UNIVERSITETET I AGDER

Estimators of long range dependence

A survey of finite samples and robustness

Lars Tjensvold Olsen

Supervisor
Jochen Jungeilges

This Master's Thesis is carried out as a part of the education at the University of Agder and is therefore approved as a part of this education. However, this does not imply that the University answers for the methods that are used or the conclusions that are drawn.

University of Agder, 2012
Faculty of Economics and Social Sciences
Department of Economics and Business Administration

Abstract

In traditional financial theory the returns of prices are assumed to be independent of each other, they are said to have *short memory*. However, it has been shown that returns in many cases are correlated and these instance are said to possess *long memory* or *long range dependence*. This phenomenon is also found in other research disciplines such as biology, economics, physics, linguistics and hydrology. Long memory can not be established on beforehand but has to be estimated. The goal of this thesis is to evaluate seven estimators of long range dependence by generating time series with varying known long memory parameters and then measure the performance of the estimators under such environments. The estimators are also evaluated when estimating a long memory time series distorted by heavy tailed noise for varying levels of corruption. The noise has similar features to what is observed in financial data. To the author's knowledge this study of estimation algorithms has the broadest coverage of long memory parameters and noise in terms of numbers of replications which make the results statistically valid. The general observation is that a heavy persistent or heavy anti-persistent series leads to less accurate estimates although some estimators are unaffected by this. There are also differences among the point estimators in how they perform under different sample sizes. When adding noise to the time series the estimation is affected little by persistent series but is affected heavily by the anti-persistent series.

Acknowledgements

This master thesis consisting of 30 ECTS credits is written as a part of my Master of Science degree in business and administration at the University of Agder (Handelshøyskolen i Kristiansand).

I am greatly indebted to my advisor, Professor Jochen Jungeilges, for guidance and inspiration throughout the course of this work. He presented me with the topic and have through several discussions introduced me with many ideas while preparing this thesis. Further I would like to thank my best friend and soul mate Camilla, my parents Kjersti and Bjørn and my sister Mari for support. Finally I would also like to thank Sylte and Kevin for the enlightening conversations concerning the topic.

Kristiansand, May 31, 2012 Lars Tjensvold Olsen

Contents

Chapter 1

Introduction

Long range dependence, *long memory* or *long range correlations* are synonyms for a property found in a type of stationary process. It differs from traditional financial theories in which returns are assumed to have short range correlations or no memory. Although this thesis will have an angle of econometrics the disciplines interested in this topic are widespread, including finance and economics [48, 30, 52, 40, 50], econometrics [62], DNA sequences [57], climates [61], pollution [74], and linguistics [1]. Within the mentioned articles a variety of issues are addressed: statistical estimation of long range dependence parameters, detection of long range dependence in data, and simulation of long range dependence among others. The majority of the studies performed on the topic of long range dependence are aimed at revealing long range dependence in empirical data whereas only a smaller part of this field consider the simulation and estimation of long range dependent data.

As long range dependent behavior cannot be assumed a priori, but requires establishment, the need for estimation procedures arises. Several studies has been conducted where the performance of estimators of long range dependence is examined [70, 4, 60]. Most of these studies rely on data generating processes which have normal distributions and uncomplicated correlation structures in the higher order moments. Many of these studies evaluate the performance of long memory estimators based on the ability to estimate independent time series with short memory. The same studies often conclude with the time series having short memory or not short memory.

The goal of this thesis is to evaluate a set of commonly used estimators of long memory. This will be done by simulating two different long range dependent data sets with known parameters and then estimate these under different parameter regimes. Then the generated data sets are then *corrupted* by a noise which has similar properties to empirical (economic and financial) data.

The experiments are carried out by first observing how the estimators perform under

Gaussian conditions for different sample sizes and intensities of long range dependence. The long range dependent time series are then corrupted by alternating levels of noise at different sample sizes and intensities of long range dependence. It is then observed for which levels of corruption the estimators break down.

The contribution to the current research will be the broadness of the simulation study and the investigation of how the estimators react to noise in the underlying process. Although similar experiments have been carried out none has been similarly performed, as far as the author is aware.

The thesis starts out by presenting definitions and concepts regarding time series analysis and stationary time series which is the foundation of long range dependent series. The second order properties, namely the autocorrelations and spectral densities, of short memory time series are covered.

The second part accounts for the concept of long range dependence and the differences between short range dependent series by investigating the second order properties of such stationary processes as these constitute most of the definitions of long memory. In particular, the second order properties are the behavior of the autocorrelations and the spectral densities where slowly decaying autocorrelations or a spectral density with a pole at the origin characterize a long range dependent process.

As follows, the third part is a scrutiny of a set of commonly used estimators of long range dependence. Each procedure is covered and the known finite sample properties are discussed.

In the fourth part the estimators are applied to pure long range dependent series with different sample sizes and different intensity of long range dependence. In economics the length of time series are often short due to low frequency measurements as opposed to financial data which can be recorded by the minute. Finite sample properties for estimating simulated long memory data is acquired for different lengths and different intensities of long memory. It is shown that the previous research which is based on fewer parameters does a poorer job describing the true performance of the estimators.

The fifth part investigates how the estimators are reacting to the presence of a noise. The goal is to observe when an estimator deviates from its baseline performance given a corrupted long memory series. This experiment is driven by the fact that empirical data often have distributions that diverge from Gaussianity since they have heavy tails. It is shown that in the presence of a noise, the estimator in general breaks down faster when the long memory series is anti-persistent.

The last part concludes the thesis by applying the introduced concepts on a set of different financial time series and a brief comparison of the higher order correlations between corrupted and empirical data. The data are from stock indexes, exchange rates and bond maturity rates.

Chapter 2

Time series analysis

2.1 Concepts of time series analysis and stochastic processes

In order to introduce the reader on the subject of long range dependence the most salient features of time series and stochastic processes are presented. The concepts introduced here is underlying the discussions that follow in later chapters.

A *time series* is a sequence of values ordered by a time parameter which can be distinguished by its measurement process. A *discrete-time series* has observations that are made at pre-determined points in time whereas *continuous-time series* are obtained when observations are written continuously over some time interval. Note that for a discrete-time series the notation x_t is employed opposite to $x(t)$ where the observations are recorded continuously. Further on, a time series is commonly accepted as a stochastic process which is observed over time. Throughout the thesis a time series will be considered as one realization of the underlying stochastic process. The time set $\mathbb{T} \in \mathbb{Z}$ denotes the number of observations whereas the sub-script t describes the exact observation.

As the nature of a time series constitutes a sequential structure where it is considered a dependence between x_t and $x_{t\pm1}$, the classical i.i.d. statistical techniques are invalid. Tests such as t, F and χ^2 will all give invalid results due to the violation of the independence assumption.

Definition 2.1.1 (Stochastic process [11])**.** *A stochastic process is a family of random variables* $\{X_t\}_{t=0}^{\mathbb{T}}$ *defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$.

Remark 1. *Observations* x_t *is a realized value of a random variable* X_t*. The time series* $\{x_t\}_{t=0}^{\mathbb{T}}$ *is a realization of the family of random variables* $\{X_t\}_{t=0}^{\mathbb{T}}$.

This framework makes it possible to reason the outcomes of an experiment in addition to calculating probabilities. The outcome space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of;

i) the *sample space* Ω which constitutes the set of possible outcomes e.g. $f : [a, b] \to \mathbb{R}$,

ii) the σ*-field* $\mathcal{F}$ that is a collection of subsets of Ω and

iii) the probability measure $\mathbb{P}$, e.g. $\mathbb{P} : \mathcal{F} \to [0, 1]$.

The concepts of weak and strong stationarity will later be addressed since its a required assumption of later models. It is imperative that there is knowledge about the stability of the observed data as this makes it possible to predict future outcomes.

Definition 2.1.2 (Weak Stationarity)**.** *A time-series process* $\{x_t\}_{t=0}^{\mathbb{T}}$ *is considered weakly stationary (or covariance stationary) if,*

i) $\mathbb{E}[x_t] = \mu < \infty \quad \forall t$

ii) $\mathbb{V}[x_t] = \mathbb{E}[x_t - \mu]^2 = \sigma_x^2 < \infty \quad \forall t$

iii) $Cov[x_t, x_{t+s}] = \gamma(t, t+s) \overset{\text{def}}{=} \gamma(s)$ *is independent for of* t *for all* s*.*

Definition 2.1.3 (Strong Stationarity)**.** *A time-series process* $\{x_t\}_{t=0}^{\mathbb{T}}$ *is strongly stationary if for each* s *the joint probability distribution of the sequence* $\{x_t, x_{t+1}, \ldots, x_{t+s}\}$ *is the same for all* s*.*

As seen in Definition 2.1.2 and Definition 2.1.3, weak stationarity is implied from strong stationarity given that the process is normally distributed. Without this condition, other moments may depend on t. Strong stationarity however, does not imply weak stationarity since $\mathbb{E}[x_t^2]$ must be finite. Nonetheless, weak stationarity has fewer restrictions as the distribution in principle can be changing in time. As pointed out by Greene [23], it will usually be difficult to find a process that is weakly but not strongly stationary. Along these lines only weak stationarity is required most of the time as it relaxes some of the excessive assumptions from strict stationarity.

Definition 2.1.4 (The Autocovariance Function)**.** *If* $\{x_t\}_{t=0}^{\mathbb{T}}$ *is a process such that* $\mathbb{V}[x_t] < \infty$ *for each* $t \in \mathbb{T}$*, then the autocovariance function* $\gamma(t, s)$ *of* x_t *is defined by*

$$\gamma(t, s) = Cov[x_t, x_{t+s}] = \mathbb{E}[(x_t - \mathbb{E}\{x_t\})(x_s - \mathbb{E}\{x_s\})], \quad t, s \in \mathbb{T} \tag{2.1.1}$$

For a stationary process the mean $\mathbb{E}[x_t] = E[x]$ and the autocovariance function $\gamma(t, s) = \gamma(s)$.

Remark 2. *Note that for stationary processes* $\gamma(s) = \gamma(-s)$*. In addition if there is no constant terms then* $\mathbb{E}[x_t] = 0$*, thus* $\gamma(s) = \mathbb{E}[x_t x_{t-s}]$ *and* $\gamma(0) = \mathbb{E}[x_t^2] = \mathbb{V}[x_t]$*.*

2.2 Basic time series processes

Throughout the thesis, a number of processes will be encountered. The more complex processes are a combination of simple ones introduced here. First of all the concept of *white noise* should be introduced. This process is also known in literature [13] as a *purely random process* and serves as an infrastructure for time series.

White Noise
A stochastic process $\{\varepsilon_t\}$ fulfilling $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t, \varepsilon_s] = 0$, $t \neq s$ (absence of serial correlation) and $\mathbb{E}[\varepsilon_t^2] = \sigma^2 < \infty$ (constant conditional variance) is called a white noise process.

The most general case, the normally distributed white-noise process, is known as a *Gaussian white noise process*

$$\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\varepsilon^2). \tag{2.2.1}$$

As a convention throughout the thesis the term ε_t will always mean white noise unless otherwise stated.

Further on, some very basic processes will be briefly introduced as these are building blocks for the more advanced ones that will later be encountered.

Random Walk
A process $\{x_t\}$ follows a *random walk* when it is represented as

$$x_t = \beta + x_{t-1} + \varepsilon_t \tag{2.2.2}$$

where β is a constant and represents the drift if non-zero. Given that the process has no drift, then it can be represented as the initial state x_0 plus the error terms:

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t \\ x_t &= x_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &\vdots \\ x_t &= x_0 + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_1 \end{aligned}$$

The mean of the random walk is then found in

$$\mathbb{E}[x_t] = \mathbb{E}[x_0 + \sum_{j=1}^{t} \varepsilon_j] = \mathbb{E}[x_0] + \sum_{j=1}^{t} \mathbb{E}[\varepsilon_j] = 0$$

Later it will be deduced that the random walk process is non-stationary as the variance is dependent on t. In financial application the random walk is known as the first stochastic process that was used to model the development of financial assets such as stock prices. Given the normally distributed increments ε_t, its application is limited.

Autoregressive (AR)
A process $\{x_t\}$ is said to be an *autoregressive* process of order p if it satisfies the linear difference equation

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \varepsilon_t, \quad \forall t \in \mathbb{Z}, \quad \phi_j \in \mathbb{R} \tag{2.2.3}$$

where $\phi_j \neq 0$ and ε_t is white noise. An AR(1) process can be written as $x_t = \phi_1 x_{t-1} + \varepsilon_t$. The term autoregression stems from the fact that the value x_t depends linearly on the last p values and because it has the appearance of a regression model. For $\phi_p = 1$ the model reduces to a random walk as in Equation (2.2.2), and becomes non-stationary. This is the case for $|\phi| > 1$, where the series becomes explosive. For $|\phi < 1|$, given the autocorrelation function $\rho(j) = \phi^s$, the process is stationary.

Moving Average (MA)
A process $\{x_t\}$ is said to be a *moving average* process of order q if it satisfies the rewritten linear difference equation

$$x_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{2.2.4}$$

where ε_t is white noise. An MA(1) process can be written as $x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$. Knowing that $\mathbb{E}[\varepsilon_t] = 0$ it can be determined that $\mathbb{E}[x_t] = 0$.

Definition 2.2.1 (Lag operator)**.** *A lag operator denoted L imposes an element of a time series to produce the previous time unit.*

The lag operator, also denoted by other authors as the backshift operator B, is a notational convenience that helps describing processes which has a lagged term. Some of the

features of the lag operator are

$$\begin{aligned} Lx_t &= x_{t-1} \\ L^k x_t &= x_{t-k} \\ L^{-k} x_t &= x_{t+k}. \end{aligned} \tag{2.2.5}$$

In addition, it can be formed polynomials of it such as

$$a(L)^p = a_0 + a_1 L + a_2 L^2 + \ldots + a_p L^p$$

and

$$a(L)^p x_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p}.$$

Remark 3. *When differencing x_t note that*

$$x_t - x_{t-1} = (1 - L)x_t \stackrel{\text{def}}{=} \Delta x_t.$$

Definition 2.2.2 (Integrated processes)**.** *A stochastic process $\{z_t\}$ is said to be integrated of order $d \in \mathbb{Z}$, or $I(d)$, if $z_t = (1-L)^d x_t$ is non-stationary and $z_t = (1-L)^{d-1} x_t$ is stationary.*

As introduced earlier, the random walk given in (2.2.2) is integrated of order 1 or $I(1)$ and white noise from (2.2.1) is $I(0)$. In most of the time series encountered here the first differences are all that it takes to produce a stationary series.

The combination of the processes AR and MA introduced above is known as a *autoregressive moving average* (ARMA) process.

Autoregressive (Integrated) Moving Average (AR(I)MA)
An ARMA process has AR and MA components and a ARMA(p,q) process is written as

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{2.2.6}$$

Using the operator notation introduced earlier, the ARMA(p,q) model can be written as

$$\phi(L)^p x_t = \theta(L)^q \varepsilon_t. \tag{2.2.7}$$

The *autoregressive integrated moving average model* ARIMA(p,d,q) is defined as

$$\phi(L)^p (1 - L)^d x_t = \theta(L)^q \varepsilon_t. \tag{2.2.8}$$

where $d \in \mathbb{Z}$ is the differencing operator.

2.3 Examination of autocorrelations I

As acknowledged later in the thesis, the *autocorrelations* of the previously introduced processes play an important role in describing the memory effects of the underlying process.

Definition 2.3.1 (The Autocorrelation Function)**.** *If* $\{x_t\}_{t=0}^{\mathbb{T}}$ *is a process such that* $\mathbb{V}[X_t] < \infty$ *for each* $t \in \mathbb{T}$*,*

$$\rho(s) = \frac{\mathbb{E}[(x_t - \mathbb{E}[x_t])]\mathbb{E}[(x_{t-s} - \mathbb{E}[x_{t-s}])]}{\sqrt{\mathbb{E}[(x_t - \mathbb{E}[x_t])^2]\mathbb{E}[(x_{t-s} - \mathbb{E}[x_{t-s}])^2]}} = \frac{\gamma(s)}{\sqrt{\mathbb{V}[x_t]\mathbb{V}[x_{t-s}]}} \quad t, s \in \mathbb{T} \tag{2.3.1}$$

which for a stationary process with mean $\mathbb{E}[x_t] = \mu_t$ *and variance* $\mathbb{V}[x_t] = \sigma^2$ *is reduced to*

$$\rho(s) = \frac{Cov[x_t, x_{t-s}]}{\sigma^2} = \frac{\gamma(s)}{\gamma(0)} \tag{2.3.2}$$

where $\rho(s) = \rho(-s)$*.*

White Noise

A white noise process has by definition no memory; $Cov[\varepsilon_t, \varepsilon_s] = 0$. Thus $\rho(0) = 1$ and $\rho(s) = 0$, $\forall s \geq 1$.

Random Walk

It is assumed for purposes of simplicity that there is no drift β and the initial value is set to zero. The model is then reduced to the sum of innovations

$$x_t = x_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

with $x_0 = 0$ in this case, the mean $\mathbb{E}[x_t] = 0$ and variance is given by

$$\mathbb{V}[x_t] = \gamma(0) = \mathbb{E}\left[\left(\sum_{j=1}^{t} \varepsilon_j\right)^2\right] = \mathbb{E}\left[\sum_{j=1}^{t}\sum_{k=1}^{t} \varepsilon_j \varepsilon_k\right]$$

$$= \mathbb{E}\left[\left(\sum_{j=1}^{t} \varepsilon_j^2 + \sum_{j=1}^{t}\sum_{k=1}^{t} \varepsilon_j \varepsilon_k\right)\right] = \sum_{j=1}^{t} \mathbb{E}[\varepsilon_j^2] + \sum_{j=1}^{t}\sum_{k=1, k\neq j}^{t} \mathbb{E}[\varepsilon_j \varepsilon_k] = \sum_{j=1}^{t} \sigma^2 = t\sigma^2$$

and $\mathbb{V}[x_{t-s}] = t\sigma^2 - s\sigma^2 = (t-s)\sigma^2$. The variance of the random walk model diverges as

time progresses and the autocovariance function is deduced

$$\gamma(s) = Cov[x_t, x_{t-s}] = \mathbb{E}\left[\sum_{k=1}^{t} \varepsilon_k \sum_{j=1}^{t-s} \varepsilon_j\right] = \sum_{j=1}^{t-s} \mathbb{E}[\varepsilon_j^2]$$

$$= \sum_{j=1}^{t-s} \sigma^2 = (t-s)\sigma^2 \quad \forall\, s > 0.$$

Since the covariance function depends on t and s the random walk process is not covariance stationary as it violates the assumptions given earlier. Finally, the autocorrelation function is given as

$$\rho(s) = \frac{\gamma(s)}{\sqrt{\mathbb{V}[x_t]\mathbb{V}[x_s]}} = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2(t-s)\sigma^2}} = \frac{(t-s)}{\sqrt{t(t-s)}} = \sqrt{1-\frac{s}{t}}.$$

Again it is clear that the process is non-stationary as ρ depends on both t and s. A random walk has *slowly decaying* autocorrelations.

AR(1)

For a stationary AR(1) process

$$x_t = \phi_1 x_{t-1} + \varepsilon_t$$

it can be proven that for any lag s the autocovariance is $\gamma(s) = \frac{\phi_1^s \gamma^2}{1-\phi_1^2}$. This gives the autocorrelation $\rho(s) = \gamma(s)/\gamma(0) = \phi_1^s$. For $|\phi_1| < 1$ the autocorrelation function decays exponentially. It can also be seen that for $\phi = 1$, where the process coincides with a *random walk* model, the autocorrelation function $\rho(s) = 1$.

MA(1)

For a MA(1) process $x_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t$ the autocovariances is given as

$$\begin{aligned}
\gamma(0) &= \mathbb{E}[x_t x_t] = \sigma^2(1+\theta_1^2) \\
\gamma(1) &= \mathbb{E}[x_t x_{t-1}] = \sigma^2(\theta_1^2) \\
\gamma(2) &= \mathbb{E}[x_t x_{t-2}] = 0 \\
\gamma(s) &= 0 \quad \text{for} \quad s \leq 2.
\end{aligned}$$

From this

$$\rho(s) = \begin{cases} 1 & \text{if} \quad s = 0 \\ \frac{\theta_1^2}{1+\theta_1^2} & \text{if} \quad s = 1 \\ 0 & \text{if} \quad s \leq 2. \end{cases}$$

It is clear that the autocorrelation decays with a cutoff at lag 2. In general, for MA(q) processes, a cutoff is seen at $q + 1$.

2.4 Spectral representation

Until now the paper has focused on representation of time series in the time domain, i.e. the relationship between x_t and $x_{t\pm s}$, but later when introducing estimators for long range dependence some will be annotated in the frequency (or spectral) domain. It is reasonable to introduce this representation here. Again, the works of Palma [56] or Brockwell and Davis [11] provide a good and readable introduction this this topic.

When analyzing economic data such as GDP, price levels, consumption and unemployment rates, the traditional methods introduced above have proven to be sufficient. A reason for this is that the data is measured in low frequency (yearly or quarterly) and is aggregated in time and by individuals (i.e. if the price level is high today it is likely to be high tomorrow). Such low frequency measurement makes the data smooth and uncomplicated to analyze. Some of the developments in econometrics, especially financial econometrics, are that data is observed at a much higher frequency and is more disaggregated than macroeconomic data [23]. The spectral analysis, which is introduced in the following, has a better ability of dealing with such processes that are measured at a high frequency.

Operating in the spectral domain makes the use of complex numbers less painful as they keep track of two real quantities by containing a real part and a imaginary part $z = \alpha + \beta i$ where $i = \sqrt{-1}$. Although it is dealt with complex numbers the solutions will always lie in the real space.

Let $\omega \in [-\pi, \pi]$ denote the *frequency* and T the *period*. The period of a cycle is the minimum time it takes for the wave to go through a complete sequence of values, thus $T = \frac{2\pi}{\omega}$.

Definition 2.4.1 (The Fourier Transform)**.** *Given a time series* $\{x_t\}$ *its* Fourier transform *is defined as*

$$d_x(\omega) = \sum_{t=-\infty}^{\infty} e^{-i\omega t} x_t. \tag{2.4.1}$$

This operation transforms the time series $\{x_t\}$, which is a function of t, into a complex valued function of ω.

Definition 2.4.2 (The Inverse Fourier Transform)**.** *Given* $d_x(\omega)$*,* x_t *can be recovered by the*

so called inverse Fourier transform

$$x_t = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} d_x(\omega) d\omega. \tag{2.4.2}$$

Proposition 1. *A transformation on a time series from time domain to frequency domain can be made without loss of information by the utilizing the* Fourier transform *and vice versa by using the* inverse Fourier transform*.*

Proof. By substituting the definition of $x(\omega)$ in the inverse Fourier transform one gets the term

$$\begin{aligned} x_t =& \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} d_x(\omega) d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\left(\sum_{s=-\infty}^{\infty} e^{-i\omega s} x_s\right) d\omega \\ =& \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} x_s \int_{-\infty}^{\infty} e^{i\omega t} e^{-i\omega s} \end{aligned}$$

then, after arranging and factoring

$$= \sum_{s=-\infty}^{\infty} x_s \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega(t-s)} d\omega$$

where the integral part can be evaluated at $t - s = 1$,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega} d\omega = 0,$$

and at $t - s = 0$,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega = 1.$$

These results follows from the fact that the integral of sin or cos from $[-\pi, \pi]$ is zero, which also is the case for any $t \neq s$. The findings can written as a function of Φ

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega(t-s)} d\omega = \Phi(t-s) = \begin{cases} 1 & \text{if } t - s = 0 \\ 0 & \text{if } t - s \neq 0 \end{cases}$$

and finally

$$\sum_{s=-\infty}^{\infty} x_s \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t-s)} d\omega = \sum_{s=-\infty}^{\infty} x_s \Phi(t-s) = x_t.$$

□

An extension of the Fourier transform is the *discrete Fourier transform* which is the foundation of many computationally efficient algorithms, including the ones introduced in Chapter 5.

Definition 2.4.3 ((Inverse) Discrete Fourier transform)**.** *Given a periodic sequence* $\{z_k\}$ *of period* N *where* $k = 0, \ldots, N-1$ *the* discrete Fourier transform *of* z_k *is given as*

$$F_n = \sum_{n=0}^{N-1} z_k \exp\left\{-2\pi i \frac{nk}{N}\right\}. \tag{2.4.3}$$

The reverse transformation is the inverse discrete Fourier transform*. This is defined as*

$$z_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n \exp\left\{2\pi i \frac{nk}{N}\right\}. \tag{2.4.4}$$

It should be mentioned that there are different notational conventions among authors regarding the factor in front of the discrete Fourier transforms. The use of either one does not influence the properties of the transform, however when operating with matrices it serves as a normalization factor.

2.4.1 Spectral densities

Using the (discrete) Fourier transform from Definition 2.4.1, the *spectral density* can be defined.

Definition 2.4.4 (Spectral density)**.** *Given* $\{x_t\}$*, a zero-mean stationary time series with autocovariance function* $\gamma(\cdot)$ *fulfilling* $\sum_{s=-\infty}^{\infty} |\gamma(s)| < \infty$*. The Fourier transform of the autocovariance function* $\gamma(s)$ *gives the relationship*

$$S(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-i\omega s} \gamma(s), \quad -\infty < \omega < \infty \tag{2.4.5}$$

where S *is defined as the* spectral density *of* $\{x_t\}$*.*

One must note that the spectral density does not necessarily exists [12]. A sufficient but strong condition for its existence is

$$\sum_{s=-\infty}^{\infty} |\gamma(s)| < \infty. \tag{2.4.6}$$

Absolutely summable autocovariances guarantees the transition from time domain to frequency domain by the fourier transform.

Remark 4. *As in most of the literature regarding time series, there is also in this area an inconsistency in notations. The usage of the factors* $\frac{1}{2\pi}$ *and* $\frac{1}{\sqrt{2\pi}}$ *can be seen amongst other authors. Some include them into the df of the inverse fourier transform as* $df = d\omega/2\pi$*. It will in the entirety of this thesis be chosen to explicitly state this term as is done in the works of Brockwell and Davis [11, 12].*

The spectral density $S(\omega)$ describes which frequencies are important in the series x_t. For a time series with quarterly data, a peak or spike in the spectral density at $\omega = \frac{1}{2}\pi$ would indicate seasonal effects since the period of one year is 2π. It should also be noted that a low frequency corresponding with to a long wavelength so that ω has more impact for a larger $S(\omega)$. In the context of examining long range dependent processes it will later be evident that their spectral densities have a spike at low frequencies thus rendering the lowest frequencies as importance in explaining the variance.

Letting the *autocovariance generating function* be

$$G(z) = \sum_{s=-\infty}^{\infty} \gamma(s) e^{i\omega s} \tag{2.4.7}$$

then the spectral density is simply (2.4.7) divided by the period 2π. Using *DeMoivre's theorem*, $e^{\pm i\omega s} = \cos(\omega s) \pm i \sin(\omega s)$, $s, t \in \mathbb{R}$ Equation (2.4.4) can be rewritten as

$$S(\omega) = \frac{1}{2\pi} \left[\gamma(0) + 2 \sum_{s=1}^{\infty} \gamma(s) \cos(s\omega) \right]. \tag{2.4.8}$$

See [11] for proof. By Equation (2.4.8) the connection between the time domain and the spectral domain is clear, and since $\cos(\omega) = \cos(-\omega)$ and $\gamma(s) = \gamma(-s)$ the spectrum is symmetric around zero. As cos has a period of 2π, in further analysis of spectra it is only necessary to consider the interval $\omega \in [0, \pi]$. By knowing $S(\omega)$, by (2.4.2) the autocovariance function can be computed

$$\gamma(s) = \int_{-\pi}^{\pi} e^{i\omega s} S(\omega) d\omega.$$

Setting $s = 0$ the variance of the process $\{x_t\}$ is found as the sum of the spectral density over the period 2π

$$\gamma(0) = \int_{-\pi}^{\pi} S(\omega) d\omega.$$

In other words, the spectral density is a decomposition of $\mathbb{V}[x_t]$ in terms of frequency. Due to this representation, it is possible to detect an existence of cyclicality as it will show as peaks or spikes in the spectral density.

Finally, by dividing the spectral density $S(\omega)$ by $\gamma(0)$ then the Fourier transform of the autocorrelation function is found and it will be defined as the *spectral distribution function*.

Definition 2.4.5 (The Spectral distribution function)**.** *Given $\{x_t\}$, a zero-mean stationary time series with a well defined autocovariance function, then the Fourier transform of the autocorrelation function $\rho(s)$ gives the relationship*

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-i\omega s} \rho(s) \tag{2.4.9}$$

where f defined as the spectral density of $\{x_t\}$.

By using the inverse Fourier transform the autocorrelation function can be recovered. As noticed in Equation (2.4.9), the spectral distribution function behaves like a probability distribution - it is positive and integrates to one

$$\rho(0) = 1 = \int_{-\pi}^{\pi} f(\omega) d\omega.$$

2.4.2 Spectral density of AR and MA processes

As previously mentioned, the spectral representation of a process is an assessment of the variance contribution for a given frequency. It will briefly be shown how the spectral densities of white noise, AR and MA processes are constructed. Note that the following results are widely known in most of the literature regarding time series analysis [12] and therefore, only the results will be presented since this chapter is purely for enlightenment. Understanding how simple processes are represented in the spectral domain plays a role in understanding methods for more advanced processes later described. The following computations can be found in [12].

White noise

It has been shown that for white noise ε_t $\gamma(0) = \sigma_\varepsilon^2$ and $\gamma(s) = 0$. Along these lines,

$$S(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma(s) e^{i\omega s} = \gamma(0) = \sigma_\varepsilon^2. \tag{2.4.10}$$

In the case of white noise all frequencies contribute the same to the variance.

Autoregressive processes

Determining the spectral density of an AR(p) process can be done by applying a linear filter

$$S(\omega) = \frac{\sigma^2}{2\pi \left| 1 - \sum_{k=1}^{p} \phi_k e^{-i\omega s} \right|^2}. \tag{2.4.11}$$

For AR(1),

$$S(\omega) = \frac{\sigma^2}{2\pi\{1 + \phi_1^2 - 2\phi_1 \cos(\omega)\}} \tag{2.4.12}$$

and AR(2)

$$S(\omega) = \frac{\sigma^2}{2\pi\{1 + \phi_1^2 + \phi_2^2 - 2\phi_1[1 - \phi_2]\cos(\omega) - 2\phi_2 \cos(\omega)\}}. \tag{2.4.13}$$

For AR processes the value of ϕ determines whether the spectral density is defined by high frequencies (negative ϕ) or low frequencies (positive ϕ) as seen in Figure 2.1a.

Moving average processes

As with AR, the MA processes are converted through a linear filter

$$S(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{q} \theta_j e^{-i\omega j} \right|^2 \tag{2.4.14}$$

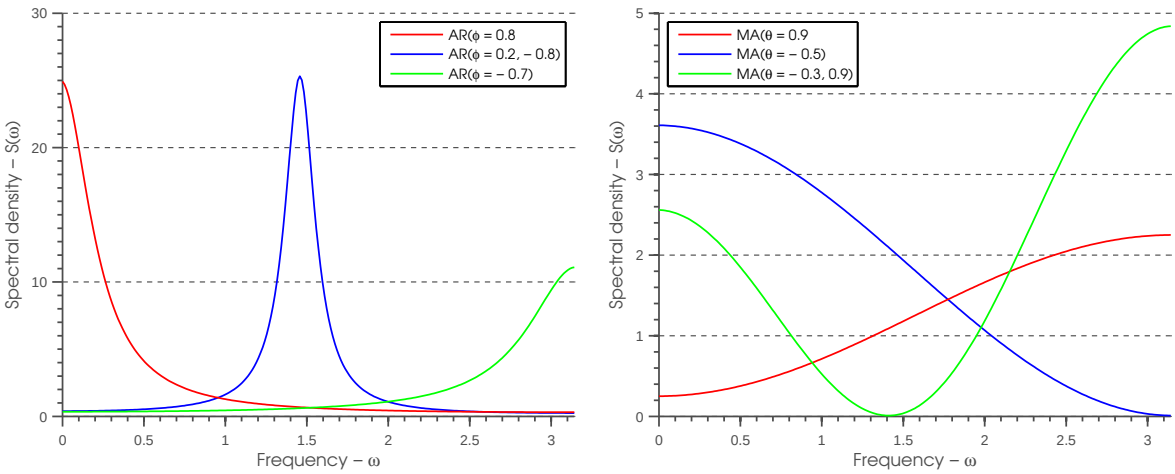
For MA(1),

$$S(\omega) = \frac{\sigma^2}{2\pi}\{1 + \theta_1^2 + 2\theta_1 \cos(\omega)\} \tag{2.4.15}$$

and MA(2)

$$S(\omega) = \frac{\sigma^2}{2\pi}\{1 + \theta_1^2 + \theta_2^2 + 2[\theta_1 + \theta_1\theta_2]\cos(\omega) + 2\theta_2 \cos(2\omega)\}. \tag{2.4.16}$$

As for AR processes, the MA spectral density is shaped by the values of θ. For positive values, the majority of the spectral density will lie in the low frequency area and vice versa.



(a) Spectral densities for a set of AR processes.

(b) Spectral densities for a set of MA processes.

Figure 2.1: Spectral densities for AR and MA processes.

Chapter 3

Long range dependence

This chapter will cover the background for long range dependence (LRD) and introduce some new processes which inhibit LRD properties. A connection will be made between series with short range correlations and long range correlations by investigating the autocorrelations and the spectral densities.

Long range dependence as a feature of time series is believed to have its beginning in the late 1960's where the phenomenon was studied by Mandelbrot and his colleagues [41, 46]. This engagement was triggered by a series of observations that were made by Hurst [28, 27] in the 1950's and had not yet been properly explained. This was not the first encounter of such an anomalous feature of a time series. In the 1940's Kolmogorov [33, 34] did research on turbulence and observed similar attributes in the data studied.

3.1 Motivation for studying long range dependence

The reasoning for studying processes with long memory is that the dependence structure plays an important role in the modeling of both economic and financial data. A definition commonly used by authors is that long range dependent processes are defined as stochastic processes with an autocorrelation function that is decaying slowly as a power law summed to infinity. This slow decay is antithetical to the quick and exponential that was found in the processes of Chapter 2.

One of the appealing features of examining memory effects in time series is that it can be described by only one parameter - the Hurst exponent. Processes with a Hurst exponent equal to 0.5 are either an independent process or a short-term dependent process [9]. Throughout the literature on the subject, there are differences in naming conventions. For $H < 0.5$ the process exhibits negative correlations and is observed to be long-range dependent with negative correlations [21]. In the seminal work of Beran [9] the case of $H < 0.5$ is termed

short-range dependent, however Mandelbrot and Van ness [46] call it an anti-persistent process. Further on, Peters [58] uses the term pink noise which stems from signal processing in physics. In the case of $H > 0.5$ the process is shows positive correlations and is said to be long-range dependent [9]. Palma [56] uses the term long-memory whereas Mandelbrot and van Ness [46] refers to it as a persistent process. Finally, Peters refers to this type of behavior as "black noise" and long range correlations are used by Peng [57].

Before further introducing the notion of long memory a set of relationships should be presented: where H is the *self-similarity parameter*, also named the *Hurst exponent* or the *scaling exponent*, d is the *differencing parameter* and D is the closely related *fractal dimension*. Note that the fractal dimension D is independent of the Hurst exponent as it is measuring the roughness of the surface [44]. D is a local property opposed to d and H which is are global properties. However, for Gaussian and closely related processes, the relationship $D = 2 - H$ is assumed to be valid [48].

type of memory	H	$d = H - 1/2$	$D = 2 - H$
short memory	$= 1/2$	$= 0$	$3/2$
anti-persistence	$< 1/2$	< 0	$< 3/2$
persistence	$> 1/2$	> 0	$> 3/2$

Table 3.1: Relationship between estimators.

3.1.1 Definitions of long range dependence

A number of different definitions for long memory exist. For instance, Guegan [24] surveyed the literature concerning the definition of the concept of long range dependence, which resulted in 11 separate definitions. It will in this case be presented some of the definitions seen in the works of Robinson [62] and Beran [9] will be presented. First definitions of long range dependence in the time domain and the spectral domain are given as follows:

Definition 3.1.1 (Long range dependence [62]). *A stochastic process $\{x_t\}_{t=-\infty}^{\infty}$ with a auto-covariance function $\gamma(0)$ possesses long range dependence if and only if*

$$\sum_{s=-\infty}^{\infty} \gamma(0) = \begin{cases} 0 & \textit{anti-persistent long range dependence} \\ \infty & \textit{persistent long range dependence} \end{cases} \tag{3.1.1}$$

Definition 3.1.2 (Long range dependence [62]). *A stochastic process $\{x_t\}_{t=-\infty}^{\infty}$ with a spectral*

density $S(\omega)$ possesses long range dependence if and only if

$$\lim_{S\to 0} S(\omega) = \begin{cases} 0 & \textit{anti-persistent long range dependence} \\ \infty & \textit{persistent long range dependence} \end{cases} \tag{3.1.2}$$

On the contrary, by Definition 3.1.2, it can be said that a stochastic process $\{x_t\}_{t=-\infty}^{\infty}$ has *short memory* if

$$0 < S(0) < \infty.$$

Two more definitions will be given which differ from the first two introduced. These definitions describe the ultimate behavior of the autocorrelations as they move towards infinity.

Definition 3.1.3. *The stationary process $\{x_t\}_{t=-\infty}^{\infty}$ possesses long range dependence if the following holds*

$$\lim_{j\to\infty} \frac{\rho(j)}{c_\rho j^{2H-2}} = 1 \tag{3.1.3}$$

where $H \in (0,1)$ and $c_\rho > 0$ is a constant.

An equivalent condition exists in the frequency domain since if the autocorrelations of a process is known, then the spectral density is also known. This was shown in Chapter 2.

Definition 3.1.4. *The stationary process $\{x_t\}_{t=-\infty}^{\infty}$ possesses long range dependence if the following holds*

$$\lim_{\omega\to 0} \frac{S(\omega)}{c_S j^{1-2H}} = 1 \tag{3.1.4}$$

where $H \in (0,1)$ and $c_S > 0$ is a constant.

These two definitions determine only the rate of convergence, not the absolute size, and do not specify the correlations for any fixed lag. They essentially say that if the correlations demonstrate slow decay then long memory is apparent. This slow decay can be somewhat difficult to detect in practice [62]. Combining the Equations (3.1.3) and (3.1.4), the following relationship can be established.

Proposition 2. *Let $S(\omega)$ be the spectral density function of a stationary process and $\rho(s)$ the autocorrelation function. Then*

$$\lim_{j\to\infty} \rho(j) \sim c_\rho j^{2H-2} \Leftrightarrow \lim_{\omega\to 0} S(\omega) \sim c_S \omega^{1-2H} \tag{3.1.5}$$

for $H \in (0,1)$ and c_s and c_ρ is positive constants.

If the autocorrelation function behaves like a power-law at large lags, then by Equation (3.1.5), the spectral density is expected to have the same behavior for very small frequencies. It can be inferred that the spectral density diverges at small frequencies.

3.2 Stationary processes with long range dependence

Here, a few processes will be introduced that will later be used in the robustness analysis of the estimators of long range dependence. As opposed to Chapter 5, where models with short memory are introduced, these processes will have features that make them long range dependent.

3.2.1 Self-similar processes

A self-similar process is said to resemble itself across a wide range of time scales. Kolmogorov wrote a paper in 1941 [33] that for the first time introduced *self-similar* processes in a theoretical setting. It was not until 1968 was this concept brought into statistics by Mandelbrot and van Ness [46]. First, let $\{x_t\}_{t=0}^{T}$ be a weakly stationary time series with autocorrelation function $\rho(s)$. Further on, a new time series is denoted as

$$x_t^{(m)} = \frac{1}{m} \sum_{\tau=(t-1)m+1}^{tm} x_\tau, \quad t = 1, 2, \ldots, [T/m] \tag{3.2.1}$$

where $x_t^{(m)}$ is the sequence obtained by averaging the original series $\{x_t\}$ by non-overlapping blocks of size $m > 1$ which describe the level of aggregation. Exact self-similarity can then be defined.

Definition 3.2.1 (Exact Self-Similarity)**.** *A strictly stationary process $\{x_t\}$ is self-similar if*

$$x_t =_d m^{1-H} x_t^{(m)} \quad \forall\, t, m \tag{3.2.2}$$

where $=_d$ means equality in distribution.

Given the self-similar process $\{y_t\}$ with the property $y_t =_d t^H y_1$ and $y_0 = 0$. The increment process is defined as $x_t = y_t - y_{t-1}$ with variance $\sigma^2 = \mathbb{E}[(y_t - y_{t-1})^2] = \mathbb{E}[y_1^2]$ and $s < t$. Then the covariance of the increment process can be computed in the manner where s is subtracted

$$\mathbb{E}[(y_t - y_s)^2] = \mathbb{E}[(y_{t-s} - y_0)^2] = \sigma^2 (t-s)^{2H}$$

and when not subtracting s, simply expanding the polynomial gives

$$\mathbb{E}[(y_t - y_s)^2] = \mathbb{E}[y_t^2] + \mathbb{E}[y_s^2] - 2\mathbb{E}[y_t y_s] = \sigma^2 t^{2H} + \sigma^2 s^{2H} - 2\gamma(s).$$

Solved for the autocovariance function $\gamma(s)$ it results in

$$\gamma(s) = \frac{1}{2}\sigma^2\{t^{2H} - (t-s)^{2H} + s^2 H\}. \tag{3.2.3}$$

Based on the fact that y_t is self-similar, the covariances $Cov[x_1, x_{s+1}]$ of the increment process $\{x_t\}$ is obtained (see [9])

$$\gamma(k) = \frac{1}{2}\sigma^2\{(s+1)^{2H} - 2s^{2H} + (s-1)^{2H}\} \tag{3.2.4}$$

for $s \geq 0$. The correlations is given as $\rho(s) = \frac{\gamma(s)}{\sigma^2}$. The equality found in Equation (3.2.2) is rather stringent and will only be used to introduce the concept of self-similarity. An assumption regarding self-similarity of a stochastic processes in terms of the autocorrelations will be made.

Definition 3.2.2 (Exact Second-Order Self-Similarity)**.** *A stationary time series* $\{x_t\}$ *is exactly second-order self-similar with self-similarity parameter* H *if, for all* m

$$\rho_s^{(m)} = \rho(s) = \frac{1}{2}\left\{(s+1)^{2H} - 2s^{2H} + |s-1|^{2H}\right\}, \quad s \geq 0, \tag{3.2.5}$$

where $\rho_s^{(m)}$ *is the autocorrelation function of* $x_t^{(m)}$*.*

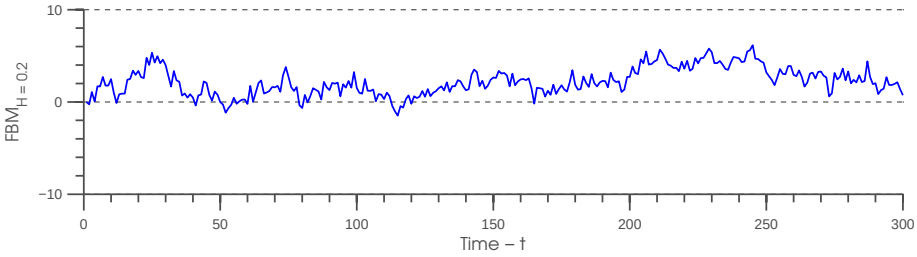
Definition 3.2.3 (Asymptotic Second-Order Self-Similarity)**.** *A stationary time series* $\{x_t\}$ *is exactly second-order self-similar with self-similarity parameter* H *if*

$$\lim_{m\to\infty} \rho_s^{(m)} \sim \rho(s). \tag{3.2.6}$$

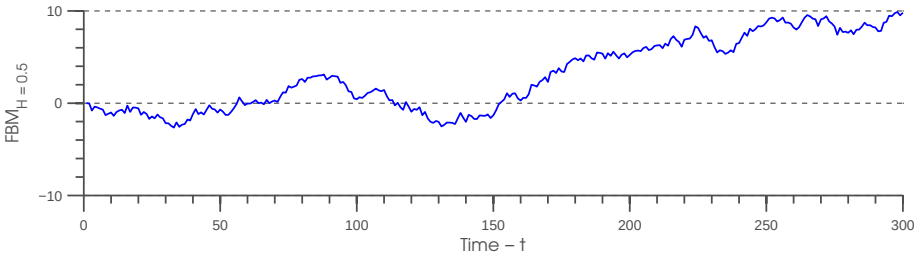
Thus, the process $\{x_t\}$ is second-order self-similar if the aggregated process $\{x_t^{(m)}\}$ is equal or becomes indistinguishable in terms of their corresponding autocorrelation functions. Finally, a process $\{x_t\}_{t\in\mathbb{R}}$ has *stationary increments* if

$$x(t+s) - x(s) =_d x(t) - x(0). \tag{3.2.7}$$

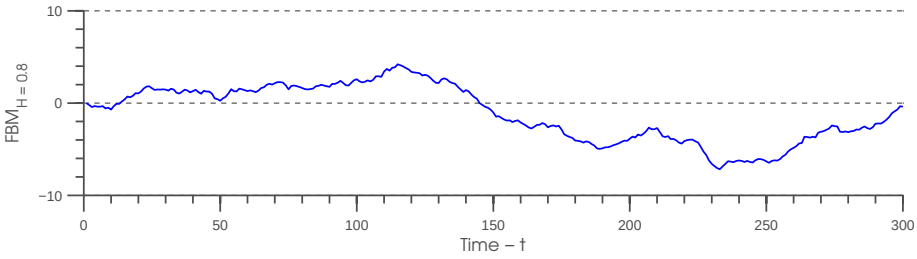
Fractional Gaussian noise(FGN) and the related *fractional Brownian motion* (FBM), which is an integrated version of FGN, will be covered. In addition the *fractional ARIMA* is used to model processes that are containing a long memory FI component in addition to the MA and AR part.



(a) Sample path of a $FBM(H = 0.2)$.



(b) Sample path of a $FBM(H = 0.5)$.



(c) Sample path of a $FBM(H = 0.8)$.

Figure 3.1: Fractional Brownian motion sample paths.

3.2.2 Fractional Brownian motion

Brownian motion [9]

Let $B(t)$ be a stochastic process with continuous sample paths and such that

i) $B(0) = 0$ almost surely,

ii) $B(t)$ has independent increments

iii) $\mathbb{E}[B(t) - B(s)] = 0$,

iv) $\mathbb{V}[B(t) - B(s)] = \sigma^2|t - s|$.

Then $B(t)$ is called a *Brownian motion.*

Brownian motion (BM) can be given a fractal component and it becomes the *fractional Brownian motion* denoted by $B_H(t)$ or $FBM(H)$. It was introduced by Kolmogorov [33]

but this representation and its name was given by Mandelbrot and van Ness [46]. As this thesis aims to investigate processes with long range dependence, it is natural to discuss the properties of fractional Brownian motions.

Fractional Brownian motion [46]
A process is called a fractional Brownian motion $B_H(t)$ if it follows the conditions

i) $B_H(0) = 0$ and $\mathbb{E}[B_H(0)] = 0$,

ii) $\mathbb{V}[B_H(t)] = \sigma^2|t|^{2H}$,

iii) $\mathbb{V}[B_H(t) - B_H(s)] = \sigma^2|t-s|^{2H}$ and

iv) $Cov[B_H(t), B_H(s)] = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$.

See [56] for proofs.

Excluding the trivial cases of $H = 0$ and $H = 1$ there will in the following only be focused on the interval $0 < H < 1$ for the case of fractional Brownian motions. Note that for $H = 0.5$ the FBM equals the Brownian motion with i.i.d. increments. The FBM is defined as *standard fractional Brownian motion* if $\mathbb{V}[B_H(1)] = 1$, $\mathbb{E}[(B_H(t))] = 0$ and

$$Cov[B_H(t), B_H(s)] = \frac{1}{2}\{t^{2H} + s^{2H} - (t-s)^{2H}\}$$

for $H \in (0, 1)$. Some observations can be made from these assumptions. It is possible to say something about the distribution of the standard fBm from the fact that the variance is given as

$$\mathbb{V}[B_H(t)] = Cov[B_H(t), B_H(t)] = t^{2H}.$$

The probability distribution is given as

$$B_H(t) \sim \mathcal{N}(0, t^{2H})$$

and for any positive constant,

$$B_H(ct) \sim \mathcal{N}(0, c^{2H}t^{2H}).$$

By Equation (3.2.2)

$$B_H(ct) \sim c^{2H} B_H(t)$$

the fractional Brownian motion is known as a self-similar process with index H and exhibits long range dependence in the sense that it is persistent for $\frac{1}{2} < H < 1$ processes and

anti-persistent for $0 < H < \frac{1}{2}$. Distinctions can be made regarding the different fractional Brownian motions by examining their respective autocovariance functions γ. In the anti-persistent case of $0 < H < 1/2$, the covariance of two adjacent increments is negative and the consecutive increments in the FBM sample path will move in opposite directions. This leads to a spiky sample path. For persistent sequences with $1/2 < H < 1$ the case is the opposite. The covariance between the adjacent increments are positive and the sample path will have a smooth appearance. Finally, for $H = 1/2$, the FBM reduces to a BM. In Figure 3.1 this behavior is shown.

3.2.3 Fractional Gaussian noise

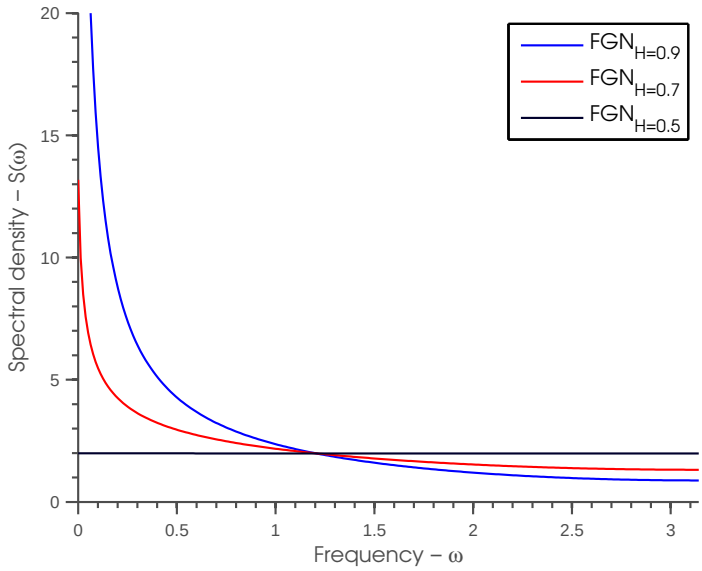


Figure 3.2: Spectral densities for a set of FGN processes.

The process referred to as *fractional Gaussian noise* is the increment process of the fractional Brownian motion introduced above.

Fractional Gaussian noise

Given the fractional Brownian motion $B_H(t)$, the increment process given as

$$z_t = B_H(t+1) - B_H(t) \tag{3.2.8}$$

where $H \in (0, 1)$ and $t \in \mathbb{T}$. The discrete time process $\{z_t\}$ is defined as fractional Gaussian noise with corresponding properties

i) $\mathbb{E}[z_t] = 0$,

ii) $\{z_t\}$ is stationary for $H \in (0,1)$,

iii) $\mathbb{V}[z_t] = \mathbb{E}[B_H(1)]^2$,

As observed in Figure 3.2, the FGN has a pole at frequency $\omega = 0$ for high intensities of long range dependence, which is in coherence with Definition 3.1.2.

3.2.4 ARFIMA

The following class of models is an extension of the ARIMA models introduced earlier. Recall that the ARIMA(p,d,q) from Equation (3.2.9) were restricted in terms of d as it could only take integer values. The *autoregressive fractional integrated moving average* process, or ARFIMA, differs from ARIMA by the fact that $d \in \mathbb{R}$.

ARFIMA

Using the lag operator introduced earlier, the ARFIMA(p,d,q) can be presented as

$$\phi(L)^p(1-L)^d x_t = \theta(L)^q \varepsilon_t \quad \text{for} \quad d \in (-0.5, 0.5)\,. \tag{3.2.9}$$

It is also possible to operate with a d outside this space, however the model will lose its invertibility and stationarity [9].

The ARFIMA model consists of a class of *fractionally integrated* processes analogous to the integrated processes defined earlier in Definition 2.2.2. The random walk encountered above was an integrated process, however in terms of fractional integration a different structure is prevailing:

$$(1-L)^d x_t = \varepsilon_t \tag{3.2.10}$$

where $d \in \mathbb{R}$ as opposed to $d \in \mathbb{N}$ is seen in Definition 2.2.2. The rationale behind the fractionally integrated processes is that an integrated process such as (3.2.10) can be integrated a fractional number of times and still maintain stationarity. Thus, a nonstationary process with $d = 1.4$ can be differenced once to obtain $d = 0.4$, which would make it stationary. Taking a closer look at the difference operator $(1-L)^d$ provides an insight on how the fractionally integrated processes work. The Taylor expansion for $d \in \mathbb{R}$ is presented as

$$\begin{aligned}(1-L)^d =& 1 - dL + \frac{1}{2!}d(d-1)L^2 - \frac{1}{3!}d(d-1)(2-d)L^3 + \dots \\ &+ \frac{(-1)^k}{k!}d(d-1)\dots(d-k+1)L^k \\ =& \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}L^k\end{aligned}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt\,, x \in \mathbb{R}^+$ is the Gamma function. The simple representation above can be represented as ARFIMA(0,d,0) and in terms of moving average it is given as

$$x_t = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}\varepsilon_{t-k}. \tag{3.2.11}$$

The ARFIMA(0,d,0), ARFIMA(d) noise or FI(d), exhibits persistence for $0 < d < 0.5$ and anti persistence for $-0.5 < d < 0$. For ARFIMA$(p, 0, q)$, or ARMA(p,q), the process has short memory as only the ARMA part remains. It is possible for d to take on values outside the interval given here but the process will no longer be stable. In fact, the ARFIMA process is stationary for $d < \frac{1}{2}$ and invertible for $d > -\frac{1}{2}$.

3.3 Examination of autocorrelations II

The earlier examination of autocorrelations in Section 2.3 displayed that they were exponentially decaying for the processes given there, which is the definition of short memory. In the previous section, processes that allegedly possess long range dependence were presented. Previously it has also defined long range dependence as slowly decaying autocorrelations (in terms of time domain) such as in Definition 3.1.1. In the following, the respective autocorrelation functions will briefly be examined.

Fractional Brownian motion

The autocorrelations is given as

$$\rho(s) = \frac{1}{2}\{(s+1)^{2H} - 2s^{2H} + (s-1)^{2H}\} \tag{3.3.1}$$

where $s \geq 0$. This is the definition of the autocorrelation of an exact second order self-similar process defined in Equation (3.2.5). In order to assess the dependence of the fractional Brownian motion it is necessary to investigate the asymptotic behavior of $\rho(s)$ which can be obtained by Taylor series expansion. By factoring out s^{2H} from Equation (3.3.1) it can be written as

$$\rho(s) = \frac{s^{2H}}{2}\left\{\left(1-\frac{1}{s}\right)^{2H} - 2 + \left(1+\frac{1}{s}\right)^{2H}\right\}.$$

The expression can be reparameterized by denoting $\{\cdot\}$ as $h(x)$ and get

$$\rho(s) = \frac{s^{2H}}{2}h\left(\frac{1}{s}\right)$$

where $h(x) = \{(1-x)^{2H} - 2 + (1+x)^{2H}\}$. In order to acquire the Taylor series the derivatives of first and second order are found, respectively

$$\frac{\partial h}{\partial x} = 2H\{(1-x)^{2H-1} - (1+x)^{2H-1}\}$$

and

$$\frac{\partial^2 h}{\partial x^2} = 2H(2H-1)\{(1-x)^{2H-2} + (1+x)^{2H-2}\}$$

The Taylor series expansion is written as

$$h(x - x_0) = h(x) = h(0) + x\frac{\partial h}{\partial 0} + \frac{x^2}{2}\frac{\partial^2 h}{\partial 0} + \dots$$

When substituting for $\frac{\partial h}{\partial x}$ and for $\frac{\partial^2 h}{\partial x^2}$ the first non zero term of $h(x)$ is found

$$h(x) = \frac{x^2}{2} 2H(2H-1)$$

and $\rho(s)$ becomes

$$\frac{s^{2H}}{j} h\left(\frac{1}{s}\right) = s^{2H}, H(2H-1)\frac{1}{s^2}$$

and when s goes to infinity

$$\lim_{s\to\infty} \rho(s) \sim H(2H-1)s^{2H-1}. \tag{3.3.2}$$

It is now appropriate to say something about the correlations in the limit. For $H \in (1/2, 1)$ the correlations decay so slowly that

$$\sum_{s=-\infty}^{\infty} \rho(s) = \infty. \tag{3.3.3}$$

For $H = 1/2$ the correlations asymptotically go to zero; the observations of the increment process is uncorrelated. For $H \in (0, 1/2)$ the correlations are summable

$$\sum_{s=-\infty}^{\infty} \rho(s) < \infty. \tag{3.3.4}$$

As follows, for $H \in (0, 1/2)$ the process has short range dependence, for $H = 1/2$ the observations are uncorrelated and for $H \in (1/2, 1)$ the process has long range dependence.

Fractional Gaussian noise

Recall that the fractional Gaussian noise process from (3.2.8) is the increment process of the

fractional Browninan motion and exhibits the same properties as a differenced FBM.

ARFIMA

Knowing that an ARFIMA(p,d,q) model can represent both short-run correlations in its AR and MA parts and long-run correlations in the FI part, the interesting case in this context is the ARFIMA(0,d,0) process defined as

$$(1-L)^d x_t = \varepsilon_t.$$

This process has an infinite autoregressive and moving average representation by Equation (3.2.11) as

$$x_t = \sum_{s=0}^{\infty} \frac{\Gamma(s+d)}{\Gamma(d)\Gamma(s+1)} \varepsilon_{t-s}. \tag{3.3.5}$$

As the white noise ε_t is independent when $t \neq s$ the autocovariance function $\gamma(s)$ must be given as [9]

$$\gamma(s) = \frac{\Gamma(1-2d)\Gamma(s+d)}{\Gamma(d)\Gamma(1-d)\Gamma(s+1-d)}. \tag{3.3.6}$$

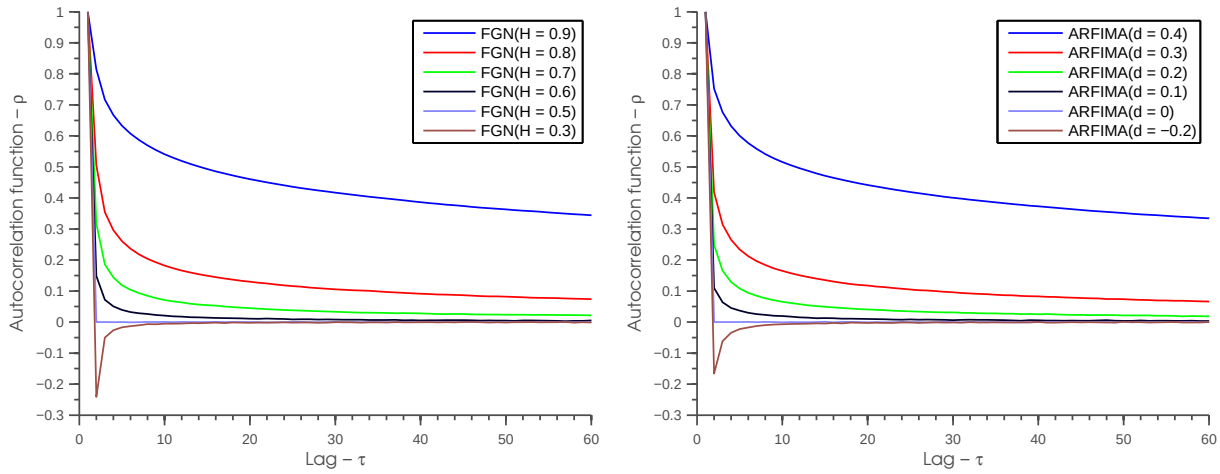
and corresponding autocorrelation function

$$\frac{\gamma(s)}{\gamma(0)} = \rho(s) = \frac{\Gamma(1-d)\Gamma(s+d)}{\Gamma(d)\Gamma(s+1-d)}. \tag{3.3.7}$$

By the use of *Stirling's formula* $\frac{\Gamma(s+c_1)}{\Gamma(s+c_2)} \approx s^{c_1-c_2}$ (3.3.7) can be approximated as

$$\lim_{s\to\infty} \rho(s) = cs^{2d-1}$$

where $c = \frac{\Gamma(1-d)}{\Gamma(d)}$ and $d \in (0, 0.5)$. When d lies in the interval $0 < d < 0.5$ the autocorrelations will decay hyperbolically at the rate of s^{2d-1} and the underlying process is long range dependent.



(a) Autocorrelations for FGN(H). (b) Autocorrelations for ARFIMA(0,d,0).

Figure 3.3: Autocorrelation functions of FGN(H) and ARFIMA(0,d,0) for different levels of H and d.

Chapter 4

Estimators of long range dependence

Thus far it has been shown that fractional Gaussian noise, fractional Brownian motion and ARFIMA models can be used to model long range dependence. In the ARFIMA(p,d,q) process, the measurement of persistence is given as the differencing operator $d \in (-0.5, 0.5)$ where long range dependence occurs in the interval $0 < d < 0.5$. For the fractional Gaussian noise, the level of dependence is measured by the Hurst exponent $H \in (0, 1)$ and persistence is found when $0.5 < H < 1$. In Table 3.1 the relationship between these parameters is given. In this chapter the problem of estimating the intensity of long range dependence in time series will be addressed. It is worth noting that there is a common estimation procedure among the estimators, including some of the LRD estimators that are not used in this context. The common regime is as follows:

i) A numerical property, say F, of the time series is computed as a function of the order δ, i.e. $F(\delta)$,

ii) a power law describing the asymptotic behavior, such as $F(\delta) \sim \delta^a$, as the value δ goes to zero or infinity is derived,

iii) where the exponent a is a function of the fractal exponent H,

iv) and finally, H is estimated by regression where the lowest values of $F(\delta) \sim \delta^a$ are emphasized.

4.1 Estimators

As was mentioned in the introduction to this chapter, the estimators of LRD introduced here follow a common procedure which is categorized with respect to each estimator in Table 4.1.

Estimator	Property	Series	Power law	Scheme
R/S	$(\mathcal{R}/\mathcal{S})_\tau$: Rescaled range	τ - lag	τ^H	$\tau \to \infty$
Periodogram	$I(\omega_j)$: Periodogram	ω - frequency	ω^{-2d}	$\omega \to 0$
Higuchi	$L(m)$: Length of curve	m - blocks	m^{-D}	$m \to \infty$
DFA	$\mathcal{F}(l)$: Fluctuation	l -block size	l^H	$l \to \infty$
Gen. Hurst	$K_q(\tau)$: q-eth moment	τ - lag	$\tau^{qH(q)}$	$\tau \to \infty$
Agg. Variance	$x_t^{(m)}$: Agg. process	m - block size	m^{2H-2}	$m \to \infty$
E.L. Whittle	$I(\omega_k)$: Periodogram	ω - frequency	ω^{-2d}	$\omega \to 0$

Table 4.1: Categorization of the main features of the LRD estimators.

In the following the procedure for each estimator and its known finite sample properties is introduced.

4.1.1 Rescaled range

First of all the widely used and pioneering method known as the *rescaled range*, or simply $\mathcal{R}/\mathcal{S}$ analysis, will be introduced. The method was first developed by Harold E. Hurst in 1951 [28] in relation to a hydrological context. He was managing a reservoir and his assignment was to determine an ideal design such that the dam would never dry out or flood. To his aid he had records of previously observed river discharges and when designing the model, he noticed that the amount of rainfall would follow a random walk. Mandelbrot and Wallis [41] introduced two terms that would describe the weather conditions. One of which being observations that were distributed according to heavy tails, namely the *Noah effect*. Long memory observations were associated with the term *Joseph effect*. The latter refers to the biblical story of Joseph who claimed that seven years of prosperity (heavy rainfall) would be followed by seven bad years.

Given a time series, $\{x_t\}_{t=1}^{\mathbb{T}}$ let $\{z_t\}_{t=1}^{\tau}$ be the logarithmic returns given by $z_t = \ln[x_{t+1}/x_t]$ and the sample mean $\bar{z} = 1/\tau \sum_t^\tau z_t$ where τ is the time range. The so called $\mathcal{R}/\mathcal{S}$ statistic is given as [9]

$$(\mathcal{R}/\mathcal{S})_\tau \stackrel{\text{def}}{=} \frac{1}{s_\tau}\left[\max_{1\le t\le\tau}\sum_{t=1}^{\tau}\{z_t-\bar{z}\} - \min_{1\le t\le\tau}\sum_{t=1}^{\tau}\{z_t-\bar{z}\}\right]. \tag{4.1.1}$$

The expression in the brackets are usually called the range which is scaled by the standard deviation s_τ defined as

$$s_\tau \stackrel{\text{def}}{=} \left[\frac{1}{\tau}\sum_{t=1}^{\tau}\{z_t-\bar{z}\}\right]^{1/2} \tag{4.1.2}$$

Using this analysis, Hurst found out that the rescaled ranges would follow a power law [48, 71]

similar to

$$\mathbb{E}[(\mathcal{R}/\mathcal{S})_\tau] \sim c\tau^H.$$

By OLS regression in a log-log plot, the Hurst exponent can be found from the equation $\log(\mathcal{R}/\mathcal{S})_\tau = \log(c) + H\log(\tau)$. For a FARIMA or a fractional Gaussian noise, process the rescaled range statistic is behaves like a power law such as τ^H as $\tau \to \infty$.

Statistical properties

In the survey carried out by Taqqu et. al [71], 9 estimators of long range dependence were tested in a simulated study. When applied to FGN and ARFIMA, the rescaled range (RS or R/S) method was overestimating given a low true H and underestimating given a high true value of H. In addition, the method displays an increasing variance for increasing values of true H. This behavior is also seen in other studies [35, 70, 73, 7]. Further on, in the extension of [71], a series of estimators were studied again [70], but in this context of heavy tailed distributions. The R/S method was shown to be robust against heavy tailed distributions and it was shown to be sensitive towards short range dependency in the underlying process. This is confirmed in the study by Kristofuek [36]. The R/S statistic has been shown to be biased for small samples of τ [16] in coherence with the study by Weron [73] which recognized that for an increasing sample size, the R/S method approximated the true Hurst exponent.

4.1.2 Periodogram method

The *periodogram method* introduced by Geweke-Porter and Hudak [22] utilizes the periodogram which is the sampling analogy of the spectral density $S(\omega)$ introduced in Equation (2.4.4). This method aims to estimate the Hurst exponent by fitting a slope in the spectral domain. Given a stationary time series $\{x_t\}$, the periodogram $I(\omega)$ is an estimator for the spectral density function $S(\omega)$ and is defined by Equation 4.1.3

$$\hat{S}(\omega_j) = I(\omega_j) = \frac{1}{2\pi}\left|\sum_{t=1}^{n}(x_t - \bar{x}_n)e^{it\omega_j}\right|^2 = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\hat{\gamma}(k)e^{ik\omega_j} \tag{4.1.3}$$

with corresponding Fourier frequencies $\omega_j = 2\pi j/n$ and $j = 1, \ldots, M$, $M = [(n-1)/2]$ being the integer part. The sample covariances $\hat{\gamma}(k)$ is defined as [11]

$$\hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(x_t - \bar{x}_n)(x_{t+|k|} - \bar{x}_n) \tag{4.1.4}$$

where the sample averages are given as $\bar{x}$. The periodogram (or GPH) estimator is found by considering the behavior of the spectral density which is given as $c|\omega|^{1-2H}$ for $\omega \to 0$ and

OLS regression gives

$$\log I(\omega_j) \simeq \log c - 1 - 2H \log \omega_j + \varepsilon_j \tag{4.1.5}$$

for a frequency ω close to zero and where $c \geq 0$.

Statistical properties

It can be shown that the periodogram is an asymptotically unbiased estimator of the spectral density [9],

$$\lim_{T\to\infty} \mathbb{E}[I(\omega)] = S(\omega). \tag{4.1.6}$$

Based on the fact that the behavior of the spectral density of a stationary time series is given $S(\omega) \sim c_s \omega^{1-2H}$ for $\omega \to 0$ as in Equation (3.1.4), the estimate of H is acquired by fitting a line trough the periodogram in a log-log plot. The periodogram has a theoretical slope of $-2d$ or $1-2H$ and since the ω^{1-2H} behavior is only valid for low frequencies, it is only necessary to utilize the lowest 10% of the observations [37] and [71].

Geweke and Porter-Hudak [22] deduced the following relationship (under a set of arguments)

$$\sqrt{M}(\hat{H} - H) \sim \mathcal{N}\left(0, \frac{\pi^2}{6}\right) \tag{4.1.7}$$

which was later shown to hold for stationary, Gaussian, zero-mean processes with $H \in (0,1)$ [63]. In terms of finite sample properties Rea et al. [60] displayed that the periodogram estimator was unbiased for FGN series and gave more precise estimates for longer series which also were the results of Taqqu et al. [71]. In other studies it has been shown that the periodogram estimator is robust to different sampling distributions as well as being insensitive to the presence of short range correlations [70]. Weron [73] found out that for small samples the periodogram method would underestimate the true H.

4.1.3 Higuchi method

In the paper *Approach to an Irregular Time Series on the Basis of the Fractal Theory* by Higuchi [69], a procedure for determining the fractal dimension $D = 2 - H$ of a self similar process was introduced. See also the survey by [71] for more coverage.

This method aims to determine the long range dependence of a time series $\{x_t\}_{t=0}^{T}$ by using its partial sums $z(t) = \sum_{j=1}^{t} x_t$, such as converting a fractional Gaussian noise into a fractional Brownian motion. Subsequently, the *normalized* length of the curve is found as

$$L(m) = \frac{T-1}{m^3} \sum_{j=1}^{m} \left[\frac{T-j}{m}\right]^{-1} \sum_{k=1}^{[(T-j)/m]} \big|z(j+km) - z(j+(k-1)m)\big|, \tag{4.1.8}$$

where T is the length of the underlying time series and m is the block sizes. Further on,

$$\mathbb{E}[L(m)] \sim c_H m^{-D}$$

where $D = 2 - H$. Finally, plotting $\ln(L(m))$ against m gives a slope of $D = 2 - H$. In this case, $\ln(L(m))$ is the dependent variable and m is the independent variable.

Statistical properties

Rea et al. [60] discovered in their finite sample study that the Higuchi method exhibited bias towards underestimating H independent of the Hurst exponent. Further on, for a H closer to one, the confidence intervals were increasing. In Taqqu et al. [71] the method were unbiased for the true H for both FGN and ARFIMA with increasing MSE for increasing levels of H.

4.1.4 Detrended fluctuation analysis

This method was proposed by Peng et al. [57] in their studies of long range dependence in DNA sequences. Given a time series $\{x_t\}_{t=1}^{\mathbb{T}}$, the detrended fluctuation analysis is composed of five steps. First, a new time series is generated

$$y_t = \sum_{k=1}^{t} x_k \tag{4.1.9}$$

which is the partial sums of the underlying time series. The time series $\{y_t\}_{t=1}^{\mathbb{T}}$ is non-stationary. Secondly, the time series is divided into $[n/l]$ non-overlapping blocks where $[\cdot]$ is the integer part. Thirdly, a least square line is fit for each block. The fourth step is to detrend the partial sums $\{y_t\}$ in the manner

$$z_t = y_t - y_t^l$$

where y_t^l is the adjusted fit for each block. Finally, the *root mean squared fluctuation* is for each $l \in \{4, 5, \ldots, g(n)\}$ given as

$$\mathcal{F}(l) = \sqrt{\frac{1}{\tilde{n}} \sum_{t=1}^{\tilde{n}} z_t^2} \tag{4.1.10}$$

where $\tilde{n} = [M * l] \leq n$ (the maximum multiple of l) with $M = [n/l]$. It has been proven [25] that $g(n) = [n/10]$ is an optimal choice. In a log-log plot, the linear relationship

$$\mathcal{F}(l) \sim c_{DFA} l^H$$

as (4.1.10) increases with the block size l. The scale of long range dependence can then be found by regressing $\ln(F(l))$ on $\ln(l)$.

Statistical properties

Along with the rescaled range method, the detrended fluctuation method (DFA) is one of the most used methods for estimating the amount of long range dependence in data. Moreover, Taqqu et al [71] showed that the DFA, although consistent for increasing values of H, underestimates the true Hurst exponent which is being diagnosed in the same article. In addition, this negative attribute has also been demonstrated in [70] and [51]. These articles executed the method on series with influence from short memory data and it appeared to be sensitive to such interference.

Further on, in the study by Crato et. al [53] it was proven that, given $\varepsilon_t \sim i.i.d\, \mathcal{N}(0, \sigma^2)$ in the regression model, the DFA method is a minimum variance unbiased and consistent estimator for H. DFA has also been shown to be sensitive to short range dependence When applied to heavy tailed data, the DFA method has precise estimates in terms of expected values but has wide confidence intervals making it difficult to make statements regarding the statistical conclusions.

4.1.5 Generalized Hurst exponent

This procedure was proposed by Barabasi and Vicsek [3], and was brought to light in recent history by Di Matteo et. al [49]. It is based on the scaling behavior of the q-th order moments of the increment process from the underlying process $\{x_t\}$. The method serves as a tool for detecting scaling behaviors as seen in [42]. The generalized Hurst exponent (GHE) has also been used to identify the level of development in a market [48, 50].

The scaling properties of a time series are given in terms of the statistic $K_q(\tau)$ which is defined as [49]

$$K_q(\tau) = \frac{1}{T-\tau} \sum_{\tau=0}^{T-\tau} |X(t+\tau) - X(t)|^q \tag{4.1.11}$$

where T is the length of the time series and $K_q(\tau)$ scales as

$$K_q(\tau) \sim c\tau^{qH(q)}.$$

On the basis of this scheme, it is possible to make statements regarding the underlying processes. In fact, two separate observations can be made which is determined by the scaling behavior of the process. A process can be observed for which the $H(q) = H$ - the Hurst exponent - is a constant independent of the moments of the process q. This is known as a

unifractal process. In this case, the process is exclusively described by the Hurst exponent and is self-similar [40]. The fractional Brownian motion $B_H(t)$ with $H \neq \frac{1}{2}$ and the ordinary Brownian motion $B(t)$ with $H = \frac{1}{2}$ are both unifractal processes. The other observation is when $H(q)$ is not constant and depends on q. The process is then known as *multifractal* and $q * H(q)$ scales differently for separate order of moments q. For $q = 1$ the scaling of the absolute increments $|x_{t+\tau} - x_t|$ is described by $H(1)$. In the context of detecting long range dependence it is essential to consider the autocorrelations of the increments which is the case for the second order of moments $q = 2$. The Hurst exponent then can be estimated from the relationship $K_2(\tau) \sim c\tau^{2H(2)}$ and will correspond to the H estimates from the other methods. In particular, the estimate of $H(q)$ can be found by regressing $K_q(\tau)$ on τ. In the analysis of Chapter 5 $H(q = 2)$ will be utilized.

Given the relationship $K_q(\tau) \sim c\tau^{q*H(q)}$, the intensity of LRD is estimated as the slope of $q * H(q)$ versus q in a plot. One will also in such a plot observe whether the process is exhibiting unifractality or multifractality if the relationship is linear or non-linear respectively. In the robustness assessment, this will be used to reveal some of the features of the data.

In the unique case of $q = q^*$ where $q^*H(q^*) = 1$ it was pointed out by Mandelbrot [42] that $K_{q^*}(\tau)$ scales linearly. Given that $qH(q)$ is a monotonic growing function for q, then all $H_q(\tau)$ with $q < q^*$ will scale slower than the time scale τ and for $q > q^*$ scale faster. It is said that q^* is the scaling threshold value and it is observed that

$$H(q = 1) = H(q = 2) = H(q = q*).$$

Statistical properties

In the study by Barunik and Kristoufek [7] it was shown that the GHE $q = 2$ estimator is robust to heavy tails in the underlying process. In the same study it was observed that GHE with $q = 2$ was outperforming R/S and DFA in terms of variance and bias.

It is also tested on random walks in [49] where it was found to be unbiased for $q = 2$, i.e. when the scaling properties of the autocorrelation function is examined.

In the works of Barunik et al. [6], the above is confirmed and in addition it reveals that the GHE method gives weakly biased results for ARFIMA signals. However, for fractional Gaussian noise and ARFIMA the estimates are equal for $GHE(q = 1 = 2 = 3)$ when the innovations come from a normal distribution. Also, when short range dependence is present the GHE will give a small upward bias. In the same article it is shown that $GHE(q = 2)$ is insensitive to heavy tails and $GHE(q = 1)$ and $GHE(q = 3)$ gives biased results under such conditions. In the study by Souza et al. [65] it was confirmed that GHE is insensitive to outliers as it is not dealing with min and max data.

4.1.6 Aggregated variance

The aggregated variance method is seen in the works of Beran [9] and more recently in Taqqu et. al. [71]. It follows from the self-similarity property of the time sample process. Recall from Equation(3.2.1) that the aggregated process is given as

$$x_t^{(m)} = \frac{1}{m} \sum_{\tau=(t-1)m+1}^{tm} x_\tau, \quad t = 1, 2, \ldots, [T/m]$$

for consecutive values of the block size m. Assuming (asymptotical) self-similarity, recall from Equation (3.2.2) that $x_t^{(m)}$ is equal in distribution to $m^{H-1}x_t$ and the variance of the aggregated series and the ordinary series is related as $\mathbb{V}[x_t^{(m)}] = m^{2H-2}\mathbb{V}[x_t]$. The sample variance of the aggregated series in Equation (3.2.1) is

$$\widehat{\mathbb{V}[x_k^{(m)}]} = \frac{1}{T/m} \sum_{t=1}^{T/m} (x_t^{(m)})^2 - \left(\frac{1}{T/m} \sum_{T/m}^{t=1} x_t^{(m)} \right)^2 . \tag{4.1.12}$$

The Hurst exponent estimate is found by plotting $\widehat{\mathbb{V}[x_k^{(m)}]}$ against $x_t^{(m)}$ in log-log plot. In other words, the estimate of H is found by linear regression where $\widehat{\mathbb{V}[x_k^{(m)}]}$ is the dependent variable and $x_t^{(m)}$ is the independent variable. If the estimated variances are equal to their true values then all of the points would lie on a straight line with slope $2H - 2$. In practice there will be small deviations from this slope and it is estimated by fitting a straight line through the points. The slope of this line is then accepted as the estimate of H. For a process with short range or no dependence the slope of $2H - 2$ in the log log plot should be equal to 1.

Statistical properties

It is pointed out by Beran [9] that this approach is very naive due to the fact that the estimator in Equation (4.1.12) is biased in the presence of non-zero correlations, which is the case for long range dependence. Beran also points out that this bias disappears for large $[T/m]$, for instance if the number of observations T is large and the block sizes m are small. However, this happens very slowly for processes with long range dependence and it can be stated that

$$\lim_{[T/m]\to\infty} \mathbb{E}\left[\widehat{\mathbb{V}[x_k^{(m)}]} \right] \sim \mathbb{V}[x_k^{(m)}]\{1 - c[T/m]^{2H-2}\}$$

where c is a constant. When the process is persistent i.e. $H > 1/2$ and c is positive, then $\widehat{\mathbb{V}[x_k^{(m)}]}$ results in a underestimation of $\mathbb{V}[x_k^{(m)}]$ and thus H as well. As shown in research conducted by Taqqu et al. [71] this underestimation is increasing for larger values of H. The

finite sample study by Rea et al. [60] confirms this for a range of sample sizes.

4.1.7 Whittle approximation

Whittle's estimator is a maximum likelihood estimator originating from the fact that the periodogram $I(\omega)$ of a covariance stationary process behaves as

$$I(\omega_k) \sim e^{S(\omega_k)^{-1}} \tag{4.1.13}$$

for large samples of T. The Whittle approximation assumes that there is a functional form for $I(\omega)$ and the goal is to minimize a set of parameters based on this. The traditional Whittle estimator requires a specification of the expected functional form, such as fractional Gaussian noise or ARFIMA, thus making misspecifications a threat. The *local* Whittle estimator [63] and [37] only assumes a functional form for the spectral density at frequencies close to zero. An exact form of the local Whittle estimator has been proposed by Philips and Shimotsu [64] which is the estimator that will be used in the analysis later.

Local Whittle estimator

In the paper *Gaussian semiparametric estimation of long-range dependence* [63], it is suggested by Robinson, by using the local Whittle estimator, that a Gaussian semiparametric estimator can be used to determine the intensity of long range dependence. This procedure relies on the fact that the spectral density can be approximated at zero frequency by the relationship $S(\omega) \sim G\omega^{-2d}$ as seen in Equation (3.1.4).

By using the relationship in Equation (4.1.13), its likelihood function for the whole sample is computed as

$$\mathcal{L}\{I(\omega)|G,d\} = \prod_{k=1}^{m} \frac{1}{S_{(G,d)}(\omega_i)} \exp\{-\frac{I(\omega_k)}{S_{(G,d)}(\omega_k)}\}.$$

The log-likelihood is found as

$$\ln \mathcal{L} = \sum_{k=1}^{m} \left\{ -\ln S_\theta(\omega_k) - \frac{I(\omega_k)}{S_\theta(\omega_k)} \right\} \tag{4.1.14}$$

where $\theta = (G, d)$ is the parameter vector. The relationship in (4.1.14) is also known as the *Whittle likelihood*. As the functional form in Equation (4.1.13) is only assumed at frequencies close to zero, the log-likelihood in the neighborhood of $\omega \approx 0$ must be computed. It is given as the objective function [37]

$$\ln \mathcal{L} \sim \sum_{k=1}^{m} \left\{ \ln(G\omega_k^{-2d}) + \frac{I(\omega_k)}{G\omega_k^{-2d}} \right\} \tag{4.1.15}$$

with corresponding likelihood equation

$$\frac{\partial \ln \mathcal{L}}{\partial G} = \sum_{k=1}^{m} \left\{ \frac{1}{G} - \frac{I(\omega_k)}{G^2 \omega_k^{-2d}} \right\} = 0. \tag{4.1.16}$$

Solved for G gives

$$\hat{G} = \frac{1}{m} \sum_{k=1}^{m} \left\{ \frac{I(\omega_k)}{\omega_k^{-2d}} \right\}$$

which is inserted into (4.1.15) yielding

$$\ln \mathcal{L} = m \ln \hat{G} - 2d \sum_{k=1}^{m} \ln(\omega_k) + m^2.$$

Subsequently, its minimization with respect to d is equal to [62]

$$\arg\min_{\hat{G},d} \mathcal{L} = \ln \hat{G} - \frac{2d}{m} \sum_{k=1}^{m} \ln(\omega_k) + m$$

which can be given as [63]

$$\hat{d}_{LW} = \arg\min_{d} \left\{ \ln \left(\frac{1}{m} \sum_{k=1}^{m} \frac{I(\omega_k)}{\omega_k^{-2d}} \right) - \frac{2d}{m} \sum_{k=1}^{m} \ln(\omega_k) \right\}.$$

Exact local Whittle estimator

As an extension of the local Whittle estimator the *exact* local Whittle estimator (ELW) [64] has been proposed as an alternative that does not rely on differencing of tapering [72] and seemingly offers a general procedure for estimating the level of long range dependence. The exact local Whittle objective function is almost deduced in the same manner as above yielding the Whittle likelihood

$$\ln \mathcal{L} \sim \sum_{k=1}^{m} \left\{ \ln(G\omega_k^{-2d}) + \frac{I_{\Delta^d \chi}(\omega_k)}{G} \right\} \tag{4.1.17}$$

However, the term $\frac{I(\omega_k)}{G\omega_k^{-2d}}$ is replaced with $\frac{I_{\Delta^d x}(\omega_k)}{G}$, which is the periodogram of the fractionally differenced series $(1-L)^d = x_t$ (ARFIMA noise) cohering to the true parameter. The ELW estimator is

$$\hat{d}_{ELW} = \arg\min_{d} \left\{ \ln \xi - \frac{2d}{m} \sum_{k=1}^{m} \ln \omega_k \right\}$$

where ξ is the averaged periodogram

$$\xi = \hat{G}(d) = \frac{1}{m}\sum_{k=1}^{m} I_{\Delta^d x}(\omega_k).$$

Statistical properties

As the intention is to use the ELW estimator only its properties are presented here. It was shown that the ELW estimator yields consistent results and have $\hat{d}_{ELW} \sim \mathcal{N}(0, \frac{1}{4})$ limit distribution [64] for $d \in (-\frac{1}{2}, 1)$, thus including the nonstationary case of $d \in (\frac{1}{2}, 1)$ [72].

4.2 Statistical properties of the estimators

As mentioned in Chapter 1, a broad field of research has been conducted on long range dependence in recent years but only a small part of this research is aimed at investigating finite sample properties of he estimators. In the following, the finite sample properties of the proposed estimators will be summarized.

	R/S	***PER***	***HIG***	***DFA***	***GHE***	***AV***	***ELW***
Unbiased[a]	N	Y	Y	Y	Y	N	Y
Small sample bias[b]	Y	Y	N	N	N	Y	N
Consistent[c]	N	N	N	Y	Y	N	Y
Heavy-tail bias[d]	N	N	-	N	N	N	N
Short-memory bias[e]	Y	N	Y	Y	N	Y	N

Table 4.2: Finite sample properties of the estimators.

In Table 4.2 a list of attributes describing the finite sample properties for each estimator is given. The unbiasedness (a) is based on the estimation of FGN and ARFIMA(0,d,0) series as in [71], [70], [7] and [73]. For the GHE estimator the unbiasedness is based on [6] for series with $H = 0.5$. The unbiased measure is this context based on a single sample size. The bias of ELW is given in [64]. For small sample bias (b), the properties are largely based on [60], [7] and [73]. Again, the estimators were tested on Brownian motions with $H = 0.5$. For ELW it is based on [64] and for GHE [7]. Regarding the consistency (c) of the estimators see [60] and [35]. The consistency was based on having smaller MSE for increasing sample sizes. In [5] this is also covered. For DFA see [53]. The heavy tailed bias (d) and the short memory bias (e) is for most of the estimators covered in [70] and [36]. For the GHE estimator see [6] and [65], and [4] for ELW.

Chapter 5

Finite sample properties

In this chapter two types of long range dependent series will be generated: fractional Gaussian noise and autoregressive fractionally integrated moving average noise (the difference of an ARFIMA signal) with corresponding known properties. Afterward the estimators are applied to the series for different levels of intensity of LRD and sample sizes in order to establish their finite sample properties.

5.1 Data generating processes

In order to construct a controlled environment, it is crucial to know the data generating process. Here, the algorithms for generating sample paths of long range dependent processes will be introduced. The processes that were introduced earlier are fractional Brownian motion, fractional Gaussian noise and ARFIMA. These algorithms rely on the discrete Fourier transforms introduced in Chapter 2. The (inverse) discrete Fourier transform is computed by the algorithm known as the *(inverse) fast Fourier transform* which requires the length of the time series to be a power of 2 in order to be efficient. Computational efficiency will be denoted by $\mathcal{O}(\cdot)$ which measures the evolution of an algorithm in terms of complexity. For example, an algorithm with complexity $\mathcal{O}(c^n)$ will increase exponentially for larger data sets.

5.1.1 Fractional Gaussian noise

This part aims to generate stationary long memory processes which are a realization $\{x_t\}_{t=1}^N$ of the underlying process $\{X_t\}_{t=1}^N$. The methods are said to be exact in that the quality of the generations rely on the white noise generator (Matlab function `randn`), specifically the one featured in Matlab [47]. The fractional Gaussian noise is simulated by the two separate methods corresponding to their level of persistence. For $H \in (0, 1/2)$, Lowen's circulant

method [39] was utilized. For $H \in (1/2, 1)$, a circulant matrix embedding method [9, 18] is used. See also Taqqu et al. [5].

Lowen's circulant method

The method for generating a fractional Brownian motion $B_H(t)$, $t = 0, \ldots, T$ for $H \in (0, 1/2)$ is given in [39, 5]. The contrasting feature of this method to other methods lies in the periodic autocorrelation function defined as

$$\tilde{\rho}(n) = \begin{cases} 2^{-1}\{1 - (n/N)^{2H}\} & \text{for} \quad 0 \leq t \leq N \\ \rho(2N - n) & \text{for} \quad N < n \leq 2N. \end{cases} \tag{5.1.1}$$

is periodic of n with period $2N$. The autocorrelation function $\tilde{\rho}(n)$ is real, nonnegative, symmetric and thus the Fourier transform exhibits the same features for all n [39]. It is therefore clear that the discrete Fourier transform F_n serves as a discrete spectral density S and will be defined as

$$F_n = S(\omega) = \sum_{n=0}^{2N-1} \tilde{\rho}(n) \exp\left\{\pi i \frac{\omega n}{N}\right\}.$$

Given the random variables

$$G_i(\omega) \sim \ i.i.d. \mathcal{N}(0, 1), \quad \omega = 1, \ldots, N - 1$$

then a function with period $2N$ can be defined

$$X(k) = \begin{cases} 0 & \text{for } \omega = 0, \\ \sqrt{\frac{S(\omega)}{2}}\{G_1(\omega) + iG_2(\omega)\} & \text{for } \omega = 1, \ldots, N - 1, \\ \sqrt{S(\omega)}\varepsilon & \text{for } \omega = N, \\ X^*(2N - \omega) & \text{for } \omega = N + 1, \ldots, 2N - 1 \end{cases} \tag{5.1.2}$$

where ε is Gaussian white noise as in Equation (2.2.1) and $X^*(\cdot)$ is the complex conjugate of $X(\cdot)$. The inverse discrete Fourier transform is computed using the inverse fast Fourier transform from (2.4.4) as

$$x(n) = \frac{1}{2N} \sum_{\omega=0}^{2N-1} X(k) \exp\left\{\pi i \frac{\omega n}{N}\right\} \text{ for } n = 0, \ldots, 2N - 1. \tag{5.1.3}$$

Bardet et al. [5] shows that $x(n)$ is stationary, Gaussian, zero mean and periodic of $2N$ computed as

$$x(n) = \frac{1}{\sqrt{2N}}(y(n) - \nu) \tag{5.1.4}$$

where y is a normally distributed an has the correlation function $\tilde{\rho}$ and ν is a zero mean, Gaussian random variable dependent on y. Then, by (5.1.4) the fractional Brownian motion is defined

$$B_H(n) = \sqrt{2N}\,\{x(n) - x(0)\}$$

Lastly, the fractional Gaussian noise is given as

$$z(n) = B_H(n) - B_H(n-1) \tag{5.1.5}$$

Circulant matrix embedding

The method proposed by Davies and Harte [18] generates FGN $H \in (0.5, 1)$ and aims to embed the covariance matrix $\mathbf{\Sigma}$ in the circulant matrix $\mathbf{C}$. The method was later refined by Wood and Chan [75] and Dietrich and Newsam [20] where the M-vector was chosen as $2N-1$. In addition to Beran [9] Dieker and Mandjes [19] give a different view of the algorithm.

The rationale behind the method is that a circulant matrix can easily be diagonalized by the DFT which skips matrix computations and drastically shortens the computational complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$. This decrease in complexity stems from the fact that most FFT softwares take advantage of a 'power-of-two' algorithm. Efficient computations take place for $N = 2^k + 1$ or $M = 2N - 1$ which will be utilized here. For this procedure, $\gamma(k)$ is defined as an slight alteration of the relationship in Equation (3.2.4)

$$\gamma(k) = \frac{\sigma^2}{2N^{2H}}\left\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right\}. \tag{5.1.6}$$

The algorithm starts with defining an N-vector as

$$\Lambda_k = \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \vdots \\ \gamma(N-1) \\ \gamma(N) \\ \gamma(N+1) \\ \vdots \\ \gamma(2N-1) \end{pmatrix} \tag{5.1.7}$$

where $\Lambda_k \stackrel{\text{def}}{=} \gamma(k)$ and the discrete Fourier transform of Λ_k is given as

$$F_p = \sum_{k=0}^{p-1} \Lambda_k \exp\left\{-2\pi i \frac{nk}{p}\right\} \tag{5.1.8}$$

for $n = (0, \ldots, p-1)$. As follows, a N-vector is simulated as $Z = U_p + iV_p$ for $1 \leq p$. Accordingly a set of independent random variables distributed as $U_p \sim \mathcal{N}(0,1)$ and $V_p \sim \mathcal{N}(0,1)$ is defined

$$\begin{aligned} Z_0 &= U_p \\ Z_p &= \frac{1}{\sqrt{2}} \{U_p + iV_p\} \\ Z_N &= U_p \\ Z_{2N-p} &= \frac{1}{\sqrt{2}} \{U_{2N-p} - iV_{2N-p}\} \end{aligned}$$

where $i = \sqrt{1}$. Further on the inverse fast Fourier transform is performed on $\sqrt{F_p} Z_p$ as

$$x_k = \frac{1}{\sqrt{2N}} \sum_{p=0}^{2N-2} \sqrt{F_p} Z_p \exp\left\{\pi i \frac{pk}{N}\right\}. \tag{5.1.9}$$

To make it more clear the process $\{x_k\}_{k=0}^{2N-1}$ can be separated into the Fourier transform of the cases [17]

$$z_k = \begin{cases} \sqrt{\frac{F_k}{2N}} U_k & \text{for} \quad k = 0, \\ \sqrt{\frac{F_k}{4N}} (U_k + iV_k) & \text{for} \quad k = 1, \ldots, N-1, \\ \sqrt{\frac{F_k}{2N}} U_k & \text{for} \quad k = N, \\ \sqrt{\frac{F_k}{4N}} (U_{2N-k} - iV_{2N-k}) & \text{for} \quad k = N+1, \ldots, 2N-1, \end{cases} \tag{5.1.10}$$

Finally the fractional Gaussian noise is produced as the real part of the first N coordinates of x_k.

5.1.2 ARFIMA

The following procedures are described by Stoev and Taqqu [68] and implement the fast Fourier transform to current well known methods for generating FARIMA sample paths.

The model defined in Equation (3.2.9) is reiterated as

$$\phi(L)^p(1-L)^d x_t = \theta(L)^q \varepsilon_t \tag{5.1.11}$$

where $d \in \mathbb{R}$. In Kokoszka and Taqqu [32] the fractionally integrated part in (5.1.11) is defined by

$$(1-L)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \tag{5.1.12}$$

with corresponding coefficients a_j from the Taylor expansion

$$a_0 = 1, \quad a_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \quad j \geq 1. \tag{5.1.13}$$

As the procedure applies a *lowpass filter* which in its simplest form is given as

$$y(t) = x(t) + x(t-1) \text{ for } t = 1, \ldots, T.$$

which in this case where the AR and MA parts are filtered on the fractional differencing term in (5.1.12) gives the expression

$$\zeta(n) = a(n) + \sum_{j=1}^{p} \theta_j \zeta(n-j) - \sum_{k=1}^{q} \phi_k b(n-k) \tag{5.1.14}$$

where $n = 1, \ldots, N$. Using (5.1.13) the series can be represented as [32]

$$x_t = \sum_{j=0}^{\infty} \zeta_j \varepsilon_{t-j} = \sum_{j=-\infty}^{t} \zeta_{t-j} \varepsilon_t. \tag{5.1.15}$$

It is now apparent that the moving average coefficients ζ_j exhibit long range dependence as conducted by the exponent d as $\zeta_j \sim j^{d-1}$ for $j \to \infty$ and then the decay is observed $\sum_j^{\infty} |\zeta_j| = \infty$.

The algorithm by Stoev and Taqqu [68] approximates (5.1.15) as a truncated moving average representation

$$x_t \approx x'_t \stackrel{\text{def}}{=} \sum_{j=0}^{N-1} \zeta_j \varepsilon_{t-j} \tag{5.1.16}$$

for $n = 1, \ldots, N-1$. Next, a set of periodic functions are defined in order to utilize the fast

Fourier transform algorithm. First $\tilde{\zeta}$ is defined as

$$\tilde{\zeta}_t = \begin{cases} \zeta_j & \text{for} \quad j = 0, \ldots, N-1, \\ 0 & \text{for} \quad j = N, \ldots, 2N \end{cases} \tag{5.1.17}$$

where $\tilde{\zeta}_{j+t(2N)} \stackrel{\text{def}}{=} \tilde{\zeta}_j$ for $j = 0, 1, \ldots, 2N-1$ making $\tilde{\zeta}_j$ $2N$-periodic. Another $2N$-periodic function is defined as

$$\tilde{y}_{j+t(2N)} \stackrel{\text{def}}{=} \varepsilon_j \quad \text{for} \quad j = 0, 1, \ldots, 2N-1. \tag{5.1.18}$$

Then the discrete Fourier transform of the periodic series is computed as

$$\mathcal{Z}_n = \sum_{n=0}^{2N-1} \tilde{\zeta}_j \exp\left\{-\pi i \frac{nj}{N}\right\} \tag{5.1.19}$$

and

$$\mathcal{Y}_n = \sum_{n=0}^{2N-1} \varepsilon_j \exp\left\{-\pi i \frac{nj}{N}\right\} \tag{5.1.20}$$

As follows, by (5.1.16), the relationship

$$x'(t) =_d \sum_{j=0}^{2N-1} \tilde{\zeta}_j \tilde{y}_{n-j} \tag{5.1.21}$$

states equality in distribution for $n = 0, 1, \ldots, N-1$. Finally the inverse discrete Fourier transform of the product $(\tilde{\zeta}\tilde{y})_t \stackrel{\text{def}}{=} \tilde{\zeta}_t \tilde{y}_t$ is yields the sample path of the ARFIMA process

$$z_t = \frac{1}{N} \sum_{n=0}^{N-1} (\tilde{\zeta}\tilde{y})_t \exp\left\{\pi i \frac{nj}{N}\right\} \quad \text{for} \quad j = 0, 1, \ldots, N-1. \tag{5.1.22}$$

5.2 Properties of the estimators

Though a series of finite sample property studies have been performed which concentrate on length [55, 73], non-Gaussian innovations [36] and short memory [70] none of these look at the baseline case for pure LRD series in the same comprehensive manner as here.

The rest of this chapter is devoted to establishing the baseline features of the estimators when presented with pure LRD signals.

5.2.1 Biasedness and standard deviation of the estimators

In order to assess the bias, standard deviation and mean squared error of the estimators a simulation study were performed on fractional Gaussian noise and ARFIMA noise type signals. The analysis was done on time series of different length and different intensity of long memory reflected in the Hurst exponent. $FGN(H)$ and $ARFIMA(0, H - 1/2, 0)$[1], $H \in \{0.1, 0.2, \ldots, 0.9\}$ for lengths of $N = 2^j$ for $j \in \{6, 7, \ldots, 14\}$ was generated. At each level of H and sample size N, $r = 1000$ replications of both signals was generated.

Three measures will be used in the plots to describe the performance of the estimators; *bias*, standard deviation and *mean squared error*. Bias is defined as

$$\mathbb{B}(\hat{H}) = \mathbb{E}[\hat{H}] - H \tag{5.2.1}$$

the standard deviation

$$\hat{\sigma} = \left(\frac{1}{r-1} \sum_{i=1}^{r} (\hat{H}_i - \mathbb{E}[\hat{H}_i])^2 \right)^{\frac{1}{2}} \tag{5.2.2}$$

and mean squared error (MSE)

$$MSE(\hat{H}) = \frac{1}{r} \sum_{i=1}^{r} (\hat{H}_i - H_i)^2 \tag{5.2.3}$$

where r is the number of replications. The mean squared error $MSE(\hat{H})$ is equal to the variance $\mathbb{V}[\hat{H}]$ if $\mathbb{B}(\hat{H}) = 0$, but not otherwise. For proof see Equation A.1.1.

As the analysis is looks at the performance of the estimators at different lengths a few terms describing the behavior is introduced. Firstly, *asymptotic unbiasedness* will be defined as

$$\mathbb{B}_n(\hat{H}_n) = \lim_{n \to \infty} \mathbb{E}[\hat{H}_n] - H \tag{5.2.4}$$

where $\hat{H}_n$ means that the estimator is based on a sample size n. Secondly, if the estimator $\hat{H}$ approaches the true value H as the sample size increases it is said to be a *consistent estimator* of H. Formally, the estimator $\hat{H}$ is said to be consistent if the probability P that the absolute value of the difference $\hat{H} - H$ is less than $\delta > 0$, an arbitrarily small value, approximates to 1,

$$\lim_{n \to \infty} P\{|\hat{H} - H| < \delta\} = 1. \tag{5.2.5}$$

A sufficient condition for consistency is that the bias $\mathbb{B}_n$ and the variance σ^2 tend to zero asymptotically (see Greene [23]). It is clear that the consistency property is a large sample

[1] ARFIMA noise which is the differenced ARFIMA series.

property whereas unbiasedness can hold for any sample size.

In order to compare the performances each plot of a measure, the same axes and reference lines are inserted at bias $\mathbb{B}(H_n) \pm |0.02|$ and at standard deviation $\sigma = 0.05$.

The plots including the standard deviations are included in Appendix A.2 as including them here would be excessive. However, as mentioned, the MSE are explained by $\mathbb{B}^2 + \sigma^2$ and therefore explains the standard deviation to a certain degree. See Appendix A.

5.2.2 Results

The graphical presentation below give an idea of how the estimators perform under conditions that are unaffected by noise. These results serve as a baseline case for each estimator and some general remarks can be made:

- All of the estimators' biases are equal or lower with increasing time series length.
- All of the estimators' standard deviation are asymptotically decreasing for increased time series length.
- For most of the estimators the intensity of long range dependence have little effect on the standard deviation, although in general the standard deviation is lower for a smaller value of H.
- There is almost always a difference in the estimators' performance on FGN versus ARFIMA type signals in terms of bias.
- The estimators relying on the periodogram have MSE's that are independent of the level of LRD intensity.

In the following discussion the main observations are discussed. In terms of referencing figures the scheme is as follows: **Estimator - Figure (Bias), (MSE), (Standard deviation)** where the pairs in the parentheses are referred to as FGN and ARFIMA noise respectively. See Figure 5.1 for a legend that is commmon for all the plots in this chapter.

Rescaled range - Figures (5.2a,5.5a),(5.2b,5.5b),(A.1a,A.1b)
The RS method suffers from from a heavy bias for small N which persist to a smaller degree for bigger N. It overestimates for intensity levels $H \in (0.1, 0.6)$, and underestimates for $H = \{0.8, 0.9\}$. For intensity levels $H = \{0.7, 0.8\}$ the estimates are within the "boundaries" at ± 0.025 for $N > 2^8$. Another observation from the simulations is that there is a pattern in the standard deviations ; for lower values of H, the standard deviation is lower. All of this holds for both FGN and ARFIMA except that the bias is for ARFIMA signals larger

for $H \in (0.1, 0.3)$ and the estimates for $H = 0.6$. Comparing the MSE plot for FGN and ARFIMA this is confirmed. The heavy bias contributes to the MSE and the method is biased for $H \in (0.1, 0.3)$.

Periodogram - Figures (5.2c,5.6a),(5.2d,5.6b),(A.2a,A.2b)

The periodogram method also suffers from bias at small values of N, but is asymptotically unbiased for large N for all levels of H (for both FGN and ARFIMA) except $H \in \{0.1, 0.2\}$ for FGN signals. In particular, at $N > 2^{10}$ the bias of the estimates are less than ± 0.025 for all levels of H except $H \in \{0.1, 0.2\}$. This method has the largest standard deviation among all estimators for large N but is independent of LRD intensity and is converging rapidly towards a low level. Given the asymptotic unbiasedness and the converging MSE and standard deviations this method seems to be consistent in the limit except for $H = 0.1$. This can be seen in the MSE plot where the MSE for $H = 0.1$ stands out .

Higuchi - Figures (5.2e,5.6c),(5.2f,5.6d),(A.2c,A.2d)

Given a FGN series, the Higuchi is performing well. Although all the estimates are within the ± 0.025 boundary at $N > 2^8$ it slightly underestimates the true value of H at all levels. At length $N > 2^{10}$ and $H \notin \{0.9\}$ the estimates are unbiased. It is also observed that the standard deviation is lower for a low intensity of LRD, especially for $H \in \{0.1, 0.2\}$. The MSE for $H = 0.1$ is one of the smallest for any intensity of LRD for any estimator.

For a ARFIMA series the story is different. Despite that the standard deviations are approximately the same the estimates are far more biased. For a low intensity of LRD it overestimates and for a high value of LRD it underestimates which persists for increasing sample sizes. Although it is close to unbiased for $H \in \{0.4, 0.5, 0.6\}$ at large sample sizes it is far from the performance given a FGN series.

Detrended Fluctuation - Figures (5.2e,5.6e),(5.3b,5.6f),(A.2e,A.2f)

For fractional Gaussian noise, the detrended fluctuation analysis shows convergence towards an asymptotically unbiased estimate for increasing N but it over- and underestimates the true value of H for low and high levels of LRD. Only at length $N = 2^{12}$ are all of the estimates within the reference lines. For ARFIMA, this pattern is also seen except the bias is larger when there is strong antipersistence in the time series. For both signals, the standard deviation depends on the intensity of LRD where it is smaller for lower values of H. It has also one of the lowest standard deviations compared to all of the methods. In terms of MSE the DFA method reveals itself to be one of the best estimators for all levels of H given a FGN series. The MSE's are the same for ARFIMA with exception of $H = 0.1$.

Generalized Hurst - Figures (5.2e,5.7a),(5.3d,5.7b),(A.3a,A.3b)

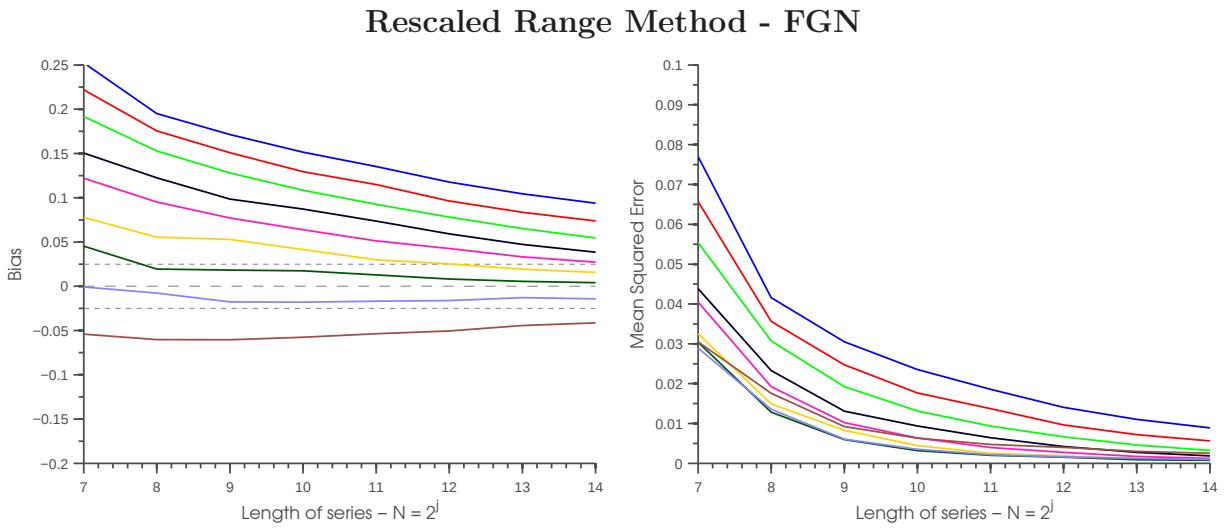
This estimation is performed with the generalized Hurst exponent for the second moment

of the incremental process, i.e. $GHE(q = 2)$. When estimating FGN series the method is asymptotically unbiased for $N > 2^{10}$ and all intensities of LRD except $H \in \{0.1, 0.2, 0.3\}$. As the standard deviation is converging towards zero as N gets larger, this method is consistent for FGN signals with $H \in (0.4, 0.9)$. For ARFIMA the results are poor in terms of bias. For any level of LRD intensity there is biasedness which is characterized by underestimating the high intensity and vice versa. Given low levels of H, the bias shows no sign of asymptotic improvement, but for high levels of H the estimates tend to underestimate. Despite this difference in bias, the standard deviations are the smallest of all estimators and the estimator seems to be consistent. The deviations are all in the range $(0.01, 0.02)$ for N^{14} and even for short series the deviations are small.

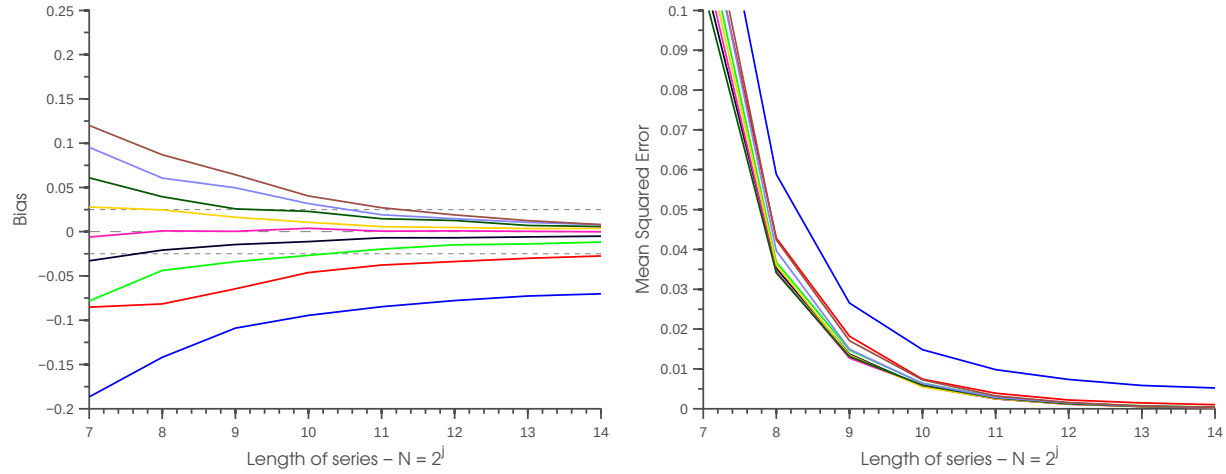
Aggregated variance - Figures (5.3e,5.7c),(5.3f,5.7d),(A.3c,A.3d)
The estimates is for this method is almost always underestimating the true value of H at small sample sizes, except for $H \in \{0.1, 0.2\}$. Apart from this underestimation the method seems to be asymptotically unbiased for $N > 2^{12}$ for $H \in (0.3, 0.7)$ and for large samples the standard deviations is converge to zero. As follows this method is consistent for large N and $H \in (0.3, 0.7)$. The standard deviations are at large sample sizes independent of H and they reach a low level at $N = 2^{11}$. In the MSE plot, it is confirmed that the estimator performs well for $H \in (0.3, 0.7)$

Exact local Whittle - Figures (5.4a,5.7e),(5.4b,5.7f),(A.3e,A.3f)
The estimator are close to unbiased for all levels of H, except $H \in (0.1, 0.3)$ given a FGN series. For large sample sizes, the same levels are asymptotically unbiased. This is reflected in the MSE plot where $H = 1$ is separated from all the other MSE's. In fact, it has nearly double the MSE for small sample sizes. For ARFIMA series all the estimates have a small bias (± 0.02) for series shorter than $N = 2^9$. For $N > 2^9$ the series is asymptotically unbiased. When estimating both types of series, the estimates are consistent as the standard deviation go to zero for increasing sample sizes.

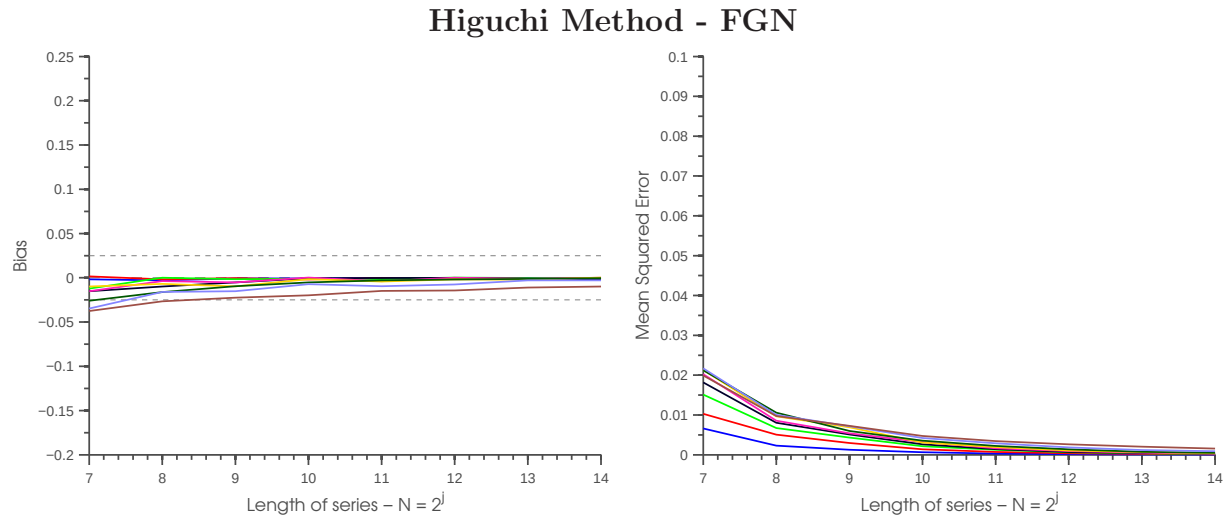
Figure 5.1: Color legend for different intensity levels of long range dependence.



(a) Bias (b) MSE

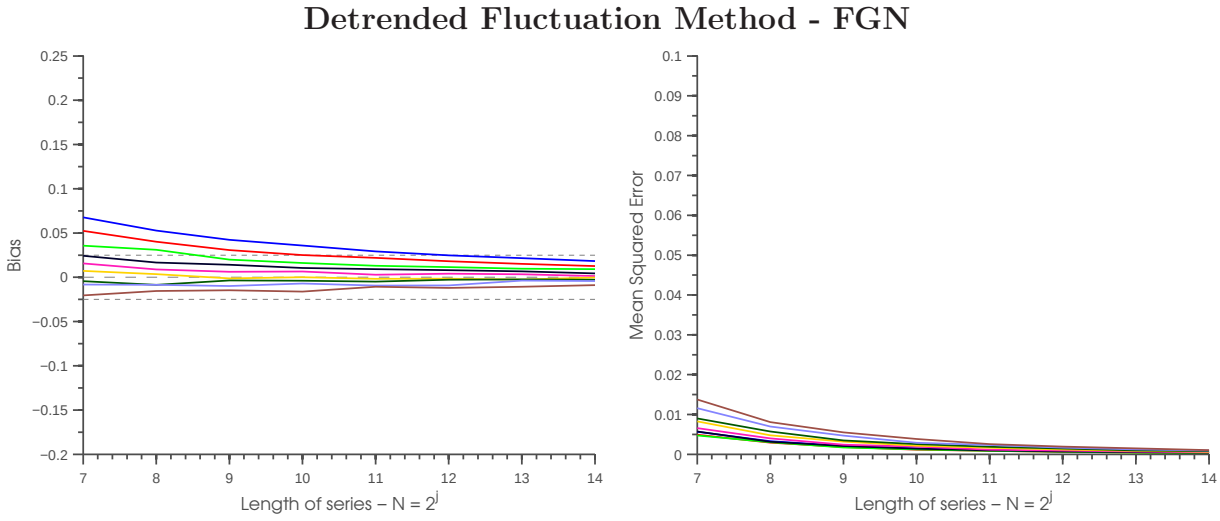


(c) Bias (d) MSE

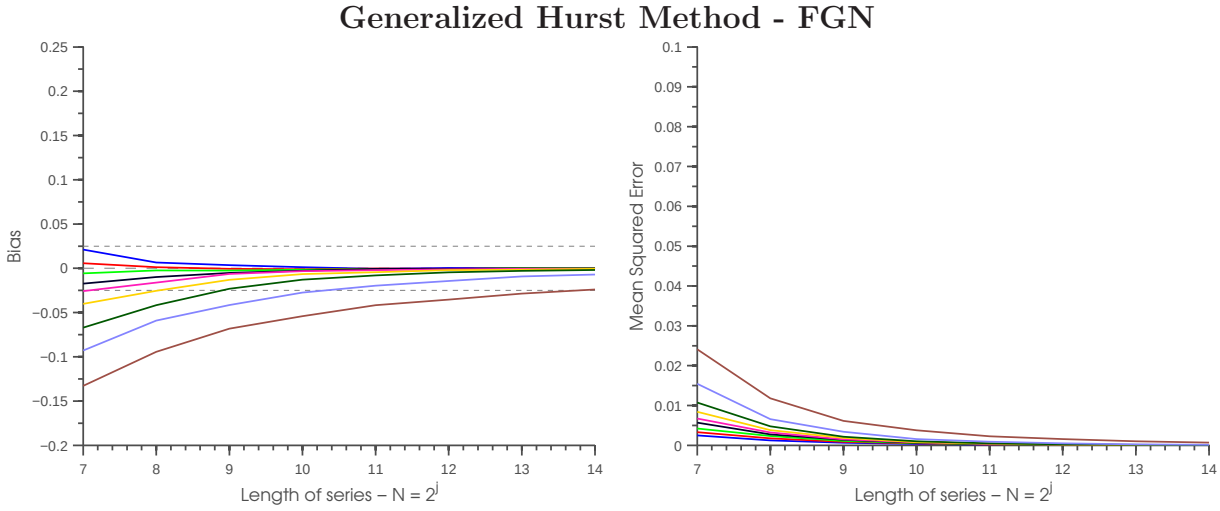


(e) Bias (f) MSE

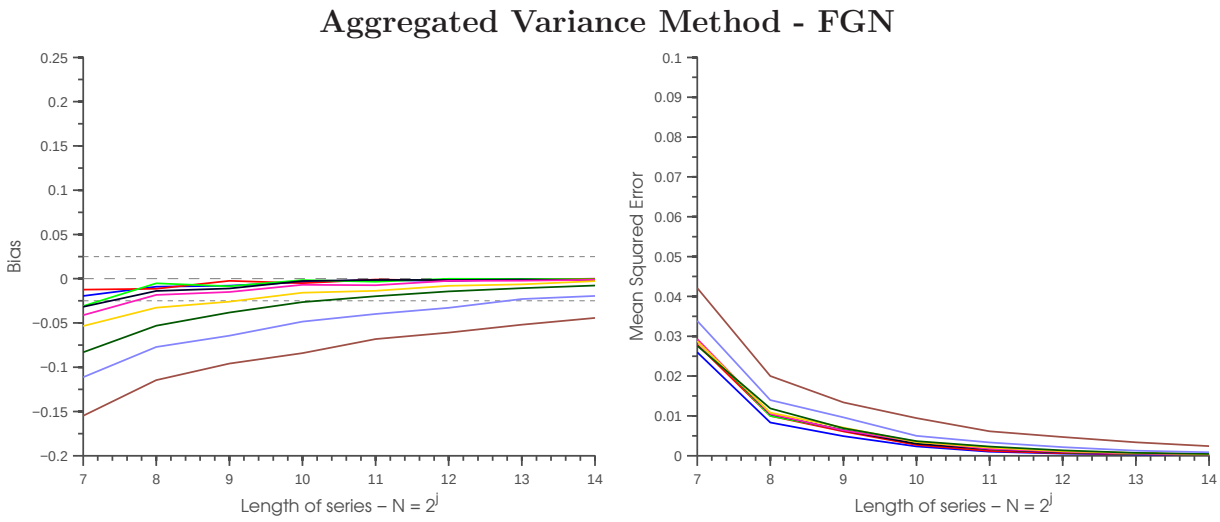
Figure 5.2: Bias and MSE for estimated FGN series.



(a) Bias (b) MSE

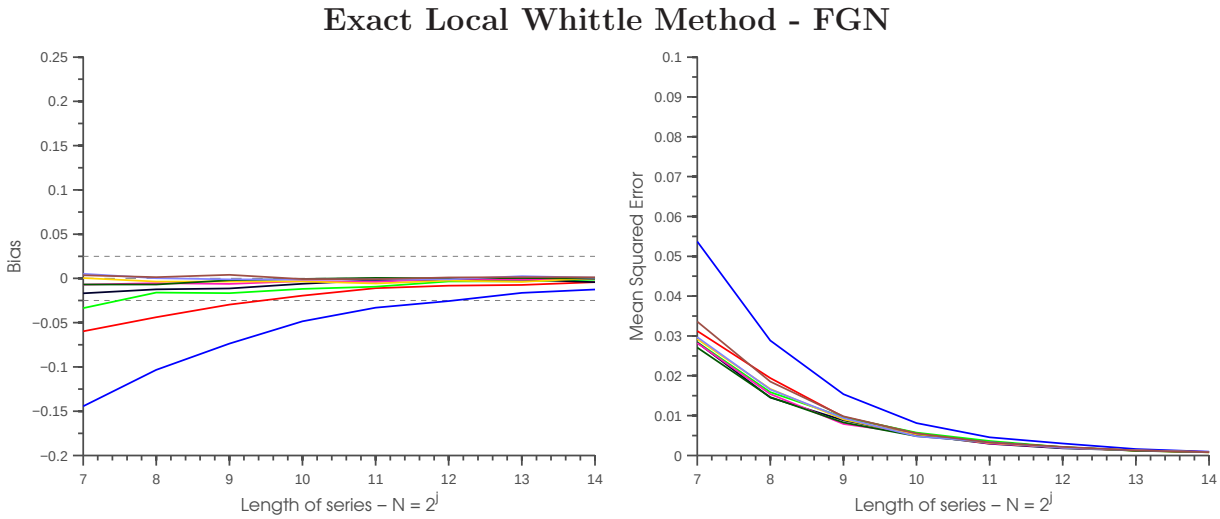


(c) Bias (d) MSE



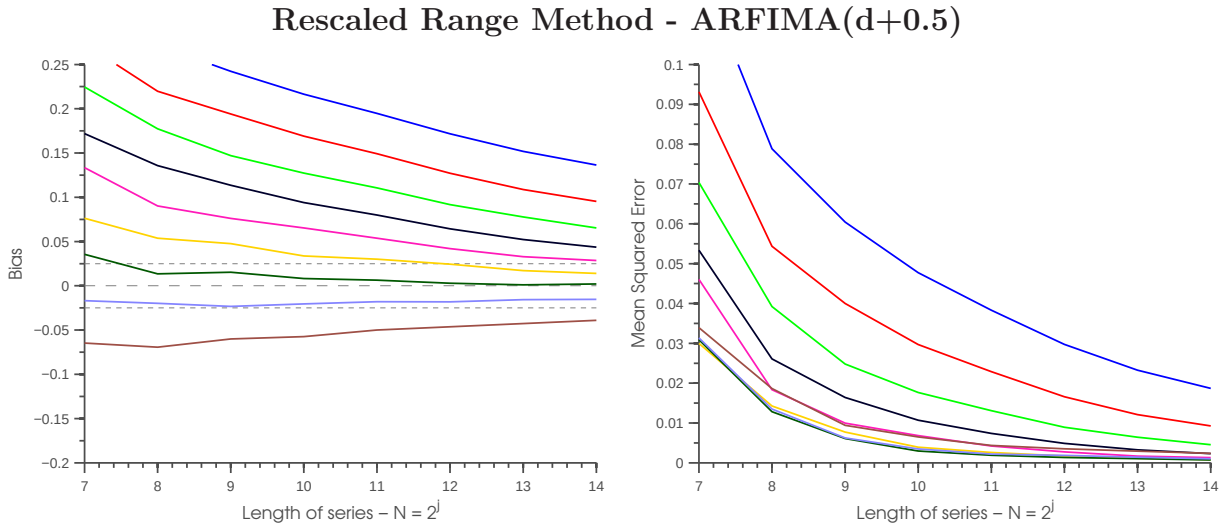
(e) Bias (f) MSE

Figure 5.3: Bias and MSE for estimated FGN series.



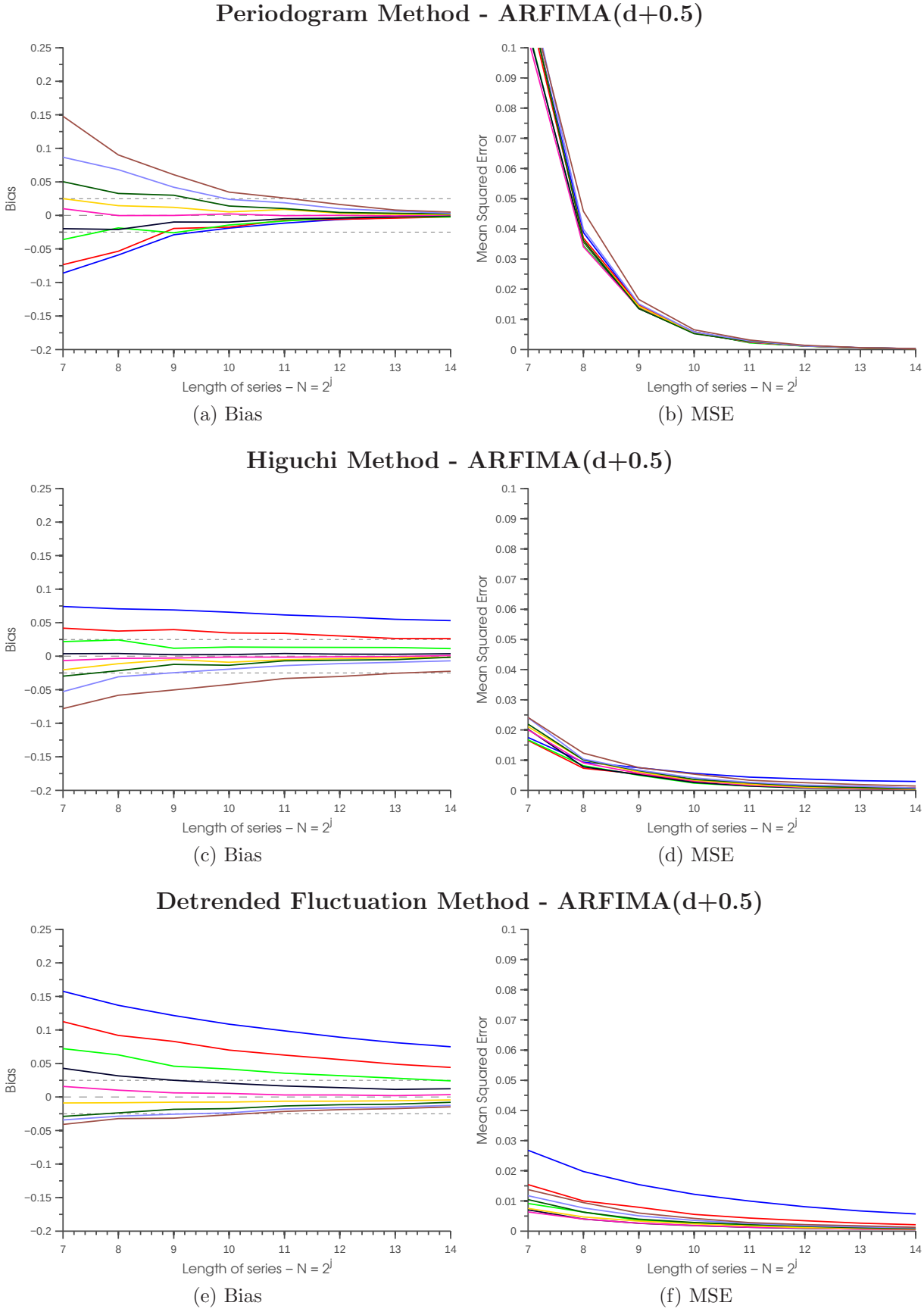
(a) Bias (b) MSE

Figure 5.4: Bias and MSE for estimated FGN series.



(a) Bias (b) MSE

Figure 5.5: Bias and MSE for estimated ARFIMA noise series.

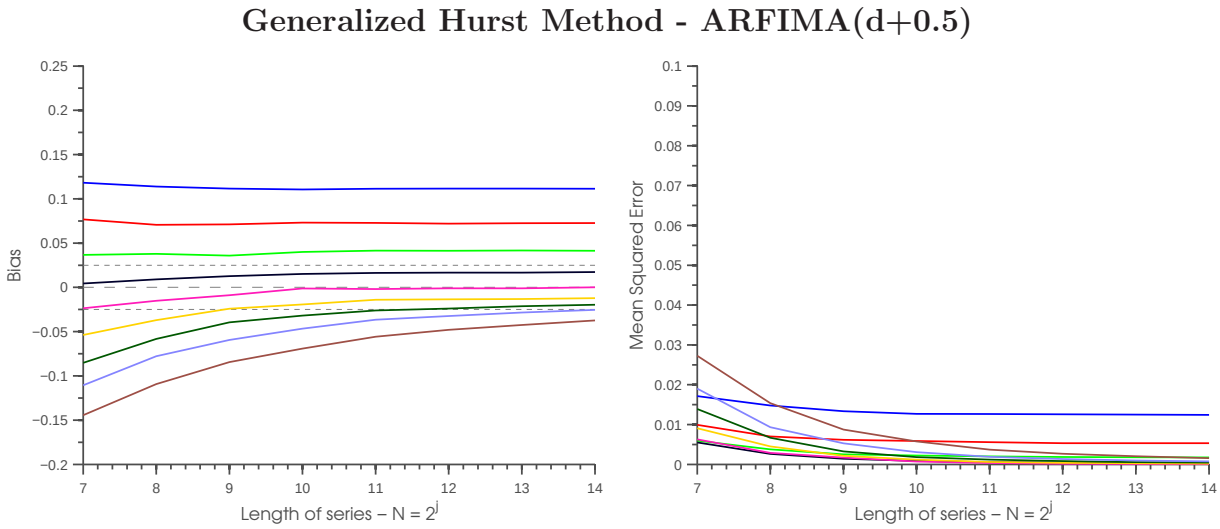


(a) Bias (b) MSE

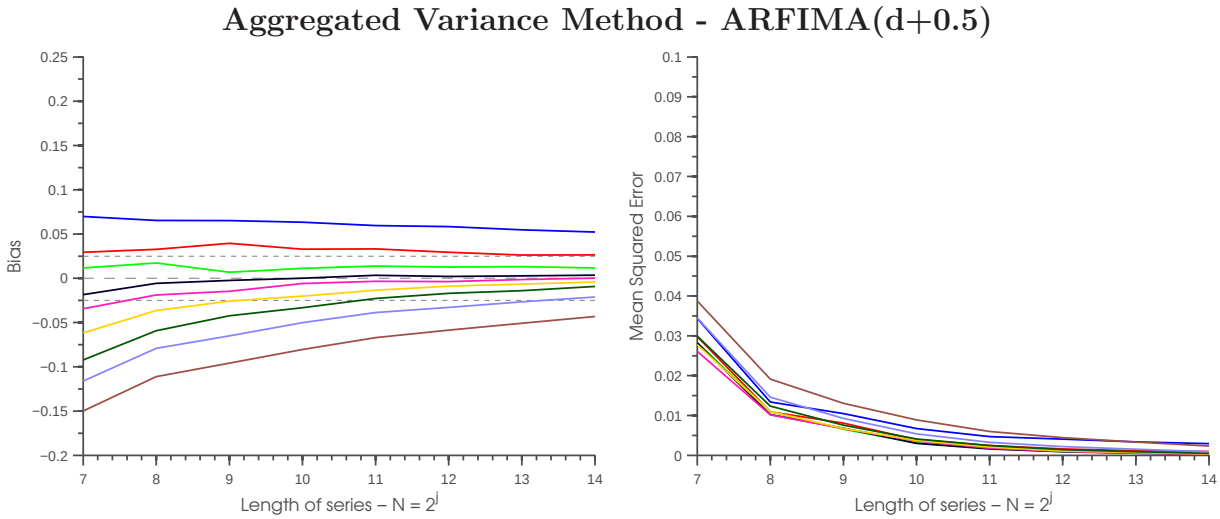
(c) Bias (d) MSE

(e) Bias (f) MSE

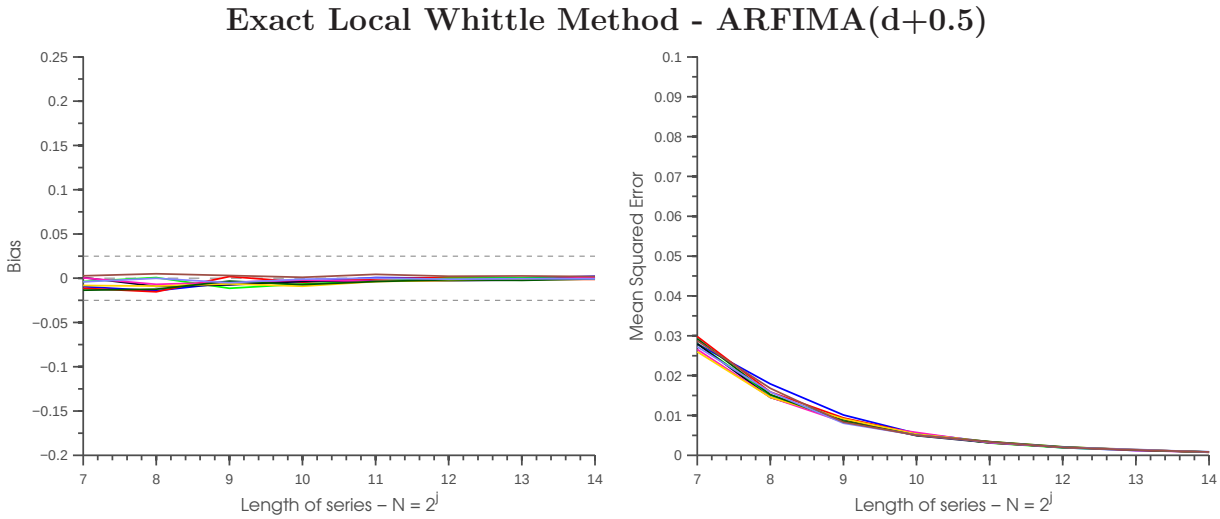
Figure 5.6: Bias and MSE for estimated ARFIMA noise series.



(a) Bias (b) MSE



(c) Bias (d) MSE



(e) Bias (f) MSE

Figure 5.7: Bias and MSE for estimated ARFIMA noise series.

Chapter 6

Robustness assessment

This chapter aims to evaluate the effects of applying a heavy tailed noise to FGN and ARFIMA noise. First the noise is described and then the corruptive effects on LRD series are described. Finally, the performance of the estimators under corrupted signals is evaluated.

6.1 Corruption of the LRD series

Research has shown that observed financial time series will have properties different than those of fractional Brownian motion or ARFIMA signals [15]. It is often observed that returns will have a large portion of its realizations close to zero and have outliers [43] and [10]. Such a distribution is similar to what is seen in the *Student's t distribution*. In the following segment the performance of the estimators under the influence of a noise coming from a t distribution will be examined.

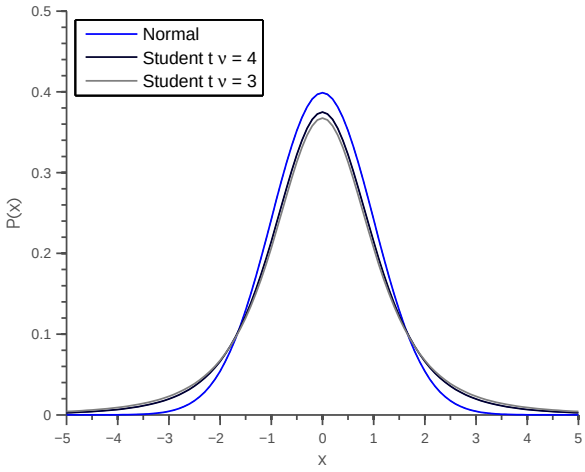
Student's t distribution

The Student's t probability distribution is given as

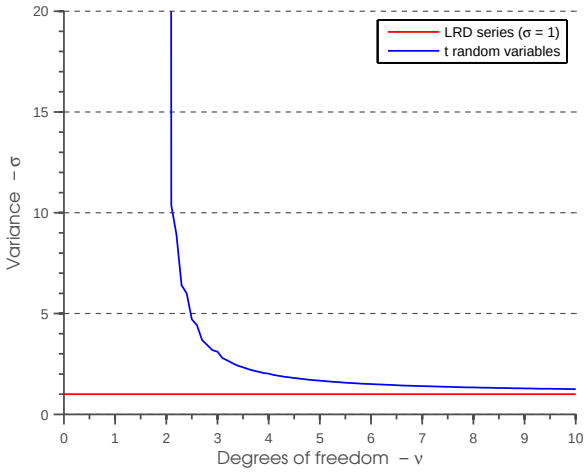
$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \tag{6.1.1}$$

where $\Gamma(\cdot)$ is the Gamma function and ν is the degrees of freedom. For low values of ν the distribution will have heavy tail whereas it approximates the normal distribution as $\nu \to \infty$.

As seen in Figure 6.1b for values of $\nu < 10$ the variance gets large and will have extreme outliers for $\nu \approx 2$ which is not applicable in this context. Also, when comparing a normal distribution with a t distribution the difference in tails is apparent as in Figure 6.1a.



(a) Gaussian pdf. vs Student's t pdf.

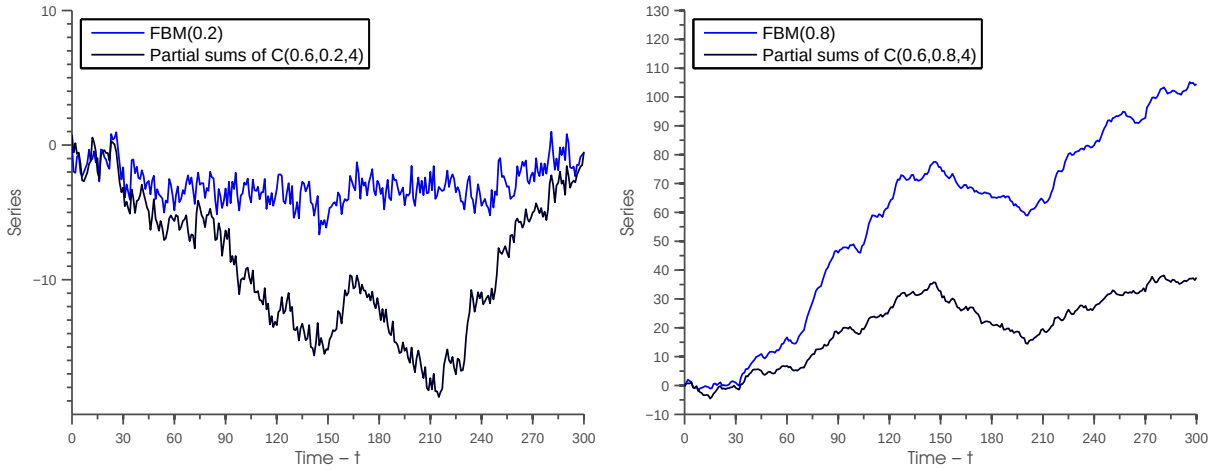


(b) Variance for levels of ν compared to LRD series.

Figure 6.1: Comparison of Student's t pdf. and normal pdf.

Moment	$\mathcal{N}(\mu, \sigma^2)$	$T(\nu)$	
Mean	μ	0 for $\nu > 1$	o.w. undefined
Variance	σ^2	$\frac{\nu}{\nu-2}$, ∞ for $1 < \nu \leq 2$	o.w. undefined
Skewness	0	0 for $\nu > 3$	o.w. undefined
Kurtosis	3	$\frac{3(\nu-2)}{\nu-4}$ for $\nu > 4$	o.w. undefined

Table 6.1: Comparison of Student's t and normal distribution.



(a) FGN(H=0.2),T($\nu = 4$), $\alpha = 0.6$ (b) FGN(H=0.8),T($\nu = 4$), $\alpha = 0.6$

Figure 6.2: Corrupting two FBM series with T($\nu = 4$) and $\alpha = 0.6$.

When using the information from Table 6.1 a t distribution $T(4)$ will have mean 0, variance 2, skewness 0 and an undefined kurtosis. In Matlab, when generating 2^{14} random numbers from a normal- and t distribution the minimum and maximum values were $[-3.87, 3.76]$ and $[-18.45, 10.32]$. These numbers confirm that the t distribution has heavy tails compared to the normal distribution.

6.1.1 Corrupting a series

The previously generated time series (FGN, ARFIMA) all have standard deviations equal to one. When adding random variables from a heavy-tailed t distribution, the combined time series will inherit some of these properties as well. A corrupted series will be a mix of a LRD series denoted LRD(H) where LRD are either FGN or ARFIMA, and a noise series denoted T(ν) which is a t distributed noise with ν degrees of freedom. The corrupted signal C will be defined by its components and the value of the signal strength variable α which determines the proportion of noise and LRD series. The relationship is given as

$$C(\alpha, H, \nu)_t = \alpha LRD(H)_t + (1 - \alpha)T(\nu)_t \tag{6.1.2}$$

where $LRD(H)$ is the long range dependent time series with fractal exponent H and T is the series of Student t random variables with ν degrees of freedom. The factor α will be named the *mixing factor*. The term *signal* regarding the LRD series will be used as opposed to noise which has been defined as Student t random variables. The partial sums of adding noise to a persistent series as opposed to a antipersistent series can be seen in Figure 6.2. Also observed, are the corrupted series in Figures 6.2a and 6.2b have a appearance closer to

an independent series compared with the FBM.

6.1.2 Signal-to-noise ratio

In evaluating the corrupted series, it would be worthwhile to state something about the relationship between the signal and the noise. The *signal-to-noise ratio* (SNR) describes the portion of variance in the signal to the portion of variance in the noise. The signal-to-noise ratio is the relationship between the signal power (P_S) and the noise power (P_ψ) defined as

$$SNR_\alpha = \frac{\alpha P_S}{(1-\alpha)P_\psi}. \tag{6.1.3}$$

where $\alpha \in (0,1)$ is the mixing factor determining the allocation of noise and signal. In the case of a zero mean stationary stochastic process, its power is given as the value of the correlation function $\gamma(s)$ at the origin which is equal to its variance yielding the relationship

$$SNR_\alpha = \frac{\alpha^2 \mathbb{V}[S]}{(1-\alpha)^2 \mathbb{V}[\psi]}. \tag{6.1.4}$$

In the literature on signal processing it is common to denote the SNR in terms of decibels (dB). Throughout the thesis all SNR's will be denoted in decibels and the transformation from nominal value to dB is

$$10 \log_{10}(SNR_\alpha) = SNR(dB)_\alpha. \tag{6.1.5}$$

As seen in Figure 6.3 a t distributed noise with $\nu = 3$, degrees of freedom leads to corrupted series which are dominated by the noise in terms of variance. For a higher degree of freedom $\nu = 4$ the SNR are more evenly distributed among the different mixing factors (as the variance of the noise and the signal are more equal). Figure 6.3 is based on the mean SNR for 1000 corrupted series. The intensity of LRD does not change the variance of the process in the generators used here so it is constant at variance $\sigma^2 = 1$. Throughout the corruption study the preferred added noise is $T(\nu = 4)$ with average signal-to-noise ratio $SNR \in (0, 38)$ for $\alpha \in (0.6, 1)$.

6.1.3 Adding a noise

Previously the noise component in Equation (6.1.1) has been introduced. Beneficial to describing the performance of the estimators under different levels of noise four plots were made. The estimation was performed on a corrupted series of altering mixture of LRD signal

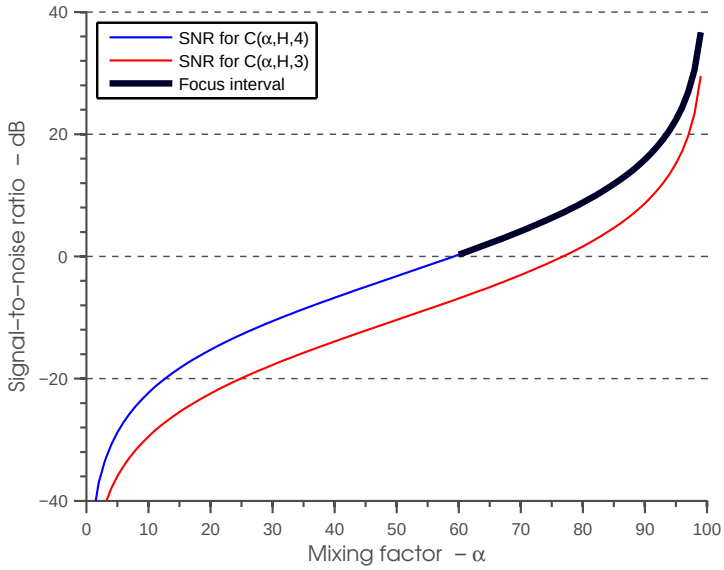


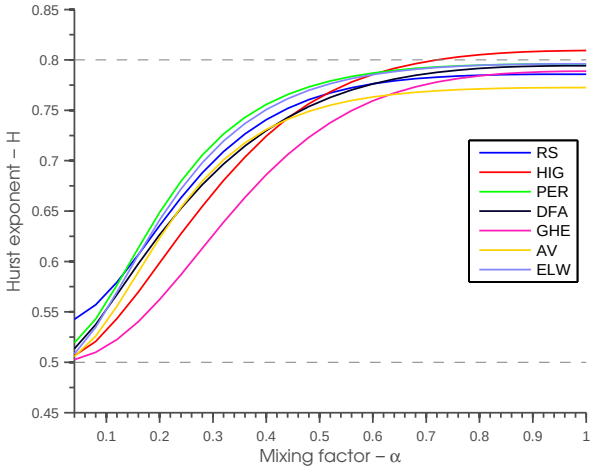
Figure 6.3: Average signal-to-noise ratio based on 1000 series of $C(\alpha, H, \nu)$ for varying α and different degrees of freedom ν.

and heavy tailed signal represented by Student t random variables. The results are seen in Figure 6.4 and Figure 6.5 where it is clear that there are differences between the estimators in the ability to deal with noise. The estimators ability do deal with noise will be addressed later in the chapter as 6.4 and Figure 6.5 is meant for illustrative purposes. The figures are also based on the same results that will be covered next.

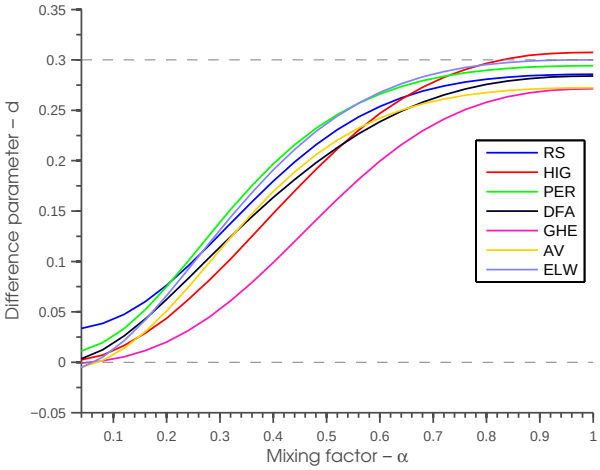
An experiment was carried out where the aim was to measure the estimators sensitivity to noise by applying them to the corrupted series $C(\alpha, H, \nu)$ for different sample sizes. The experiment was performed at levels of $\alpha \in \{1, 0.9, \ldots, 0.6\}$, for intensities of LRD $H \in \{0.1, 0.2, \ldots, 0.9\}$ and for sample sizes N^j for $j \in 8, 10, 12, 14$. For each estimate 500 replications was estimated. The results are given in Appendix A.2. In the results are commented in the next section and summarized in terms of breakdown points in Table 6.2 and Table 6.3

6.1.4 Breakdown points

It is possible to visually detect a *breakdown point* of an estimator by using the tables in Appendix A.2. By arbitrarily setting a limit for which the estimates of the corrupted series are deviating from the estimates of the uncorrupted series, it is possible to determine the breakdown at this limit. The finite breakdown point of an estimator is formally defined as the minimal fraction of deviant observations rendering the estimate meaningless [26]. In

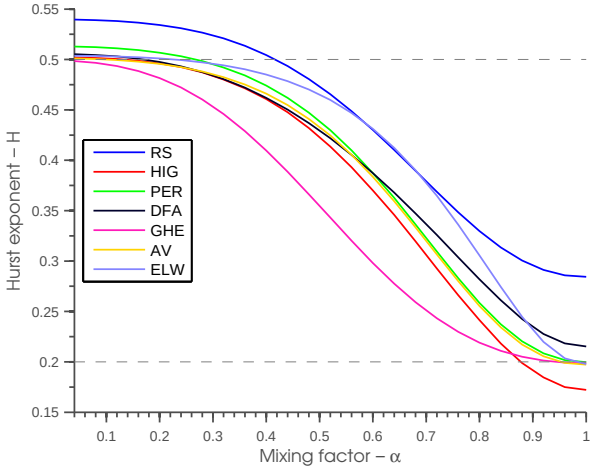


(a) FGN($H = 0.8$), T($\nu = 4$), N $= 2^{14}$.

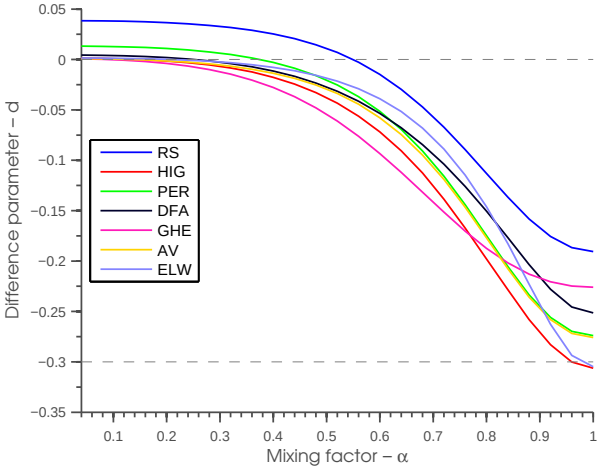


(b) ARFIMA($d = 0.3$), T($\nu = 4$),N$= 2^{14}$.

Figure 6.4: Mean estimates of H based on 500 replications for different levels of $\alpha \in (0, 1]$.



(a) FGN($H = 0.2$), T($\nu = 4$), N $= 2^{14}$.



(b) ARFIMA($d = -0.3$), T($\nu = 4$),N$= 2^{14}$.

Figure 6.5: Mean estimates of H based on 500 replications for different levels of $\alpha \in (0, 1]$.

this thesis the measure of robustness is given in terms of α, for which the estimates deviate increasingly from the uncorrupted series. In the context of Equation 6.1.2 a breakdown point with a high α (small portion of noise) suggests that the estimator is sensitive to the changes in the observations. On the other side, an estimator which breaks down at a low level of α (large portion of noise) will be less sensitive to the changes in the observations. Thus, an estimator breaking down at low levels of α is said to have a high breakdown point and vice versa. Intuition tells that a estimator cannot have a breakdown point higher than $\alpha = 0.5$. For a corrupted series with $\alpha < 0.5$ the portion of noise is larger than the portion of original signal and it makes it unobtainable to separate the signal and the noise. In terms of the problem under scrutiny the breakdown point can in this case be given as the relationship

$$BP(\alpha) = |\hat{C}(1, H, 4) - \hat{C}(\alpha, H, 4)| \leq \varphi \tag{6.1.6}$$

where $\hat{C}$ is the estimate of the corrupted series with a constant LRD intensity H, mixing factor α and where φ is a arbitrary value. In other words, for a given H, the breakpoint $BP(\alpha)$ is the level of mixing where the bias adjusted estimate deviate more than the arbitrary value φ. In the analysis below, the boundary for breakdown is set at $\varphi = 0.035$.

In Table 6.2 and Table 6.3 the breakdown points for the estimators at each length for both FGN and ARFIMA signals of varying LRD intensity, separated by persistent and anti-persistent series, is compiled. The general observation is that the estimators break down earlier at lower intensities of LRD; they are more sensitive to noise as the value of H goes down. There is rarely any notable difference in breakdown points between the FGN and ARFIMA signals except for the case of $H = 0.2$ where all estimators except the ELW break down earlier for FGN signals. Also, for persistent LRD signals, the breakdown points are more affected by the sample size than for anti-persistent signals. It was decided not to include the changes in standard deviations for $\hat{C}$ since they are outside of the scope of this discussion. The only pattern in the changes of standard deviation is that the standard deviations were larger for smaller samples, which is also known from the baseline case.

Rescaled range - Corrupted

The RS method has the one of the smallest deviations from the pure LRD signal for increasing levels of α among the estimators, i.e. it has robust properties. This is seen for all lengths N and for both LRD signals. This observation is opposing what is known about the estimators robustness which is given in Chapter 4. By looking at Table A.1 - A.8 the deviations are almost always the lowest which hold for different sample sizes. Also, the RS estimator shows little or no change in robustness for different sample sizes but it is more sensitive to the type of LRD signal than the other estimators.

Periodogram - Corrupted

The periodogram is ranked at the bottom in terms of robustness among the estimators. It is only showing signs of robust properties in the persistent case at $H \in \{0.6, 0.9\}$ which is only at large sample sizes. This is shown in Table 6.3. Also by looking at Table A.1 - A.16 it also shows that the deviations stand out from the rest. In the anti-persistent case, the periodogram breaks down at α-levels generally higher than the others but is invariant of the type of LRD signal underlying the estimation.

Higuchi - Corrupted

As with the periodogram estimator the Higuchi estimator shows weaknesses in terms of robustness in the case of anti-persistent series. Although it seems to break down earlier than the other estimators in this case, the difference is not as significant as with the periodogram estimator. In the case of persistent series, the estimator performs the best compared to the rest, in terms of smallest deviation for increasing levels of α. In particular, the Highuchi estimator has the most instances of breakdown points at $\alpha < 0.6$ given $H \geq 0.6$. In addition, the estimator shows little sensitivity with respect to the length of the time series and to the type of LRD signal.

Detrended Fluctuation Analysis - Corrupted

The DFA method correlates with the Higuchi method in terms of breakdown points; it breaks down early for the anti-persistent series and shows robustness tendencies in the case of persistent series. The method breaks down at $\alpha < 0.6$ for all lengths and both LRD signals given $H = 0.6$ as the intensity of LRD. For all the other persistent series, i.e. LRD$H > 0.6$, the lowest breakdown point of the estimator is $\alpha = 0.6$. The DFA method is sensitive to the sample sizes but is unaffected by the type of LRD series.

Generalized Hurst - Corrupted

The GHE method has the lowest breakdown points in terms of α for the anti-persistent series. It has the highest breakpoint at $\alpha = 0.8$ for LRD series with $H = 0.1$ which is exclusive in comparison with the other estimators. For the persistent LRD series, the breakdown points of this method are generally higher compared to the ones with the lower breakpoints. There are also few instances where the breakpoint is less than $\alpha = 0.6$. The breakdown points of the GHE estimator are insensitive to the type of LRD series and the sample sizes.

Aggregated variance - Corrupted

The aggregated variance estimator shows signs of breaking down at a high level of α given anti-persistent series of LRD signals. The breakdown points are mediocre relative to the rest of the estimators. For persistent series, the estimator has breakdown points that are among

the lowest for all levels of $H \geq 0.6$. The AV method is insensitive to sample sizes and LRD signal type in the anti-persistent case but for persistent LRD series it is sensitive to sample sizes in terms of breakdown point.

Exact local Whittle - Corrupted

When estimating corrupted anti-persistent series, the exact local Whittle estimator gives the poorest results in terms of breakdown point, compared to all the other estimators. It breaks down for every level of LRD intensity and sample size at a level of α equal to or larger than the other estimators. Opposing the results from the anti-persistent signals, presented with corrupted persistent signals the breakdown points all are at level with the others. In particular, the breakdown points are all at $\alpha \lesssim 0.6$ for both FGN and ARFIMA signals at The ELW estimator is insensitive to type of LRD signal and sample size.

6.1.5 A closer look at breakdown levels

Taking a closer look at Table 6.2 and 6.2 it is apparent that the persistence is a deciding factor regarding the breakdown of the estimator. When estimating a corrupted LRD series with $H = 0.1$ all of the the estimators break down at levels of $\alpha > 0.8$ as opposed to the estimation of a LRD series with $H = 0.6$ where all estimators break down at $\alpha < 0.6$.

It was shown in Chapter 5 that, in general, the estimators perform worse for series with high level of anti-persistence than for series with heavy persistence when no noise is added.

As the t distributed noise used here is independently distributed, and therefore has no memory, its theoretical intensity of LRD is equal to the one of a $LRD(H = 0.5)$ series. When estimating a $LRD(H = 0.5)$ series mixed with a $LRD(H \neq 0.5)$ series the expected result is that the estimate is reflected in the portion of the series that are mixed. For a large portion of $LRD(H = 0.5)$ the estimates will be close to 0.5 and vice versa. This is not the result in this study. If it was then the breakdown point for $C(\alpha, 0.3, 4)$ and $C(\alpha, 0.7, 4)$ would be equal as for $C(\alpha, 0.1, 4)$ and $C(\alpha, 0.9, 4)$. Figure 6.6 justifies the statement that in general, the performance is worse when estimating a corrupted anti-persistent series.

The difference in breakdown levels can easily be seen when comparing the plots in Figure 6.6. Given 500 replications for 25 levels[1] of α the estimates were applied to corrupted series $C(\alpha, H, 4)$. Then the series was adjusted for its uncorrupted estimate at $\alpha = 1$ by subtracting the estimate at $\alpha = 0.1$ from the entire series, i.e. the breakdown point $BP(\alpha)$ is computed. The series were multiplied by -1 so that deviations appear as negative numbers which represents the break*down* in a better way. The axes are different in this case to aid the identification of each estimate. Comparing Figure 6.6 with the results in the previous

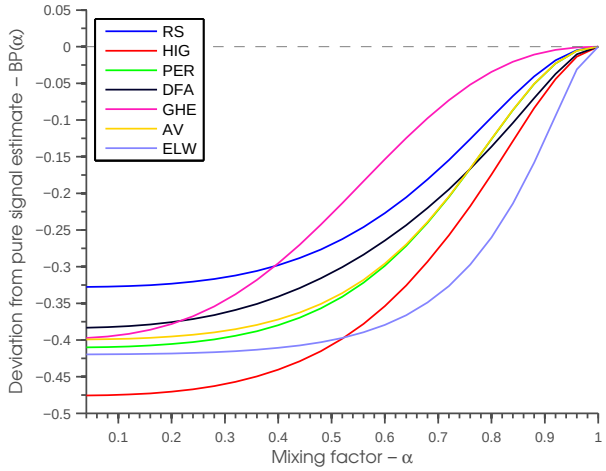
[1] As only 25 levels of α was computed, excluding $\alpha = 0$, the first point of each plot is at $\alpha = 100/25 = 4$.

		RS	PER	HIG	DFA	GHE	AV	ELW
		Breakdown point in terms of α.						
FGN(H = 0.1) ARFIMA(d = -0.4)	2^8	0.8	0.9	0.9	0.8	0.8	0.9	0.9
		0.8	0.9	0.8	0.8	0.8	0.8	0.9
	2^{10}	0.9	0.9	0.9	0.9	0.8	0.9	0.9
		0.8	0.9	0.9	0.8	0.8	0.9	0.9
	2^{12}	0.9	0.9	0.9	0.9	0.8	0.9	0.9
		0.8	0.9	0.9	0.9	0.8	0.9	0.9
	2^{14}	0.9	0.9	0.9	0.9	0.8	0.9	0.9
		0.9	0.9	0.9	0.9	0.8	0.9	0.9
FGN(H = 0.2 ARFIMA(d = -0.3	2^8	0.8	0.9	0.8	0.8	0.8	0.8	0.9
		0.7	0.8	0.8	0.8	0.7	0.8	0.9
	2^{10}	0.8	0.9	0.8	0.8	0.8	0.8	0.9
		0.8	0.8	0.8	0.8	0.8	0.8	0.9
	2^{12}	0.8	0.9	0.9	0.9	0.8	0.9	0.9
		0.8	0.8	0.8	0.8	0.8	0.8	0.9
	2^{14}	0.9	0.9	0.9	0.9	0.8	0.9	0.9
		0.8	0.8	0.8	0.9	0.8	0.8	0.9
FGN(H = 0.3) ARFIMA(d = -0.2)	2^8	0.7	0.8	0.8	0.7	0.8	0.8	0.8
		0.7	0.8	0.8	0.7	0.7	0.8	0.8
	2^{10}	0.7	0.8	0.8	0.8	0.7	0.8	0.8
		0.7	0.8	0.8	0.7	0.7	0.8	0.8
	2^{12}	0.8	0.8	0.8	0.8	0.8	0.8	0.8
		0.7	0.8	0.8	0.8	0.7	0.8	0.8
	2^{14}	0.8	0.8	0.8	0.8	0.8	0.8	0.8
		0.8	0.8	0.8	0.8	0.7	0.8	0.8
FGN(H = 0.4) ARFIMA(d = -0.1)	2^8	0.6	0.8	0.6	0.6	0.6	0.6	0.6
		•	0.7	0.6	0.6	0.6	0.6	0.7
	2^{10}	0.6	0.7	0.7	0.6	0.6	0.7	0.7
		•	0.7	0.7	0.6	0.6	0.6	0.7
	2^{12}	0.6	0.7	0.7	0.7	0.6	0.7	0.7
		0.6	0.7	0.7	0.7	0.6	0.7	0.7
	2^{14}	0.7	0.7	0.7	0.7	0.6	0.7	0.7
		0.6	0.7	0.7	0.7	0.6	0.7	0.7

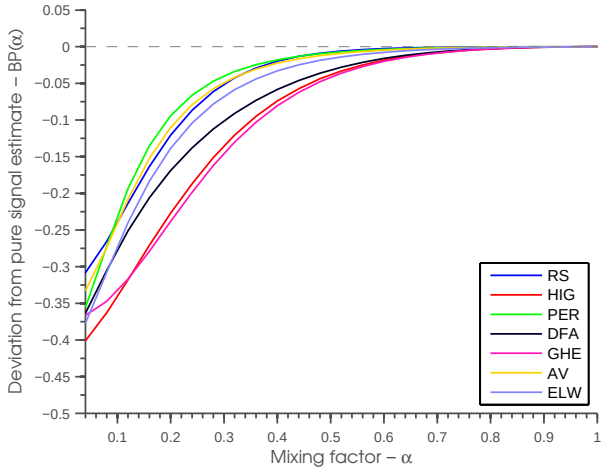
Table 6.2: Breakdown points where the breakpoint is set as a deviation larger than 0.035 from the pure signal. For example, a breakdown at 0.6 means that the estimator broke down in the interval $\alpha \in (0.6, 0.7)$. The symbol • indicates a breakdown level at $\alpha < 0.6$.

		RS	PER	HIG	DFA	GHE	AV	ELW
		Breakdown point in terms of α.						
FGN(H = 0.6 ARFIMA(d = 0.1)	2^8	•	0.6	•	•	0.6	•	0.6
		•	0.6	•	•	•	•	•
	2^{10}	•	0.6	•	•	0.6	•	•
		•	0.6	•	•	•	•	•
	2^{12}	•	•	•	•	0.6	•	•
		•	0.6	•	•	0.6	•	•
	2^{14}	•	•	•	•	0.6	•	•
		•	0.6	•	•	0.6	•	•
FGN(H = 0.7 ARFIMA(d = 0.2)	2^8	0.6	0.6	0.6	0.6	0.6	0.6	0.6
		0.6	0.6	0.6	0.6	0.6	0.6	0.6
	2^{10}	0.6	0.6	0.6	0.6	0.6	0.6	0.6
		0.6	0.6	0.6	0.6	0.6	0.6	0.6
	2^{12}	0.6	0.6	0.6	0.6	0.7	0.6	0.6
		0.6	0.6	0.6	0.6	0.6	0.6	0.6
	2^{14}	•	0.6	0.6	0.6	0.7	0.6	•
		•	0.6	0.6	0.6	0.6	0.6	•
FGN(H = 0.8) ARFIMA(d = 0.3)	2^8	0.7	0.7	0.6	0.7	0.7	0.6	0.7
		0.6	0.7	0.7	0.7	0.7	0.7	0.7
	2^{10}	0.6	0.7	0.6	0.7	0.7	0.6	0.6
		0.6	0.6	0.6	0.6	0.7	0.6	0.6
	2^{12}	0.6	0.6	•	0.6	0.7	0.6	0.6
		0.6	0.6	•	0.6	0.6	•	0.6
	2^{14}	•	0.6	0.6	0.6	•	•	•
		•	0.6	•	0.6	0.6	•	•
FGN(H = 0.9) ARFIMA(d = 0.4)	2^8	0.7	0.7	0.6	0.7	0.7	0.6	0.7
		0.6	0.7	0.6	0.7	0.7	0.6	0.7
	2^{10}	0.6	0.7	0.6	0.7	0.7	0.6	0.7
		0.6	0.7	•	0.7	0.7	•	0.6
	2^{12}	0.6	0.6	•	0.7	0.7	0.6	0.6
		•	0.6	•	0.6	0.6	•	•
	2^{14}	•	0.6	•	0.6	0.6	•	0.6
		•	•	0.6	0.6	•	0.6	•

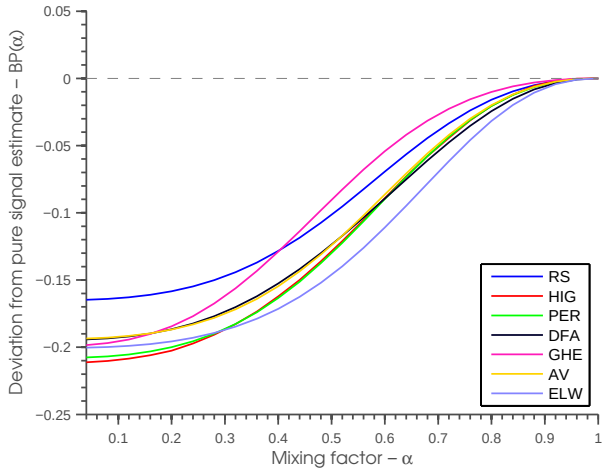
Table 6.3: Breakdown points where the breakpoint is set as a deviation larger than 0.035 from the pure signal. For example, a breakdown at 0.6 means that the estimator broke down in the interval $\alpha \in (0.6, 0.7)$. The symbol • indicates a breakdown level at $\alpha < 0.6$.



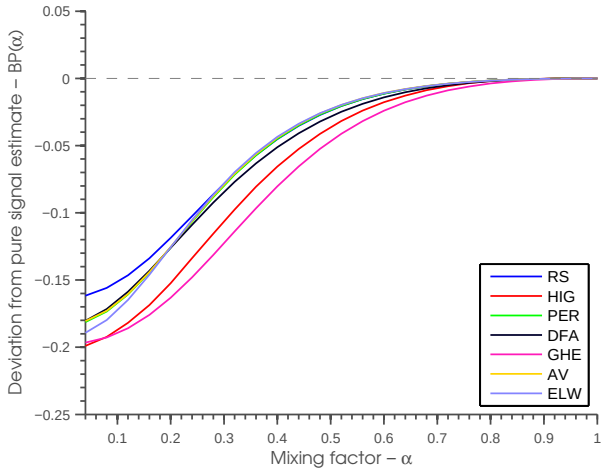
(a) $BP(\alpha)$ for $C(\alpha, 0.1, 4)$.



(b) $BP(\alpha)$ for $C(\alpha, 0.9, 4)$.



(c) $BP(\alpha)$ for $C(\alpha, 0.3, 4)$.



(d) $BP(\alpha)$ for $C(\alpha, 0.7, 4)$.

Figure 6.6: Breakdown points $BP(\alpha)$ for different levels of $\alpha \in (0, 1]$.

section it can be seen that there is a pattern in the breakdown points when comparing the anti-persistent C in Figures 6.6a and 6.6c with the persistent C in Figure 6.6d and 6.6b. In the anti-persistent case the GHE method have a breakdown point distinctively higher than the other estimators. On the contrary, the ELW method has the lowest breakdown point. This is the case for both anti-persistent series. In the persistent case the periodogram performs well in estimating both of the corrupted, persistent, series as opposed to the GHE and HIG which comparatively breaks down sooner than the other estimators. In addition, it can be observed that the differences in breakdown levels are smaller for a persistent series. In Figure 6.6 the comparison displays this. For the antipersistent series the estimators are more separated in terms of breakdown.

	Asym. Unbiased	***Unbiased***	***Small N bias***	***Consistent***
RS	$H \in \{0.7\}$	N	Y	$H \in \{0.7\}$
PER	$H \in (0.5, 0.7)$	$H \in \{0.5\}$	$H \notin \{0.5\}$	$H \in (0.5, 0.7)*$
HIG	$H \in (0.1, 0.8)$	$H \in (0.1, 0.2)$*	$H \notin (0.1, 0.2)*$	$H \in (0.1, 0.3)*$
DFA	$H \in (0.5, 0.6)*$	N*	Y*	$H \in (0.5, 0.6)$
GHE	$H \in (0.1, 0.7)$	$H \in (0.2, 0.3)*$	$H \notin (0.2, 0.3)$	$H \in (0.1, 0.7)$
AV	$H \in (0.1, 0.3)$	N	Y	$H \in (0.1, 0.3)*$
ELW	$H \in (0.3, 0.9)$	$H \in (0.6, 0.8)$	N	$H \in (0.3, 0.9)$

Table 6.4: Results contributing to the known finite samples, given a FGN signal. A $*$ indicates a difference from previous research.

	Asym. Unbiased	***Unbiased***	***Small N bias***	***Consistent***
RS	$H \in \{0.7\}$	N	Y	$H \in \{0.7\}$
PER	$H \in (0.5, 0.7)$	$H \in \{0.5\}$	$H \notin \{0.5\}$	$H \in (0.5, 0.7)*$
HIG	$H \in (0.4, 0.7)$	$H \in \{0.4\}$	$H \notin \{0.4\}$	$H \in (0.4, 0.7)*$
DFA	$H \in 0.5$	N*	Y*	$H \in \{0.5\}$
GHE	$H \in \{0.5\}$	N*	Y*	$H \in \{0.5\}$
AV	$H \in (0.4, 0.5)$	N	Y	$H \in (0.4, 0.5)$*
ELW	$H \in (0.1, 0.9)$	N*	N	$H \in (0.1, 0.9)$

Table 6.5: Results contributing to the known finite samples, given a ARFIMA signal. A $*$ indicates a difference from previous research.

6.1.6 Comparison with previous research

In Table 4.2 the known finite sample properties of the estimators were listed. As mentioned in Chapter 4, some of the finite sample properties in this table are based on restricted experiments such as few replications of each series or small sample sizes. However, a refined summary of finite sample properties are presented in Table 6.4.

In Table 6.4 the results are based on FGN series and in Table 6.5 the results are based on ARFIMA series. The differences between previous results and the ones from this experiment are marked with a $*$. In Table 6.4 it can be seen that the differences that are found from previous research stems from the fact that the study performed in this thesis covered 9 levels of LRD intensity and different sample sizes. When measuring consistency it was always the case , given a H, that a asymptotically unbiased estimator was also consistent. Mainly the differences from previous research was from the fact that these levels of LRD intensity had not been studied before. The results from the DFA method is contradicting what is found in [7] where DFA was shown to be unbiased. In Table 6.5 the same argumentation holds for

the differences in previous research regarding the consistency of the estimators. In addition to DFA, the results from the GHE method contradicts the ones found in [6].

6.2 Fractality of a time series

"The notion of fractality, or scaling, is an assessment further describing a sequence in terms of distributions and functions of rescaling" [15]. As many of the concepts introduced in this thesis, fractality is adapted from physics into other disciplines including economics and finance [48].

Remark 5. *In [45] fractality a process* $X(t)$, $t \in \{-\infty, \infty\}$ *is defined as* fractal *if*

$$X(ct) =_d c^{H(c)} X(t) \tag{6.2.1}$$

for a positive constant c *and for a non-negative, random function* $H(c)$ *where* $=_d$ *is representing equality in distribution. The following relationship can be stated*

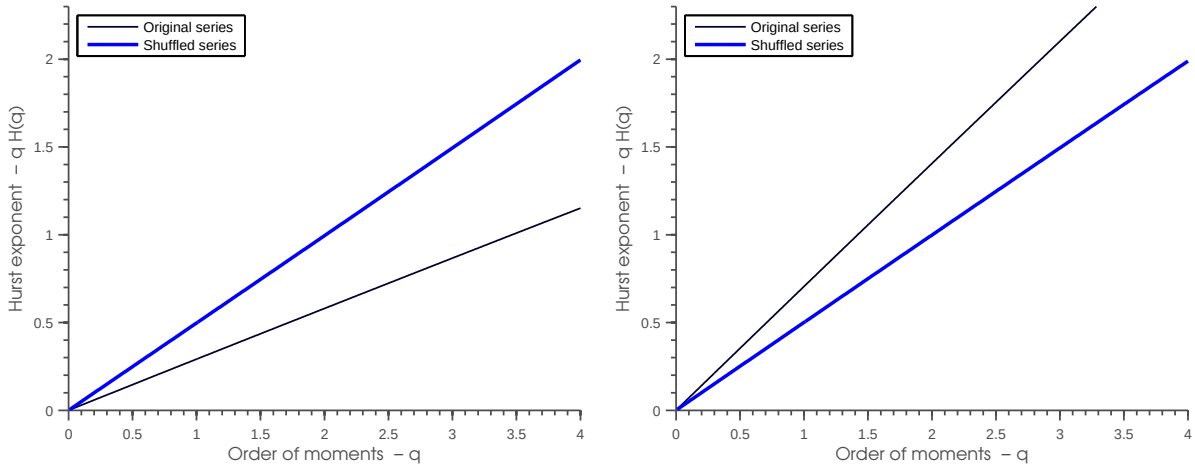
$$H(c) = \begin{cases} H & \textit{Unifractal} \\ H(q) & \textit{Multifractal} \end{cases} \tag{6.2.2}$$

In other words if $H(q)$ *is dependent on the* $q-eth$ *moment then the process is said to exhibit multifractality.*

It has been proven [6] that FGN and ARFIMA(0,d,0) are unifractal by the use of the *generalized Hurst exponent*, i.e. $H(q=1) = H(q=2) = H(q=3) = H$. It has also been shown [31] that there are two sources of multifractality in time series; the first is due to heavy tailed density functions such as a Levy distribution and the second is caused by differences in the correlations of the small and large fluctuations, such as AR noise. This is the case when extreme events are correlated on a different level than the regular events and the data may come from a finite moment distribution such as a Gaussian distribution. In the case of two sources of multifractality, a shuffling of the data will weaken the effects since it is dissolving the long range correlations.

In terms of the estimators introduced here, it is only possible to use the general Hurst exponent in order to display the fractality of a time series.

To visualize the fractal evolution of a time series (or any other sequence) it is necessary to compare the estimated intensity of long range dependence $H(q)$ for a range of q which represents the order of the moment. This is done by using the generalized Hurst exponent. In Figure 6.7 the relationship is linear. The fact that $H(q) = H$ for all q renders the series



(a) FGN(H=0.3) and its shuffled series.

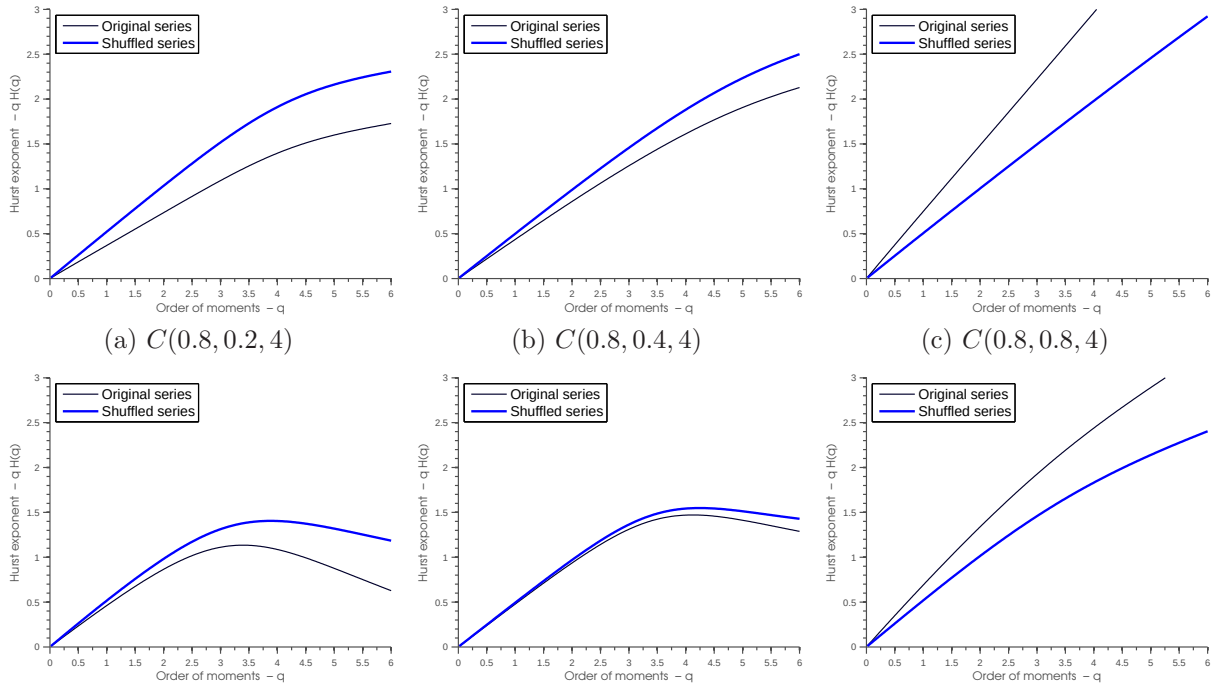
(b) FGN(H=0.7) and its shuffled series.

Figure 6.7: Fractality $qH(q)$ as a function of q for two FGN series.

unifractal. In both cases the fractality can be explained by a single variable. Notice also how the shuffled series appear (blue line); it has a slope equal to white noise or a Brownian motion.

6.2.1 Fractality of the corrupted series

As mentioned, research [31] has shown that multifractality can be caused by two features of the series: heavy tails and/or short range correlations between small and large fluctuations. Naturally, as the original series has been corrupted by a heavy tailed signal, it is necessary to investigate for multifractality. In order to decide the source of multifractality the general Hurst method will be used to display the Hurst exponent for different order of the moments. By plotting $qH(q)$ versus q the fractality will reveal itself: if the relationship is linear, the underlying series is unifractal. If it is a non-linear relationship, the underlying signal is multifractal. By comparing the series to a shuffled version of itself one can decide if the fractality stems from heavy tails or short range correlations in the fluctuations. A shuffled series should have no correlation and a non-linear relationship in this case indicates that the multifractality is explained by heavy tails. When there is difference between the shuffled and original series, this difference is explained by the correlations. In Figure 6.8, the fractality of six different corrupted series are estimated. It is clear that the more corrupted a series is, the stronger are the degree of multifractality. By comparing Figures 6.8a - 6.8c and in Figures 6.8d - 6.8f this can be seen. Another observation that can be made from Figure 6.8 is that the anti-persistent series are more affected by the noise corruption in that they deviate more from the unifractal case. For the series $C(0.6, 0.2, 4)$ in Figure 6.8a, the shuffled series and the original series has a similar slope for $q \in (0, 1.5)$. On the other hand, $C(0.6, 0.8, 4)$



(a) $C(0.8, 0.2, 4)$ (b) $C(0.8, 0.4, 4)$ (c) $C(0.8, 0.8, 4)$

(d) $C(0.6, 0.2, 4)$ (e) $C(0.6, 0.4, 4)$ (f) $C(0.6, 0.8, 4)$

Figure 6.8: Fractality of corrupted FGN series with different H and α.

in Figure 6.8c the original series exhibit a clear deviation from the shuffled series for nearly all q's, which is what is seen in the unifractal case. As the slope of the original and shuffled series in Figure 6.8a are similar, it suggest that the corrupted series is more similar to the noise than the original anti-persistent series. If the series would have been unaffected by the noise its slope would be equal to 0.2. This effect is not as strong in Figure 6.8c where the original series still has the appearance of a persistent series at $\alpha = 0.6$. This leads back to the discussion in Section 6.1.5 where it was explained that the expected slope of the noise was equal to $H = 0.5$.

Chapter 7

Empirical application

As a concluding exhibition, the estimators are applied to a set of financial time series to display some of their applications. The purpose of this section is to show some of the similarities between the corrupted signal C and the empirical data. In the application, six different financial time series are studied which all are derived from the Federal Reserve Economic Data - FRED website [54]. The data was retrieved at May 23.

DJIA The Dow Jones Industrial Average. The series ranges from 1896-05-26 to 2012-05-23 and has a sample size of $N = 29070$. The series contains daily values.

SP500 Standard and Poors 500 Index. The series contains daily values from 1957-01-02 to 2012-05-23. Sample size $N = 13947$

NORUS Norway / U.S. Foreign Exchange Rates. The series contains daily units of NOK to one USD. The series ranges from 1971-01-04 to 2012-05-18. Sample size $N = 10387$

SWEUS Sweden / U.S. Foreign Exchange Rates. Daily units of SEK to one USD. The series ranges from 1971-01-04 to 2012-05-18 and has $N = 10387$ data points.

DGS05 5-Year Treasury Constant Maturity Rate. Daily units in percentages. The data set ranges from 1962-01-02 to 2012-05-23 and has $N = 13147$ points.

DGS10 10-Year Treasury Constant Maturity Rate. The set contains daily percentage values from 1962-01-02 to 2012-05-23 and has a sample size of $N = 13147$.

The returns are computed as $r_t = p_t - p_{t-1}$ in the case of bonds as they are given in percentages and use logarithmic returns $R_t = \log(p_t) - \log(p_{t-1})$ for the other instruments.

From Table 7.1 it is clear that the returns of the empirical data deviate significantly from the normal distribution by comparing the four moments. Recall the moments of a normal

	Mean	Variance	Skewness	Kurtosis
DJIA	1.9688e-04	1.3435e-04	-0.8342	27.5420
SP500	2.4676e-04	9.8579e-05	-1.1459	36.1686
DGS05	7.2088e-05	1.7143e-04	-0.6674	28.9857
DGS10	9.9751e-05	1.1827e-04	-0.7163	24.8114
NORUS	8.9838e-05	1.0780e-04	-0.8220	29.1261
SWEUS	1.0719e-04	1.0865e-04	-0.6780	30.5526

Table 7.1: Moments of the returns r_t or R_t of empirical data.

distribution and t distribution in Chapter 6. By the transformation of r_t and R_t the data has zero mean and stationarity is assumed. Accordingly, the data is suitable for the estimators of long range dependence.

7.1 Estimation

By using 1000 estimates based on $FGN(H = 0.5)$ and $ARFIMA(d = 0)$ noise for lengths matching the financial series, it was possible to create confidence intervals for the short range dependent case of $H = 0.5$. By comparing this with the estimated intensity of LRD for the financial data it is possible to state whether or not the series is long range dependent. When estimating 1000 $FGN(H = 0.5)$ series, the densities for the estimators resulted in Figure 7.1. This is the same result as seen in Chapter 5 but presented differently.

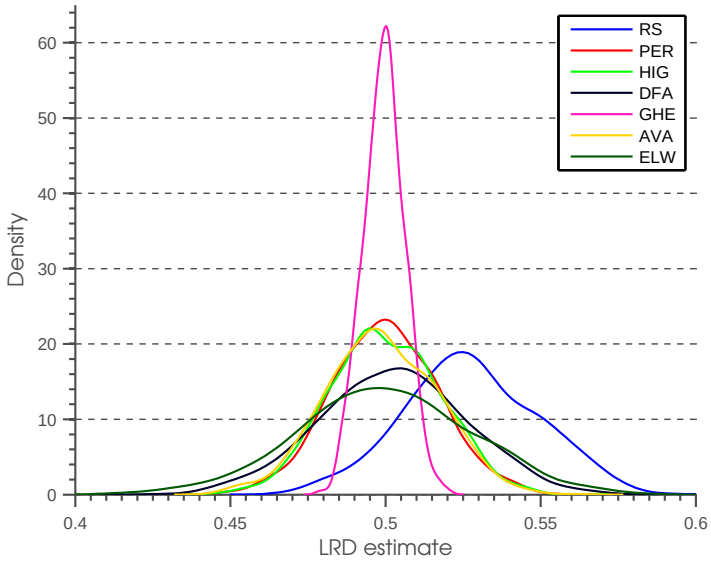


Figure 7.1: Kernel density estimates of $FGN(H = 0.5)$ (white noise).

	RS	PER	HIG	DFA	GHE	AV	ELW
DJIA	0.5656	0.4713	0.5187	0.4732	**0.4496**	0.4977	0.4830
SP500	**0.5742**	0.4685	0.5287	**0.3733**	**0.4390**	0.5149	0.5081
DGS05	**0.5902**	**0.5673**	**0.5626**	**0.5265**	**0.5456**	0.5293	0.5290
DGS10	**0.5616**	**0.5367**	**0.5396**	0.5242	**0.5346**	0.5208	0.5124
NORUS	0.5419	0.5191	0.5404	0.5037	0.4871	0.5321	0.5253
SWEUS	**0.5779**	0.5245	**0.5636**	0.5146	0.4958	0.5338	0.5579

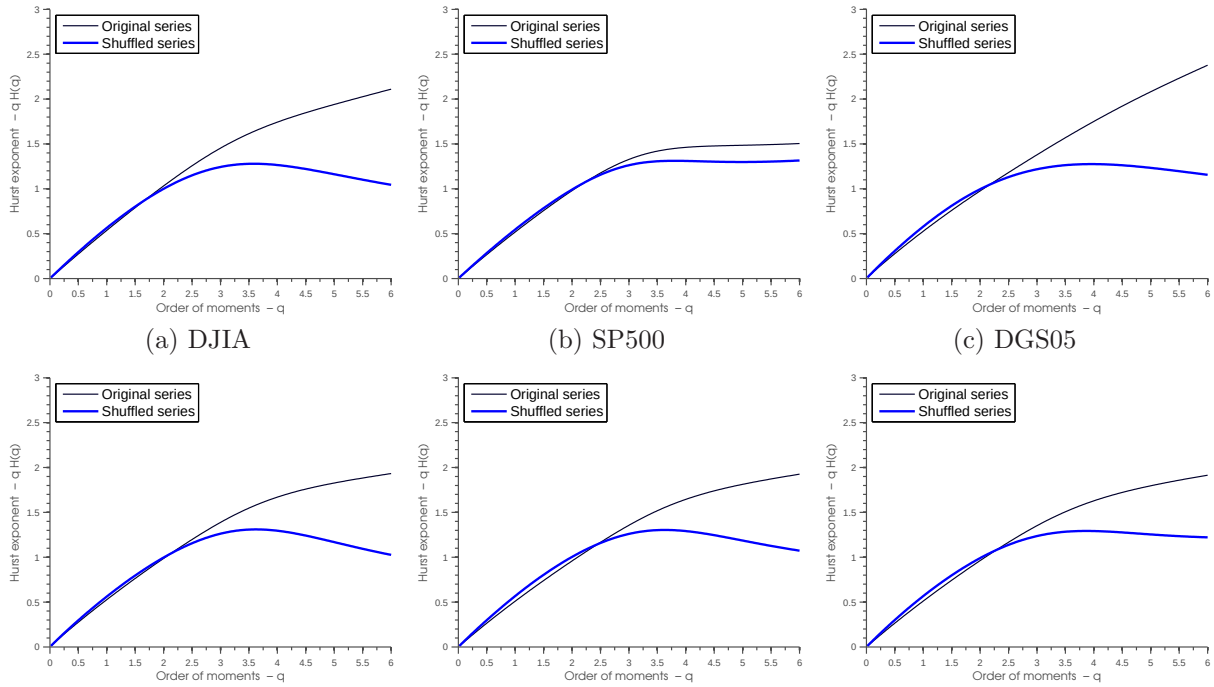
Table 7.2: Estimates for financial data. Bold numbers are significantly different at level 0.95 from $FGN(H = 0.5)$ based on confidence intervals from 1000 samples.

In Table 7.2 the estimates for each series are given. The bold numbers denote a deviation from a standard Brownian motion. The statements of deviation are based on the confidence interval for 1000 estimates of $FGN(H = 0.5)$ series for each estimator and at corresponding sample sizes to the financial data. In this case, when employing the entire data set in the estimation, there are a few instances where there are evidence for persistence. The empirical series with the most instances of deviation from a Brownian motion is the 10-year and the 5-year treasury bonds. It is also worthwhile noting that the RS and HIG estimator is constantly estimating high numbers opposed to the other estimators have estimates that are suggesting both anti-persistence and persistence.

7.1.1 Multifractality of financial time series

As introduced in Remark 5, the fractality of a series describes the correlation of the moments of the series' increments. For instance, the second moment, $q = 2$, of the increments is the autocorrelation function which is used by the estimators to estimate the intensity of long range dependence. In Chapter 6, the fractality of the corrupted series was given which turned out to be multifractal. By comparing Figure 6.8 and Figure 7.2, one can detect similarities. In Figure 6.8, the series $C(0.8, H, 4)$ are all displaying different levels of multifractality since the relationship between $qH(q)$ and q is non-linear. The same series $C(0.8, H, 4)$, in Figure 6.8, are displaying a point q^* where the slope is changing. This is also the case for the empirical series in Figure 7.2 where the relationships are all non-linear and the slope has a kink at $q^* = 3$.

In the plots of Figure 7.2, the blue line represents the shuffled data. In the unifractal case $qH(q)$ is linear and the slope represents the intensity of LRD - the Hurst exponent H. Also in the unifractal case the blue lines would all have a slope of 0.5 as it is estimating shuffled and independent data. However, when taking a closer look at the plots in Figure 7.2 there are clear evidence of multifractality in all of the series as $qH(q)$ is non-linear. This was



(a) DJIA (b) SP500 (c) DGS05

(d) DGS10 (e) NORUS (f) SWEUS

Figure 7.2: Multifractality of financial instruments.

also seen in Figure 6.8. As mentioned earlier, there are there two sources of multifractality: short range dependence and heavy tails. The shuffled series removes any correlation and the multifractality is solely explained by heavy tails. The case for all of the financial time series is that the shuffled series has a curve that is more "bent" than the original one. According to the statements regarding the multifractality of a series, the observation in Figure 7.2 suggests that the shuffled series have a higher degree of fractality than the original. The original series differs from the shuffled one due to its correlations.

7.2 Link between the corrupted and empirical series

As mentioned earlier in this chapter, financial data do not have the same properties as fractional Brownan motion or ARFIMA processes, which are unifractal processes. The returns of financial data have most of its values close to zero and have outliers and are multifractal. In this section, the similarity of corrupted signals and empirical data will briefly be discussed.

A comparison between Figure 7.2 and Figure 6.8 will be made, which depicts the fractality of the empirical and simulated data respectively. Here, the correlations of the q-eth order moments will be used as a measure of comparing a corrupted series and a financial time series. It is possible through visual examination to state that there are similarities between

the two. The pattern in the empirical data shows that $q* \approx 3$ represents a kink in $qH(q)$. The same can be said about the antipersistent series in Figure 6.8 where the graphs between the original and shuffled series differ at $q^* \approx 3$.

Based on the plots in Figure 7.2 and Figure 6.8 it can be concluded that a similarity in the features of the corrupted and empirical series exists. As the fractality of a time series is a vast research field that only has been briefly discussed here, it is not possible to make any other statement regarding the similarities of the higher order moments.

Chapter 8

Summary & outlook

In this thesis, a thorough approach to the notion of long range dependence in time series and its estimation has been made. This approach is done with connection to financial time series and how the estimators manage under the presence of a heavy tailed noise. The estimators that were examined was the rescaled range analysis, periodogram regression, Higuchi method, detrended fluctuation analysis, generalized Hurst exponent, aggregated variance and the exact local whittle method. Two simulation studies and an empirical application was carried out in order to gain knowledge of the behavior of estimators of long range dependence under different conditions. The following topics have been carefully analyzed:

- how estimators perform on time series of different length and different intensities of long memory given a pure signal,
- how estimators perform when time series of different length and different long range dependence intensity are corrupted with a heavy-tailed noise with varying level of corruption and,
- how the correlations of the higher order moments react to the corruption of heavy-tailed noise.

The results from these experiments are summarized in the following:

The uncorrupted case
When presented with a pure long memory series, the findings in general are that for a higher absolute value of d (i.e. levels of H deviating from the level of standard Brownian motion / white noise $H = 0.5$) the estimates have a higher bias which is also noticeably higher in smaller sample sizes. The standard deviations of the estimates are in few instances independent of the intensity of long range dependence. When compared to previous research, the the same results are found except for the case of generalized Hurst exponent and detrended

fluctuation analysis. It was also discovered complements to previous research which was due to the broadness of the experiment conducted here. The results in the experiments here tell with more precision at which levels of H and sample sizes N the estimators are asymptotically acquiring certain statistical properties. The estimators that displayed the most satisfactory performances in terms of bias and standard deviation was the Higuchi method, which was close to unbiased for all levels of H, and partially the generalized Hurst method given a fractional Gaussian noise signal. Given an fractionally integrated noise the exact local Whittle estimator was close to unbiased for all levels of H.

The corrupted case

The long range dependent signal was corrupted with signal-to-noise ratio $[0, 38]$. In general, when the series are corrupted with a noise, the breakdown point is highly dependent on the intensity of long memory. For lower intensities, namely the anti-persistent cases, the breakdown of the estimators happens with only a small portion of noise added. For higher persistent series the breakdown point of the estimators are all robust to changes in the underlying series. For the highest level of persistence, the estimators only break down close to the point where 50 percent of the signal is corrupted, which is the maximum breakdown point an estimator can have. Differences were also found in the performance of the estimators when presented with the same corrupted series. A pattern was found - the estimators which performed well in the persistent case performed equally poor in the anti-persistent case. In particular the generalized Hurst method performed well in the anti-persistent case and the periodogram method in the persistent. The rescaled range method performed comparatively well under both conditions.

Properties of the corrupted signal

When a long range dependent time series was corrupted with a heavy tailed noise, it had a larger effect on the higher order moments of the correlations for anti-persistent series. When comparing the corrupted anti-persistent series with empirical data similarities are found in the higher order moments. In particular both the empirical series and the corrupted, anti-persistent series had different fractality at moments higher than 3.

Combining the baseline properties and the robustness of the long memory estimators in order to create a method for dealing with long range dependence under the influence of heavy tails remains a task for future studies. Investigating difference in performance between persistent and antipersistent corrupted series is also something that future research could pursue.

Appendix A

Proofs, tables and codes

A.1 Mean Squared Error

Theorem A.1.1. *The mean squared error $MSE(\hat{H})$ is equal to $\mathbb{V}[\hat{H}] + \mathbb{B}^2(\hat{H})$.*

Proof.

$$\begin{aligned} MSE(\hat{H}) &= \frac{1}{r}\sum_{i=1}^{r}(\hat{H}_i - H_i) = \mathbb{E}[(\hat{H} - H)^2] \\ &= \mathbb{E}[(\hat{H} - \mathbb{E}[\hat{H}]) + (\mathbb{E}[\hat{H}] - H)^2] \\ &= \mathbb{E}\{(\hat{H} - \mathbb{E}[\hat{H}])^2 + (\mathbb{E}[\hat{H}] - H)^2 + (\hat{H} - \mathbb{E}[\hat{H}])(\mathbb{E}[\hat{H} - H])\} \\ &= \mathbb{V}[\hat{H}] + \mathbb{B}^2[\hat{H}] + \mathbb{E}[(\hat{H} - \mathbb{E}[\hat{H}])(\mathbb{E}[\hat{H}] - H)] \\ &= \mathbb{V}[\hat{H}] + \mathbb{B}^2[\hat{H}] + \mathbb{E}[\hat{H}\mathbb{E}[\hat{H}] - (\hat{H})^2 - H\hat{H} + \mathbb{E}[\hat{H}]] \\ &= \mathbb{V}[\hat{H}] + \mathbb{B}^2[\hat{H}] + (\mathbb{E}[\hat{H}])^2 - (\mathbb{E}[\hat{H}])^2 - H\mathbb{E}[\hat{H}] + H\mathbb{E}[\hat{H}] \\ &= \mathbb{V}[\hat{H}] + \mathbb{B}^2[\hat{H}]. \end{aligned} \tag{A.1.1}$$

□

A.2 Tables with estimates of C

The tables containing the estimates of corrupted series C, such as in Table A.1 were computed in the manner:

$$\begin{pmatrix} \hat{C}_1(1,H,4) & |\hat{C}_1(1,H,4) - \hat{C}_1(0.9,H,4)| & \dots & |\hat{C}_1(1,H,4) - \hat{C}_1(0.6,H,4)| \\ \hat{C}_2(1,H,4) & |\hat{C}_2(1,H,4) - \hat{C}_1(0.9,H,4)| & \dots & |\hat{C}_1(1,H,4) - \hat{C}_2(0.6,H,4)| \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}_7(1,H,4) & |\hat{C}_7(1,H,4) - \hat{C}_7(0.9,H,4)| & \dots & |\hat{C}_7(1,H,4) - \hat{C}_7(0.6,H,4)| \end{pmatrix}$$

where $\hat{C}_i$ is estimate by estimator i. Each table consists of four arrays containing this information for 4 different lengths. The use of absolute values is justified since all the deviations are in the same direction - i.e. for $H < 0.5$ all the changes are positive and for $H > 0.5$ all the changes are negative (as the LRD intensity of i.i.d. t distributed noise is equal to 0.5).

(Continued on next page.)

FGN(H=0.1) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.2922	0.0258	0.0884	0.1636	0.2189	0.2497	0.0393	0.1262	0.1993	0.2508
PER	0.0461	0.1273	0.3167	0.4269	0.4863	0.0053	0.1463	0.3068	0.3985	0.4485
HIG	0.0963	0.051	0.1701	0.2753	0.3362	0.0977	0.0732	0.2083	0.3035	0.3587
DFA	0.1529	0.0336	0.1162	0.2005	0.2582	0.134	0.0648	0.1646	0.2432	0.2967
GHE	0.1053	0.0224	0.0933	0.1833	0.2543	0.0992	0.0273	0.1071	0.1962	0.2749
AV	0.0889	0.0508	0.1662	0.2596	0.3143	0.0959	0.0714	0.2056	0.2966	0.346
ELW	0.0158	0.1426	0.3041	0.4002	0.4448	0.0485	0.1741	0.3213	0.3931	0.4212
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.2181	0.0650	0.1674	0.2395	0.2813	0.1934	0.1052	0.2158	0.2735	0.3061
PER	0.0233	0.1516	0.3049	0.3915	0.4351	0.0302	0.1582	0.3068	0.3888	0.4327
HIG	0.0996	0.0948	0.2354	0.3218	0.3675	0.1000	0.1180	0.2614	0.3372	0.3756
DFA	0.1245	0.1006	0.2059	0.2746	0.3206	0.1182	0.1416	0.2432	0.3003	0.3393
GHE	0.0999	0.0283	0.1077	0.1999	0.2792	0.0997	0.0293	0.1086	0.2019	0.2811
AV	0.0985	0.0941	0.2325	0.3163	0.3589	0.1000	0.1182	0.2598	0.3323	0.3673
ELW	0.0755	0.2171	0.3396	0.3864	0.4064	0.0885	0.2571	0.3589	0.3879	0.4014

Table A.1: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d = -0.4) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.3713	0.0158	0.0516	0.0845	0.1383	0.3135	0.0221	0.0667	0.1261	0.1751
PER	0.0434	0.0519	0.1734	0.2685	0.3408	0.0714	0.0712	0.1745	0.2742	0.3411
HIG	0.1670	0.0322	0.0977	0.1687	0.2407	0.1622	0.0368	0.1114	0.2032	0.2660
DFA	0.2340	0.0189	0.0663	0.1211	0.1787	0.2081	0.0310	0.0943	0.1612	0.2129
GHE	0.2082	0.0134	0.0539	0.1026	0.1615	0.2094	0.0120	0.0522	0.1123	0.1730
AV	0.1657	0.0267	0.0903	0.1545	0.2163	0.1606	0.0363	0.1109	0.1970	0.2548
ELW	0.0815	0.0593	0.1572	0.2410	0.3114	0.0956	0.0837	0.1931	0.2850	0.3370
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.2710	0.0301	0.0959	0.1625	0.2103	0.2358	0.0531	0.1412	0.2079	0.2476
PER	0.0928	0.0679	0.1745	0.2706	0.3310	0.0987	0.0720	0.1808	0.2696	0.3288
HIG	0.1590	0.0435	0.1353	0.2248	0.2819	0.1526	0.0560	0.1621	0.2467	0.3001
DFA	0.1891	0.0509	0.1329	0.2015	0.2448	0.1744	0.0811	0.1701	0.2313	0.2724
GHE	0.2103	0.0130	0.0531	0.1149	0.1750	0.2112	0.0125	0.0536	0.1142	0.1745
AV	0.1577	0.0431	0.1355	0.2225	0.2757	0.1520	0.0561	0.1622	0.2443	0.2942
ELW	0.1031	0.1086	0.2371	0.3149	0.3521	0.0999	0.1532	0.2805	0.3430	0.3753

Table A.2: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H=0.2) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.3697	0.0124	0.0549	0.1065	0.1554	0.3290	0.0174	0.0716	0.1289	0.1728
PER	0.1291	0.0412	0.1401	0.2393	0.2912	0.1564	0.0479	0.1481	0.2293	0.2846
HIG	0.1950	0.0293	0.0968	0.1832	0.2364	0.1987	0.0340	0.1166	0.1962	0.2493
DFA	0.2395	0.0182	0.0697	0.1354	0.1837	0.2259	0.0288	0.0973	0.1585	0.2110
GHE	0.1974	0.0156	0.0589	0.1259	0.1840	0.1975	0.0176	0.0635	0.1323	0.1930
AV	0.1859	0.0306	0.0923	0.1732	0.2230	0.2000	0.0325	0.1138	0.1838	0.2356
ELW	0.1595	0.0393	0.1329	0.2217	0.2650	0.1849	0.0554	0.1625	0.2271	0.2708
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.2984	0.0276	0.0960	0.1559	0.1952	0.2735	0.0459	0.1269	0.1845	0.2172
PER	0.1684	0.0532	0.1518	0.2310	0.2749	0.1729	0.0573	0.1546	0.2298	0.2752
HIG	0.2000	0.0415	0.134	0.2102	0.2558	0.1994	0.0522	0.1510	0.2231	0.2638
DFA	0.2178	0.0459	0.1225	0.1849	0.2254	0.2126	0.0659	0.1509	0.2048	0.2416
GHE	0.1991	0.0167	0.0656	0.1335	0.1952	0.1998	0.0161	0.066	0.1335	0.1954
AV	0.1999	0.0414	0.1327	0.2059	0.2476	0.1989	0.0524	0.1499	0.2199	0.2582
ELW	0.1930	0.0793	0.1851	0.2463	0.2730	0.1950	0.1035	0.2109	0.2600	0.2816

Table A.3: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d = -0.3) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.4239	0.0075	0.0222	0.0625	0.0935	0.3706	0.0114	0.0403	0.0854	0.1250
PER	0.1519	0.0340	0.0979	0.1755	0.2440	0.1829	0.0296	0.0947	0.1738	0.2357
HIG	0.2369	0.0251	0.0567	0.1189	0.1791	0.2352	0.0211	0.0703	0.1390	0.1979
DFA	0.2918	0.0145	0.0427	0.0858	0.1316	0.2704	0.0197	0.0610	0.1141	0.1627
GHE	0.2671	0.0117	0.0330	0.0738	0.1216	0.2705	0.0111	0.0355	0.0797	0.1301
AV	0.2335	0.0209	0.0512	0.1123	0.1628	0.2352	0.0197	0.0663	0.1309	0.1851
ELW	0.1848	0.0327	0.0861	0.1615	0.2214	0.1976	0.0371	0.1086	0.1802	0.2349
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.3261	0.0197	0.0629	0.1170	0.1589	0.2955	0.0276	0.0902	0.1461	0.1852
PER	0.1949	0.0342	0.1022	0.1784	0.2308	0.1991	0.0348	0.1066	0.1769	0.2289
HIG	0.2297	0.0271	0.0887	0.1604	0.2121	0.2261	0.0310	0.1040	0.1731	0.2231
DFA	0.2548	0.0286	0.0871	0.1439	0.1854	0.2443	0.0431	0.1114	0.1670	0.2036
GHE	0.2724	0.0094	0.0370	0.0838	0.1312	0.2729	0.0087	0.0383	0.0838	0.1326
AV	0.2286	0.0263	0.0873	0.1582	0.2067	0.2256	0.0309	0.1038	0.1711	0.2182
ELW	0.1977	0.0518	0.1393	0.2125	0.2545	0.1990	0.0695	0.1675	0.2305	0.2638

Table A.4: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H = 0.3) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.4505	0.0036	0.0182	0.0499	0.0949	0.4095	0.0057	0.0301	0.0680	0.1044
PER	0.2572	0.0179	0.0665	0.1170	0.1754	0.2784	0.0093	0.0572	0.1133	0.1627
HIG	0.3015	0.0051	0.0411	0.0866	0.1367	0.3004	0.0102	0.0498	0.1026	0.1478
DFA	0.3291	0.0086	0.0371	0.0711	0.1130	0.3177	0.0101	0.0449	0.0862	0.1261
GHE	0.2966	0.0030	0.0308	0.0653	0.1079	0.2978	0.0071	0.0345	0.0746	0.1178
AV	0.2957	0.0013	0.0362	0.0802	0.1268	0.2960	0.0126	0.0494	0.1002	0.1430
ELW	0.2851	0.0090	0.0513	0.0992	0.1466	0.2904	0.0163	0.0636	0.1141	0.1563
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.3769	0.0137	0.0453	0.0842	0.1190	0.3560	0.0154	0.0570	0.0992	0.1309
PER	0.2847	0.0172	0.0667	0.1139	0.1585	0.2890	0.0177	0.0647	0.1159	0.1555
HIG	0.2990	0.0167	0.0617	0.1110	0.1557	0.3007	0.0171	0.0665	0.1181	0.1580
DFA	0.3125	0.0148	0.0585	0.0987	0.1360	0.3092	0.0227	0.0699	0.1148	0.1475
GHE	0.2992	0.0084	0.0359	0.0749	0.1188	0.2997	0.0085	0.0357	0.0761	0.1188
AV	0.2990	0.0162	0.0607	0.1085	0.1480	0.3008	0.0165	0.0651	0.1151	0.1516
ELW	0.2961	0.0206	0.0795	0.1256	0.1621	0.2987	0.0291	0.0894	0.1381	0.1687

Table A.5: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d = -0.2) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.4708	0.0030	0.0225	0.0387	0.0640	0.4281	0.0093	0.0242	0.0532	0.0815
PER	0.2757	0.0060	0.0429	0.0932	0.1434	0.2826	0.0184	0.0497	0.0989	0.1446
HIG	0.3173	0.0048	0.0404	0.0721	0.1150	0.3145	0.0121	0.0387	0.0850	0.1262
DFA	0.3606	0.0071	0.0258	0.0531	0.0855	0.3434	0.0076	0.0300	0.0681	0.0984
GHE	0.3359	0.0046	0.0189	0.0473	0.0777	0.3401	0.0038	0.0223	0.0510	0.0848
AV	0.3065	0.0072	0.0404	0.0667	0.1085	0.3149	0.0097	0.0361	0.0783	0.1139
ELW	0.2936	0.0097	0.0451	0.0834	0.1320	0.2985	0.0157	0.0466	0.0948	0.1322
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.3914	0.0088	0.0343	0.0688	0.1021	0.3638	0.0147	0.0468	0.0871	0.1188
PER	0.2959	0.0138	0.0480	0.0987	0.1387	0.2978	0.0147	0.0511	0.1000	0.1380
HIG	0.3123	0.0126	0.0465	0.0941	0.1349	0.3102	0.0151	0.0531	0.1014	0.1402
DFA	0.3311	0.0121	0.0446	0.0848	0.1186	0.3246	0.0155	0.0555	0.0958	0.1290
GHE	0.3411	0.0054	0.0236	0.0540	0.0870	0.3414	0.0054	0.0240	0.0546	0.0878
AV	0.3110	0.0141	0.0466	0.0920	0.1297	0.3095	0.0155	0.0526	0.0997	0.1363
ELW	0.2964	0.0196	0.0631	0.1157	0.1530	0.2985	0.0235	0.0747	0.1248	0.1582

Table A.6: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H=0.4) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.5178	0.0097	0.0092	0.0265	0.0422	0.4853	0.0024	0.0146	0.0274	0.0443
PER	0.3677	0.0021	0.0391	0.0494	0.0886	0.3869	0.0024	0.0224	0.0407	0.0684
HIG	0.3966	0.0011	0.0131	0.0336	0.0644	0.3977	0.0001	0.0186	0.0393	0.0655
DFA	0.4173	0.0002	0.0123	0.0283	0.0497	0.4106	0.0014	0.0143	0.0322	0.0540
GHE	0.3858	0.0057	0.0132	0.0297	0.0532	0.3952	0.0042	0.0159	0.0305	0.0529
AV	0.3873	0.0013	0.0120	0.0285	0.0542	0.3962	0.0002	0.0168	0.0377	0.0583
ELW	0.3840	0.0030	0.0205	0.0349	0.0680	0.3941	0.0042	0.0163	0.0418	0.0631
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.4583	0.0005	0.0133	0.0318	0.0502	0.4386	0.0037	0.0172	0.0390	0.0561
PER	0.3913	0.0049	0.0205	0.0440	0.0661	0.3949	0.0048	0.0194	0.0426	0.0644
HIG	0.3985	0.0012	0.0181	0.0412	0.0669	0.3989	0.0038	0.0197	0.0451	0.0669
DFA	0.4054	0.0026	0.0188	0.0376	0.0588	0.4042	0.0052	0.0211	0.0430	0.0628
GHE	0.3987	0.0034	0.0136	0.0321	0.0530	0.3997	0.003	0.0133	0.0322	0.0535
AV	0.3963	0.0027	0.0191	0.0406	0.0626	0.3981	0.0043	0.0198	0.0439	0.0638
ELW	0.3965	0.0006	0.0219	0.0427	0.0641	0.3973	0.0044	0.0240	0.0493	0.0677

Table A.7: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d=-0.1) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.5345	0.0066	0.0006	0.0188	0.0304	0.4943	0.0006	0.0105	0.0225	0.0334
PER	0.3875	0.0029	0.0178	0.0430	0.0542	0.3920	0.0006	0.0153	0.0438	0.0658
HIG	0.4074	0.0036	0.0090	0.0310	0.0518	0.4023	0.0028	0.0141	0.0379	0.0575
DFA	0.4302	0.0023	0.0120	0.0269	0.0421	0.4191	0.0057	0.0100	0.0307	0.0502
GHE	0.4069	0.0010	0.0080	0.0226	0.0365	0.4144	0.0019	0.0103	0.0249	0.0416
AV	0.4002	0.0063	0.0112	0.0251	0.0413	0.3987	0.0046	0.0141	0.0342	0.0545
ELW	0.3896	0.0036	0.0138	0.0407	0.0579	0.3933	0.0037	0.0145	0.0388	0.0603
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.4628	0.0034	0.0139	0.0305	0.0454	0.4425	0.0046	0.0158	0.0344	0.0527
PER	0.3970	0.0047	0.0167	0.0419	0.0581	0.3988	0.005	0.0181	0.0385	0.0595
HIG	0.4034	0.0042	0.0160	0.0399	0.0597	0.4027	0.0044	0.0188	0.0405	0.0625
DFA	0.4138	0.0052	0.0158	0.0359	0.0517	0.4102	0.0063	0.0190	0.0386	0.0574
GHE	0.4167	0.0022	0.0103	0.0258	0.0423	0.4168	0.0031	0.0118	0.0263	0.0435
AV	0.4029	0.0033	0.0152	0.0361	0.0543	0.4027	0.0047	0.0187	0.0381	0.0581
ELW	0.3988	0.0037	0.0159	0.0426	0.0594	0.3983	0.0058	0.0237	0.0449	0.0651

Table A.8: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(0.6) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.6566	0.0016	0.0011	0.0015	0.0319	0.6386	0.0028	0.0045	0.0101	0.0257
PER	0.6238	0.0028	0.0083	0.0154	0.0532	0.6101	0.0035	0.0109	0.0173	0.0370
HIG	0.5923	0.0056	0.0086	0.0100	0.0290	0.5944	0.0005	0.0066	0.0113	0.0266
DFA	0.5979	0.0016	0.0066	0.0093	0.0338	0.5989	0.0019	0.0083	0.0152	0.0318
GHE	0.5701	0.0003	0.0064	0.0127	0.0378	0.5913	0.0004	0.0096	0.0195	0.0392
AV	0.5673	0.0051	0.0093	0.0089	0.0324	0.5815	0.0002	0.0067	0.0092	0.0252
ELW	0.5936	0.0023	0.0042	0.0023	0.0386	0.5928	0.0037	0.0077	0.0101	0.0275
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.6226	0.0024	0.0049	0.0082	0.0252	0.6158	0.0006	0.0063	0.0128	0.0281
PER	0.6019	0.0013	0.0042	0.0125	0.0345	0.6014	0.0014	0.0068	0.0169	0.0342
HIG	0.5966	0.0013	0.0065	0.0108	0.0281	0.5995	0.0001	0.0065	0.0142	0.0289
DFA	0.5981	0.0001	0.0072	0.0150	0.0319	0.5982	0.0004	0.0046	0.0150	0.0298
GHE	0.5971	0.0015	0.0091	0.0202	0.0411	0.5993	0.0017	0.0093	0.0219	0.0409
AV	0.5904	0.0010	0.0061	0.0109	0.0291	0.5964	0.0000	0.0064	0.0145	0.0305
ELW	0.5960	0.0009	0.0079	0.0107	0.0277	0.5991	0.0009	0.0067	0.0137	0.0291

Table A.9: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d=0.1) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.6583	0.0059	0.0047	0.0141	0.0308	0.6340	0.0003	0.0041	0.0132	0.0238
PER	0.6141	0.0022	0.0003	0.0339	0.0405	0.6077	0.0036	0.0111	0.0212	0.0411
HIG	0.5851	0.0009	0.0071	0.0191	0.0336	0.5936	0.0009	0.0072	0.0161	0.0310
DFA	0.5871	0.0008	0.0065	0.0174	0.0286	0.5921	0.0006	0.0079	0.0179	0.0317
GHE	0.5584	0.0006	0.0070	0.0172	0.0315	0.5796	0.0011	0.0086	0.0181	0.0336
AV	0.5572	0.0018	0.0052	0.0171	0.0286	0.5808	0.0010	0.0053	0.0157	0.0313
ELW	0.5852	0.0021	0.0032	0.0220	0.0318	0.5946	0.0031	0.0077	0.0174	0.0342
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.6225	0.0008	0.0052	0.0128	0.0287	0.6142	0.0031	0.0052	0.0133	0.0281
PER	0.6030	0.0015	0.0068	0.0205	0.0383	0.6015	0.0029	0.0069	0.0194	0.0381
HIG	0.5958	0.0001	0.0067	0.0138	0.0302	0.5982	0.0022	0.0065	0.0147	0.0294
DFA	0.5930	0.0006	0.0074	0.0141	0.0317	0.5959	0.0019	0.0075	0.0155	0.0296
GHE	0.5857	0.0019	0.0077	0.0192	0.0352	0.5877	0.0020	0.0074	0.0196	0.0362
AV	0.5904	0.0002	0.0061	0.0143	0.0326	0.5960	0.0024	0.0068	0.0156	0.0308
ELW	0.5961	0.0013	0.0075	0.0140	0.0320	0.5998	0.0031	0.0081	0.0145	0.0284

Table A.10: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H=0.7) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.7274	0.0001	0.0002	0.0152	0.0520	0.7164	0.0001	0.0061	0.0249	0.0533
PER	0.7452	0.0011	0.0016	0.0239	0.0855	0.7213	0.0011	0.0083	0.0255	0.0655
HIG	0.6855	0.0025	0.0091	0.0194	0.0508	0.6916	0.0006	0.0042	0.0169	0.0444
DFA	0.6988	0.0034	0.0106	0.0348	0.0703	0.6958	0.0035	0.0120	0.0270	0.0609
GHE	0.6596	0.0032	0.0089	0.0328	0.0672	0.6850	0.0015	0.0119	0.0337	0.0697
AV	0.6507	0.0017	0.0016	0.0245	0.0571	0.6702	0.0007	0.0047	0.0168	0.0474
ELW	0.7014	0.0028	0.0020	0.0314	0.0717	0.6961	0.0010	0.0037	0.0202	0.0539
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.7078	0.0034	0.0073	0.0184	0.0405	0.7016	0.0007	0.0020	0.0128	0.0349
PER	0.7098	0.0034	0.0105	0.0271	0.0588	0.7063	0.0033	0.0116	0.0273	0.0562
HIG	0.6977	0.0056	0.0076	0.0185	0.0377	0.6991	0.0015	0.0055	0.0167	0.0369
DFA	0.6969	0.0067	0.0112	0.0261	0.0525	0.6976	0.0021	0.0084	0.0224	0.0438
GHE	0.6947	0.0039	0.0144	0.0366	0.0700	0.6979	0.0033	0.0143	0.0364	0.0702
AV	0.6856	0.0059	0.0079	0.0195	0.0422	0.6919	0.0013	0.0058	0.0170	0.0375
ELW	0.6997	0.0057	0.0092	0.0186	0.0423	0.6993	0.0019	0.0050	0.0153	0.0344

Table A.11: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d = 0.2) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.7180	0.0019	0.0118	0.0178	0.0462	0.7090	0.0032	0.0076	0.0213	0.0423
PER	0.7267	0.0046	0.0028	0.0277	0.0574	0.7147	0.0014	0.0105	0.0287	0.0579
HIG	0.6881	0.0051	0.0138	0.0329	0.0549	0.6872	0.0048	0.0036	0.0169	0.0386
DFA	0.6794	0.0030	0.0120	0.0309	0.0605	0.6794	0.0055	0.0063	0.0254	0.0488
GHE	0.6392	0.0018	0.0101	0.0291	0.0583	0.6659	0.0001	0.0107	0.0326	0.0589
AV	0.6443	0.0002	0.0103	0.0259	0.0511	0.6665	0.0028	0.0053	0.0203	0.0425
ELW	0.6882	0.0005	0.0053	0.0262	0.0545	0.6939	0.0024	0.0078	0.0267	0.0460
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.7017	0.0032	0.0037	0.0197	0.0398	0.7009	0.0002	0.0066	0.0153	0.0343
PER	0.7019	0.0010	0.0058	0.0258	0.0525	0.7009	0.0003	0.0110	0.0251	0.0523
HIG	0.6900	0.0067	0.0024	0.0157	0.0351	0.6959	0.0009	0.0063	0.0171	0.0350
DFA	0.6823	0.0042	0.0006	0.0177	0.0414	0.6904	0.0003	0.0075	0.0197	0.0416
GHE	0.6752	0.0002	0.0106	0.0310	0.0597	0.6798	0.0021	0.0123	0.0315	0.0614
AV	0.6777	0.0047	0.0006	0.0163	0.0370	0.6897	0.0003	0.0072	0.0169	0.0365
ELW	0.6933	0.0047	0.0003	0.0174	0.0391	0.6995	0.0005	0.0080	0.0157	0.0343

Table A.12: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H=0.8) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.7921	0.0091	0.0170	0.0432	0.0681	0.7866	0.0042	0.0101	0.0272	0.0576
PER	0.8795	0.0218	0.0328	0.054	0.1099	0.8322	0.0067	0.0107	0.0386	0.0837
HIG	0.7736	0.0109	0.0004	0.0226	0.0419	0.7912	0.0021	0.0046	0.0244	0.0411
DFA	0.7873	0.0015	0.0158	0.0449	0.0932	0.7918	0.0036	0.0118	0.0403	0.0785
GHE	0.7341	0.0025	0.0156	0.0430	0.0846	0.7714	0.0046	0.0153	0.0445	0.0895
AV	0.7239	0.0014	0.0130	0.0284	0.0607	0.7527	0.0052	0.0066	0.0233	0.0484
ELW	0.8035	0.0084	0.0202	0.0440	0.0940	0.8019	0.0099	0.0089	0.0322	0.0668
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.7817	0.0002	0.0069	0.0153	0.0409	0.7839	0.0027	0.0019	0.0106	0.0274
PER	0.8135	0.0005	0.0124	0.0329	0.0708	0.8072	0.0022	0.0129	0.0331	0.0696
HIG	0.7936	0.0045	0.0068	0.0157	0.0326	0.7975	0.0003	0.0047	0.0149	0.0317
DFA	0.7902	0.0041	0.0093	0.0291	0.0585	0.7964	0.0037	0.0099	0.0282	0.0532
GHE	0.7837	0.0005	0.0156	0.0426	0.0875	0.7926	0.0038	0.0176	0.0453	0.0894
AV	0.7658	0.0046	0.0044	0.0143	0.0350	0.7799	0.0008	0.0049	0.0148	0.0324
ELW	0.7970	0.0061	0.0053	0.0181	0.0414	0.7998	0.0006	0.0061	0.0147	0.0320

Table A.13: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(d=0.3) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.7809	0.0000	0.0126	0.0275	0.0548	0.7796	0.0065	0.0094	0.0260	0.0516
PER	0.8669	0.0019	0.0027	0.0633	0.0948	0.8189	0.0015	0.0062	0.0307	0.0655
HIG	0.7785	0.0011	0.0055	0.0395	0.0577	0.7857	0.0029	0.0055	0.0219	0.0389
DFA	0.7695	0.0005	0.0133	0.0493	0.0832	0.7722	0.0011	0.0095	0.0303	0.0594
GHE	0.7163	0.0001	0.0112	0.0412	0.0731	0.7501	0.0024	0.0118	0.0363	0.0697
AV	0.7189	0.0014	0.0048	0.0372	0.0602	0.7451	0.0010	0.0040	0.0198	0.0391
ELW	0.7973	0.0066	0.0104	0.0535	0.0819	0.7927	0.0007	0.0040	0.0243	0.0507
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.7797	0.0007	0.0015	0.0150	0.0356	0.7863	0.0027	0.0043	0.0120	0.0263
PER	0.8078	0.0023	0.0061	0.0267	0.0628	0.8026	0.0021	0.0104	0.0281	0.0605
HIG	0.7903	0.0002	0.0040	0.0128	0.0325	0.7937	0.0018	0.0056	0.0132	0.0285
DFA	0.7834	0.0019	0.0090	0.0250	0.0502	0.7883	0.0015	0.0084	0.0214	0.0421
GHE	0.7664	0.0024	0.0117	0.0344	0.0711	0.7742	0.0028	0.0136	0.0357	0.0715
AV	0.7665	0.0001	0.0037	0.0143	0.0325	0.7797	0.0009	0.0057	0.0135	0.0278
ELW	0.7974	0.0009	0.0021	0.0135	0.0362	0.7994	0.0002	0.0038	0.0126	0.0263

Table A.14: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

FGN(H=0.9) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.8438	0.0049	0.0093	0.0439	0.0987	0.8391	0.0002	0.0047	0.0249	0.0682
PER	0.9715	0.0152	0.0083	0.0386	0.1171	0.9359	0.0034	0.0137	0.0506	0.1119
HIG	0.869	0.0006	0.0041	0.0225	0.0539	0.8803	0.0048	0.0047	0.0137	0.0388
DFA	0.8780	0.0036	0.0214	0.0653	0.1352	0.8865	0.0054	0.0250	0.0582	0.1100
GHE	0.7991	0.0032	0.0202	0.0554	0.1139	0.8445	0.0045	0.0230	0.0599	0.1161
AV	0.7825	0.0021	0.0045	0.0220	0.0675	0.8186	0.0027	0.0080	0.0222	0.0511
ELW	0.8942	0.0028	0.0164	0.0549	0.1169	0.9012	0.0005	0.0096	0.0375	0.0815
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.8502	0.0021	0.0040	0.0169	0.0458	0.8589	0.000	0.0005	0.0075	0.0225
PER	0.9186	0.0012	0.0206	0.0504	0.1059	0.9105	0.0051	0.0208	0.0533	0.1020
HIG	0.8878	0.0018	0.0021	0.0100	0.0315	0.8921	0.0015	0.0021	0.0088	0.0263
DFA	0.8906	0.0036	0.0175	0.0473	0.0867	0.8952	0.0047	0.0181	0.039	0.074
GHE	0.8648	0.0041	0.0221	0.0575	0.1124	0.8770	0.0047	0.0227	0.0574	0.1118
AV	0.8415	0.0003	0.0054	0.0175	0.0394	0.8580	0.0011	0.0061	0.0145	0.0323
ELW	0.9012	0.0002	0.0076	0.0251	0.0547	0.9031	0.0011	0.0064	0.0173	0.0379

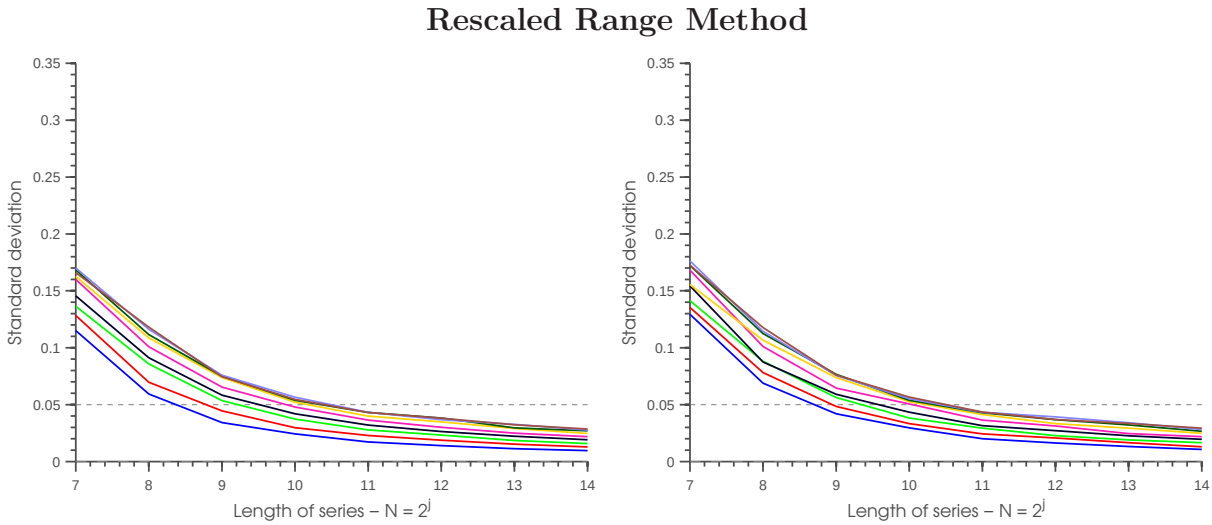
Table A.15: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

ARFIMA(0.4) corrupted with T($\nu = 4$)										
	ΔH $N = 2^8$					ΔH $N = 2^{10}$				
	α=1	α=0.9	α=0.8	α=0.7	α=0.6	α=1	α=0.9	α=0.8	α=0.7	α=0.6
RS	0.8368	0.0027	0.0147	0.0285	0.0496	0.8448	0.0038	0.0077	0.0214	0.0426
PER	0.9913	0.0035	0.0206	0.0544	0.0952	0.9445	0.0131	0.0209	0.0442	0.0808
HIG	0.8688	0.0045	0.0163	0.0240	0.0423	0.8707	0.0029	0.0016	0.0132	0.0257
DFA	0.8625	0.0056	0.0210	0.0473	0.0833	0.8729	0.0053	0.0177	0.0364	0.0654
GHE	0.7867	0.0044	0.0159	0.0363	0.0703	0.8303	0.0047	0.0152	0.0351	0.0692
AV	0.7865	0.0050	0.0122	0.0235	0.0425	0.8186	0.0041	0.0074	0.0174	0.0301
ELW	0.9058	0.0121	0.0266	0.0479	0.0848	0.9055	0.0119	0.0173	0.0287	0.0527
	ΔH $N = 2^{12}$					ΔH $N = 2^{14}$				
RS	0.8521	0.0032	0.0014	0.0119	0.0225	0.8617	0.0003	0.0026	0.0044	0.0127
PER	0.9128	0.0010	0.0101	0.0258	0.0607	0.9047	0.0003	0.0100	0.0272	0.0599
HIG	0.8731	0.0034	0.0021	0.0081	0.0162	0.8765	0.0010	0.0029	0.0049	0.0144
DFA	0.8777	0.0016	0.0091	0.0249	0.0467	0.8841	0.0015	0.0064	0.0178	0.0391
GHE	0.8493	0.0004	0.0113	0.0307	0.0635	0.8624	0.0015	0.0112	0.0292	0.0628
AV	0.8394	0.0015	0.0030	0.0108	0.0194	0.8568	0.0010	0.0029	0.0065	0.0164
ELW	0.9002	0.0009	0.0049	0.0135	0.0269	0.9027	0.0002	0.0058	0.0112	0.0224

Table A.16: Differences in H estimates for different scaling factors α and lengths N. For $\alpha = 1$ the estimates of $C(1, H, 4)$ are given, whereas for $\alpha = \{0.6, 0.7, 0.8, 0.9\}$ the difference between $C(\alpha, H, 4)$ and $C(1, H, 4)$ is given.

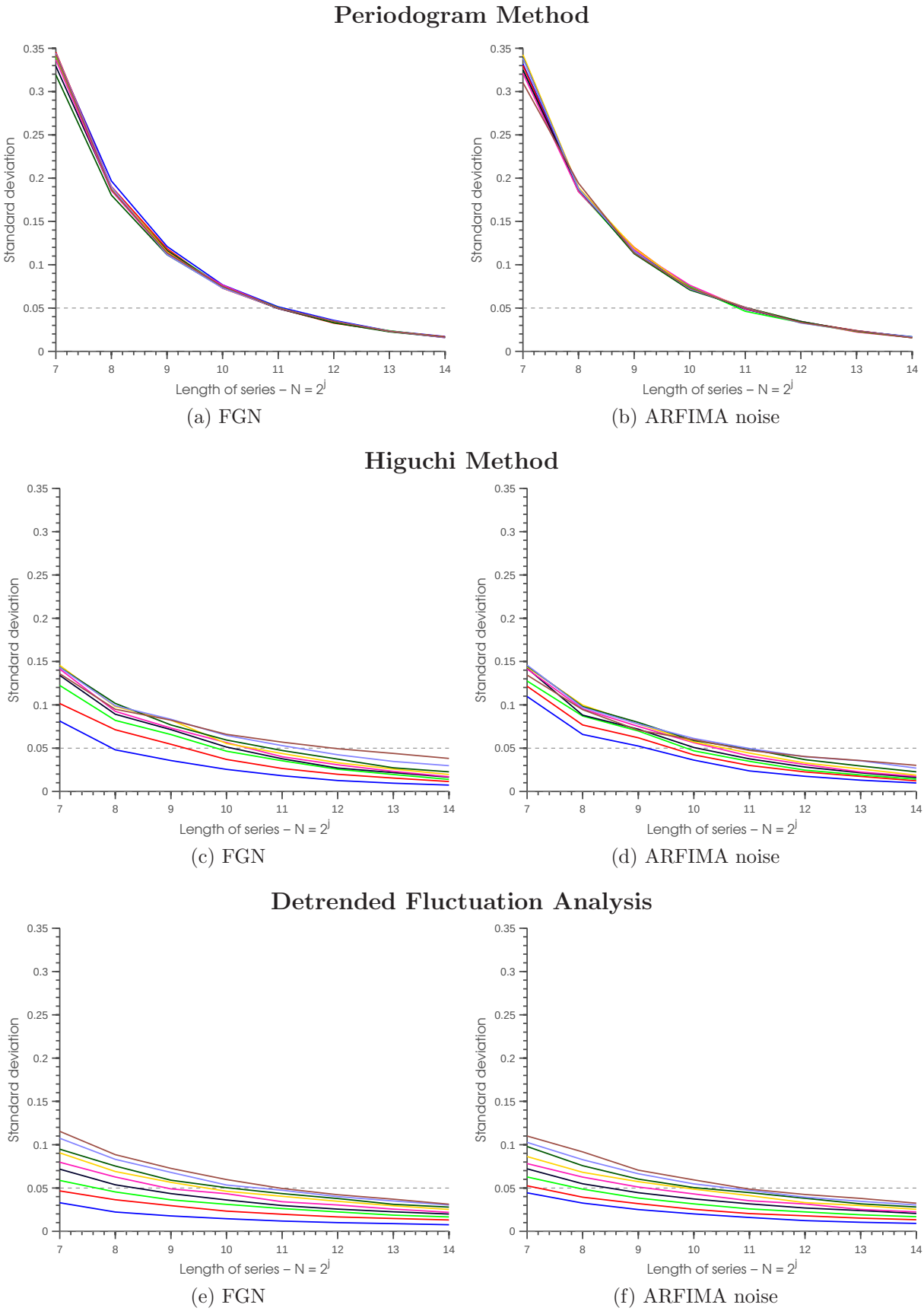
A.3 Standard deviations of the estimates of C

Included are the standard deviations of the $\hat{C}(\alpha, H, \nu)$ estimates from Chapter 5.



(a) FGN (b) ARFIMA noise

Figure A.1: Standard deviations of $\hat{H}$ estimates of a LRD series.

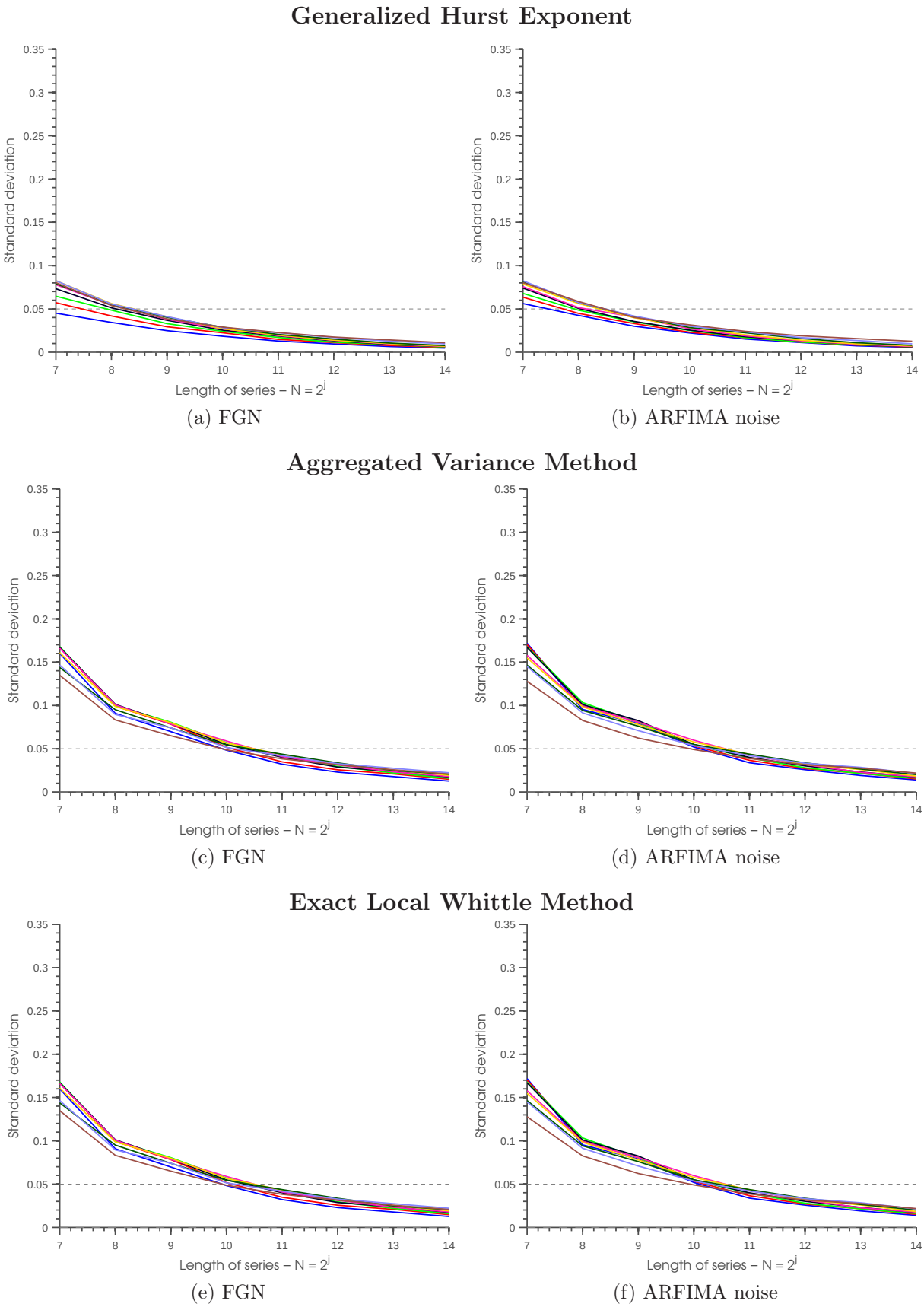


(a) FGN (b) ARFIMA noise

(c) FGN (d) ARFIMA noise

(e) FGN (f) ARFIMA noise

Figure A.2: Standard deviations of $\hat{H}$ estimates of a LRD series.



(a) FGN (b) ARFIMA noise

(c) FGN (d) ARFIMA noise

(e) FGN (f) ARFIMA noise

Figure A.3: Standard deviations of $\hat{H}$ estimates of a LRD series.

Spectral densities	R[59] package `longmemo` based on programs written by Jan Beran. Found at R-project.org [8]. See Table 2.1 and 3.2.
Rescaled range analysis	MATLAB[47] program by Chu Chen found at the File Exchange [14].
Periodogram method	MATLAB program program by Chu Chen found at the File Exchange[14].
Higuchi method	MATLAB program program by Chu Chen found at the File Exchange[14].
Detrended fluctuation analysis	MATLAB program program by Max Little found at the File Exchange [38].
Generalized Hurst exponent	MATLAB program program by Tomaso Aste found at the File Exchange [2].
Aggregated Variance	MATLAB program program by Chu Chen found at the File Exchange[14].
Exact local Whittle	MATLAB program program by György Inzelt found at the File Exchange [29].
Fractional Gaussian noise	MATLAB program program by Stilian Stoev and Yingchun Zhou found at the File Exchange [67].
ARFIMA	MATLAB program program by Stilian Stoev found at the File Exchange [66].
t **distributed noise**	Built in MATLAB program `trnd`.

Table A.17: Programs used in this thesis.

A.4 Credits

Throughout the thesis a number of computations have been performed in different coding environments. Most of the codes that are written by the author in connection with this thesis are for ease of computation. It is possible to generate all of the results in this thesis without the aid of these programs. Included are only the codes that are necessary to generate the identical results. In Table A.17 the use of the main programs, which other authors have written, is justified. Programs that perform standard tasks are not included as they are common features to statistical programs. The table is in chronological order. Names in caps, such as MATLAB, are denoting the program used.

Bibliography

[1] E. Alvarez-Lacalle, B. Dorow, J. P. Eckmann, and E. Moses. Hierarchical structures induce long-range dynamical correlations in written texts. *PNAS*, 103(21):7956–7961, 2006.

[2] T. Aste. Matlab program: Generalized hurst exponent. http://www.mathworks.com/matlabcentral/fileexchange/30076-generalized-hurst-exponent. Accessed: 05/29/2012.

[3] A. Barabasi and T. Vicsek. Multifractality of self-affine fractals. *Physical Review A*, 44(4), 1991.

[4] J. Bardet, G. Bardet, G. Oppenheim, A. Philippe, S. Stoev, and M. S. Taqqu. Semi-parametric estimation of the long-range dependence parameter: a survey. *Birkhauser*, pages 579–623, 2003.

[5] J. M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev, and M. S. Taqqu. Generators of long-range dependence processes: A survey. In P. Doukhan, G. Oppenheim, and M. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, pages 579–623. Birkhauser, New York, 2003.

[6] J. Barunik, T. Aste, T. D. Matteo, , and R. Liu. Understanding the source of multifractality in financial markets. *Physica A: Statistical Mechanics and its Applications*, 2012. Article in Press.

[7] J. Barunik and L. Kristoufek. On hurst exponent estimation under heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications*, 389(18):3844 – 3855, 2010.

[8] J. Beran. R-package: longmemo. http://cran.r-project.org/web/packages/longmemo/index.html. Accessed: 05/29/2012.

[9] J. Beran. *Statistics for long-memory processes*. Monographs on statistics and applied probability. Chapman & Hall, 1994.

[10] B. O. Bradley and M. S. Taqqu. Chapter 2 - financial risk and heavy tails. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 35 – 103. North-Holland, Amsterdam, 2003.

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

[12] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting*. Springer-New York, 1996.

[13] P. Broersen. *Automatic autocorrelation and spectral analysis*. Springer, 2006.

[14] C. Chen. Matlab program: Long range depependence estimators. http://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate. Accessed: 05/29/2012.

[15] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[16] M. Couillard and M. Davidson. A comment on measuring the hurst exponent of financial time series. *Physica A*, 348:404–418, 2005.

[17] P. F. Craigmile. Simulating a class of stationary gaussian processes using the davies and harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24(5):505–511, 2003.

[18] R. B. Davies and D. S. Harte. Tests for hurst effect. *Biometrika*, 74(1):pp. 95–101, 1987.

[19] A. B. Dieker and M. Mandjes. On spectral simulation of fractional brownian motion. *Probability in the Engineering and Informational Sciences*, 17:417–434, 2003.

[20] 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.*, 18(4):1088–1107, 1997.

[21] P. Embrechts and M. Maejima. *Self-similar Processes*. Princeton University Press, 2002.

[22] J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238, 1983.

[23] W. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 2003.

[24] D. Guégan. How can we define the concept of long memory? an econometric survey. Technical Report 178, School of Economics and Finance, QUT, 2004.

[25] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. Eugene Stanley. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E*, 64, 2001.

[26] P. J. Huber. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience, 1981.

[27] H. H. Hurst. Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers*, pages 519 –577.

[28] H. H. Hurst. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–799, 1951.

[29] G. Inzelt. Matlab program: Whittle estimator. http://www.mathworks.com/matlabcentral/fileexchange/30238-arfimapdq-estimator. Accessed: 05/29/2012.

[30] S. Jamdee and C. A. Los. Long memory options: Lm evidence and simulations. *Research in International Business and Finance*, 21:260–280, 2007.

[31] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(14):87 – 114, 2002.

[32] P. Kokoszka and M. Taqqu. Parameter estimation for infinite variance fractional arima. *Annals of Statistics*, 24(5):1880–1913, 1996.

[33] A. Kolmogorov. Dissipation of energy in a locally isotropic turbulence. *Dokl. Akad. Nauk SSSR*, 32(141), 1941.

[34] A. Kolmogorov. On degeneration of isotropic turbulence in an incompressible viscous liquid. *Dokl. Akad. Nauk SSSR*, 31:358, 1941.

[35] L. Kristoufek. Rescaled range analysis and detrended fluctuation analysis: Finite sample properties and confidence intervals. *Czech Economic Review*, 4(3):315–329, 2010.

[36] L. Kristoufek. How are rescaled range analyses affected by different memory and distributional properties? a monte carlo study. *Physica A: Statistical Mechanics and its Applications*, 2012.

[37] H. Künsch. Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society, Vol. 1*, pages 67–74. VNU Sci. Press, 1987.

[38] M. Little. Matlab program: Fast dfa algorithm. http://web.media.mit.edu/~maxl/software/index.html. Accessed: 05/29/2012.

[39] S. B. Lowen. Efficent generation of fractional brownian motion for simulation of infrared focal-plane array calibration drift. *Methodology and Computing in Applied Probability*, 1:445–456, 1999.

[40] T. Lux. Turbulence in financial markets: the surprising explanatory power of simple cascade models. *Quantitative Finance*, 1:632, 2001.

[41] B. Mandelbrot and J. Wallis. Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5):909 – 918, 1968.

[42] B. B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer Verlag New York, 1997.

[43] B. B. Mandelbrot. Chapter 1 - heavy tails in finance for independent or multifractal price increments. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 1 – 34. North-Holland, Amsterdam, 2003.

[44] B. B. Mandelbrot. *Fractals and Scaling In Finance: Discontinuity, Concentration, Risk*. Springer, 1st edition. edition, Nov. 2010.

[45] B. B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. Cowles Foundation Discussion Papers 1164, Cowles Foundation for Research in Economics, Yale University, 1997.

[46] B. B. Mandelbrot and J. W. V. Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422 – 437, 1968.

[47] Mathworks. Matlab webpage. http://www.mathworks.se/products/matlab/. Accessed: 05/29/2012.

[48] T. D. Matteo. Multi - scaling in finance. *Quantitative Finance*, 7(1):21–36, 2007.

[49] T. D. Matteo, T. Aste, and M. Dacorogna. Scaling behaviors in differently developed markets. *Physica A*, (324):183–188, 2003.

[50] T. D. Matteo, T. Aste, and M. M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, 29:827–851, 2005.

[51] S. Michalski. Blocks adjustmentreduction of bias and variance of detrended fluctuation analysis using monte carlo simulation. *Physica A: Statistical Mechanics and its Applications*, 387(1):217 – 242, 2008.

[52] R. F. Mulligan and R. Koppl. Monetary policy regimes in macroeconomic data: An application of fractal analysis. *The Quarterly Review of Economics and Finance*, 51, 2011.

[53] N.Crato, R. Linhares, and S. Lopes. Statistical properties of detrended fluctuation analysis. *Journal of Statistical Computation and Simulation*, 80(6):625–641, 2001.

[54] F. R. B. of St. Louis. Federal reserve economic data. http://research.stlouisfed.org/fred2/. Accessed: 05/23/2012.

[55] J. C. R. Pacheco and D. T. Roman. Accuracy of time-domain algorithms for self-similarity: An empirical study. *International Conference on Computing*, 0:379–384, 2006.

[56] W. Palma. *Long-memory time series - Theory and method*, volume 1. Wiley, 2007.

[57] C. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A. Goldberger. Mosaic organization of dna nucleotides. *Physical Review E*, 49(2):1685–1689, 1994.

[58] E. E. Peters. *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. Wiley, New York, 1994.

[59] R-project. R-project webpage. http://www.r-project.org/. Accessed: 05/29/2012.

[60] W. Rea, L. Oxley, M. Reale, and J. Brown. Estimators for long range dependence: an empirical study. *Electronic journal of statistics*, 0, 2009. PREPRINT.

[61] S. Rehman. Study of saudi arabian climatic conditions using hurst exponent and climatic predictability index. *Chaos, Solitons and Fractals*, 39(2):499 – 509, 2009.

[62] P. Robinson. *Time Series with Long Memory*. Oxford University Press, 2003.

[63] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):pp. 1630–1661, 1995.

[64] K. Shimotsu and P. C. B. Phillips. Exact local Whittle estimation of fractional integration. *Ann. Statist.*, 33(4):1890–1933, 2005.

[65] S. Souza, B. Tabak, and D. Cajueiro. Long-range dependence in exchange rates: The case of the european monetary system. *International Journal of Theoretical and Applied Finance*, 11(2):199–223, 2008.

[66] S. Stoev. Matlab program: Arfima simulation. http://www.stat.lsa.umich.edu/~sstoev/. Accessed: 05/29/2012.

[67] S. Stoev. Matlab program: Fgn simulation. http://www.mathworks.com/matlabcentral/fileexchange/19797-simulation-of-fractional-gaussian-noise-exact. Accessed: 05/29/2012.

[68] S. Stoev and M. Taqqu. Simulation methods for linear fractional stable motion and farima using the fast fourier transform. *Fractals*, 12(1):95–121, 2004.

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

[70] M. Taqqu and V. Teverovsky. On estimating the intensity of long-range dependence in finite and infinite variance time series. 1996. Article featured in the book 'Heavy Tails: Statistical Techniques and Applications'.

[71] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: An empirical study. *Fractals*, 3:785–798, 1995.

[72] C. Velasco. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20:87–127, 1999.

[73] R. Weron. Estimating long-range dependence: Finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, 312(1-2):285–299, 2002.

[74] H. Windsor and R. Toumi. Scaling and persistence of UK pollution. *Atmospheric Environment*, 35(27):4545–4556, 2001.

[75] A. T. A. Wood and G. Chan. Simulation of stationary gaussian processes in [0,1] d. *Journal of Computational and Graphical Statistics*, 3(4):pp. 409–432, 1994.