Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

RIGORÓZNÍ PRÁCE

Multifractal nature of financial markets and its relationship to market efficiency

Author: Mgr. Jakub Jeřábek

Supervisor: Prof. Ing. Miloslav Vošvrda, CSc.

Academic Year: 2009/10

Děkuji svému školiteli Prof. Miloslavu Vošvrdovi za cenné rady a připomínky a rovnež kolegovi Ladislavu Krištoufkovi za připomínky k první verzi této práce.

Prohlašuji, že jsem svou diplomovou práci napsal(a) samostatně a použil pouze uvedené prameny a literaturu

V Praze dne 15. února 2010 Jakub Jeřábek

Bibliographic Evidence

Jeřábek, Jakub, Multifractal nature of financial markets
and its relationship to market efficiency, Prague: Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, 2010, 84 pages.
Supervisor: Prof. Ing. Miloslav Vošvrda, CSc.

Abstrakt

Práce ukazuje souvislost mezi perzistencí ve výnosech finančních trhů a jejich efektivitou. Interpretuje hypotézu efektivních trhů a představuje různé modely časových řad použitelné k analýze finančních trhů. Důkladně vysvětluje koncept dlouhé paměti a analyzuje dva hlavní typy metod k odhadu dlouhé paměti – metody v časové a frekvenční doméně. Pomocí Monte Carlo studie srovnává kvalitu jednotlivých metod a vybrané metody následně používá na data z reálného světa – směnné kurzy a akciové trhy. Práce nenachází žádné doklady o dlouhé paměti ve výnosech, nicméně volatilita akciových trhů vykazuje jasné známky perzistence.
Klíčová slova: tržní efektivnost, perzistence, metody odhadu dlouhé paměti

Abstract

The thesis shows the relationship between the persistence in the financial markets returns and their efficiency. It interprets the efficient markets hypothesis and provides various time series models for the analysis of financial markets. The concept of long memory is broadly presented and two main types of methods to estimate long memory are analysed – time domain and frequency domain methods. A Monte Carlo study is used to compare these methods and selected estimators are then used on real world data – exchange rate and stock market series. There is no evidence of long memory in the returns but the stock market volatilities show clear signs of persistence.
Keywords: market efficiency, persistence, long memory estimation methods

Contents

List of Figures

Chapter 1

Introduction

Market efficiency problem

A finance professor and a student come across a $100 bill lying on the ground. As the student stops to pick it up, the professor says, "Don't bother – if it were really a $100 bill, it wouldn't be there."

The recent financial crisis has brought many economists and even more bankers and businessmen to distrust about the efficiency of financial markets. How could the markets be efficient if they allow for such long term deviations from their "equilibrium" and origin of persistent "bubbles"? In this thesis I will present the theoretical background for the market efficiency analysis by explaining the *Efficient market hypothesis (EMH)*, its different forms and proposed alternatives. However, the main focus will be given on specific methods to test for some forms of market efficiency. The *long memory* property, also known as *persistence*, *fractional integration* or *fractality* is a set of interrelated statistical properties that may indicate the level of market efficiency. You can basically understand it as a presence of long non-periodical deviations from the equilibrium and therefore *long memory* in market returns can indicate lower efficiency.

Interestingly, *fractional integration* is closely related to another financial time series phenomena called *scaling* or *self-similarity*. If the probability distribution of a *random process* sampled in different timescales is the same after a power-law transformation, we can say that it shows *scaling* or that it is a *self-similar* process. It shows up that this similarity of market movements on different timescales coincides with the *long memory* property and the *scaling coefficient* is the same as the coefficient of *long memory*, the famous *Hurst exponent*. *Multifractality* or *multiscaling* is then the situation when the scaling coefficient for higher moments of the series is different from the basic one.

There has been extensive research in this field during the last two decades and therefore the classical and recent theoretical papers will be presented in the literature review. Besides the papers aimed at the methodology to estimate *long memory* there will be also shown main empirical results concerning efficiency of real world equity, exchange rate and commodity markets.

In the theoretical section emphasis will be given on two themes - efficient markets theory and the theoretical basis for long memory modelling. It will be also briefly shown how the presence of long memory can be linked to inefficiency in the market.

Since the theory of *self-similar* and *long memory* processes is based on solid statistical ground, the models section has to start with the basics of random processes and its properties. There are two basic types of models used to approximate the behaviour of financial markets - continuous time and discrete time models. The most basic continuous time model is the *Brownian motion*, which was proposed for financial markets by the legendary Bachelier (1900). Its long memory counterpart, the *fractional Brownian motion*, will be also discussed and recent developments in this area will be mentioned.

Main focus will be given to the discrete time models - the presentation will start with the basic AR, MA, ARMA and ARIMA models, which all do not possess the property of long memory. The long memory model ARFIMA follows. ARCH and GARCH, models for short memory in volatilities are mentioned and FIGARCH, their long memory counterpart will be also presented.

Following chapter is the crucial part of my thesis where the methodology for estimation of long memory is presented. There are two main types of estimation methods - time domain and frequency domain. Time domain methods either use the property of long memory series that their autocorrelation function goes down very slowly with increasing Δt or cut the series into blocks and analyse how some type of spread in the block increases with the size of the block. The most well known time domain methods are the *R/S-analysis*, which is connected with the founder of the long memory analysis, H. E. Hurst (Hurst, 1951) and the *DFA*, detrended fluctuation analysis (Peng *et al.*, 1994).

Frequency domain methods compute the spectral density using *Fourier* transform and the estimation is based on certain features of the spectral density. Periodogram methods, such as *GPH* by Geweke and Porter-Hudak (1983) and its modifications use the periodogram to estimate the sample spectral density and then regress logarithm of the periodogram. On the other hand the local *Whittle* estimator is semiparametric and imposes some restrictions on the underlying distribution of the time series. Within this restrictions is it very efficient since it uses the local maximum likelihood estimation technique.

The methods chapter is concluded with a summary of bootstrap techniques. The estimators of long memory are mostly complicated with unknown small sample distribution and asymptotic properties. Therefore advanced methods are needed to compute correct confidence intervals for our estimations. Bootstrap techniques enable us to provide confidence intervals by resampling the original series we use to estimate the long memory parameter.

Some of the presented methods of long memory estimation are then compared in a small *Monte Carlo* study. On simulated ARFIMA series with various parameters we test the efficiency of the estimators, their bias and variance. We can compare the estimators by their *MSE*, mean square error, to show which one is the most suitable to determine if there is a long memory present in a market.

The best estimators chosen in the *Monte Carlo* study are then used to estimate long memory in returns and volatilities of real life financial markets - four major world exchange rates and two stock market indices, US index DJIA that represents advanced markets and Czech PX50 (now PX) as a representative of emerging markets. The *Hurst exponent* for the returns and volatilities is estimated and bootstrapped confidence intervals tell us if the estimates are significantly different from the no long memory hypothesis.

Chapter 2

Literature Review

In the literature review, I will shortly present the most relevant papers and monographs to the topic of long memory estimation. I will start with the idea of long memory and fractional integration as an extension of the short memory time series models. Then classical and modern papers introducing time domain and spectral domain estimators of long memory are presented, together with their extensions and generalisations and followed by empirical papers comparing their accuracy and efficiency. The last section of literature review will be dedicated to the applications of long memory estimation in real financial markets data, reaching from stock markets to exchange rates and commodities.

The fractional integration concept was proposed independently by Hosking (1981) and Granger and Joyeux (1980), both of them following the ideas of Mandelbrot (1965). It can be seen as an intermediate step between stationary ($I(0)$) and integrated ($I(1)$) time series. Whereas Hosking starts from the *white noise* and the *Brownian motion* as its cumulative sum, Granger's approach comes as an extension of the ARMA model halfway towards the ARIMA. Eventually they both arrive at the fractional differencing operator $(1 - B)^d$ as a generalisation of the differencing operator B defined through the Taylor expansion. Very good summary of the topic is provided by Baillie

(1996), he summarises the theoretical properties of both the fractional Brownian motion and the fractional ARIMA model, shows early extensions and tests for long memory. Also the book by Beran (1994) gives an excellent treatment of long memory processes.

The R/S analysis was first proposed by Hurst (1951) to capture non-periodic cyclical behaviour. Its based on a computationally simple method, but his further research in this field found out that there is a high risk of false positivity of his tests. The R/S statistics is sensitive to the presence of short memory and such presence is mistaken for a long memory result. Lo (1991) therefore proposed a modified R/S statistic depending on the supposed number of non-zero autocorrelations q. Teverovsky *et al.* (1999) on the other hand show that Lo's modified R/S statistics for finite samples is biased in the other direction - as the truncation lag q increases, the method shifts towards accepting the null hypothesis of no long-range dependence.

The discussion was further developed by Hall *et al.* (2000), who stresses the use of correct confidence intervals and states that the asymptotic values cannot be used for small sample estimation. Jin and Frechette (2004) presents a different correction for the t-test critical values. Alfi *et al.* (2006) also calls attention to the finite size properties of the estimator and finally, Ellis (2006), Ellis (2007) and Couillard and Davison (2005) further analyse the mis-specification of the R/S analysis and bring further ways to enhance the estimation. Finally, Mielniczuk and Wojdyllo (2007) brings one of the latest tips to enhance this classical estimator.

The *DFA*, detrended fluctuation analysis is somewhat more recent method, introduced by Peng *et al.* (1994) and originally used for completely different purposes. Its basic setup is the same, but it seems to be less prone to contamination by short memory effects. Bashan *et al.* (2008) compares different detrending methods used for this type of estimator and recommends using more types of trend-correction procedures to ensure higher quality of results.

Another estimation procedure using detrending is the *DMA*, detrended moving average proposed by Arianos and Carbone (2007). It is based on generalised variance over a moving window which is then plot in a log-log graph and a regression gives the final value of Hurst exponent. Carbone *et al.* (2004) highly values its independence on probability distribution. Alessio *et al.* (2002) compared this method to the DFA and R/S analysis and found very interesting results concerning the ability to show the scaling of a time series. Serletis and Rosenberg (2007) then use this method to compute time dependent Hurst exponent of major US stock market series and their conclusion is anti-persistence (negative long memory) of these series.

The Rescaled Variance V/R estimator was introduced by the paper of Giraitis *et al.* (2003), who shows its similarity to the Lo's modified R/S analysis and to the Kwiatkowski *et al.* (1992) KPSS test. The KPSS test is primarily used for testing against unit root, but it can be also use against fractional integration. The newly introduced V/R statistics replaces the range in R/S with variance and the scaling is accordingly readjusted. Giraitis *et al.* (2003) derives the asymptotic theory for this estimator and shows its slight superiority to the R/S and KPSS. Cajueiro and M. (2005) use this newly discovered statistics in their paper to perform a Monte Carlo study of efficiency and analyse Pacific Basin stock markets. Long range dependence is observed in volatilities, but not in returns of the markets. Lima and Tabak (2007) use this technique in their survey of emerging markets stock markets and combine it with the scaling analysis.

The seminal paper of Calvet and Fisher (2002) started a whole new approach. Their MMAR model is a generalisation of the previous work and at the same time new estimation method - the scaling analysis was introduced. The scaling function approach estimates not just the Hurst exponent but its extension, the generalised Hurst exponent which can differ from the ordinary Hurst exponent for higher moments of the analysed series. Fillol and Tripier (2003) and Fillol and Tripier (2004) provided more detailed analysis of the qualities of the novel approach. Finally, di Matteo (2007) builds on the previous papers a synthetic work with systematic explanation of theory and large number of applications. Liu *et al.* (2007) further develops this approach by introducing the Markov switching model *MSM* as an extension of the MMAR and shows that it better explains the stylized facts of the financial markets.

The seminal paper for the periodogram estimation branch was Geweke and Porter-Hudak (1983) and this type of estimator become very popular since then. Geweke and Porter-Hudak (1983) already developed the asymptotical properties of the estimator and also computed the empirical confidence intervals. Further authors seeked to improve this estimator, as for example Andrews and Guggenberger (2003) who introduced the bias-reduced form. Molinares and Reisen (2009) advise to use robust sample autocorrelation to get an estimator which is robust against additive outliers in ARFIMA models. From the newer improvements the paper of Andersson (2002) presents two modification that decrease the MSE thus objectively improving the estimator. Fillol (2007) again compares the GPH estimator to the scaling function and concludes that it is outperformed. Lopes and Mendes (2006) analyses the problem of bandwidth selection which is crucial for this type of estimators - too wide bandwidth brings excessive bias but narrow increases the variance.

The Whittle estimator stays a bit apart from the mentioned methods since it is not a graphical procedure, but it is based on the maximum likelihood estimation. Robinson (1995) came with the theoretical basis for this estimation method and many contemporary papers as took the lead. Shimotsu and Phillips (2006) presents variations of the estimator whereas Shao and Wu (2007) and Tabak and Cajueiro (2006) use it directly to assess inefficiency in stock and exchange rate markets. Wang *et al.* (2007) compares the Whittle estimator to the GPH and R/S estimators and concludes that depending on the underlying process every estimator has its strength and that is desirable to always use combination to make the final decision about the presence of long memory.

Most of the applied papers do not restrain themselves to one estimation method and combine and asses more methods to get robust results. Taqqu *et al.* (1995) provide one of the first such comparison. Clegg (2006) on the other hand gives a comprehensible summary of estimation methods and compares them on very long series (100.000 points). Grau-Carles (2005) introduces the post-blackening bootstrap approach to enhance the testing for long memory by using more precise data based confidence intervals and critical values. Bisaglia *et al.* (2006) show us another use of the bootstrap using it to make not the confidence intervals, but the estimates themselves more precise.

There are so many applied papers using long memory to characterise and compare diverse financial markets that we will mention only few as a sample from those hundreds. Matteo *et al.* (2005) compute scaling properties of markets for foreign exchange, stocks and fixed income instruments and try to differentiate their stage of development. Sadique and Silvapulle (2001) give an international survey of long memory in stock markets and finds evidence of inefficiency in some asian stock market data. Finally, Weron (2002) successfully uses R/S analysis, DFA and periodogram estimation in his search for long memory in electricity markets.

Chapter 3

Theoretical Concepts

3.1 Efficient Markets Hypothesis

The efficient market hypothesis, *EMH* is a well known theoretical concept brought to life by Eugene Fama (1970) and Paul Samuelson that is until these days constant source of conflict and even misconception. The original idea was very simple – a financial market is effective only then if it is not possible to attain higher earnings than the average of the market on a long term basis. Famas argument was that all the available information about the market, its subjects, participants and its history will be used to make profit and every new piece of information will be almost immediately incorporated into the prices, such that there is no place for arbitrage. For example the long memory about that we spoke in the introduction is not possible in such market since even the slightest long-term trend in the prices will be exploited and effectively neutralised by the traders using it.

There are three versions of the EMH - weak, semi-strong and strong (Lo, 2008). The weak form states that the market immediately uses all historical information. Therefore no trend and seasonality should exist in the market. The semi-strong form of the EMH states that in addition to the historical data, any public information about the subjects in the market is immediately used and that you can't make profit by using publicly available information. This could mean that in the same moment when the last years profit is announced the price of the firm changes accordingly.

The strong form of the EMH states that even any private (i.e. insider) information cannot be used to make extraordinary profit in an efficient market. This is really quite strong assumption and in many markets is it officially impossible to attain this because of the laws prohibiting insider trading.

What is more interesting than those three forms is the extension of this concept that has much stronger implications on the behaviour of the market, but is in fact not a conclusion of the original EMH. The public and also many economists interpret the EMH such that an efficient market must follow a martingale or even a random walk. These simplification come from the problem that the EMH is by itself very difficult to be tested. The only test possible is a test of joint hypotheses (Timmermann and Granger, 2004) about the efficiency of the market and at the same time rationality of the traders, since otherwise you couldn't guess the adequate reaction on the available information. As Malkiel (2003) argues, if you have agents that are not homogenous and completely rational then they may exist bubbles in the market because of the collective misinterpretation of the available information and nobody knows when these bubbles will burst. Thus even if the market is shown not following martingale or a random walk path, it could be the result of badly selected model of agent's behaviour, not of the inefficiency of the market.

Peters (1994) presents an alternative concept to the EMH, the fractal market hypothesis, *FMH* that consists of many types of traders with different investment horizon that perceive the information flow differently and act accordingly. The FMH does not constrain the market to any specific distribution in contrast the EMH that allegedly forces normal distribution (that would be only true under very specific restrictions, but as we said the true EMH imposes no such restriction). As we previously stated his hypothesis may contradict the misinterpretation of the EMH (that we may call RMH, rational market hypothesis) where all the agents are identical and completely rational. The real EMH does not postulate any such restriction. Peters creates a model of a financial market that can produce time series with long memory, but it has nothing to do with inefficiency of the market itself, more with the assumptions about the agents in the market.

On the other hand, there is a possible way to indicate *relative* market inefficiency. If one market behaves differently from the others it may have two reasons. Either are the agents in this market so much different from the agents in the "more efficient" markets, have different preferences and risk aversion or the market itself is different. In such case the difference may be in the efficiency level, since for example transition markets may not be attractive enough to gather the needed liquidity for smooth and efficient functioning, price manipulations may be present and above standard returns collected. Then such market may present e.g. long memory properties in contrast to the developed markets and thus enabling to prove market inefficiency. The long memory properties would mirror the fact that the events driving the price movements do not affect the prices instantaneously but over longer period of time, which allows arbitrage and is generally considered inefficient. We will now focus on the theoretical basis of long memory and methods how to test for long memory. Afterwards we will try to prove the mentioned difference in market efficiency empirically.

3.2 Notion of Long Memory and Self-similarity

The notion of a long memory in a time series was firstly formalised by British hydrologist Harold Edwin Hurst in his famous article about Nile river minima (Hurst, 1951). He found out that long term patterns are present in the water levels of the Nile river. Series of years with high water levels and high fertilising floods is followed by a series of years with low water level which leads to low crop and starvation. Mandelbrot called this phenomenon the

"Joseph effect" after the Old Testament prophet the foretold "seven years of great abundance" and "seven years of famine". Hurst proposed a theoretical treatment to statistically identify this type of dependence.

He presented the "rescaled range analysis" (R/S-analysis) as a tool to compute the optimal capacity of a water reservoir. He found out that standardised range of a long memory time series over a time interval follows a straight line in a log-log plot. The slope of this line is bigger than 0.5 for a time series with true long memory and exactly 0.5 for a series with only short memory. The coefficient is named after him the *Hurst exponent* and the value 0.5 separates long and short memory processes.

Similar dependence can even be found in data that do not have the time series character. (Smith, 1938) An interesting example of long memory can be found for spatial data. In agricultural uniformity trials, the variances of yields for individual plots depend on the Euclidian distance between them. But the decay with the distance is not exponential, but rather hyperbolic, which results in slower rate of convergence of the average variance.

Granger (1966) in his article *"The typical spectral shape of an economic variable."* shows that we can find similar relationship for many economic time series. He finds a different implication of the same effect in estimates of the spectral density. The typical shape for economic time series is a function with a pole at the origin, i.e. for the lowest frequencies, even after the seasonal and cyclical pattern is removed. He states that *"the same basic shape is found regardless of the length of the data available"*.

This is a related phenomenon called *self-similarity*, *self-affinity* or simply *scaling*. A self-similar process is always similar regardless the time scale. If the scale undergoes a power-law transformation, the probability distribution of the process stays the same. For a short memory processes, the power-law coefficient is 0.5. The well known example of this phenomenon is the *Brownian motion*, also called *random walk*. Its probability distribution differs with increasing time t only by the factor $t^{0.5}$. Long memory processes show similar behaviour, but the power of the factor is different from 0.5.

The *fractional integration*, introduced by Granger and Joyeux (1980) and Hosking (1981) is practically the same effect as long memory. The only difference is in the notation. The fractional integration coefficient, d, is equal to 0 for short memory processes and is bigger then 0 for fractionally integrated (or in other words long memory) processes. There is a simple relationship between the d and the Hurst exponent H: $H = d + 0.5$. Third related phenomenon is the *fractal dimension* of the process trajectory, hence *fractality*. The fractal dimension can be understood as a generalisation of the common dimension notion, but without further technicalities we will just state that the fractal dimension $D = 2 - H$ for uni-fractal processes.

There are several equivalent definitions of the long-memory in a time series. The definitions in the time domain (in the original series dependent on t) are based on the autocorrelation function. For a series with just a short-term memory (e.g. $ARIMA$)the autocorrelation function decays exponentially to zero with the distance between the two time points. As a result, the sum of the absolute values of autocorrelation function is finite. For a time series with long memory, the opposite is true. The autocorrelation function decays slower than exponentially, e.g. as a hyperbole. Effect of this is that the sum of absolute autocorrelations is infinite.

Alternative definition of long memory can be based on the frequency domain and the spectral density function. The periodogram (squared absolute fourier coefficients) of a long-memory series has a pole in the origin and the spectral density of a long-memory time series is proportional to ω^{-2H-1}

In terms of long memory we can divide all stationary processes into one of three classes: Persistent processes (also know and mean-reverting) with $H > 0.5$, non-persistent or short memory processes with $H = 0.5$ and anti-persistence processes which have $H < 0.5$ and instead of reverting to the mean they fluctuate even more. Example of the three types of processes can be seen in the Figure 3.1.

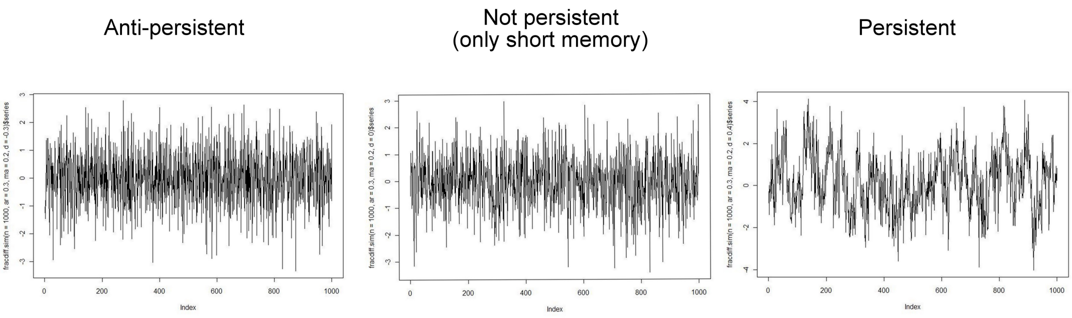

Figure 3.1: Anti-persistent, non-persistent and persistent process

If we examine more than one moment of the series (e.g. mean and variance as the first two), we can identify different persistence characteristics. This is not consistent with the model of an uni-fractal process and therefore such processes that have different Hurst exponents for different moments are known as *multifractal*.

Chapter 4

Time Series Models

The entities and theoretical concepts in this are defined according to field specific monographs (e.g. Brockwell and Davis (2006)). We start with a abstract probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ with Ω being the sample space of elementary events, $\mathcal{F}$ a sigma algebra defined on the sample space and $\mathbb{P}$ being a probability measure on Ω. We introduce an index T that can be either discrete or continuous, e.g. $T = \mathbb{R}$ or $T = \mathbb{Z}$. Now, we can state the formal definition:

Definition 1. *(Stochastic process)*
Let $(\Omega, \mathcal{F}, P)$ *be a probability space. Then a family of random variables* $\{X_t, t \in T\}$ *defined on this space is called a* ***Stochastic process****. Functions* $\{X(\omega), \omega \in \Omega\}$ *of* T *are known as the realisations of this process, or* time series.

In reality, we have just one realisation from a certain stochastic process, i.e. ω is fixed and we typically only have finite number of points, i.e. T is finite. Therefore we usually suppress the ω and simply write $\{X_t\}$. We will use the term *time series* for both the process $\{X_t\}$ and its realisation.

Now we introduce the autocovariance function as an extension of the covariance matrix for a random vector.

Definition 2. *(The Autocovariance function)*
If $\{X_t, t \in T\}$ *is a stochastic process such that* $\mathrm{Var}(X_t) < \infty$ *for each* $t \in T$*, then the autocovariance function* $\gamma_X(\cdot,\cdot)$ *of* $\{X_t\}$ *is defined by*

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \qquad r, s \in T \qquad (4.1)$$

It would be useful, if we could easily describe time series that model repetitive events. For this task, we first define the two types of stationarity - weak and strict (strong).

Definition 3. *(Stationarity - weak stationarity)*
The time series $\{X_t, t \in \mathbb{Z}\}$ *with index set* $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$*, is said to be stationary if*

(i) $E|X_t| < \infty$ *for all* $t \in \mathbb{Z}$,

(ii) $E(X_t) = m$ *for all* $t \in \mathbb{Z}$,
and

(iii) $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ *for all* $r, s, t \in \mathbb{Z}$.

This form of stationarity is in the literature often referred to as *covariance stationarity* or *weak stationarity* in contrast to strict stationarity.

Definition 4. *(Strict stationarity)*
The time series $\{X_t, t \in \mathbb{Z}\}$ *is said to be strictly stationary if the joint distributions of* $(X_{t_1}, \ldots, X_{t_k})'$ *and* $(X_{t_1+h}, \ldots, X_{t_k+h})'$ *are the same for all positive integers* k *and for all* $t_1, \ldots, t_k, h \in \mathbb{Z}$.

We can easily see that strictly stationary time series with finite second moments is also weakly stationary. The converse is not necessarily true. A weakly stationary process does not have to follow the restrictions on higher moments.

If $\{X_t, t \in \mathbb{Z}\}$ is stationary then $\gamma_X(r,s) = \gamma_X(r-s, 0)$ and we can write the autocovariance function as a function of one variable:
$\gamma_X(h) \equiv \gamma_X(h, 0) = \mathrm{Cov}(X_{t+h}, X_t)$ for all $t, h \in \mathbb{Z}$. Analogously we define the autocorrelation function (ACF) as

$$\rho_X \equiv \gamma_X(h)/\gamma_X(0) = \mathrm{Corr}(X_{t+h}, X_t) \quad \forall\ t, h \in \mathbb{Z}.$$

If $\gamma(\cdot)$ and $\rho(\cdot)$ are autocovariance and autocorrelation functions of a random process $\{X_t\}$, following properties can be proved:

- $\gamma(0) = \mathrm{Var}(X_t) \geq 0$, $\rho(0) = 1$
- $|\gamma(k)| \leq \gamma(0)$ and $\rho(k) \leq 1$ for all $k \in \mathbb{Z}$
- $\gamma(k) = \gamma(-k)$ for all $k \in \mathbb{Z}$

We further introduce the partial autocorrelation function as a correlation of residuals from an orthogonal projection on the intermediate observations:

Definition 5. *The partial autocorrelation function (PACF) $\alpha(\cdot)$ of a stationary time series $\{X_t\}$ is defined by*

$$\begin{aligned} \alpha(1) &= \mathrm{Corr}(X_2, X_1) = \rho(1) \quad \textit{and} \\ \alpha(k) &= \mathrm{Corr}(X_{k+1} - P_{\{1,X_2,\dots,X_k\}}X_{k+1}, X_1 - P_{\{1,X_2,\dots,X_k\}}X_1), \;\; k \geq 2 \end{aligned} \tag{4.2}$$

where P is the projection of X on the intermediate variables in Hilbert space $L^2\{\Omega, \mathcal{F}, \mathbb{P}\}$.

For finite samples P has the form of $a_1 + \sum_{j=2}^{k} a_j X_j$, where the coefficients a_j come from a least squares regression $X_{k+1} = a_1 + \sum_{j=2}^{k} a_j X_j + \epsilon$, ϵ independent of $X_2, \dots, X_k$. PACF is therefore a correlation of X_1 and X_{k+1} with exclusion of the influence of the intermediate variables.

Let us now define two special types of processes.

Definition 6. *(White noise) White noise is a special case of a weakly stationary process. $\{\epsilon_t\}$ is a white noise if $E(\epsilon_t) = 0$ and $\gamma_\epsilon(i,j) = \sigma^2\delta_{ij}$ where δ_{ij} is the* Kronecker symbol. *It can be written $\epsilon_t \sim WN(0, \sigma^2)$.*

Definition 7. *(Gaussian time series) The process $\{X_t\}$ is a Gaussian time series if and only if the distribution functions of $\{X_t\}$ are all multivariate normal. A stationary Gaussian process is automatically strictly stationary.*

4.1 Spectral Analysis

Sometimes it is useful to decompose a stationary time series to its individual frequencies. Using the Fourier analysis, we are able to write the series as a sum of of sinusoidal components with uncorrelated random coefficients. This is often referred as *spectral analysis* or *frequency domain anaysis* in contrast to time domain analysis which takes into account the individual observations X_t. The spectral analysis can show movements of low frequencies which would normally be covered by the high frequency fluctuations. We will now introduce spectral density of a stationary process X_t and its sample estimate, the periodogram.

Theorem 1. *(Herglotz's theorem)*
A complex-valued function $\gamma(\cdot)$ defined on the integers is non-negative definite if and only if

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\nu} dF(\nu) \quad \textit{for all } k = 0, \pm 1, \ldots, \tag{4.3}$$

where $F(\cdot)$ is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ and $f(-\pi) = 0$. The function is called ***the spectral distribution function*** *of γ and if $f(\lambda) = \int_{-\pi}^{\lambda} f(\nu)\mathrm{d}\nu$, $-\pi \leq \lambda \leq \pi$, then f is called a* ***spectral density*** *of $\gamma(\cdot)$.*

Proof: See (Brockwell and Davis, 2006), p. 118.

Since necessary and sufficient condition for an autocovariance function is that it is even and non-negative definite, we have a spectral representation for every possible autocovariance function. From the evenness of the autocovariance function we get the following property of $F(\cdot)$: for any $0 \leq a \leq b \leq \pi$, we have $F(b) - F(a) = F(-a) - F(-b)$, i.e. F has symmetric increments.

The spectral density function $f(\cdot)$ exists if F is everywhere continuous and differentiable and the equation 4.3 simplifies to

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda \quad \text{for all } k = 0, \pm 1, \ldots, \tag{4.4}$$

In terms of an autocovariance function the following theorem can be proven:

Theorem 2. *An absolutely summable complex valued function $\gamma(\cdot)$ defined on the integers is the autocovariance function of a stationary process if and only if*

$$f(\omega) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \gamma(k) \geq 0 \quad \text{for all } \omega \in [-\pi, \pi], \tag{4.5}$$

in which case $f(\cdot)$ is the spectral density of $\gamma(\cdot)$.

Proof: See (Brockwell and Davis, 2006), p. 120.

If we need a sample estimation of the spectral density function to perform the spectral analysis of a time series, we can start by constructing the periodogram. Though we must be careful with its use, since some of its asymptotic properties are undesirable.

Definition 8. *(Periodogram) Let $X_1, \ldots, X_n$ be a sequence from a stationary time series $\{X_t\}$ with meanμ and absolutely summable autocovariance function $\gamma(\cdot)$. The periodogram of $\{X_1, \ldots, X_n\}$ is defined at the Fourier frequencies $\omega_j = 2\pi j/n$, $\omega_j \in [-\pi, \pi]$, by*

$$I_n(\omega_j) = n^{-1} \left| \sum_{t=1}^{n} X_t e^{-it\omega_j} \right|^2. \tag{4.6}$$

b The periodogram is a discrete Fourier transform of the vector of observations X and it decomposes $\|X\|^2$ into sum a of components associated with n Fourier frequencies ω_j, $j \in F_n \equiv \{-[(n-1)/2], \ldots, [n/2]\}$, where $[x]$ denotes the integer part of x. Thus $\|X\|^2 = \sum_{j \in F_n} I(\omega_j)$. The raw periodogram is an unbiased, but not a consistent estimator of the spectral density function.

Theorem 3. *If X_t is stationary with mean μ and absolutely summable autocovariance function $\gamma(\cdot)$, then*

$$\begin{aligned} &\text{(i) } EI_n(0) - n\mu^2 \to 2\pi f(0) \\ \text{and} \\ &\text{(ii) } EI_n(\omega) \to 2\pi f(\omega) \text{ if } \omega \neq 0. \end{aligned} \tag{4.7}$$

(If $\mu = 0$ then $EI_n(\omega)$ converges uniformly to $2\pi f(\omega)$ on $[-\pi, \pi]$.

PROOF: See (Brockwell and Davis, 2006), p. 343.

Further analysing the convergence of the periodogram, Brockwell and Davis (2006), p. 350 prove that $I_n(\lambda)/2\pi$ is not a consistent estimator of $f(\lambda)$. But we can average the periodogram ordinates over a small neighbourhood of λ to get a consistent estimator. This procedure is called *smoothing* or *tapering*.

4.2 Brownian Motion

Brownian motion or *Wiener process* $\mathcal{W}_t$ is defined as a continuous time stochastic process with independent Gaussian increments with a distribution $\mathcal{W}_t - \mathcal{W}_s \sim \mathcal{N}(0, t-s)$. The increments process, the Gaussian noise is a stationary process, but the brownian motion itself is not. The Brownian motion has some very nice mathematical properties and it is the simplest continuous time process that can be used as a basis to create continuous time long memory process, the *fractional Brownian motion*. However we will now turn to the discrete time models and design a long memory model based on the well know Box-Jenkins ARMA methodology.

4.3 The Box-Jenkins Methodology

The most popular approach to the modelling of time series is the ARIMA methodology of Box and Jenkins (1970). It uses autoregressive and moving average sequences to supply simple and efficient models to wide class of time series. We will discuss their methodology as a short memory basis for our approach to long memory modelling. Before we start with discussing the of short memory time series modelling, we introduce a a useful simplification of our notation - the backshift operator B:

Definition 9. *(Backshift and identity operator) Let X_t be a time series. The backshift of X_t is defined as*

$$BX_t := X_{t-1},\ B^2X = B(BX_t) = X_{t-2},\ \ldots, B^kX_t = X_{t-k} \tag{4.8}$$

and the identity operator I is defined as $IX_t = X_t$.

Introducing the backshift operator B and identity operator I enables us to write a time series model in a compact form that is easy to manipulate and deal with.

4.3.1 Autoregressive (AR) Models

Suppose that the current observed value is Y_t and the p past values are available, $Y_{t-1}, \ldots, Y_{t-p}$. An AR model with order p can be expressed as follows:

$$Y_t = a_1Y_{t-1} + a_2Y_{t-2} + \cdots + a_pY_{t-p} + \epsilon_t, \tag{4.9}$$

where Y_t is weakly stationary, $a_1, \ldots, a_p$ are constants and $a_p \neq 0$. Unless otherwise stated, we assume that $\{\epsilon_t\}$ is a white noise series with zero mean and constant variance.

Using the backshift operator B, equation 4.9 can be rewritten as

$$\left(I - \sum_{r=1}^{p} a_rB^r\right) Y_t = \epsilon_t \tag{4.10}$$

or in more compact notation as $a(B)Y = \epsilon$, where $a(B)$ is a p^{th} order polynomial

$$a(B) = 1 - a_1B + a_2B^2 + \cdots + a_pB^p. \tag{4.11}$$

The process Y_t is weakly stationary if the polynomial $a(B)$ has all roots outside the unit circle (Box and Jenkins, 1970).

The ACF plot of an AR(p) process has p initial spikes and then damps out as a mixed exponential decay of order p (never 0). The PACF plot has p initial spikes and then cuts off. It makes perfect sense that in terms of ACF and PACF plots, correlations between two events become smaller and smaller as the time interval becomes larger and larger. Figure 4.1 shows an example of the ACF and PACF plots of a simulated AR(2) process with $a_1 = 0.7$ and $a_2 = 0.2$ based on 1000 observations.

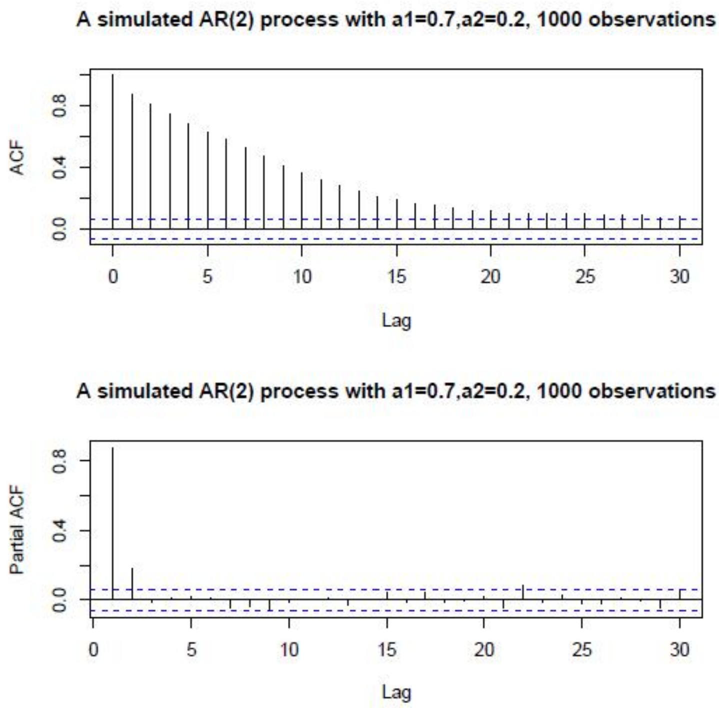

Figure 4.1: The ACF and PACF of a simulated AR(2) process

4.3.2 Moving Average (MA) Models

The AR processes are defined as a modified sum of past observations. On the other hand, the moving average (MA)is based on the past error terms, which are unobservable and cannot be accounted for by the autoregressive component. A zero mean stationary process Y_t is called a moving average process of order q if Y_t satisfies

$$Y_t = \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \cdots + b_q\epsilon_{t-q}, \tag{4.12}$$

where ϵ_t is a white noise. Again, this equation can be rewritten as

$$Y_t = (I + b_1 B + b_2 B^2 + \cdots + b_q B^q)\epsilon_t, \tag{4.13}$$

or in the more compact form as $Y = b(B)\epsilon$, where $b(B)$ is the polynomial in backshift operator B of degree q. There is a duality between the moving average process and the autoregressive process (see Box and Jenkins (1970)), that is, the moving average equation above can be inverted into an autoregressive form of infinite order. However, analogous to the stationarity condition described above, this can only be done if the moving average parameters follow certain conditions, i.e. the model is invertible.

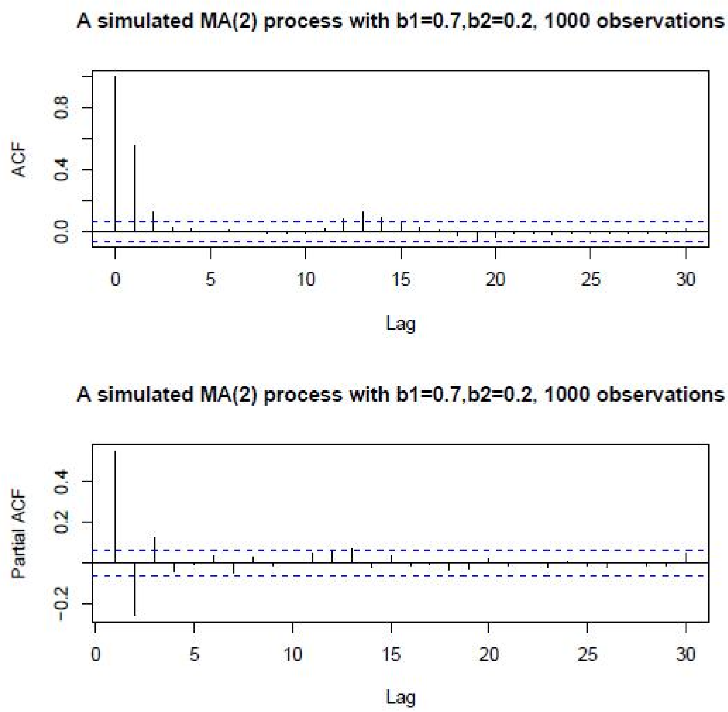

Figure 4.2: The ACF and PACF of a simulated MA(2) process

Due to the duality between MA and AR models, we can easily find that for a MA(q) model, the PACF plot has q initial spikes and then damps out as an exponential decay of order q (never 0). The ACF plot has q initial spikes and then cuts off. The sample ACF and PACF plots of a simulated MA(2) process with $b_1 = 0.7$ and $b_2 = 0.2$ based on 1000 observations are shown in figure 4.2.

4.3.3 Autoregressive Moving Average (ARMA) Models

An ARMA$(p; q)$ process is a mixture of p autoregressive components and q moving average components. In short, it can be expressed as $a(B)Y_t = b(B)\epsilon_t$; where ϵ_t is white noise, and $a(B)$ and $b(B)$ are the polynomials of degree p and q respectively. In a mixed model ARMA$(p; q)$ process, neither the ACF nor the PACF cut off at a certain lag. Both the ACF and PACF exhibit mixed exponential decay. This happens because AR(p) component brings mixed exponential decay into the ACF, while the MA component brings mixed exponential decay into the PACF. Figure 4.3 is an example of ACF and PACF plots of a simulated ARMA(1,1) process with $a_1 = 0.5$ and $b_1 = 0.2$ based on 1000 observations.

4.3.4 Autoregressive Integrated Moving Average (ARIMA) Model

In the above sections we discussed ARMA models that require weak stationarity of time series data. However, weak stationarity is not always achieved in real life series, and non-stationarity can be caused by unit roots existing in the AR component of an ARMA model. Such phenomena can be dealt with ARMA models including an extra integrating (or differencing) parameter d, i.e., ARIMA models. We define Y_t to be an ARIMA$(p; d; q)$ process if $a(B)Y_t = b(B)\epsilon_t$, where:

- p is the number of autoregressive terms,
- d is the number of differences and d takes positive integer values,
- q is the number of moving-average terms.

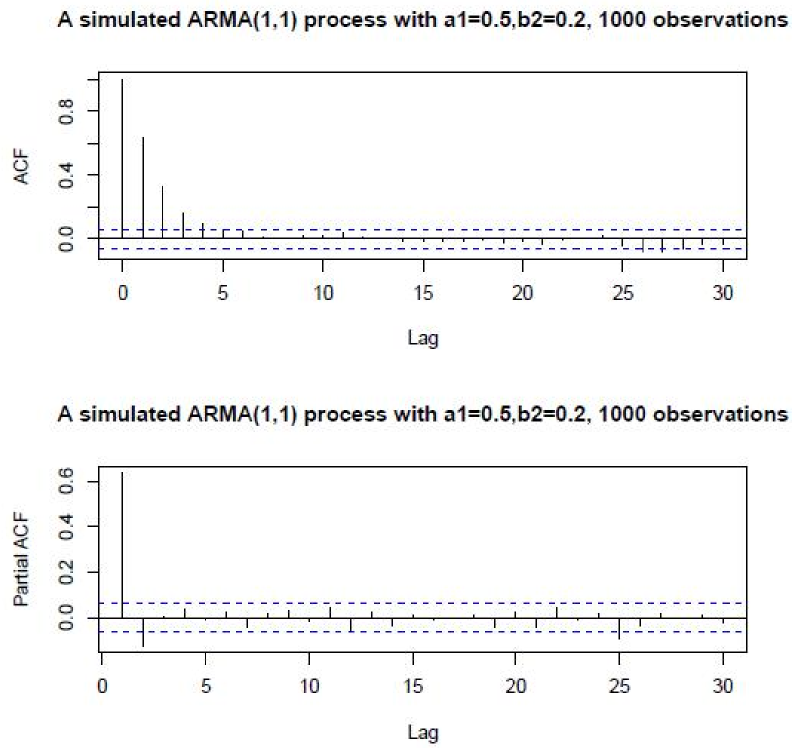

Figure 4.3: The ACF and PACF of a simulated ARMA(1,1) process

To identify the appropriate ARIMA model for a time series, we start by identifying the order of differencing needed to stabilise the series and remove the seasonality, perhaps in conjunction with a variance-stabilising transformation such as logarithmic transformation of dividing by a deflator. After stabilising the data, we can start looking for an appropriate ARMA model to fit the data. Therefore, fitting an ARIMA model is basically a combination of differencing the data and fitting an ARMA model. Figure 4.4 shows time series plots of a simulated ARIMA(1; 1; 1) process with $a_1 = 0.7$ and $b_1 = 0.2$ based on 1000 observations and its single differenced transformation. The time series plot of raw data does not exhibit stationarity. It decreases first until the lag 390 and then increases thereafter. However, after one differencing, the process seems to be more stationary with zero mean and constant variance. Hence, an appropriate ARMA model can be fitted to the differenced data.

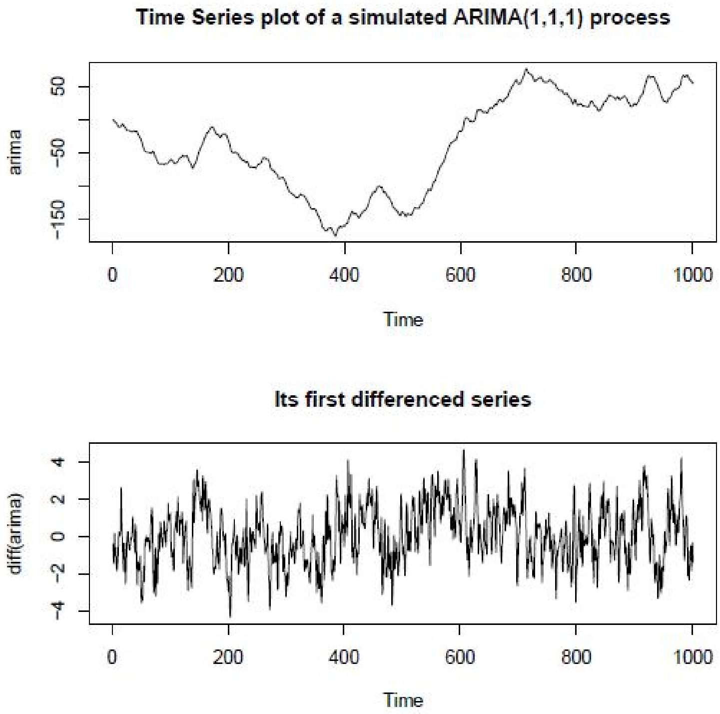

Figure 4.4: Sample path and first difference series of a simulated ARIMA(1,1,1) process

4.4 Long Memory

There are two basic theoretical models for long memory processes - Fractional Brownian motion *fBm* and Autoregressive fractionally integrated moving average (*ARFIMA*). Fractional Brownian motion, introduced by Mandelbrot (1965) is a generalisation to a standard Brownian motion. It can be perceived as a fractional derivative of the standard Brownian motion, *B(t)*.

In this thesis we will be mainly concerned with discrete time series and hence with the second model – ARFIMA. Before discussing the ARFIMA model into detail, we first introduce the formal definition of long memory in both the time and frequency domains.

Let $\{X_t\}$ be a stationary process with autocorrelation function $\gamma(\cdot)$ and spectral density $f(\cdot)$.

Definition 10. *(Time domain) $\{X_t\}$ is called a stationary process with long memory property if there exist a real number $H \in (0.5, 1)$ and a constant $c_t > 0$ such that*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_t k^{2(H-1)}} = 1, \tag{4.14}$$

where H is called the Hurst parameter and d = H - 0.5 is called the long memory parameter or fractional differencing parameter in ARFIMA$(p; d; q)$ processes.

Definition 11. *(Frequency domain) $\{X_t\}$ is called a stationary process with long memory property if there exits a constant $c_f > 0$ such that*

$$\lim_{\nu\to 0} \frac{f(\nu)}{c_f|\nu|^{1-2H}} = \lim_{\nu\to 0} \frac{f(\nu)}{c_f|\nu|^{-2d}} = 1 \tag{4.15}$$

We can see in Figure 4.5 how the long memory properties translate into the autocovariance function (slow hyperbolic decay) and into the spectral density (pole in the origin).

4.5 Autoregressive Fractional Integrated Moving Average (ARFIMA) model

Long memory processes are detected in many applications and plays an increasingly important role in time series analysis. Autoregressive Fractional Integrated Moving Average (ARFIMA) models were introduced independently by Granger and Joyeux (1980) and Hosking (1981) to deal with long memory series in discrete time. An ARFIMA$(p; d; q)$ process is defined by

$$a(B)(1-B)^d Y_t = b(B)\epsilon_t, \tag{4.16}$$

where $\{Y_t\}$ is the process of interest and $\epsilon_t \sim WN(0, \sigma^2)$; B is the backward shift operator; $a(B)$, $b(B)$ polynomials with degrees p, q respectively. The

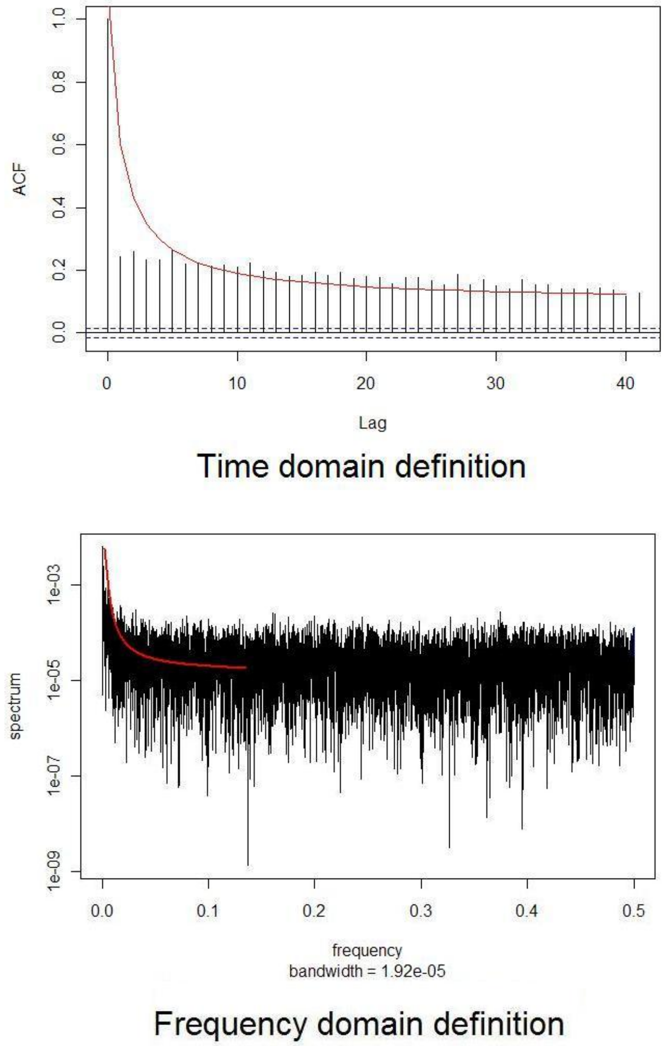

Figure 4.5: Time and spectral definition of long memory

operator $(1-B)^d$ is the fractional differencing operator defined by

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)B^k}{\Gamma(k+1)\Gamma(-d)} \tag{4.17}$$

with $\Gamma(\cdot)$ being the gamma function. ARFIMA$(p; d; q)$ is stationary and invertible if $|d| \leq 0.5$ and the roots of the $a(B)$ and $b(B)$ lie outside the unit circle. Note that the ARMA and ARIMA models can be thought of as particular cases of ARFIMA models having $d = 0$ and $d = 1, 2, \ldots$ respectively. ACF and PACF of an ARFIMA$(p; d; q)$ process never cut off. Moreover, because of the long memory property, it has slower decay of autocorrelation than those of short memory models, e.g ARMA. This is shown in Figure 4.6 by comparison of their ACF and PACF.

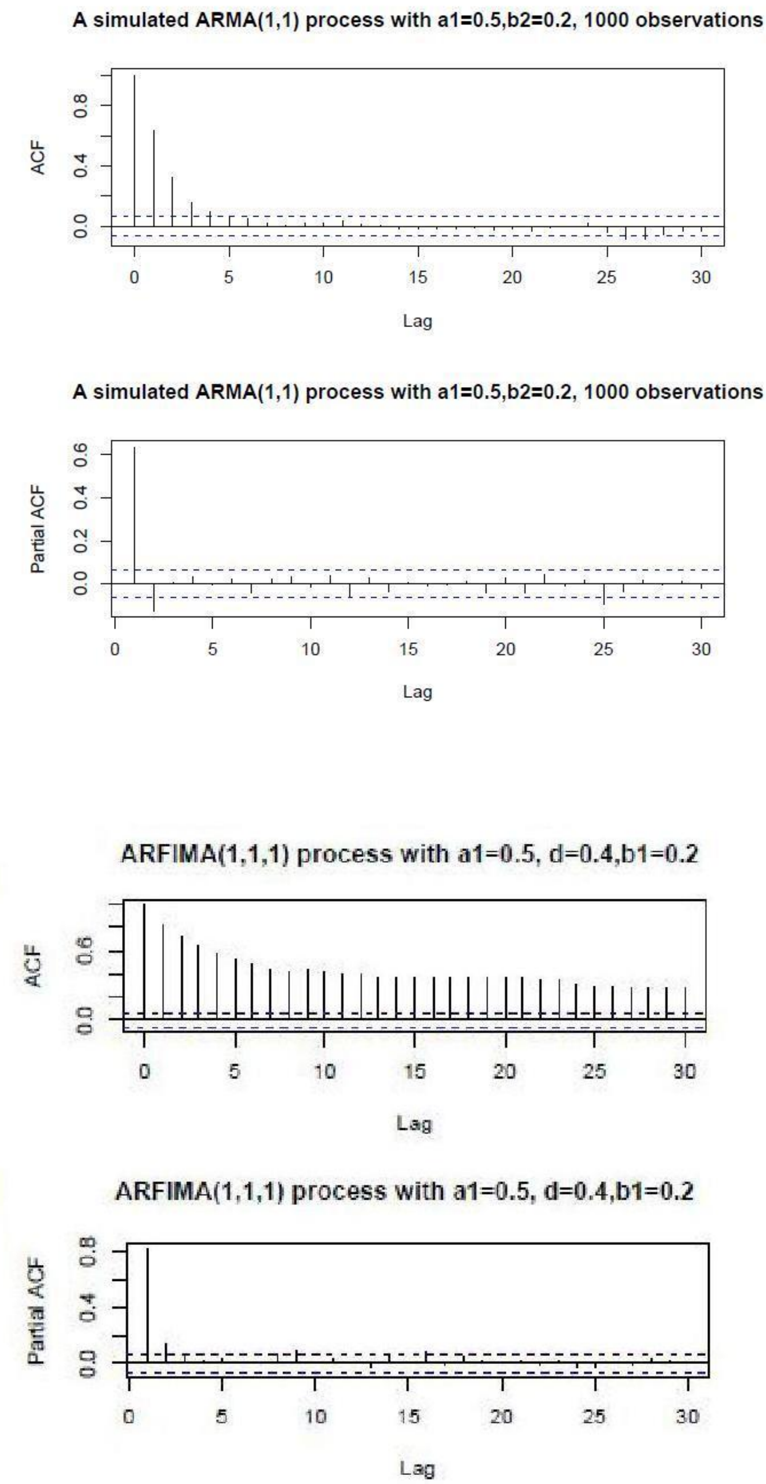

Figure 4.6: Comparison of the ACF and PACF of a simulated ARMA(1,1) and ARFIMA(1,1,1) processes

Chapter 5

Estimation Methods

The two classes of ARFIMA estimation methods are developed, i.e. two-step and one-step procedures. The majority of applications utilize the two-step approach. In the first step, an estimate of the long memory parameter d is obtained (usually in the frequency domain). In second, the data is fractionally differenced using the estimated d and a standard ARMA estimation procedure is applied to the adjusted data.

The alternative one-step method is to simultaneously estimate the long memory parameter d using the maximum likelihood procedure in either time or frequency domain. The main drawbacks of the one-step approach are the need for accurate initial guess parameter values, the potential existence of local maxima in the likelihood function and very high computational complexity. We will therefore focus our attention on separate estimators of d.

Many methods are available for detecting the existence of long-memory and estimating the fractional differencing parameter d. Some of them are well described in the monograph of Beran (1994). These techniques include graphical methods (e.g., classic rescaled adjusted range analysis, i.e., R/S analysis; aggregated variance method etc.), parametric methods (e.g., Whittle maximum likelihood estimation method) and semi-parametric method (e.g., GPH method and local Whittle method). Graphical methods are useful to heuristically test if there exists a long-range dependence in the data and to find a first estimate of d, but they generally are inaccurate and sensitive to short range serial correlations. The parametric methods obtain consistent estimates of d via maximum likelihood estimation of parametric long-memory models. They give a more accurate estimate of d, but generally require knowledge of the true model which is in fact always unknown. Semi-parametric methods, such as the GPH method of Geweke and Porter-Hudak (1983), seek to estimate d under few prior assumptions concerning the spectral density of a time series and, in particular, without specifying a finite parameter model for the d-th difference of the time series.

5.1 Time Domain Methods

5.1.1 R/S analysis

In classical R/S analysis, for a given time series $\{X_t\}, t = 1, 2, \ldots, N$, with the n-th partial sum $Y_j = \sum_{i=1}^{n} X_i, n = 1, 2, \ldots, N$ and the sample variance $S_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - n^{-1}Y_n)^2, n = 1, 2, \ldots, N$, the rescaled adjusted range statistic or R/S-statistic is defined by

$$R/S(n) = \frac{1}{S_n}\left[\max_{0\le t\le n}\left(Y_t - \frac{t}{n}Y_n\right) - \min_{0\le t\le n}\left(Y_t - \frac{t}{n}Y_n\right)\right],\ n = 1, 2, \ldots, N \tag{5.1}$$

The original specification of the classical rescaled adjusted range provided by Hurst (1951) was such that the exponent was estimated for the whole sample length N. The procedure was later modified by Mandelbrot and Wallis (1969) to incorporate OLS regression techniques where the exponent (denoted H) was estimated over several subseries, $n \le N$ as

$$\log(R/S)_n = \alpha + H\log(n) + \epsilon \tag{5.2}$$

where $\log(R/S)_n$ is the logarithm of the mean rescaled range for a subseries of length n, $\log(n)$ the logarithm of the subseries length and H the series Hurst exponent. The Hurst exponent is the slope coefficient in a regression of the individual R/S points in a log-log graph - see Figure 5.1

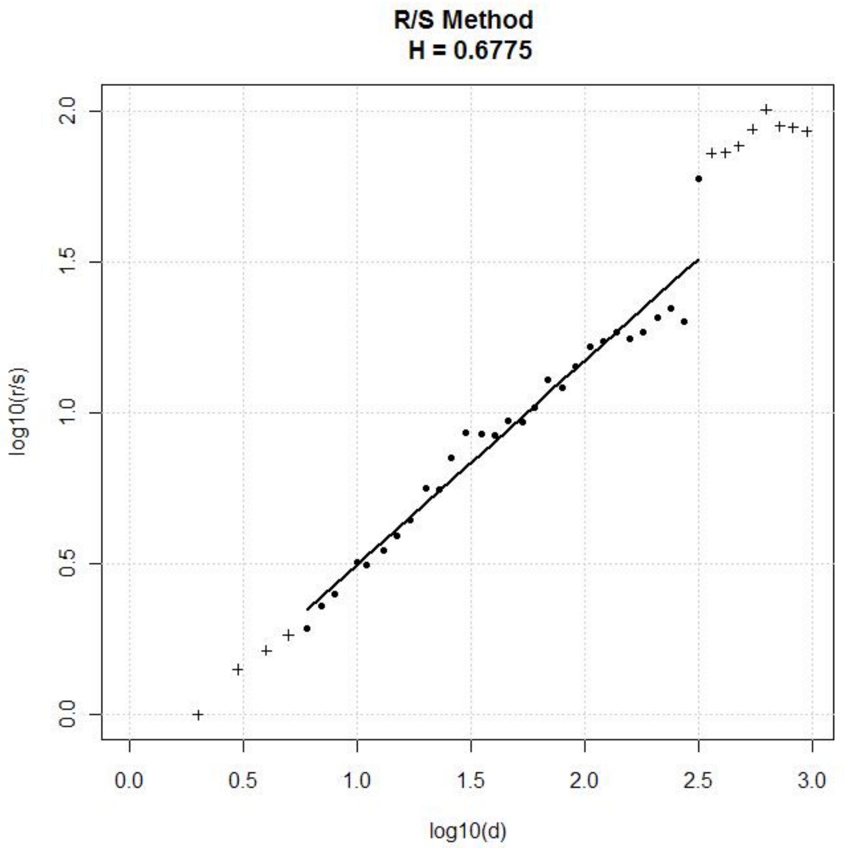

Figure 5.1: Estimation of the Hurst exponent by R/S analysis: log-log plot

The OLS procedure described in 5.2 is now the standard for estimation of the Hurst exponent by R/S-analysis. It is established that for a Gaussian series given long subseries lengths the value of the exponent will tend to its asymptotic $H = 0.5$. Yet when n is small the classical R/S-analysis may exhibit a variety of biases. The choice of n is therefore crucial for the precision of the estimator. Not less important is also the choice of individual subseries - it can be either contiguous or overlapping.

The use of alternative techniques for the division of the series length into subseries is analysed i. e. by Ellis (2007) who identifies the manner in which the series length was divided into subseries significant to the estimation of the Hurst exponent. Two alternative techniques based on overlapping subseries are mentioned - the *F-Hurst* and *G-Hurst*. Whereas F-Hurst uses full set of subseries, G-Hurst is satisfied with only a subgroup. Another approach is described by Peters (1994), *P-Hurst*, which is claimed to be superior to both previous possibilities. It uses non-overlapping subseries of lengths being powers of 2. The choice of optimal subseries (scales) is beyond the scope of this thesis, for a detailed references see di Matteo (2007) and Weron (2002).

The classical R/S analysis is sensitive to the presence of explicit short-range dependence structures, and lacks a distribution theory for the underlying statistic. To overcome these shortcomings, Lo (1991) proposed a modified R/S statistic that is obtained by replacing the denominator S_j in Eq. 5.1, i.e., the sample standard deviation, by a modified standard deviation S_q which takes into account the autocovariances of the first q lags, in order to discount the influence of the short-range dependence structure that might be present in the data. Instead of considering multiple lags as in Eq. 5.1, only focus on lag $j = n$. The S_q is defined as

$$S_q = \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2 + \frac{2}{n} \sum_{j=1}^{q} w_j(q) \left[\sum_{i=j+1}^{n} (X_i - \overline{X}_n)(X_{i-j} - \overline{X}_n) \right] \right)^{1/2} \tag{5.3}$$

where $\overline{X}_n$ denotes the sample mean of the time series, and the weights $w_j(q)$ are given by $w_j(q) = 1 - j/(q+1), q < n$. Then the Lo's modified R/S statistic is defined by

$$R/S(n,q) = \frac{1}{S_q} \left[\max_{0 \leq i \leq n} \sum_{j=1}^{i} (X_j - \overline{X}_n) - \min_{0 \leq i \leq n} \sum_{j=1}^{i} (X_j - \overline{X}_n) \right]. \tag{5.4}$$

If a series has no long-range dependence, Lo (1991) showed that given the right choice of q, the distribution of $n^{-1/2}Q_{n,q}$ is asymptotic to that of

$$W = \max_{0\leq r\leq 1} V(r) - \min_{0\leq r\leq 1} V(r). \tag{5.5}$$

where V is a standard Brownian bridge, that is $V(r) = B(r) - rB(1)$, where B denotes standard Brownian motion. Since the distribution of the random variable W is known as

$$P(W \leq x) = 1 + 2\sum_{j=1}^{\infty}(1 - 4x^2j^2)e^{-2x^2j^2}, \tag{5.6}$$

he gave the critical values of x for hypothesis testing at sixteen significance levels using Eq. 5.6, which can be used for testing the null hypothesis H_0 that there is only short-term memory in a time series at a significance level α.

5.1.2 Aggregated-variance Method

A characteristic trait of long-memory processes is that the variance of an N-member sample mean decreases more slowly than N^{-1} (Beran, 1994). He also shows that given N data points $X_i, i = 1, \ldots, N$

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = N^{2d-1}, \text{ as } N \to \infty. \tag{5.7}$$

This suggests the following method for estimating d. Divide the series into $k = N/m$ blocks of size m and compute the mean for each block

$$x_k(m) = \frac{1}{m}\sum_{(k-1)m+1}^{km} X_i, \text{where } k = 1, \ldots, N/m. \tag{5.8}$$

Variance of the block means

$$s^2(m) = \frac{1}{N/(m-1)}\sum_{k=1}^{N/m}(x_k(m) - \overline{x})^2, \text{where } \overline{x} \text{ is the overall mean.} \tag{5.9}$$

Now a log-log plot of $s^2(m)$ against m should yield a straight line with a slope of $2d - 1$. This is known as aggregated variance method.

A drawback of this method is that inhomogeneity in the data can produce a positive value of d even in the absence of long memory. A modification of the above method is called the differenced variance method, which avoids this problem.

5.1.3 Differenced-variance Method

The main idea of the differenced variance method is to study the first-order difference of the above variances

$$\nabla s^2(m) = s^2(m+1) - s^2(m). \tag{5.10}$$

Teverovsky *et al.* (1999) show that a log-log plot of this quantity against m will again asymptotically produce a straight line with slope $2d - 1$ and the value of d is not affected by the inhomogeneity of the data.

5.1.4 Detrended Fluctuation Analysis

The method of Detrended Fluctuation Analysis (DFA) by Peng *et al.* (1994) is an improvement of classical fluctuation analysis (FA), which is similar to Hurst's rescaled range R/S analysis (Hurst, 1951). They allow determining the correlation properties on large time scales. All three methods are based on random walk theory. One first calculates the 'profile'

$$X(n) = \sum_{i=1}^{n} (x_i - \langle x \rangle) \tag{5.11}$$

of a time $(x_i), i = 1, \ldots, N$ (with mean $\langle x \rangle$), which can be considered as the position of a random walker on a linear chain after n steps. Then the profile is divided into $N_s \equiv [N/s]$ non-overlapping segments of equal length ('scale') s. The mean-squared fluctuation function of the FA method is given by

$$F^2(s) = \frac{1}{N_s} \sum_{\nu=1}^{N_s} [X((\nu - 1)s) - X(\nu s)]^2 \tag{5.12}$$

In Hurst's R/S analysis, one calculates in each segment ν the range R of $X(n)$ given by the difference between maximal and minimal value, $R(s) = X_{max} - X_{min}$. The "rescaling of range" is done by dividing $R(s)$ by the corresponding standard deviation $S(s) = \sigma(X(n))$ of the same segment ν. The mean of all quotients at a particular scale s is equivalent to $F(s)$ (except for multi-fractal data) and usually shows a power-law scaling relationship with s. While both, FA and Hurst's method fail to determine correlation properties if linear or higher order trends are present in the data, DFA explicitly deals with monotonous trends in a detrending procedure. This is done by estimating a piecewise polynomial trend $y_s^{(p)}(n)$ within each segment ν by least-square fitting. I.e., $y_s^{(p)}(n)$ consists of concatenated polynomials of order p which are calculated separately for each of the segments. The detrended profile function $\tilde{X}_s(n)$ on scale s is determined by ('detrending'):

$$\tilde{X}_s(n) = X(n) - y_s^{(p)}(n). \tag{5.13}$$

The degree of the polynomial can be varied in order to eliminate linear ($p = 1$), quadratic ($p = 2$) or higher order trends of the profile function. Conventionally the DFA is named after the order of the fitting polynomial (DFA1, DFA2,...). Note that DFA1 is equivalent to Hurst's analysis in terms of detrending. The variance of $\tilde{X}_s(n)$ yields the fluctuation function on scale s

$$F(s) = \left[\frac{1}{N}\sum_{n=1}^{N}\tilde{X}_s^2(n)\right]^{1/2}. \tag{5.14}$$

This function, which has to be calculated for different scales s, corresponds to the trend-eliminated root mean square displacement of the random walker mentioned above and is related to the auto-correlation function by an integral expression.

If $F(s)$ increases for increasing s asymptotically as

$$F(s) \sim s^H, \tag{5.15}$$

with $0.5 < H < 1$, one finds Hurst exponent, H. If the type of trends in given data is not known beforehand, the fluctuation function $F(s)$ should be calculated for several orders p of the fitting polynomial. If p is too low, $F(s)$ will show a pronounced crossover to a regime with larger slope for large scales s. The maximum slope of $\log F(s)$ versus $\log s$ is $p + 1$. The crossover will move to larger scales s or disappear when p is increased, unless it is a real crossover in the intrinsic fluctuations and not due to trends. Hence, one can find p such that detrending is sufficient. However, p should not be larger than necessary, because deviations on short scales s increase with increasing p.

5.2 Frequency Domain Methods

5.2.1 Periodogram Method

The periodogram method is based on the equation 4.6. In particular, the power spectral density of a long memory process obeys a power law near the origin, i.e. $f(\nu) \sim c_f|\nu|^{-2d}$ as $\nu \to 0$. Thus, by taking logarithm on both sides, we get $\log f(\nu) \sim -2d \log(|\nu|)$, as $\nu \to 0$.

Since the spectral density $f(\nu)$ is the Fourier transform of the autocorrelation function, an estimate of the spectral density can be obtained by taking the inverse Fourier transform of the estimate of the autocorrelation function. This estimator is referred to as a periodogram $I(\nu)$. Therefore, the long memory parameter d can be estimated from the least squares regression

$$\log(I(\nu_j)) = c - 2d \log(\nu_j) + \eta_j,\ j = 1, 2, \ldots, n \tag{5.16}$$

where $\nu_j = 2\pi j/T, j = 1, \ldots, T-1, n = g(T) \ll T$, and T is the sample size. $I(\nu_j)$ is the periodogram of the series at frequency ν_j as defined by the equation 4.6.

The periodogram plot is the graph of $\{\log(\nu_j), \log I(\nu_j)\}, j = 1, 2, \ldots, n$. The typical threshold value utilised in detection of d is $n = T^{0.5}$. Theoretically the log-log plot should provide a straight line with a slope of $-2d = 1 - 2H$. An example of such regression is shown in Figure 5.2

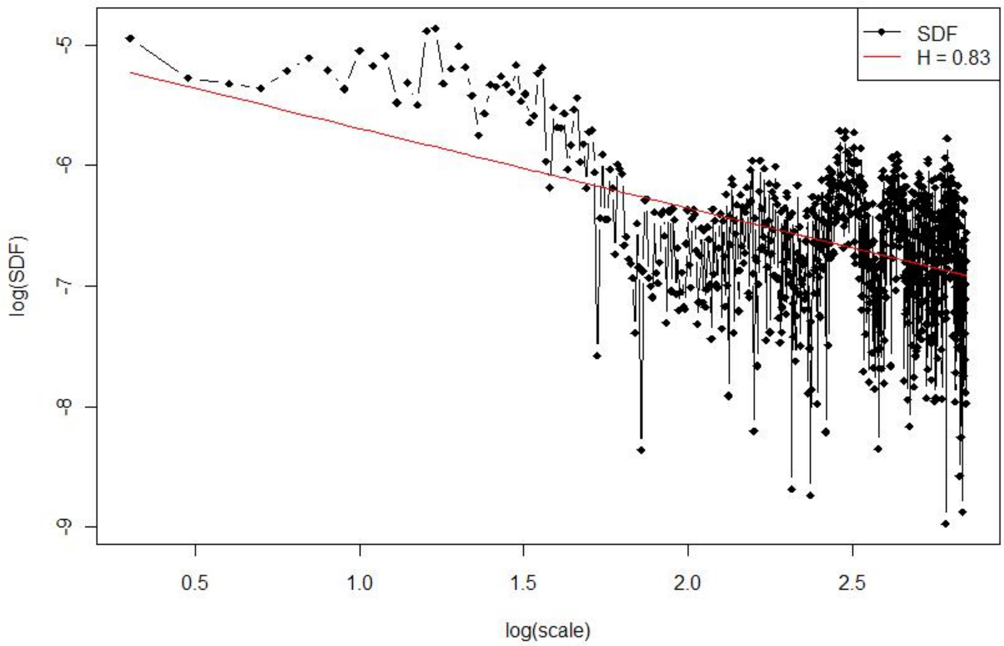

Figure 5.2: Estimation of the Hurst exponent by the periodogram method: log-log plot

5.2.2 GPH Estimator

Geweke and Porter-Hudak (1983) proposed a semi-parametric approach to the testing for long-memory. Given a fractionally integrated process $\{X_t\}$, its spectral density is given by:

$$f(\omega) = [2\sin(\omega/2)]^{-2d} f_u(\omega) \tag{5.17}$$

where ω is the Fourier frequency, $f_u(\omega)$ is the spectral density corresponding to u_t, and u_t is a stationary short memory disturbance with a zero mean. Consider the set of harmonic frequencies $\omega_j = (2\pi j/n), j = 0, 1, \ldots, n/2$, where n is the sample size. By taking the logarithm of the spectral density $f(\omega)$, we have

$$\ln f_u(\omega_j) = \ln f_u(\omega) - d\ln[4\,\mathrm{semi}^2(\omega_j/2)] \tag{5.18}$$

which may be written in the alternative form

$$\ln f_u(\omega_j) = \ln f_u(\omega) - d \ln \left[4 \sin^2(\omega_j/2)\right] + \ln[f_u(\omega_j)/f_u(0)] \tag{5.19}$$

The fractional difference parameter d can be estimated by the regression equations constructed from Eq. 5.19. Geweke and Porter-Hudak (1983) showed that using a periodogram estimate of $f(\omega_j)$, if the number of frequencies used in the regression Eq. 5.19 is a function $g(n)$ (a positive integer) of the sample size n where $g(n) = n^\alpha$ with $0 < \alpha < 1$, the least squares estimate $\hat{d}$ using the above regression is asymptotically normally distributed in large samples:

$$\hat{d} \sim N(d, \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (U_j - \overline{U})^2}) \tag{5.20}$$

where $U_j = \ln[4 \sin^2(\omega_j/2)]$ and $\overline{U}$ is the sample mean of $U_j, j = 1, \ldots, g(n)$. Under the null hypothesis of no long-memory ($d = 0$), the t-statistic

$$t_{d=0} = \hat{d} \cdot \left(\frac{\pi^2}{6 \sum_{j=1}^{g(n)} (U_j - \overline{U})^2} \right)^{-1/2} \tag{5.21}$$

has a limiting standard normal distribution. The GPH estimator is similar to the Periodogram estimator, they both use the periodogram as the left side regression variable.

5.3 Whittle Estimator

We begin with the formula for maximum likelihood estimation (MLE) of fractional differencing parameter d. The Gaussian log-likelihood of a long-memory ARFIMA process X defined by Eq. 4.16 is given by

$$\log L(\eta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} X' \Sigma^{-1} X \tag{5.22}$$

where $\eta = (a_1, \ldots, a_p; d; b_1, \ldots, b_q)$ is the parameter vector; Σ denotes the $n \times n$ covariance matrix of X depending on η and σ^2, and $|\Sigma|$ denotes the determinant of Σ. The maximum likelihood estimators $\hat{\eta}$ and $\hat{\sigma}^2$ can be found by maximising $\log L(\eta, \sigma^2)$ with respect to η and σ^2.

Due to computation problems (Beran, 1994), p. 108, an approximate MLE's are needed. According to (Wang *et al.*, 2007), the Whittle's estimator is the result of minimisation of the function:

$$L_W(\eta) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda;\eta)} d\lambda + \int_{-\pi}^{\pi} \log f(\lambda;\eta) d\lambda \tag{5.23}$$

where the subscript W stands for Whittle; $f(\lambda;\eta)$ is the spectral density and $I(\lambda)$ is the periodogram of the process. Instead of using the full maximum likelihood, we concentrate our attention just on the frequencies near to 0.

5.4 Bootstrap

The bootstrap technique can be used to estimate the small-sample distribution of a statistic. Introduced by Efron (1979), the bootstrap enables correction of size distortions for tests or data-based confidence intervals for estimations. For details, see Davison and Hinkley (1997). The original test, designed for iid observations, fails for dependent observations. There are many improvements that are aimed at dependent data. The moving-block bootstrap and model-based resampling methods perform better for short-range dependence. Short- and long-memory processes are examined using the post-blackened moving-block bootstrap method studied by Srinivas and Srinivasan (2000), which is an approach intermediate between both and appears to capture the dependence structure of the data, even using a small number of bootstrap replications. The idea underlying the block-resampling is that if the blocks are long enough, the original dependence is preserved in the resampled series.

The procedure works as follows for a given time series $x_t, t = 1, \dots, T$:

1. Compute the estimate (R/S, periodogram or DFA) and obtain $\hat{H}$
2. "Pre-whiten" the time series, fitting an $AR(p)$ model with a suitably large number of lags and obtain the estimated residuals e_t and the centered

residuals $et - \bar{e}$. The order of the autorregresive model is estimated using the Schwartz criterion.

3. Resample blocks of the centered residuals from the estimated model using the moving-block bootstrap to generate B bootstrap samples.
4. "Post-blacken" the resampled centered residuals using the estimated parameters of the AR model to generate B bootstrap samples of x denoted x^b.
5. For each bootstrapped sample, compute the statistic estimate H^b.
6. Check, if the estimated statistics is in the $p\%$ confidence interval for the null hypothesis of no long memory. Following Davison and Hinkley (1997) the estimated bootstrap p-value for a two tailed confidence interval is defined by

$$p^*(\hat{H}) = \frac{1}{B} \sum_{b=1}^{B} I(|H^b| \geq |\hat{H}|) \tag{5.24}$$

where $I(\cdot)$ equals one if the inequality is satisfied and zero otherwise. The null hypothesis is rejected when the selected significance level exceeds $p^*(\hat{H})$.

Chapter 6

Comparison of estimators

6.1 *Monte Carlo* Study

The advantage of working with simulated datasets in the context of long memory processes, i.e. ARFIMA(p; d; q), is that we know the true long memory parameters. Thus, for each simulated series, four long memory estimation methods are applied to detect the degree of persistence, which enables to carry out the detailed comparative study and evaluate the performance of each method. In our case, several parameters may influence the estimation of parameter d in ARFIMA$(p; d; q)$ models. For examples, the order of autoregressive part, i.e. p, and moving-average part, i.e. q. In particular, the aim of this simulation is to provide the insights regarding the parameter estimation of d from the four estimation methods with various p and q.

6.2 Simulation Description

The performance of each method is measured with respect to two criterions, bias and variance, that are calculated on Monte Carlo (MC) simulations of a long memory process from a given ARFIMA$(p; d; q)$ model. Bias is the average difference between the MC estimates and the true d, i.e. mean$(\hat{d} - d)$. Variance reflects the spread of the MC estimates from different simulated time series around the estimated d, i.e. var$(d_i - \hat{d})$. The method that provides the smallest variance and bias is considered to be the most preferable procedure. We will then combine these two criteria together to get the MSE, mean square error which is defined as the variance around the true value of d, i.e. var$(d_i - d)$. The MSE can be computed as $MSE = bias^2 + var$ and gives an overall measure of estimator performance.

We simulate time series processes from a selection of ARFIMA$(0; d; 0)$ and $(1; d; 1)$ models and compare the results by variance, bias ad MSE provided by each method. For each simulated data set, we use the four methods that are the most used in current literature. We detect the persistence by classical periodogram method, denoted by PER in Table 6.1, Peng's detrended fluctuation analysis denoted by DFA, classical R/S analysis denoted by R/S and Whittle estimator denoted by WHIT.

6.3 Summary of results

We generate 100 Monte Carlo simulations of an ARFIMA$(p; d; q)$ process with a sample size of 1000 observations and summarise our findings below for each considered ARFIMA model by comparing bias and variance and MSE, mean square error of the estimated d. We also show the empirical 95% confidence intervals for for the Hurst exponent H to see if it can be seen statistically significantly different from 0.5.

We have done two series of tests – first against the "pure" long memory (ARFIMA(0,d,0) and second to test for robustness against short memory contamination for ARFIMA(1,d,1) with d and MA coefficient constant of 0.3 and AR increasing from 0.2 up to 0.8. The length of the simulated series was 1000 observation which seems a reasonable number considering it is about 4 years of daily financial market data.

All time series were simulated by the classical Durbin-Levinson algorithm as presented in Durbin (1960). The computational time was in range of seconds for all the estimators except Whittle which was a bit slower and particularly DFA, which took several minutes for computation of those hundreds of estimates. For the R/S analysis we use the subseries starting from 16 up to 512 and the P-Hurst modification. The threshold for the periodogram estimator was set to $t^{0.5}$ as advised by Lobato and Robinson (1996) and Baillie (1996) and the DFA polynomial is of degree 1 - i. e. linear.

ARFIMA$(0; d; 0)$

The first series of tests shows the raw performance of the estimators in an ideal case without the presence of any jamming. In the table 6.1 we can see sample bias, sample variance and total MSE of the estimators and also the 95% empirical quantiles for all five cases. (The Whittle estimator could not be computed for the Gaussian noise.) We can conclude that the biases are low for lower levels of H but for $H = 0.8$ and 0.9 the Whittle's estimator bias is so big that the confidence interval does not cover the true value. In terms of MSE none of the estimators is uniformly better than the others. Although Whittle seems to take the lead for the lower values of H, it is beaten in the $H = 0.8$ case by all the other estimators. Second in performance may be the DFA followed by the PER etimator. The R/S performs almost consistently the worst.

The confidence intervals in the table show us that Whittle estimator has a tendency to underestimate and in the last two cases the true value is completely out of the empirical 95% quantile. The rest of the estimators seem to cover the true value quite well although the confidence intervals are pretty wide. Only the Whittle estimator can reject the null hypothesis of no long dependence for $H = 0.6$ which is for 1000 observations a weak result, but all the other three lower bounds are very close to rejecting.

ARFIMA$(1; d; 1)$

In this subsection we consider the multiple-parameter case ARFIMA$(1; d; 1)$. Those models allow to combine the long and short memory components. Additional part of the estimation error arises here through the difficulty in discerning between the AR part and the long memory associated with H, since the AR part and long memory parameter H can imply similar patterns of autocorrelation for the first few lags. This agrees with our simulation results presented in table 6.2. Our findings reveal that MSEs of all estimators are higher than that in ARFIMA $(0; d; 0)$. The results for one of the models can be analysed also in the form of histograms for all four estimators in Figure 6.1

If we compare the results for individual estimators we the the R/S performs consistently without being affected by the AR component. the DFA is again one of the best, having bad performance only for the highest AR-contamination. The periodogram estimator performs very good in the first two cases but then shows very high bias and also MSE. Finally the Whittle estimator completely breaks down for the last two models, gravely overestimating the true value of the Hurst exponent H. Overall we can say the for the highest AR contamination of the data (AR coefficient $= 0.8$) are the estimator completely unusable, three of four not covering the true value of estimated parameter in the empirical confidence intervals.

The boxplot diagrams from all estimates in this chapter are present in the Appendix.

Method	BIAS	VAR	MSE	$Q_{2.5\%}$	$Q_{97.5\%}$
Gaussian noise - H=0.5					
R/S	0.0569	0.0080	0.0112	0.3984	0.7317
DFA	-0.0174	0.0019	0.0022	0.4008	0.5767
PER	-0.0044	0.0060	0.0060	0.3624	0.6713
ARFIMA(0,0.1,0) - H=0.6					
R/S	0.0233	0.0062	0.0068	0.4894	0.7730
DFA	-0.0284	0.0017	0.0025	0.4901	0.6525
PER	0.0080	0.0051	0.0051	0.4617	0.7451
WHIT	-0.0193	0.0004	0.0008	0.5486	0.6255
ARFIMA(0,0.2,0) - H=0.7					
R/S	0.0115	0.0096	0.0097	0.5024	0.8789
DFA	-0.0326	0.0018	0.0028	0.5877	0.7390
PER	0.0121	0.0048	0.0050	0.5839	0.8334
WHIT	-0.0376	0.0004	0.0018	0.6262	0.7039
ARFIMA(0,0.3,0) - H=0.8					
R/S	-0.0407	0.0083	0.0088	0.5991	0.9155
DFA	-0.0369	0.0023	0.0085	0.6846	0.8514
PER	0.0279	0.0057	0.0079	0.6822	0.9500
WHIT	-0.0595	0.0005	0.0107	0.7059	0.7801
ARFIMA(0,0.4,0) - H=0.9					
R/S	-0.0737	0.0061	0.0115	0.6692	0.9660
DFA	-0.0368	0.0035	0.0048	0.7695	0.9725
PER	0.0372	0.0062	0.0076	0.7426	1.0576
WHIT	-0.0681	0.0004	0.0050	0.7908	0.8663

Table 6.1: Comparative analysis of the four methods for estimation of the long memory parameter d in ARFIMA$(0; d; 0)$ processes with 1000 observations and 100 Monte Carlo simulations per model.

Method	BIAS	VAR	MSE	$Q_{2.5\%}$	$Q_{97.5\%}$
ARFIMA(0.2,0.3,0.3) - H=0.8					
R/S	-0.0407	0.0099	0.0115	0.6111	0.9567
DFA	-0.0655	0.0027	0.0070	0.6229	0.8382
PER	0.0082	0.0057	0.0058	0.6782	0.9456
WHIT	-0.1158	0.0004	0.0138	0.6402	0.7209
ARFIMA(0.4,0.3,0.3) - H=0.8					
R/S	-0.0227	0.0077	0.0082	0.6115	0.9313
DFA	-0.0134	0.0029	0.0031	0.6881	0.8871
PER	0.0447	0.0047	0.0067	0.7019	0.9510
WHIT	0.0103	0.0005	0.0006	0.7599	0.8539
ARFIMA(0.6,0.3,0.3) - H=0.8					
R/S	0.0304	0.0074	0.0084	0.6352	0.9906
DFA	0.0782	0.0023	0.0084	0.7787	0.9697
PER	0.1389	0.0061	0.0254	0.7817	1.0925
WHIT	0.1558	0.0003	0.0245	0.9237	0.9802
ARFIMA(0.8,0.3,0.3) - H=0.8					
R/S	0.0847	0.0055	0.0127	0.7121	0.9917
DFA	0.2408	0.0030	0.0610	0.9393	1.1475
PER	0.3585	0.0066	0.1350	1.0070	1.3161
WHIT	0.1976	$3 \cdot 10^{-7}$	0.0390	0.9965	0.9984

Table 6.2: Comparative analysis of the four methods for estimation of the long memory parameter d in ARFIMA$(1; d; 1)$ processes with AR coefficient increasing from 0.2 to 0.8, 1000 observations and 100 Monte Carlo simulations per model.

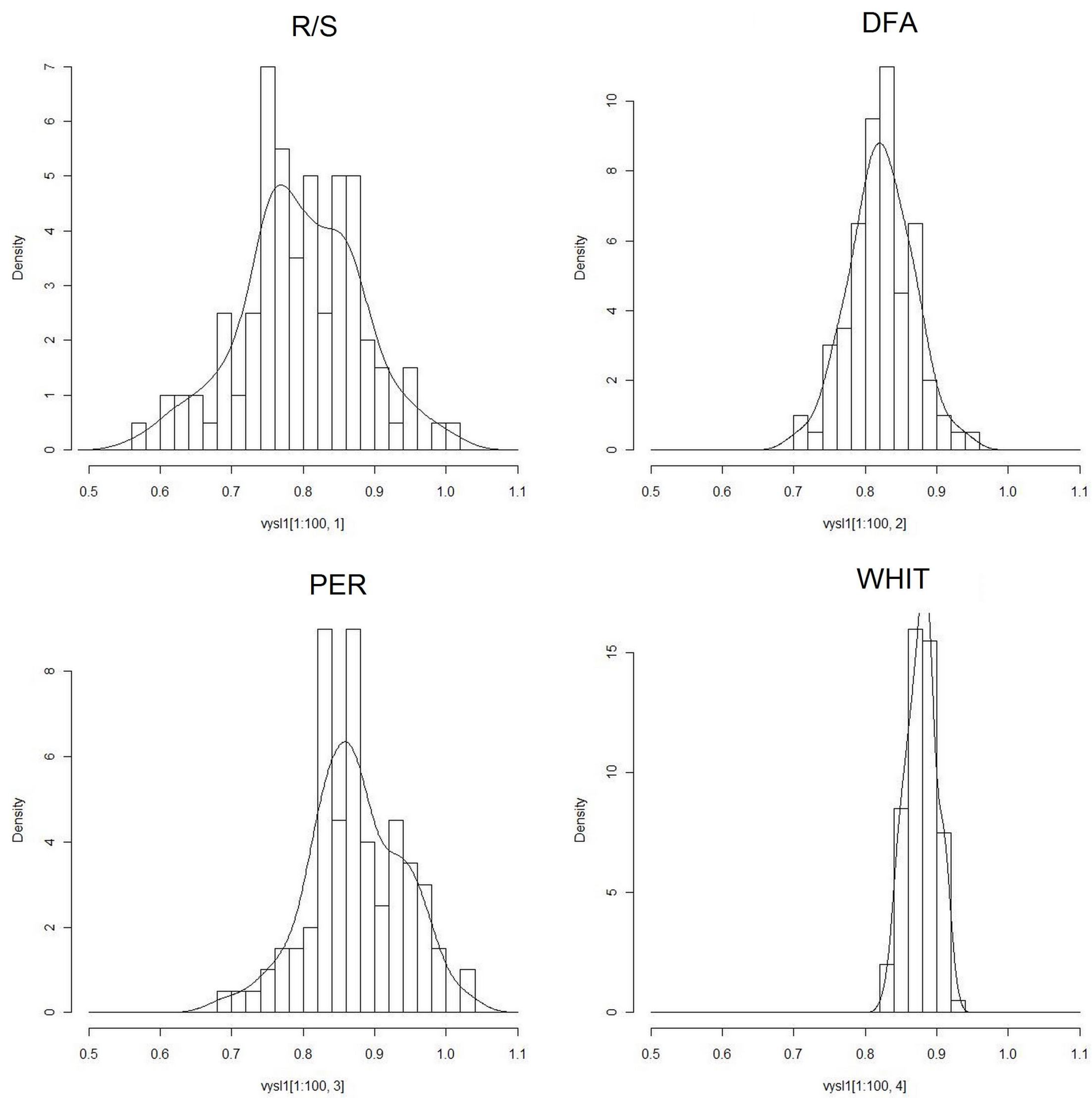

Figure 6.1: Histograms of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0.4,0.3,0.3), 1000 observations, 100 samples

Chapter 7

Efficiency of Financial Markets

7.1 Data Description

We examine the long range dependence of four exchange rate series and returns and volatilities of two stock market index series. Firstly, we consider the nominal dollar spot rates per US Dollar for British pound (GBP), Euro (EUR), Swiss frank (SFR), and Japanese yen (JPY). Daily exchange rates from January 1974 to December 1987 are obtained from the FXHistory website at: http://www.oanda.com/convert/fxhistory. The Euro data are only available from December 15, 1998 to December 30, 2004. Time series plots of the data are presented in Fig. 7.1.

The two chosen stock markets indices are Dow Jones Industrial Average (DJIA) from NYSE (New York Stock Exchange) as a representant of a developed market and PX50, the index of Prague Stock Exchange as a typical transition country stock market. The data are obtained from Yahoo Finance, http://finance.yahoo.com. The DJIA series is a daily index value and reaches from January 1950 to December 2008 having 14825 observations. The PX50 series also has a daily timescale and reaches from July 1997 until December 2008 having 2825 observations. The time series plots are in Fig. 7.2.

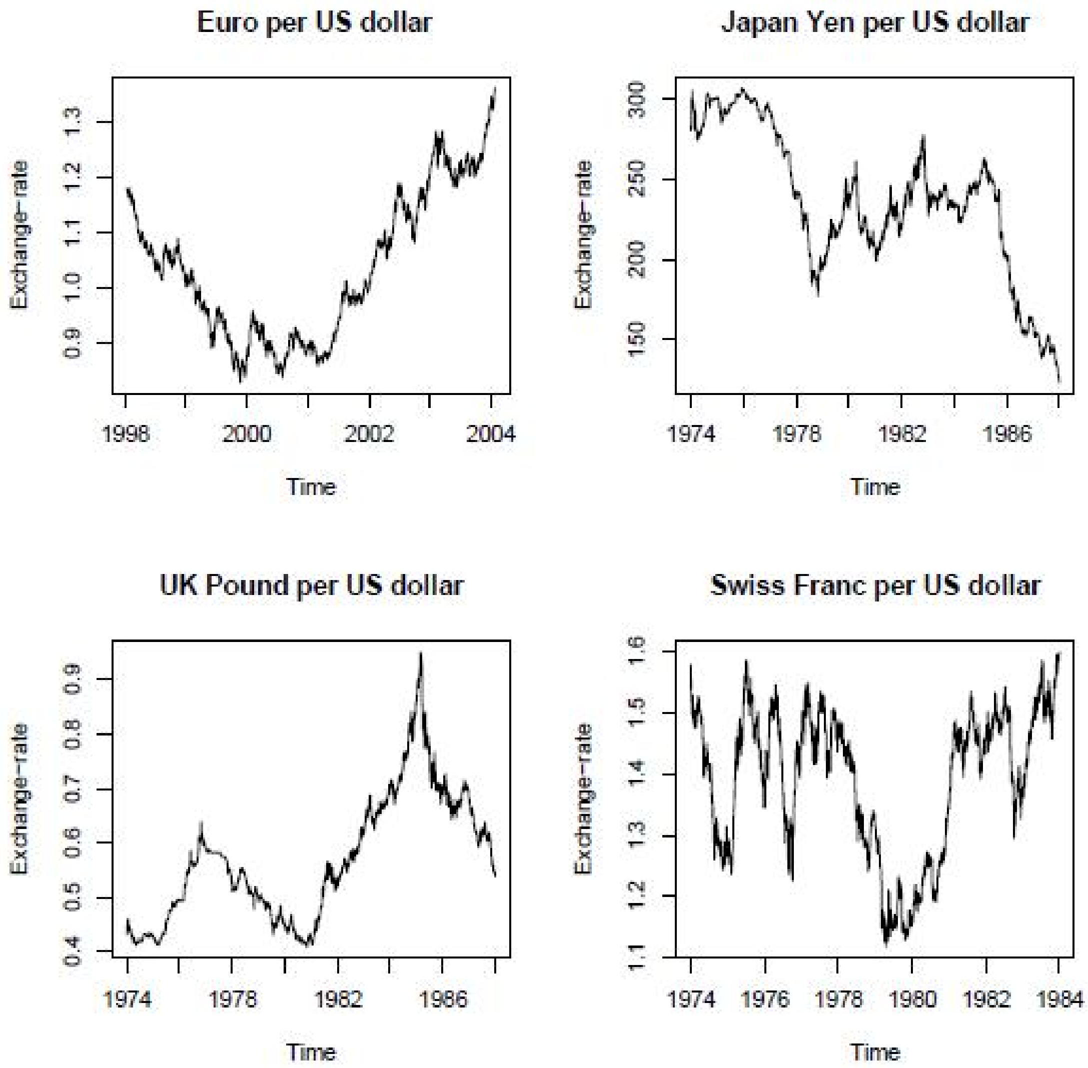

Figure 7.1: Time series plots of four major currency exchange rates

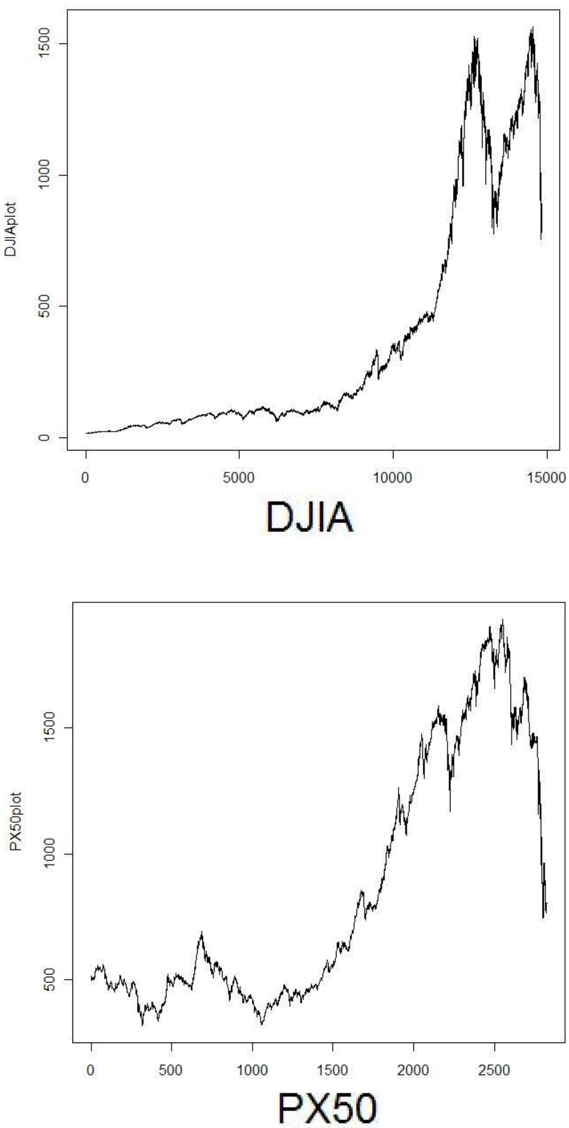

Figure 7.2: Time series plots of two stock market return series

7.2 Preliminary Analysis

We start from the preliminary transformations of the raw data in order to make the time series being weakly stationary. The logarithmic transformation and classical differencing techniques are applied to all data sets. Graphs of the first-differenced log data are presented in Figure 7.3 Exchange-rate changes appear to have zero mean and constant variance over the sample period, which implies that the filtered (transformed) data are approximately weakly stationary.

Indeed, the augmented Dickey-Fuller test for a unit root on each individual exchange-rate change series yields p-values of less than 0.1, which along with the empirical assessment of Figure 7.3 indicates the likely weak stationarity. One should be careful though in relying solely on the Dickey-Fuller test, since this standard unit-root test has low power for long memory processes.

The analysis of the stock market series is similar. The log-differenced data are presented in Figure 7.4. Also the stock market data appear to have zero mean, but the variance does not seem to be constant. We will for now settle with this mild heteroskedasticity and analyse it later when we will be testing for long memory in volatilities.

7.3 Summary of Results

The following methods are applied to estimate the long memory parameter H: R/S analysis (R/S), Detrended fluctuation analysis (DFA), classical periodogram method (PER) and Local Whittle estimator (WHIT). The point estimates will be supplemented by bootstrapped confidence intervals computed according to the algorithm in 5.4 with blocks of size 32.

We can say there exists an evidence of long memory if the estimate of H is significantly larger than 0.5. The results from various methods are compared, and our conclusion is that we can not solely rely on a single method. The results supported by most methods can be considered more reliable. Table

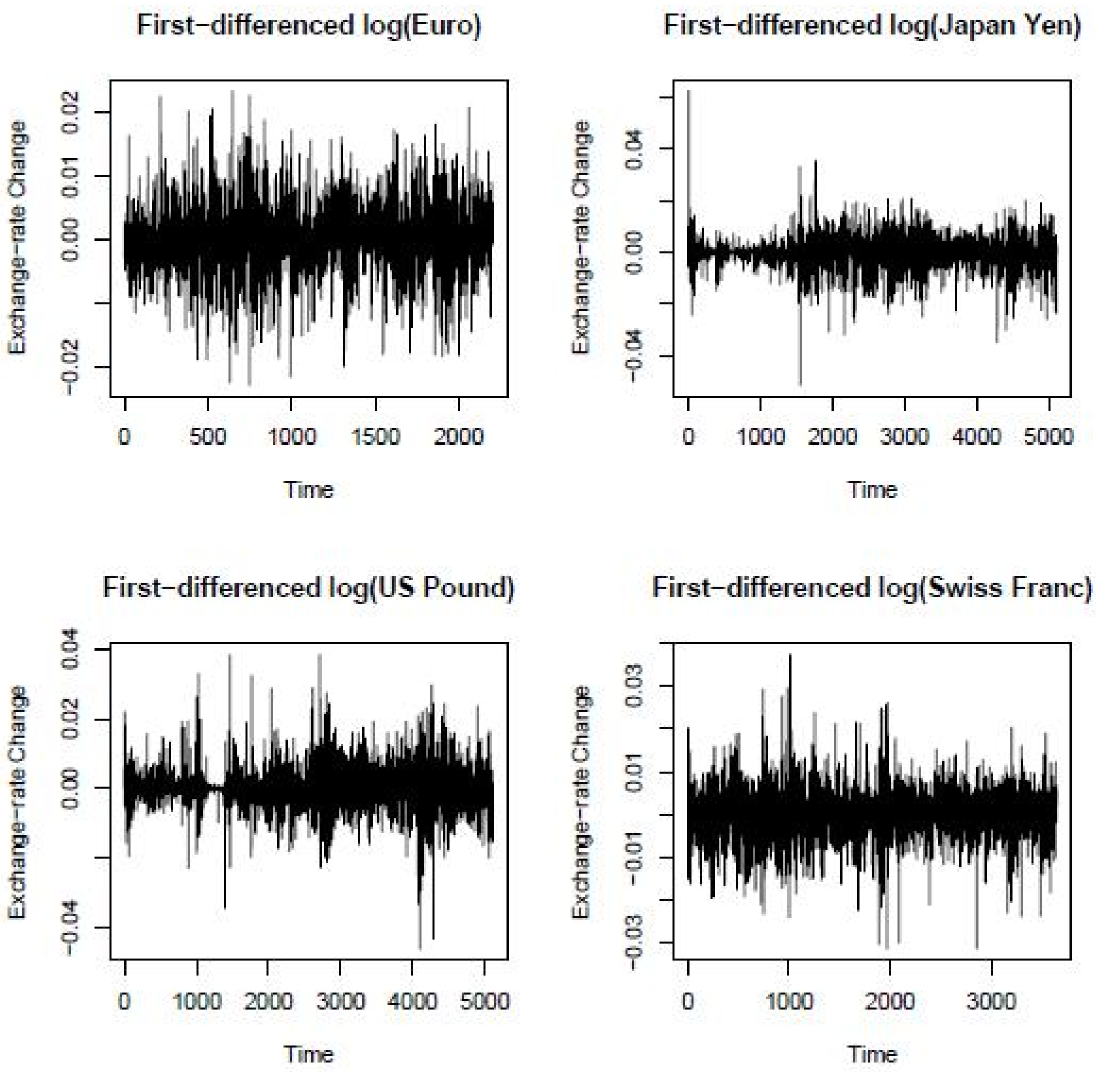

Figure 7.3: Four major currency exchange rates changes

7.1 presents the estimates of H from various methods for four major foreign exchange-rate change series and the respective bootstrapped confidence intervals. Following table 7.2 presents the estimates of H for DJIA and PX50 returns and volatilities. Volatilities are estimated as absolute returns.

For Euro most of the methods suggest that H is only slightly over 0.5 and the bootstrapped confidence intervals do not reject the null hypothesis

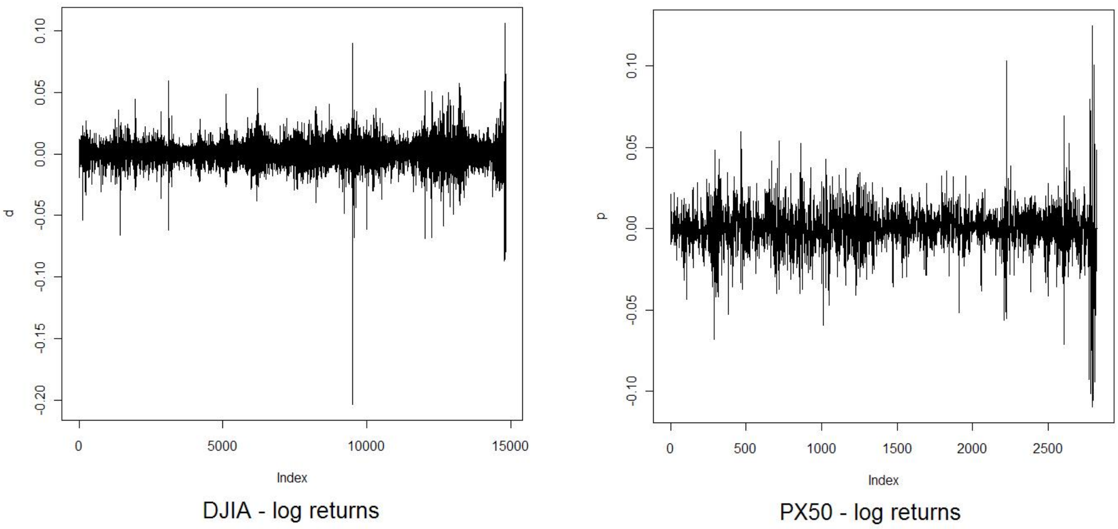

Figure 7.4: Stock market log-returns for DJIA and PX50

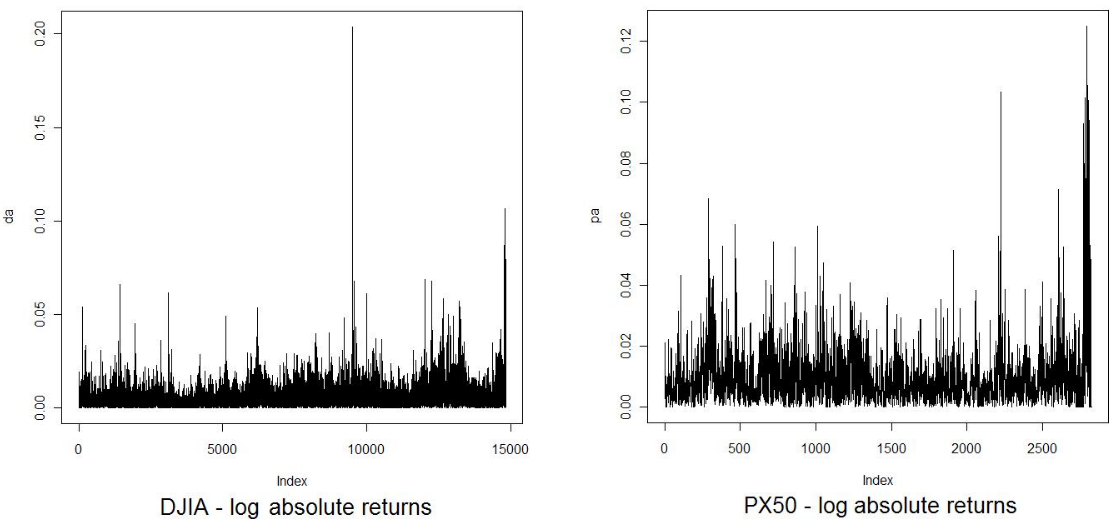

Figure 7.5: Stock market log-absolute returns for DJIA and PX50

of no long memory for all methods except the Whittle estimator. Seeing the performance of Whittle estimator in the Monte Carlo study we must conclude that we cannot reject the null hypothesis of no long range dependence.

When testing the GBP series, DFA and Whittle estimator according to the bootstrapped confidence intervals reject the null hypothesis of only short memory, but the other two methods do not. Three of the estimates are over 0.65 and the lower confidence bound of the Per estimate being 0.49 gives us GBP as the most likely candidate for a long memory presence.

For the last two exchange rate series, the JPY and SFR, only the Whittle method suggests that the value of H is elevated. It seems to be an outlier, since all the other methods do not report any higher values. Also the confidence intervals do not reject the null hypothesis. Of the four exchange rate series, the most plausible is the existence of long memory in the GBP time series and even there it is questionable.

For the stock market time series, we estimate the H for returns and as a proxy for market volatility we use the absolute returns. Use of squared returns is also mentioned in the literature, but absolute returns perform as well and are asymptotically equivalent. In the Figure 7.6 the plots of the periodogram estimator are shown for comparison for both the DJIA and PX50 returns and volatilities.

The DJIA returns give for two estimation methods the value of H over 0.5 and for the other two below 0.5. In combination with the bootstrapped confidence intervals we conclude that there is no long memory present in the DJIA returns, either positive or negative. In the second series, the PX50 returns we can see that three of four of the H estimates are over 0.5, but only slightly and the confidence intervals lead us the the following verdict – no evidence of long memory.

On the other hand both of the series of market volatilities, DJIA and PX50 show strong evidence of long memory in volatility. Also the bootstrapped confidence intervals suggest that the null hypothesis of no long memory should be rejected. Therefore our final statement is that both DJIA and PX50 volatility is strongly persistent. Standard volatility modelling techniques such as ARCH

Method	$\hat{H}$	$Q_{2.5\%}$	$Q_{97.5\%}$
EUR			
R/S	0.5503	0.3784	0.6874
DFA	0.5575	0.3812	0.7132
PER	0.5250	0.3622	0.6636
WHIT	0.7877	0.5231	0.9656
GBP			
R/S	0.5633	0.4182	0.7044
DFA	0.6824	0.5257	0.8192
PER	0.6656	0.4912	0.7875
WHIT	0.6907	0.5437	0.8593
JPY			
R/S	0.5933	0.4732	0.7530
DFA	0.5515	0.4214	0.6637
PER	0.5631	0.4439	0.6902
WHIT	0.7619	0.5682	0.8876
SFR			
R/S	0.5191	0.4088	0.6269
DFA	0.5239	0.3774	0.6691
PER	0.6306	0.4941	0.7244
WHIT	0.6602	0.5369	0.8005

Table 7.1: Estimates of H for four exchange-rate change series

and GARCH would then produce biased results which could be significantly improved by the use of FIGARCH, fractionally integrated GARCH model.

We can say that we have proven the presence of multifractality in both DJIA and PX50 stock market indices. The market returns (first moments) show no indication of persistence ($H = 0.5$) while the variance is persistent ($H > 0.5$). Thus the moments of the stock market returns series (both developed and emerging) show different degree of persistence (different Hurst exponents) which is the definition of multifractality.

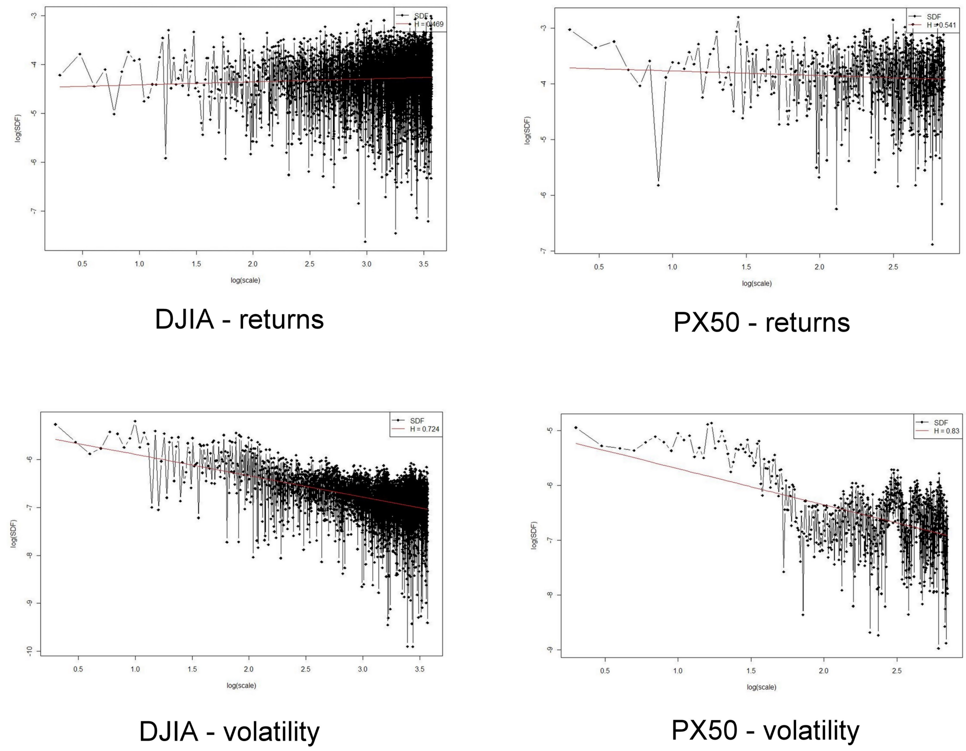

Figure 7.6: Comparison of the periodogram estimator for DJIA and PX50

Method	$\hat{H}$	$Q_{2.5\%}$	$Q_{97.5\%}$
	DJIA		
R/S	0.5667	0.4882	0.6340
DFA	0.4647	0.3480	0.5547
PER	0.4769	0.3605	0.6010
WHIT	0.5210	0.4118	0.6037
	PX50		
R/S	0.5586	0.4223	0.7055
DFA	0.4872	0.3901	0.6397
PER	0.5235	0.3778	0.6859
WHIT	0.5249	0.3632	0.6781
	DJIA – volatility		
R/S	0.8014	0.6907	0.9447
DFA	0.7330	0.5279	0.8351
PER	0.8686	0.6736	0.9755
WHIT	0.6601	0.5230	0.7709
	PX50 – volatility		
R/S	0.6774	0.5189	0.8306
DFA	0.7176	0.5822	0.8633
PER	0.9850	0.7723	1.0959
WHIT	0.6563	0.5465	0.7977

Table 7.2: Estimates of H for stock market returns and volatilities

Chapter 8

Conclusion

The main goal of this thesis was to present the possible relationship between persistence in returns or volatilities of financial markets and their efficiency. We have thoroughly analysed the theoretical basis of the *Efficient market hypothesis* and its variant versions. The loose formulation of the EMH allows for apparent irregularities in the market as long as they are not permanent and are coming from the preferences of the agents using the market. One possible way to indicate market inefficiency would be to show that a market inhabited by essentially similar agents as other markets shows consistently abnormal behaviour. The reason for the abnormality would then be in the market itself, not in the behaviour of the agents. Long memory in the market returns could eventually be interpreted as a signal of such anomalous development in a financial market.

We have presented the theoretical basis for several long memory models and many corresponding estimators. Some of them we tested in a Monte Carlo study and compared their performance on simulated artificial long memory time series. The results of the study strengthened the importance of correct bootstrapped confidence intervals due to often large bias of the selected methods. Finally we tried to prove long memory presence in the PX50 returns time series, an impersonator of an emerging market. The results differed slightly from those of the DJIA index, a representative of the developed markets group. Whereas the former showed light tendency to persistence, the latter tended more to anti-persistent behaviour. Unfortunately none of the results was significant, which would enable us to prove our point.

Despite the momentarily inconclusive results is the question of long memory in returns still tempting and attractive. And our second empirical result, the common presence of long memory in volatility in both the developed and emerging markets tends to be almost equally important and interesting and urging for future research. We have at least proven the multifractal nature of the chosen stock markets, since the Hurst exponent for the second moment (variance) is significantly different from the Hurst exponent of the first moment (returns), which is typical for multifractal series.

Bibliography

ALESSIO, E., CARBONE, A., CASTELLI, G. and FRAPPIETRO, V. (2002). Second-order moving average and scaling of stachastic time series. *The European Physical Journal B*, **27**, 197–200.

ALFI, V., COCCETTI, F., MAROTTA, M., PETRI, A. and PIETRONERO, L. (2006). Exact results for the roughness of a finite size random walk. *Physica A: Statistical Mechanics and its Applications*, **370** (1), 127–131.

ANDERSSON, J. (2002). An improvement of the gph estimator. *Economics Letters*, **77**, 137–146.

ANDREWS, D. and GUGGENBERGER, P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica*, **71** (2), 675–712.

ARIANOS, S. and CARBONE, A. (2007). De trending moving average algorithm: A closed-form approximation of the scaling law. *Physica A*, **382**, 9–15.

BACHELIER, L. (1900). Theorie de la speculation. *Annales de l'Ecole normale superiure.*

BAILLIE, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5–59.

BASHAN, A., BARTSCH, A., KANTELHARDT, J. and HAVLIN, S. (2008). Comparison of detrending methods for fluctuation analysis. *Physica A*, **384** (21), 5080–5090.

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

Bisaglia, L., Bordignon, S. and Cecchinato, N. (2006). A new bootstrap approach for long memory gaussian time series. *W.P. del Dip. di Sc. Statistiche*, **6**.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco.

Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and methods*. New York: Springer, 2nd edn.

Cajueiro, D. and M., T. B. (2005). The rescaled variance statistic and the determination of the hurst exponent. *Mathematics and Computers in Simulation*, **70**, 172–179.

Calvet, L. and Fisher, A. (2002). Multifractality in asset returns: Theory and evidence. *Review of Economics and Statistics*, **84** (3), 381–406.

Carbone, A., Castelli, G. and Stanley, H. (2004). Time-dependent hurst exponent in financial time series. *Physica A: Statistical Mechanics and its Applications*, **344** (1-2), 267–271.

Clegg, R. (2006). A practical guide to measuring the hurst parameter. *International Journal of Simulation Systems, Science and Technology*, **7** (2), 3–14.

Couillard, M. and Davison, M. (2005). A comment on measuring the hurst exponent of financial time series. *Physica A: Statistical Mechanics and its Applications*, **348**, 404–418.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge Univesity Press.

di Matteo, T. (2007). Multi-scaling in finance. *Quantitative Finance*, **7** (1), 21–36.

Durbin, J. (1960). The fitting of time series models. *Review of the International Institute of Statistics*, **28**, 233–244.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Ellis, C. (2006). The mis-specification of the expected rescaled adjusted range. *Physica A: Statistical Mechanics and its Applications*, **363**, 469–476.

— (2007). The sampling properties of hurst exponent estimates. *Physica A: Statistical Mechanics and its Applications*, **375** (1), 159–173.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, 383–417.

Fillol, J. (2007). Estimating long memory: Scaling function vs andrews and guggenberger gph. *economics letters*, **95**, 309–314.

— and Tripier, F. (2003). The scaling function-based estimator of the long memory parameter: a comparative study. *Economics Bulletin*, **3** (23), 1–7.

— and — (2004). The scaling function-based estimator of the long memory in the presence of a short-term component. *Economics Letters*, **84**, 49–54.

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

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssiere, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, **112**, 265–294.

Granger, C. W. J. (1966). The typical spectral shape of an economic variable. *Econometrica*, **34**, 150–161.

— and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1** (1), 15–29.

Grau-Carles, P. (2005). Tests of long memory: A bootstrap approach. *Computational Economics*, **25**, 103–113.

Hall, P., Härdle, W., Kleinow, T. and Schmidt, P. (2000). Semiparametric bootstrap approach to hzpothesis tests and confidence intervals for the hurst coefficient. *Statistical Inference for Stochastic Processes*, **3** (3), 263–276.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68** (1), 165–176.

Hurst, H. E. (1951). The long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770–799.

Jin, H. J. and Frechette, D. L. (2004). A new t-test for the r/s analysis and long memory in agricultural commodity prices. *Applied Economics Letters*, **11**, 661–667.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**, 159–178.

Lima, E. and Tabak, B. M. (2007). Testing for inefficiency in emergine markets exchange rates. *Chaos, Solitons and Fractals*, **33**, 617–622.

Liu, R., Di Matteo, T. and Lux, T. (2007). True and apparent scaling: The proximitz of the markov-switching multifractal model to long-range dependence. *Physica A: Statistical Mechanics and its Applications*, **383** (1), 35–42.

Lo, A. (2008). *The New Palgrave: A Dictionary of Economics*, Palgrave MacMillan, chap. Efficient Markets Hypothesis.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, **59** (5), 1279–1313.

Lobato, I. and Robinson, P. (1996). Averaged periodogram estimation of long memory. *Journal of Econometrics*, **73**, 303–324.

Lopes, S. and Mendes, B. (2006). Bandwith selection in classical and robust estimation of long memory. *International Journal of Statistics and Systems*, **1** (1), 167–190.

Malkiel, B. G. (2003). Efficient market hypothesis and its critics. *Journal of Economic Perspectives*, **17** (1), 59–82.

Mandelbrot, B. and Wallis, J. (1969). Robustness of the rescaled range r/s in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, **5**, 967–988.

Mandelbrot, B. B. (1965). Une classe de processus stochastiques homothetiques a soi: Application a la loi climatologique de h. e. hurst. *Compte Rendus Academie Science*, **260**, 3274–3277.

Matteo, T. D., Aste, T. and Dacorogna, M. (2005). Using the scaling analysis to characterize financial markets. *Journal of Banking and Finance*, **29**, 827–851.

Mielniczuk, J. and Wojdyllo, P. (2007). Estimation of hurst exponent revisited. *Computational Statistics and Data Analysis*, **51**, 4510–4525.

Molinares, F. A. F. and Reisen, F., V. A.and Cribari-Neto (2009). Stationary long-memory process in the presence of additive outliers. a robust model estimation. *Journal of Statistical Planning and Inference*, **139** (8), 2511–2525.

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

Peters, E. E. (1994). *Fractal Market Analysis*. New York: Wiley.

Robinson, P. (1995). Gaussian semiparametric estimation of long-range dependence. *The Annals of Statistics*, **23** (5), 1630–1661.

Sadique, S. and Silvapulle, P. (2001). Long-term memory in stock market returns: International evidence. *International Journal if Finance and Economics*, **6**, 59–67.

Serletis, A. and Rosenberg, A. (2007). Mean reversion in the us stock market. *Chaos, Solitons and Fractals*, **In Press, Corrected Proof**.

Shao, X. and Wu, W. (2007). Local asymptotic powers of nonparametric and semiparametric tests for fractional integration. *Stochastic processes and their Applications*, **117**, 251–261.

Shimotsu, K. and Phillips, P. (2006). Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, **130**, 209–233.

Smith, R. L. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Sciences*, **28**, 1–23.

Srinivas, V. and Srinivasan, K. (2000). Post-blackening approach for modeling dependent annual streamflows. *Journal of Hydrology*, **230**, 86–126.

Tabak, B. and Cajueiro, D. O. (2006). Assessing inefficiency in euro bilateral exchange rate. *Physica A*, **367**, 319–327.

Taqqu, M., Teverovsky, V. and Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, **3** (4), 785–788.

Teverovsky, V., Taqqu, M. S. and Willinger, W. (1999). A critical look at Lo's modified r/s statistic. *Journal of Statistical Planning and Inference*, **80**, 211–227.

Timmermann, A. and Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, **20**, 15–27.

Wang, W., Van Gelder, P., Wrijling, J. K. and Chen, X. (2007). Detecting long-memory: Monte carlo simulations and application to daily streamflow processes. *Hydrology and Earth System Sciences*, **11**, 851–862.

Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, **312**, 285–299.

Appendix

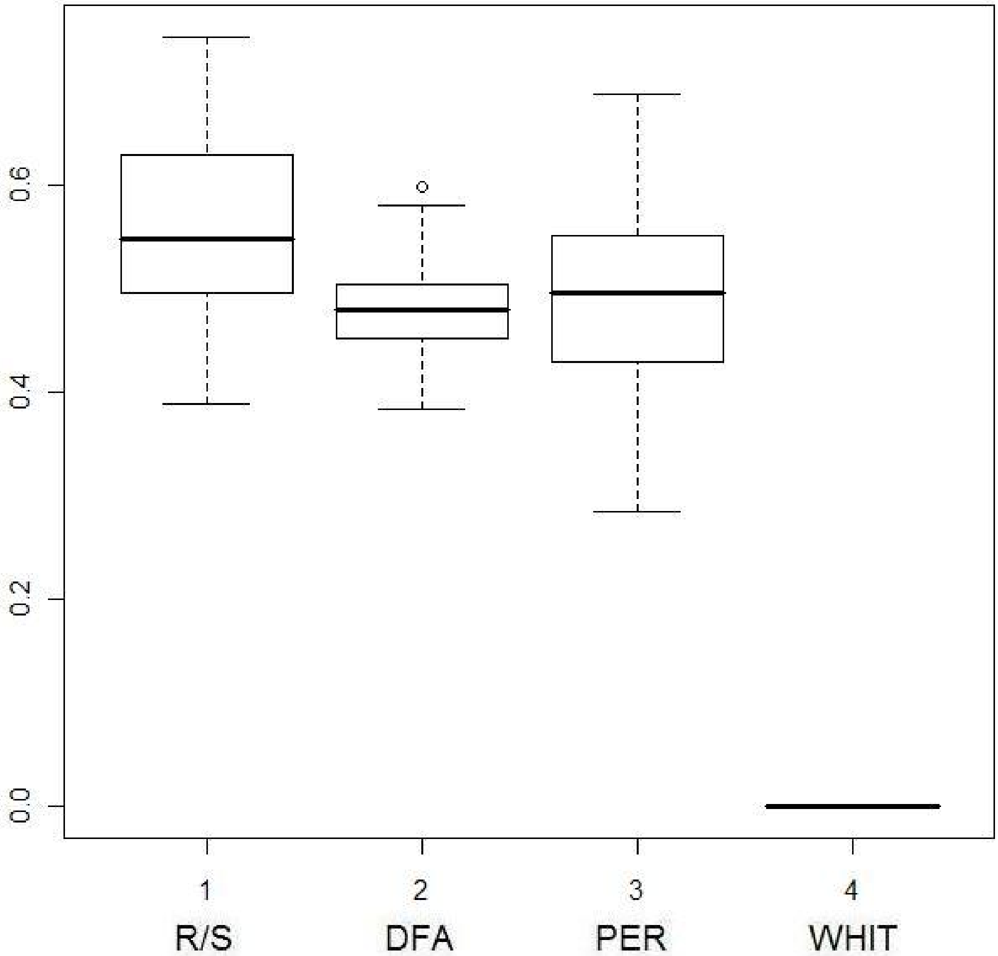

Figure 8.1: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is *Gaussian noise* with $H = 0.5$, 1000 observations, 100 samples

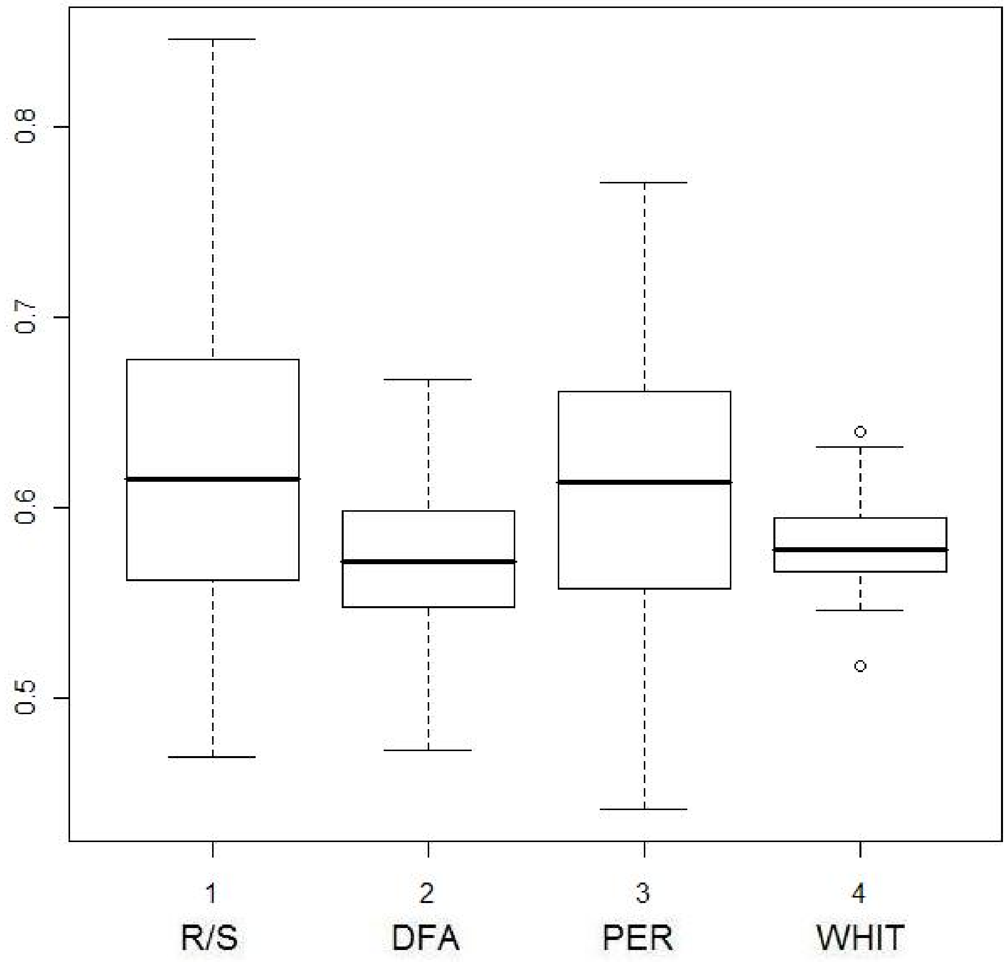

Figure 8.2: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0,0.1,0), 1000 observations, 100 samples

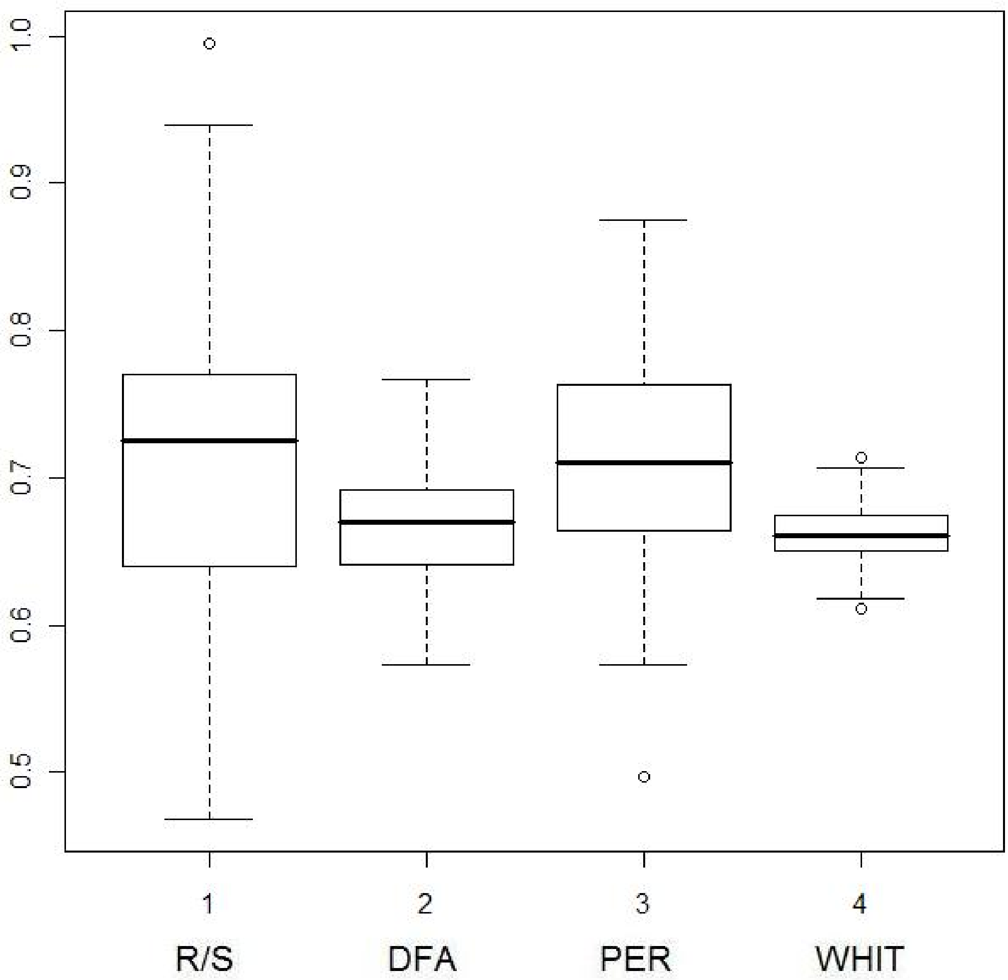

Figure 8.3: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0,0.2,0), 1000 observations, 100 samples

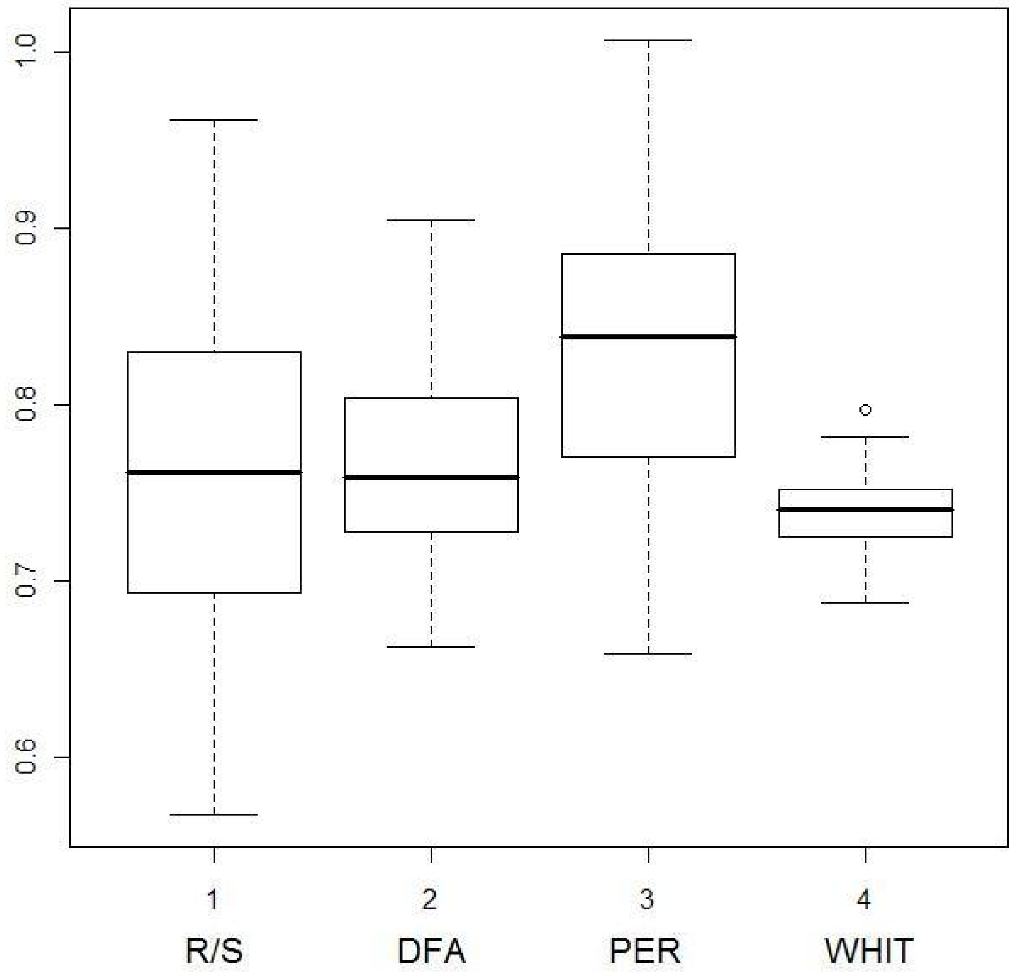

Figure 8.4: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0,0.3,0), 1000 observations, 100 samples

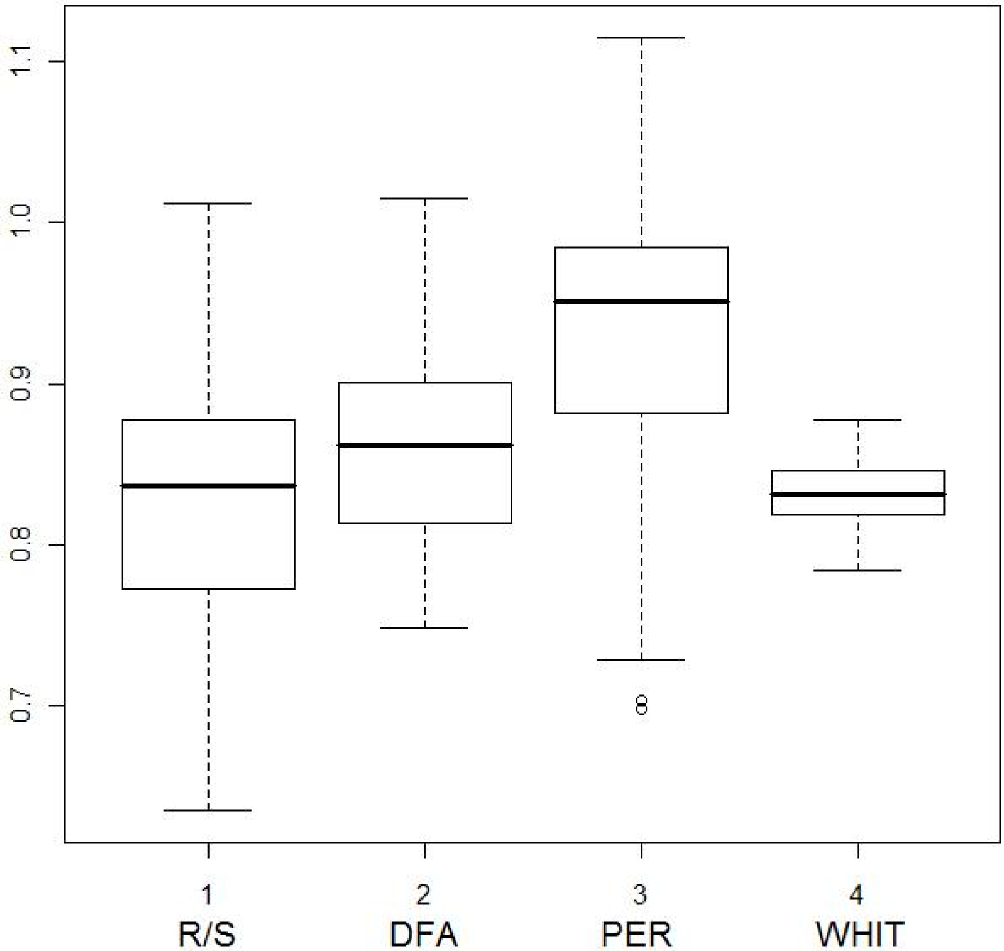

Figure 8.5: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0,0.4,0), 1000 observations, 100 samples

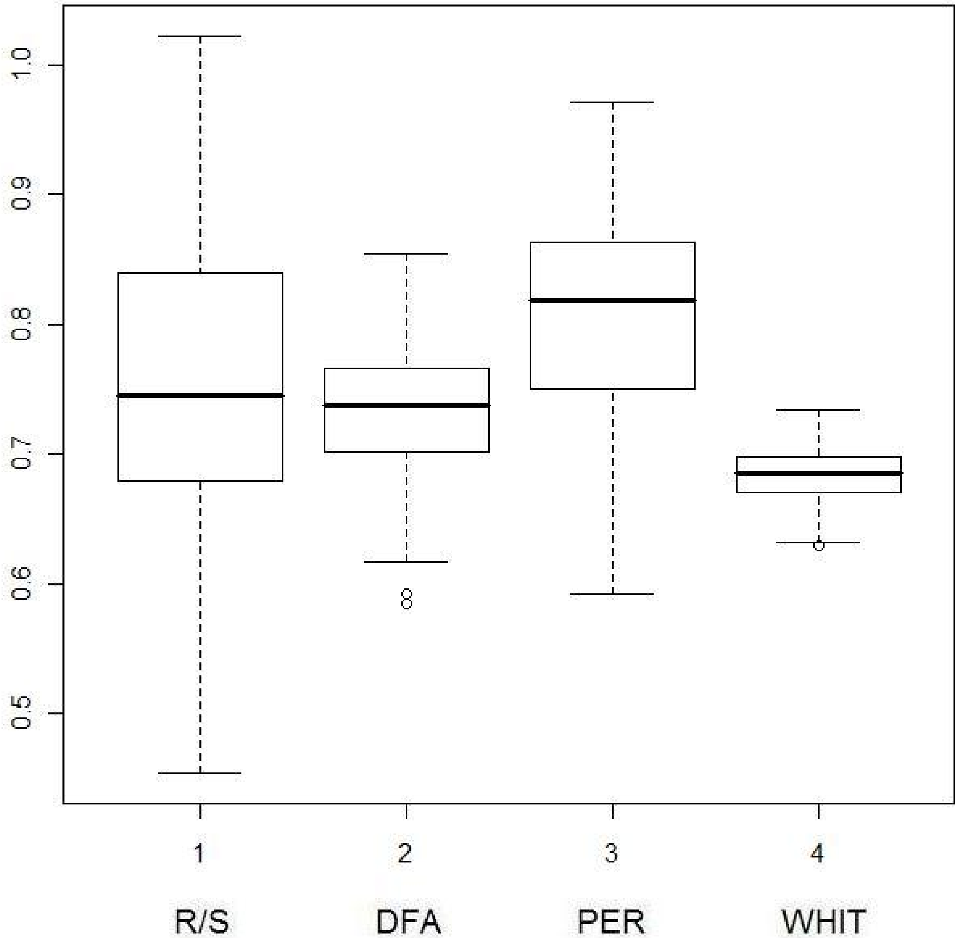

Figure 8.6: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0.2,0.3,0.3), 1000 observations, 100 samples

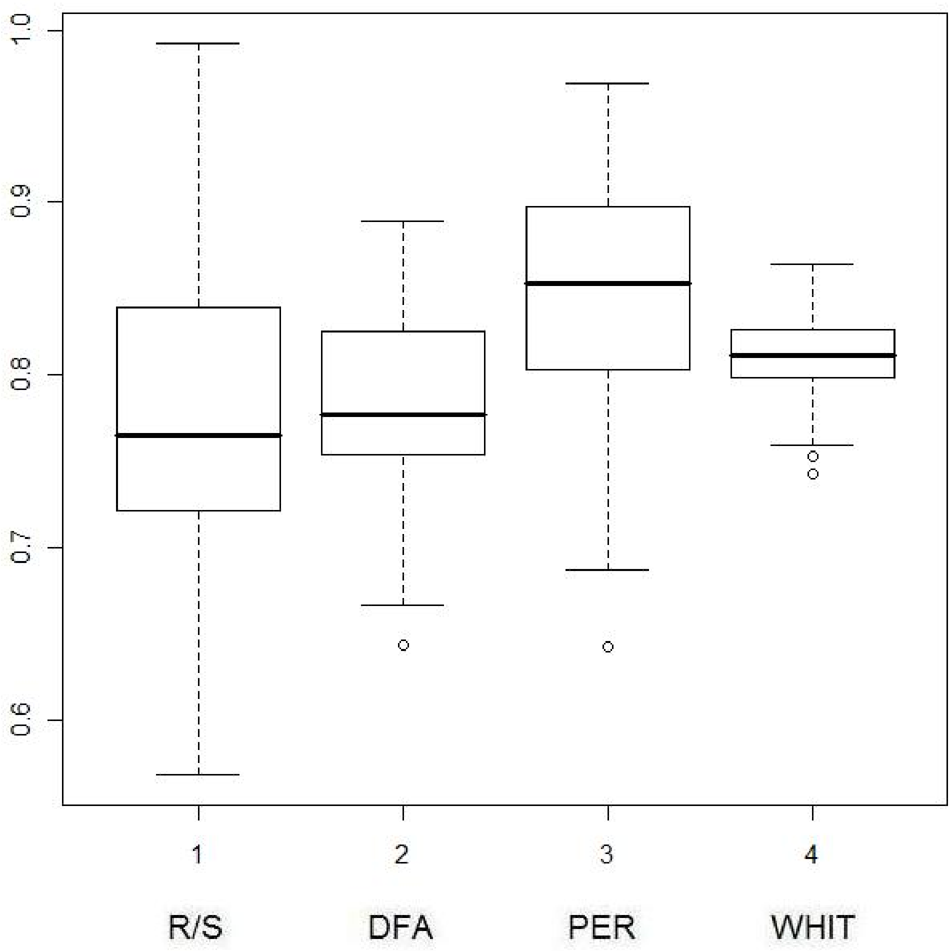

Figure 8.7: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0.4,0.3,0.3), 1000 observations, 100 samples

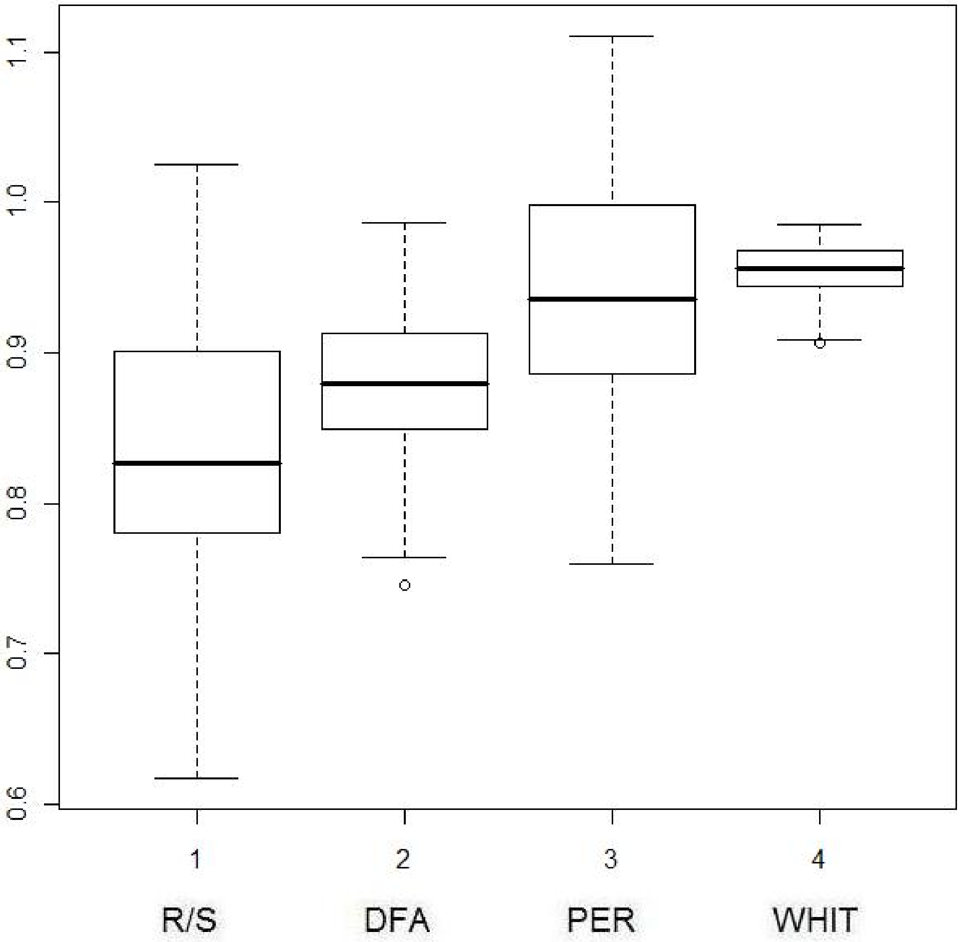

Figure 8.8: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0.6,0.3,0.3), 1000 observations, 100 samples

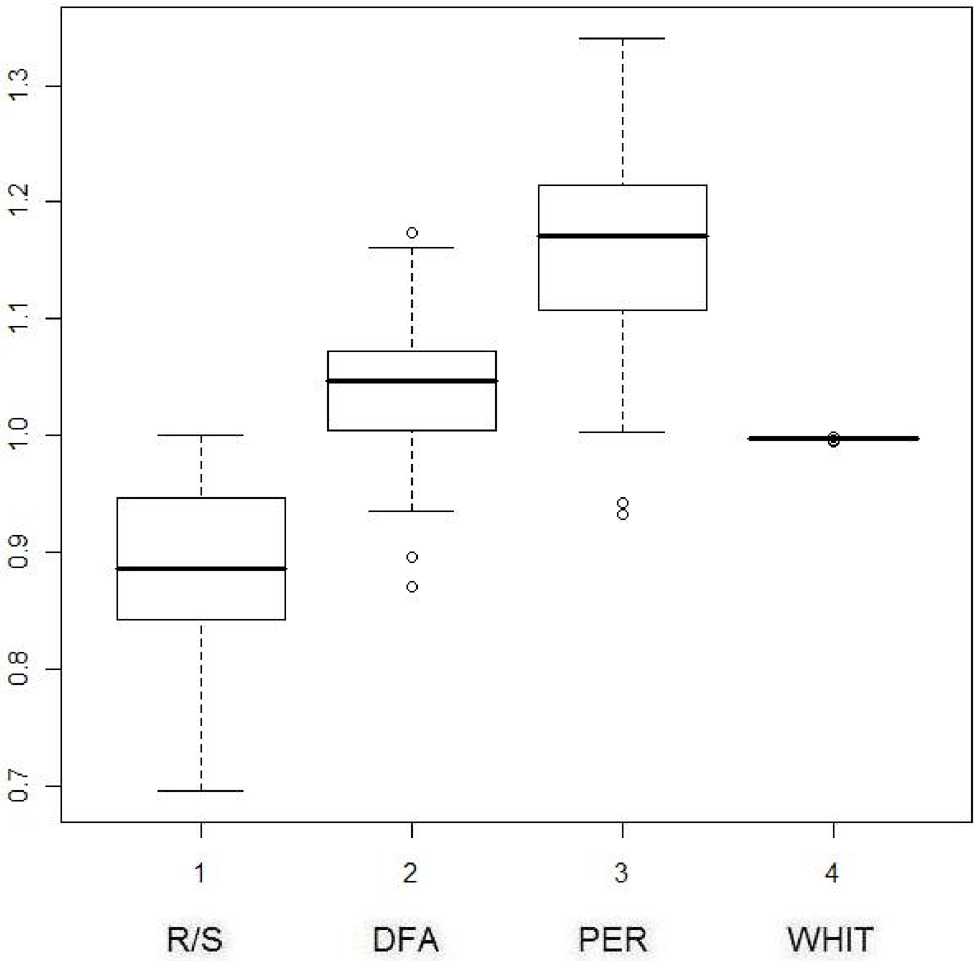

Figure 8.9: Boxplot of the empirical distributions of the four estimators in Monte Carlo study. Underlying distribution is ARFIMA(0.8,0.3,0.3), 1000 observations, 100 samples