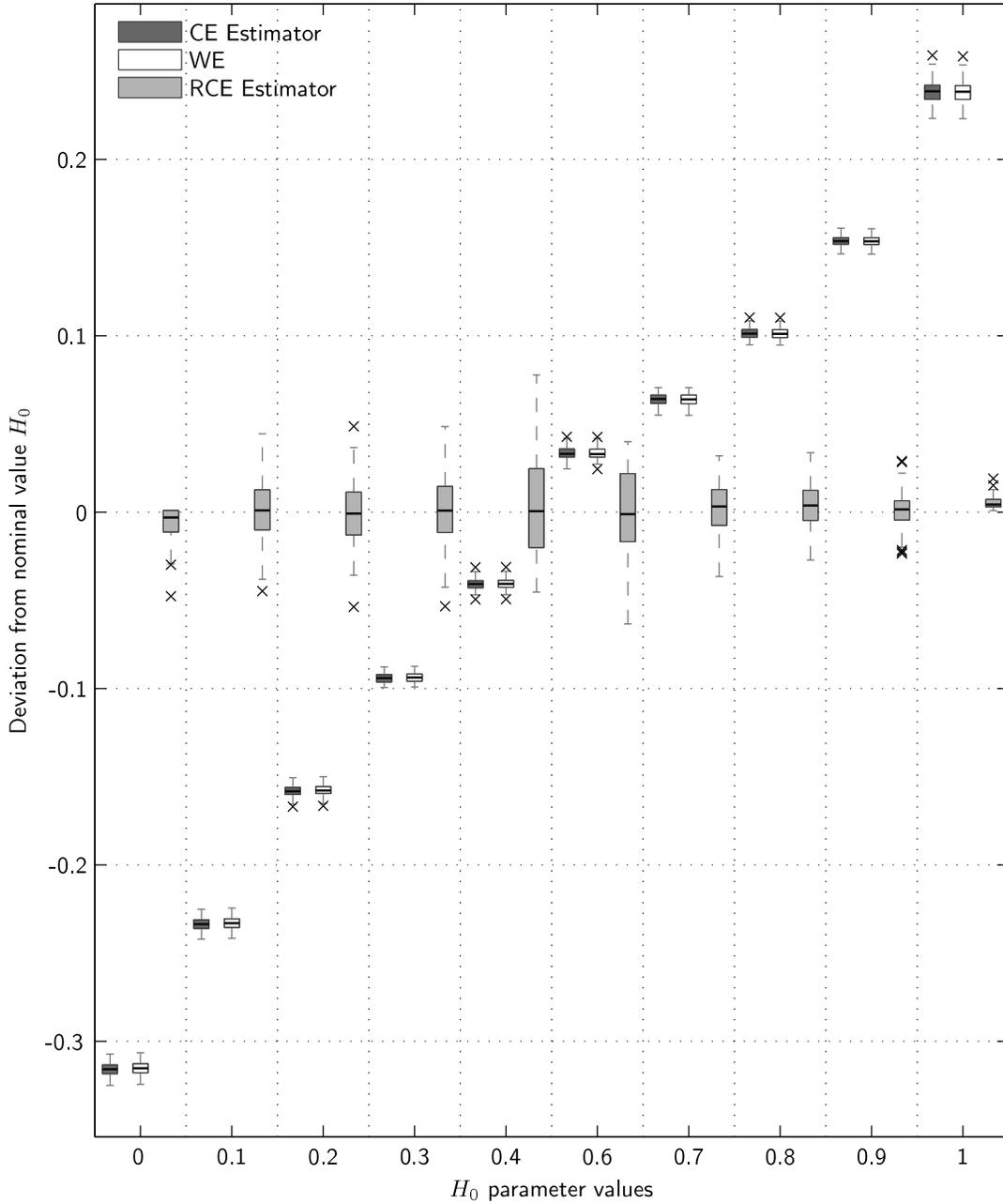


Figure 5.27: Comparative results between the likelihood based estimators and their regularised versions. The box-plot of the estimates have lines at the lower quartile, median and upper quartile values. The whiskers extend 1.5 times the interquartile range from the ends of the box. The outliers are also displayed. It can be observed that the bias of the regularised estimates is dramatically reduced at an expense of a slight increase on the estimate’s variance. The estimators were applied to 100 realisations of fGN with different nominal coefficient d_0 . The realisations had a length of $N = 32768$ data points and they were corrupted at a SNR of 5 dB.

Chapter 6

Conclusions & Future Work

6.1 General Conclusions

Given the wide applicability of long-range dependent processes, there is great interest in the estimation of model parameters that allow their understanding. Maximum likelihood is a method that yields estimates with desirable statistical properties [24, 110]. However, it has been criticised as the computational implementation of this method is impractical in terms of efficiency [79].

This thesis presents a set of computationally efficient estimators for finding the parameters of long-range dependent processes. The estimators presented in this thesis produce accurate results under error-free and error-corrupted measurements.

For the error-free measurements scenario, the concepts of maximum likelihood estimation and circulant embedding were employed to develop the CE Estimator. This estimator is a computationally efficient alternative to maximum likelihood due to the attractive eigenstructure of the circulant embedding matrix.

Theorems 3.3.6 and 3.3.7 and corollary 3.3.8 justify the use of the CE Estimator for LRD process. In particular, theorem 3.3.6 implies that as the matrices grow large, they behave similarly. Corollary 3.3.8 shows that the CE Estimator is asymptotically normal, consistent and efficient.

The Monte-Carlo experiments in section 5.1.1 show that for fGN and FDN and small data samples, the performance of the CE Estimator is superior to that of the WE. Furthermore, for the case of fGN, the CE Estimator is computationally more efficient than the WE. For the case of ARFIMA, the CE Estimator is computationally less efficient than the WE. Finally, for the case of FDN, the

computational efficiency of both estimators is similar.

Theorems 3.3.9, 3.3.10 and 3.3.11 show that the likelihood based estimators are highly sensitive to the effect of measurement errors. This is due to the ill-conditioning of the respective matrices involved in their computation.

For the error-corrupted measurements scenario, the concept of regularisation was employed to improve the robustness of the likelihood based methods to the presence of errors in the measurements. The choice of regularisation structure ensures that the regularised versions of the approximations to the MLE, the RWE and the RCE Estimator, remain with low computational complexity. Under the assumption of Gaussian errors, theorem 4.4.4 shows that the estimates obtained by the regularisation methods are related to the MLE of a model which accounts for the measurement errors.

6.2 Future Work

There are three major research paths that follow after the work presented in this thesis: *a)* extensions to systems with exogenous inputs, *b)* multifractal processes and *c)* the estimator robustness and model validation.

Extension to Systems with Exogenous Inputs

The approach of fitting a mathematical model to input-output relations obtained from an experiment is known as system identification [4]. One of the basic approach to finding the model parameters to such input-output relations is given by the prediction-error method [65]. This method has the advantage that it can be applied to very general model structures within the Box-Jenkins family of models. Its asymptotic accuracy is optimal when the “true” system (or process) can be represented within the model structure. Conversely, when the “true” system cannot be represented within the model structure, the method has reasonable approximation properties [65].

The convergence properties mentioned above hold when the correlations of the process decay exponentially [64, 65]. Thus, when the correlations decay at a much slower rate such convergence results are no longer guaranteed. Therefore, the inclusion of LRD processes into this framework is an interesting and challenging problem.

Extension to Multifractal Processes

In some applications, it has been shown that LRD is not sufficient to properly characterise the set of observations. Particularly, the stationary condition on the increments of fBM can be too restrictive to some applications [7]. Introduced by Benassi et al. [7], multifractional Brownian Motions (mBM) are a generalisation of fBM by letting the LRD coefficient H be a time varying function $H(t)$.

Multifractional Brownian Motion no longer possesses stationary increments. Thus there is no reason to believe that the methods presented in this thesis would effectively estimate the parameters of these kind of processes. Nevertheless, by combining the estimators presented in this thesis with a segmenting framework such as the one presented in [section 4 in 105] the user might be able to differentiate between a multifractal processes and the processes presented in this thesis. This is the subject of further investigation.

Estimator Robustness and Model Validation

The models presented in this thesis are a simple approximation to real world applications. In rare cases, the data obtained from a real life application satisfies all the assumptions employed to develop the theory behind the models and the estimators. For example, the ocular accommodation datasets of [61] have been reported to possess heavy-tailed distributions [62]. The estimators developed in this thesis must be tested under different assumption departures such that their robustness can be assessed.

Another important aspect is that of model validation. While the model structure selection can be performed by the Akaike's and Bayes' Information Criteria [9], the crucial question is whether the identified model is good enough for the intended purpose. Beran [8] provides a goodness of fit test based on the frequency response of the estimates and the periodogram of the time-series. However, in the case of non-stationarities and non-Gaussian time-series the validity of the estimated models remains an open question.

Bibliography

- [1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, 9th ed., M. Abramowitz, Ed. Dover Publications, 1965.
- [2] J. Aldrich, “R. A. Fisher and the Making of Maximum Likelihood 1912-1922,” *Statistical Science*, vol. 12, no. 3, pp. 162–176, 1997.
- [3] K. J. Åström, *Introduction to Stochastic Control Theory*. Academic Press, 1970.
- [4] —, “Maximum likelihood and prediction error methods,” *Automatica*, vol. 16, no. 5, pp. 551–574, 1980.
- [5] J.-M. Bardet and I. Kammoun, “Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2041–2052, 2008.
- [6] J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M. Taqqu, “Generators of long-range dependent processes: A survey,” in *Theory and applications of long-range dependence*. Birkhäuser, 2003, pp. 579–623.
- [7] A. Benassi, S. Cohen, and J. Istas, “Identifying the multifractional function of a Gaussian process,” *Stat Probabil Lett*, vol. 39, no. 4, pp. 337–345, 1998.
- [8] J. Beran, “A goodness-of-fit test for time series with long range dependence,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 54, no. 3, pp. 749–760, 1992.
- [9] —, *Statistics for long-memory processes*. Chapman-Hall, 1994.
- [10] S. Bertelli and M. Caporin, “A note on calculating autocovariances of long-memory processes,” *Journal of Time Series Analysis*, vol. 23, no. 5, pp. 503–508, 2002.

- [11] G. Box and N. Draper, *Empirical Model-Building and Response Surfaces*. Wiley, 1987.
- [12] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, 2nd ed. Holden Day, 1976.
- [13] D. R. Brillinger, *Time Series: Data Analysis and Theory*. SIAM, 2001.
- [14] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer, 1991.
- [15] T. G. Buchman, “The community of the self,” *Nature*, vol. 420, no. 6912, pp. 246–51, 2002.
- [16] ———, “Nonlinear dynamics, complex systems, and the pathobiology of critical illness,” *Curr Opin Crit Care*, vol. 10, no. 5, pp. 378–82, 2004.
- [17] M. Calder and R. Davis, “Introduction to Whittle (1953): The analysis of multiple stationary time series,” *Breakthroughs in Statistics, Springer-Verlag*, vol. 3, pp. 141–148, 1997.
- [18] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. Duxbury Press, 2001.
- [19] N. Chan and W. Palma, “State space modeling of long-memory processes,” *Annals of Statistics*, vol. 26, no. 2, pp. 719–740, 1998.
- [20] W. W. Chen, C. M. Hurvich, and Y. Lu, “On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series,” *J. Amer. Statist. Assoc.*, vol. 101, no. 474, pp. 812–822, 2006.
- [21] Z. Chen, P. C. Ivanov, K. Hu, and H. E. Stanley, “Effect of nonstationarities on detrended fluctuation analysis,” *Physical Review E*, vol. 65, no. 4, pp. 1–15, 2002.
- [22] Z. Chen, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. C. Ivanov, “Effect of nonlinear filters on detrended fluctuation analysis,” *Physical Review E*, vol. 71, no. 1, pp. 1–11, 2005.

- [23] P. F. Craigmile, “Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes,” *J. Time Ser. Anal.*, vol. 24, no. 5, pp. 505–511, 2003.
- [24] R. Dahlhaus, “Efficient parameter estimation for self-similar processes,” *The Annals of Statistics*, vol. 17, no. 4, pp. 1749–1766, 1989.
- [25] R. B. Davies and D. S. Harte, “Tests for hurst effect,” *Biometrika*, vol. 74, no. 1, pp. 95–101, 1987.
- [26] P. J. Davis, *Circulant Matrices*, 2nd ed. AMS Chelsea Publishing, 1994.
- [27] D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, and G. Ninot, “Fractal analyses for ‘short’ time series: a re-assessment of classical methods,” *Journal of Mathematical Psychology*, vol. 50, no. 6, pp. 525–544, 2006.
- [28] A. Dembo, C. L. Mallows, and L. A. Shepp, “Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation,” *Information Theory, IEEE Transactions on*, vol. 35, no. 6, pp. 1206–1212, 1989.
- [29] M. Deriche and A. H. Tewfik, “Maximum likelihood estimation of the parameters of discrete fractionally differenced gaussian noise process,” *Signal Processing, IEEE Transactions on*, vol. 41, no. 10, pp. 2977–2989, 1993.
- [30] C. R. Dietrich and G. N. Newsam, “Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix,” *SIAM J. Sci. Comput.*, vol. 18, no. 4, pp. 1088–1107, 1997.
- [31] H. Ding, S. Crozier, and S. Wilson, “A new heart rate variability analysis method by means of quantifying the variation of nonlinear dynamic patterns,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1590–1597, 2007.
- [32] J. A. Doornik and M. Ooms, “Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models,” *Computational Statistics and Data Analysis*, vol. 42, no. 3, pp. 333–348, 2003.

- [33] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers, 2000.
- [34] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed. Wiley, 1968, vol. 1.
- [35] ———, *An Introduction to Probability Theory and its Applications*, 2nd ed. Wiley, 1971, vol. 2.
- [36] P. Ferreira, “Localization of the eigenvalues of Toeplitz matrices using additive decomposition, embedding in circulants, and the Fourier transform,” *Proceedings of the 10th IFAC Symposium on System Identification*, pp. 271–276, 1994.
- [37] R. Fischer and M. Akay, “A comparison of analytical methods for the study of fractional brownian motion,” *Annals of Biomedical Engineering*, vol. 24, no. 4, pp. 537–543, 1996.
- [38] R. Fischer, M. Akay, P. Castiglioni, and M. Di Rienzo, “Multi-and monofractal indices of short-term heart rate variability,” *Medical and Biological Engineering and Computing*, vol. 41, no. 5, pp. 543–549, 2003.
- [39] R. Fox and M. S. Taqqu, “Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series,” *The Annals of Statistics*, vol. 14, no. 2, pp. 517–532, 1986.
- [40] N. Gache, P. Flandrin, and D. Garreau, “Fractal dimension estimators for fractional brownian motions,” *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, pp. 3557–3560 vol.5, 1991.
- [41] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C. K. Peng, and H. E. Stanley, “Fractal dynamics in physiology: alterations with disease and aging,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 1, pp. 2466–2472, 2002.
- [42] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

- [43] C. W. J. Granger and A. Andersen, “On the invertibility of time series models,” *Stochastic Processes and their Applications*, vol. 8, no. 1, pp. 87–92, 1978.
- [44] C. W. J. Granger and R. Joyeux, “An introduction to long-memory time series models and fractional differencing,” *Journal of Time Series Analysis*, vol. 1, no. 1, pp. 15–29, 1980.
- [45] R. M. Gray, “Toeplitz and circulant matrices,” *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [46] J. A. Gubner, “Theorems and fallacies in the theory of long-range-dependent processes,” *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 1234–1239, 2005.
- [47] E. J. Hannan, “The asymptotic theory of linear time-series models,” *Journal of Applied Probability*, vol. 10, no. 1, pp. 130–145, 1973.
- [48] E. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. John Wiley & Sons, 1988.
- [49] M. A. Hauser, “Semiparametric and nonparametric testing for long memory: A Monte Carlo study,” *Empirical Economics*, vol. 22, no. 2, pp. 247–271, 1997.
- [50] J. R. M. Hosking, “Fractional differencing,” *Biometrika*, vol. 68, no. 1, pp. 165–176, 1981.
- [51] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, “Effect of trends on detrended fluctuation analysis,” *Physical Review E*, vol. 64, no. 1, pp. 1–19, 2001.
- [52] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, and S. Havlin, “From 1/f noise to multifractal cascades in heartbeat dynamics,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 11, no. 3, pp. 641–652, 2001.
- [53] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, “Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics,” *AJP - Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–1084, 1996.

- [54] T. Kailath, *Linear Systems*. Prentice-Hall, 1980.
- [55] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, “Detecting long-range correlations with detrended fluctuation analysis,” *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3-4, pp. 441–454, 2001.
- [56] R. L. Kashyap and A. R. Rao, *Dynamic Stochastic Models from Empirical Data*. Academic Press, 1976.
- [57] M. S. Keshner, “1/f noise,” *Proceedings of the IEEE*, vol. 70, no. 3, pp. 212–218, 1982.
- [58] M. E. Kilmer and D. P. O’Leary, “Choosing regularization parameters in iterative methods for ill-posed problems,” *J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1204–1221, 2001.
- [59] D. Koutsoyiannis, “The Hurst phenomenon and fractional Gaussian noise made easy,” *Hydrological Sciences Journal*, vol. 47, no. 4, pp. 573–596, 2002.
- [60] D. Krawczyk-Stańdo and M. Rudnicki, “Regularization parameter selection in discrete ill-posed problems - the use of the U-Curve,” *International Journal of Applied Mathematics and Computer Science*, vol. 17, no. 2, pp. 157–164, 2007.
- [61] C. Leahy, C. Leroux, C. Dainty, and L. Diaz-Santana, “Temporal dynamics and statistical characteristics of the microfluctuations of accommodation: dependence on the mean accommodative effort,” *Opt Express*, vol. 18, no. 3, pp. 2668–81, 2010.
- [62] C. Leahy, “Temporal dynamics and statistical characteristics of ocular wavefront aberrations and accommodation,” Ph.D. dissertation, National University of Ireland, Galway, 2010.
- [63] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, “On the self-similar nature of ethernet traffic (extended version),” *Networking, IEEE/ACM Transactions on*, vol. 2, no. 1, pp. 1–15, 1994.
- [64] L. Ljung, “Convergence analysis of parametric identification methods,” *Automatic Control, IEEE Transactions on*, vol. 23, no. 5, pp. 770–783, 1978.

- [65] ———, *System Identification - Theory for the User*, 2nd ed. Prentice-Hall, 1999.
- [66] M. Loève, *Probability Theory II*, 4th ed. Springer-Verlag, 1977.
- [67] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, “Fractional brownian motion: A maximum likelihood estimator and its application to image texture,” *Medical Imaging, IEEE Transactions on*, vol. 5, no. 3, pp. 152–161, 1986.
- [68] B. B. Mandelbrot, “Some noises with $1/f$ spectrum, a bridge between direct current and white noise,” *Information Theory, IEEE Transactions on*, vol. 13, no. 2, pp. 289–298, 1967.
- [69] B. B. Mandelbrot and J. W. V. Ness, “Fractional Brownian motions, fractional noises and applications,” *SIAM Review*, vol. 10, no. 4, pp. 422–437, Oct., 1968.
- [70] B. B. Mandelbrot and J. R. Wallis, “Noah, Joseph, and operational hydrology,” *Water Resour. Res.*, vol. 4, no. 5, pp. 909–918, 1968.
- [71] ———, “Computer experiments with fractional Gaussian noises. part 2: Rescaled ranges and spectra,” *Water Resources Research*, vol. 5, no. 1, pp. 242–259, 1969.
- [72] A. I. McLeod, “Derivation of theoretical autocovariance function of autoregressive-moving average time series,” *Applied Statistics*, vol. 24, no. 2, pp. 255–256, 1975.
- [73] A. I. McLeod and K. W. Hipel, “Preservation of the rescaled adjusted range 1. A Reassessment of the Hurst Phenomenon,” *Water Resources Research*, vol. 14, no. 3, pp. 491–508, 1978.
- [74] J. L. Melsa and A. P. Sage, *An Introduction to Probability and Stochastic Processes*, T. Kailath, Ed. Prentice-Hall, 1973.
- [75] G. M. Molchan, “Historical comments related to fractional brownian motion,” in *Theory and applications of long-range dependence*. Birkhäuser, 2003, pp. 39–42.

- [76] R. M. Neal, “Bayesian Learning for Neural Networks,” Ph.D. dissertation, University of Toronto, 1995.
- [77] A. Neumaier, “Solving ill-conditioned and singular linear systems: A tutorial on regularization,” *SIAM Review*, vol. 40, no. 3, pp. 636–666, 1998.
- [78] B. Ninness, “Strong laws of large numbers under weak assumptions with application,” *Automatic Control, IEEE Transactions on*, vol. 45, no. 11, pp. 2117–2122, 2000.
- [79] —, “Estimation of 1/f noise,” *Information Theory, IEEE Transactions on*, vol. 44, no. 1, pp. 32–46, 1998.
- [80] W. Palma, *Long-Memory Time Series: Theory and Methods*. Wiley-Interscience, 2007.
- [81] E. Panas, “Estimating fractal dimension using stable distributions and exploring long memory through ARFIMA models in athens stock exchange,” *Applied Financial Economics*, vol. 11, pp. 395–402, 2001.
- [82] V. Paxson, “Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic,” *Computer Communications Review*, vol. 27, no. 5, pp. 5–18, 1997.
- [83] C. Peng, S. Havlin, H. Stanley, and A. Goldberger, “Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 5, no. 1, pp. 82–87, 1995.
- [84] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, “Mosaic organization of dna nucleotides,” *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.
- [85] E. Perrin, R. Harba, R. Jennane, and I. Iribarren, “Fast and exact synthesis for 1-d fractional Brownian motion and fractional Gaussian noises,” *Signal Processing Letters, IEEE*, vol. 9, no. 11, pp. 382 – 384, 2002.
- [86] B. Pilgram and D. T. Kaplan, “A comparison of estimators for 1/f noise,” *Physica D: Nonlinear Phenomena*, vol. 114, no. 1-2, pp. 108–122, 1998.

- [87] M. M. Platisa, S. Mazic, Z. Nestorovic, and V. Gal, “Complexity of heart-beat interval series in young healthy trained and untrained men,” *Physiological Measurement*, vol. 29, pp. 439–450, 2008.
- [88] P. M. Robinson, “Long-memory time series,” in *Time Series with Long Memory*, P. M. Robinson, Ed. Oxford University Press, 2003, pp. 1–48.
- [89] C. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, “Robust optimal experiment design for system identification,” *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.
- [90] G. Samorodnitsky, “Long memory and self-similar processes,” *Annales-Faculte des Sciences Toulouse Mathematiques*, vol. 15, no. 1, pp. 107–123, 2006.
- [91] J. Scargle, “Studies in astronomical time series analysis. II- Statistical aspects of spectral analysis of unevenly spaced data,” *The Astrophysical Journal*, vol. 263, pp. 835–853, December 15 1982.
- [92] D. T. Schmitt and P. C. Ivanov, “Fractal scale-invariant and nonlinear properties of cardiac dynamics remain stable with advanced age: a new mechanistic picture of cardiac control in healthy elderly,” *AJP: Regulatory, Integrative and Comparative Physiology*, vol. 293, no. 5, pp. R1923–R1937, 2007.
- [93] S. D. Silvey, *Statistical Inference*. Chapman-Hall, 1995.
- [94] T. Söderström, *Discrete-time Stochastic Systems*, 2nd ed. Springer-Verlag, 2002.
- [95] T. Söderström, J. Jezek, and V. Kucera, “An efficient and versatile algorithm for computing the covariance function of an ARMA process,” *Signal Processing, IEEE Transactions on* DOI - 10.1109/78.678473, vol. 46, no. 6, pp. 1591–1600, 1998.
- [96] V. Solo, “Intrinsic random functions and the paradox of 1/f noise,” *SIAM Journal on Applied Mathematics*, vol. 52, no. 1, pp. 270–291, Feb., 1992.
- [97] F. Sowell, “Maximum likelihood estimation of stationary univariate fractionally integrated time series models,” *Journal of Econometrics*, vol. 53, no. 1-3, pp. 165–188, 1992.

- [98] —, “Modeling long-run behavior with the fractional arima model,” *Journal of Monetary Economics*, vol. 29, no. 2, pp. 277–302, 1992.
- [99] M. Spivak, *Calculus*, 3rd ed. Addison Wesley, 1994.
- [100] M. S. Taqqu, “Fractional brownian motion and long-range dependence,” in *Theory and applications of long-range dependence*. Birkhäuser, 2003, pp. 5–38.
- [101] M. S. Taqqu and V. Teverovsky, “Robustness of Whittle-type estimators for time series with long-range dependence,” *Comm. Statist. Stochastic Models*, vol. 13, no. 4, pp. 723–757, 1997.
- [102] M. S. Taqqu, V. Teverovsky, and W. Willinger, “Estimators for long-range dependence: an empirical study,” *Fractals*, vol. 3, no. 4, pp. 785–798, 1995.
- [103] K. Torre, D. Delignieres, and L. Lemoine, “ $1/f^\beta$ fluctuations in bimanual coordination: an additional challenge for modeling,” *Experimental Brain Research*, vol. 183, pp. 225–234, 2007.
- [104] —, “Detection of long-range dependence and estimation of fractal exponents through arfima modelling,” *British Journal of Mathematical and Statistical Psychology*, vol. 60, pp. 85–106, 2007.
- [105] O. Vivero and W. P. Heath, “On MLE methods for dynamical systems with fractionally differenced noise spectra,” *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pp. 1842–1847, 2009.
- [106] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *Audio and Electroacoustics, IEEE Transactions on*, vol. 15, no. 2, pp. 70–73, 1967.
- [107] B. J. West and N. Scafetta, “Nonlinear dynamical model of human gait,” *Physical Review E*, vol. 67, no. 5, 2003.
- [108] P. Whittle, “Estimation and information in stationary time series,” *Arkiv för matematik*, vol. 2, no. 23, pp. 423–434, 1953.

- [109] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Statistical Science*, vol. 10, no. 1, pp. 67–85, 1995.
- [110] Y. Yajima, "On estimation of long-memory time series models," *Australian & New Zealand Journal of Statistics*, vol. 27, no. 3, pp. 303–320, 1985.