Tel Aviv University
The Raymond and Beverly Sackler
Faculty of Exact Sciences

Maximum Likelihood Estimation in Fractional Gaussian Stationary and Invertible Processes

Thesis submitted in partial fulfillment of graduate requirements for the degree M.Sc. from Tel Aviv University

School of Mathematical Sciences
Department of Statistics and Probability

by
Roy Rosemarin

This work was carried out under the supervision of
Professor Offer Lieberman
Professor Isaac Meilijson

August 2008

TABLE OF CONTENTS

ABSTRACT

A stationary and invertible time series with a spectral density $f(\omega) \sim |\omega|^{-\alpha} L(\omega)$ as $\omega \to 0$ where $|\alpha| < 1$ and $L(\omega)$ positive and varies slowly at $\omega = 0$ is said to have long memory if $0 < \alpha < 1$, short memory if $\alpha = 0$ and negative memory if $-1 < \alpha < 0$. Maximum likelihood estimates for Gaussian time series were shown to be consistent and asymptotically normal by Hannan (1973) in the short memory case and by Dahlhaus (1989) in the long memory case. The main objective of this work is to generalize these results to include possibly long, short or negative memory, without a priori knowledge of the memory of the time series. We adapt the proof technique of Dahlhaus (1989) essentially based on the asymptotic behaviour of Toeplitz matrices, but many of Dahlhaus's arguments are extended and simplified. The applicability of the results for fractional Gaussian noise and fractional ARMA processes is shown, and the performance of the estimates on simulated fractional ARMA data is illustrated.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank some of the people who have helped me bring this thesis to its completion. My first words of gratitude go to my thesis supervisors, Prof. Offer Lieberman and Prof. Issac Meilijson.

I am deeply indebted to Prof. Offer Lieberman for introducing me to the exciting topic of time series analysis and directing me throughout this research. He always took the time to read the endless versions of this thesis and to provide a detailed and thoughtful feedback. Thanks to his willingness to share from his own personal experiences, as well as his thoroughness and diligence, I have learnt a lot about academic research.

I would like to express my gratitude to Prof. Issac Meilijson, whose wide experience and insightful probabilistic perspective have inspired me during my studies in Tel-Aviv University and encouraged me to set higher standards for this thesis. I am grateful to him for his invaluable and constructive criticism during my thesis writing.

I would like to thank Mr. Uriel Cohen for the ongoing personal support and encouragement over the years. In addition, I very much appreciate his effort and kind assistance in performing the Monte Carlo simulations for this thesis.

I am thankful to Mr. Ishay Tubi for many stimulating discussions about the main results of the thesis. His perceptive comments helped me to organize and refine my ideas.

Finally, special thanks go to my parents, Yehuda and Margalit Rosemarin, for their unconditional love and belief in me. They are always around to encourage and support me and my choices, both mentally and financially, and for this I am most grateful.

Roy Rosemarin

August 2008

Chapter 1

INTRODUCTION

1.1 Background

Stationary time series with long range dependence have been considered in many fields as diverse as hydrology, astronomy, biology, computer networks, chemistry, agriculture, geophysics and economics. The property of long range dependence (or long memory) implies that the time series is characterized by a slow (hyperbolic) decay of the correlations between observations that become farther away for each other (An accurate definition and a discussion on this property are given in the Chapter 2 of the thesis).

Many of the classical results that are typical for time series with short range dependence (or short memory) does not hold anymore under long memory. For instance, the variance of the sample mean at a time series with long memory converges to zero in a slower rate than the classical rate of $O\left(N^{-1}\right)$. Generally, in the framework of long-range dependence, it turns out that most point estimates and test statistics have a slower rate of convergence than in the case of short range dependence. Equivalently, one may also consider the possibility of negative-memory, in which some of these estimates may converge faster than in the short memory case.

According to Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Sowell (1992), Beran (1994a) and others, many common Gaussian likelihood-based parametric methods perform efficiently under long memory with the same first order properties as under short memory. Particularly, Dahlhaus (1989, 2005) proved that under some mild regularity conditions the exact Gaussian MLE of stationary Gaussian long memory time series is consistent, asymptotically normally distributed and efficient in the sense of Fisher.

However, due to a troublesome balance between a slow numerical computation in large sample sizes on the one hand, and relatively weak performances in small sample sizes on the other hand (cf. Cheung and Diebold 1994), it seems that a relatively few attention in the field of parametric inference was given to exploration of further theoretical properties of the Gaussian MLE for time series with long memory. While there are some published simulations studies that analyze the Gaussian MLE performances in both long and short-range dependence (see Sowell 1992, Cheung and Diebold 1994, Hauser 1999 and Nielsen and Frederiksen 2005), to our knowledge, there is no available theoretical result in the literature that generalizes Dahlhaus's (1989, 2005) results to hold over these cases, without a priori knowledge of the dependence range.

In contrast to that, the vast literature of long memory time series is dedicated to expansion and refinement of other methods of estimation, and particularly of methods in the frequency domain, usually endowed with a simpler implementation and a relatively intuitive theory.

1.2 Objectives and Motivation

The Gaussian MLE, fairly perceived as a fundamental estimation technique, is of major importance in the theory and practice of estimation. Indeed, Chapter 3 shows that many of the most popular methods of estimation for long memory are derived as approximations to the Gaussian MLE.

Moreover, some recent empirical results show that the Gaussian MLE may be more efficient than other approximation-based parametric methods of estimation for some parameter ranges, particularly in medium sample sizes (see Nielsen and Frederiksen 2005). Some other theoretical work on higher order asymptotics suggests adding correction terms for the Gaussian MLE that may significantly improve its performance (see Lieberman 2005, Lieberman and Phillips 2005). Finally, bearing in mind today's powerful computing resources, the computational burden involved in the procedure even for large sample sizes is still relatively mild. Particularly, Doornik and Ooms (1999) report that their ARFIMA package for Ox system works fast even for very long time series (see also Doornik and Ooms 2003).

Overall, as mentioned in the last section, it seems that there is a theoretical gap that needs to be filled, concerning the statistical properties of the Gaussian MLE. This thesis attempts to make a contribution in this respect. We prove in Chapter 4 that Dahlhaus's (1989, 2005) conclusions holds if the parameter space is extended to contain short and negative memory, without a priori knowledge of the memory of the series. Such an expansion resorts to limit theorems of symmetric Toeplitz and inverse-Toeplitz matrices, motivated by the key fact that the covariance matrix of a stationary process is a Toeplitz matrix. The slow decay of the correlations requires careful handling of many of the matrix manipulations needed for the establishment of the result. Luckily, some work of Fox and Taqqu (1987), Avram (1988), Dahlhaus (1989, 2005) as well as Lieberman and Phillips (2004) shed light on some crucial asymptotic properties of symmetric Toeplitz matrices that may be applied to refine Dahlhaus's (1989, 2005) result.

1.3 Chapters Outline

The rest of the thesis is organized as follows. In Chapter 2 the reader is presented with background knowledge on the unique characteristics of time series with long memory, and introduced to some possible ways to model these memory characteristics within a parametric framework. Chapter 3 surveys some prominent popular methods of estimation of stationary long memory time series. Chapter 4 extends Dahlhaus's (1989, 2005) results to include possibly short-memory or anti-persistent time series. Chapetr 5 presents some simulation results of the Gaussian MLE of several time series with long memory. Chapter 6 briefly concludes and summarizes the achieved results and possible directions for future research which arise directly out of the thesis.

Chapter 2

PRELIMINARIES

2.1 Introduction

In the field of time series analysis, it is sometimes taken for granted that stationary time series are weakly correlated in the sense that the effect of the correlations between far away observations is negligible. Long-range dependence, on the other hand, is sometimes mistakenly viewed as a property that implies nonstationary, as in the case of random walk. In this chapter we introduce to the reader some of the last century's developments in the theory of time series that led statisticians and econometricians to realize the possibility and necessity of modeling stationary time series that exhibit behaviour of long memory.

The rest of Chapter 2 is organized as follows. In Section 2.2 we the reader is presented to some basic definitions essentially concerning with the L_2 structure of a stationary time series. In section 2.3 we introduce the notions of long memory and anti-persistent time series and we provide some references from the literature for evidence of time series that seem to exhibit long memory behavior. Section 2.4 presents the fractional Gaussian noise process, following essentially the line of Beran's (1994b) presentation of the topic, which also deals with it in the context of long memory time series. Section 2.5 deals with an extension of the known Box and Jenkins' ARIMA models to fractionally integrated ARMA models, called the ARFIMA models.

2.2 Stationary Time Series

Consider a real-valued discrete time series $X = \{X_t\}_{t\in\mathbb{Z}}$. A common assumption is that the time series has some particular characteristics of statistical equilibrium, in the following sense.

Definition 2.2.1 (Strict Stationarity) *A time series is said to be strictly stationary* *if the joint probability distributions of* $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ *and* $(X_{t_1+h}, X_{t_2+h}, ..., X_{t_k+h})$ *are identical for any positive integer* k *and* $t_1, t_2, ..., t_k, h \in \mathbb{Z}$.

Strict stationarity, however, is quite a strong assumption and in most application only a weaker form of stationarity is assumed. This weaker form (henceforth stationary) assumes finite variances but restricts the time-homogeneous requirement to means, variances and covariances.

Definition 2.2.2 (Stationarity) *A time series* X *in* $L_2(\mathbb{R})$ *is said to be stationary (or weak stationary, second-order stationary or covariance stationary) if* $E(X_t)$ *does not depend on* t *and* $Cov(X_s, X_t)$ *is a function of* $|s-t|$ *for all* $s, t \in \mathbb{Z}$.

Example 2.2.1 (Gaussian Time Series) *Let* X *be a time series, all of whose finite-dimensional joint distributions are multivariate normal. Because Gaussian distributions completely determined by its mean and covariances, then if* X *is stationary, then it is strictly stationary.*

Example 2.2.2 (White noise) *In many respects, the simplest kind of time series,* X*, is one in which the observations are iid with zero mean and finite variance* σ^2*. From a second order point of view, i.e., ignoring all properties of the joint distributions of* X *except those which can be deduced from the means, covariances and variances, such time series are identified with the class of all stationary time series having mean zero, and autocovariance function*

$$\gamma(k) = \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} . \tag{2.1}$$

If X *has zero mean and autocovariance function (2.1), then the time series is called a white noise, denoted by*

$$X \sim WN\left(0, \sigma^2\right).$$

Figure 2.1 presents a $WN(0,1)$ *series of length 400. The disorderly ragged path of the series is a consequence of the fact that each two observations at different times are uncorrelated.Define the covariance matrix of the first* N *observations of a time series* X *by*

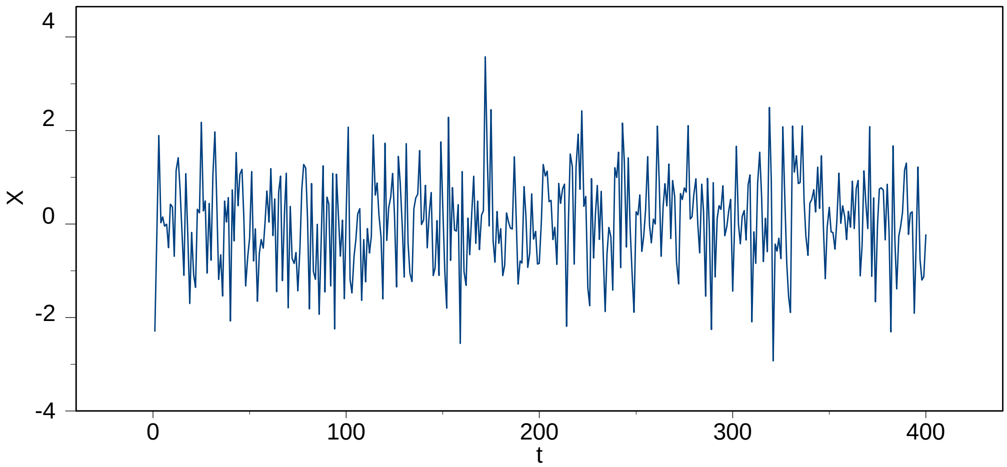


Figure 2.1: Simulated series of WN(0,1).

$$\Sigma_N = \left[Cov\left(X_s, X_t\right)\right]_{s,t=1,\ldots,N}.$$

If X is stationary then Σ_N has the form,

$$\Sigma_N = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \cdots & \gamma(N) \\ \gamma(1) & \gamma(0) & \gamma(1) & \ddots & & \vdots \\ \gamma(2) & \gamma(1) & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \gamma(1) & \gamma(2) \\ \vdots & & \ddots & \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(N) & \cdots & \cdots & \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix},$$

where

$$\gamma(k) = Cov\left(X_s, X_t\right) \quad \textit{for all } s,t \in \{1,2,...,N\} \textit{ such that } |s-t| = k. \qquad (2.2)$$

Definition 2.2.3 (Toeplitz Matrix) *A Toeplitz matrix Σ is a matrix in which each descending diagonal from left to right is constant. If a Toeplitz matrix Σ is symmetric then we call Σ a symmetric Toeplitz matrix.*

A covariance matrix of a stationary time series is therefore a symmetric Toeplitz matrix. We will recall this fact again in Chapter 4 when we derive the asymptotic properties for the Likelihood function of a stationary time series.

2.3 The Memory of a Time Series

There are several possible definitions of the property of "long memory" and they are not necessarily identical (cf. Guégan 2005). In order to cover all types of memory, it is probably best to describe the memory of a series in terms of the spectral density structure in case of stationary time series. The following definition is of Robinson (2003).

Definition 2.3.1 (Memory of a Time Series) *A stationary time series X with spectral density $f(\omega)$ has long memory (or long-range dependence) if*

$$f(0) = \infty, \tag{2.3}$$

so that $f(\omega)$ has a pole at frequency zero. In the opposite situation of a zero at $\omega = 0$,

$$f(0) = 0, \tag{2.4}$$

X is said to be anti-persistent or to have negative memory. We then said that X has short memory (or short-range dependence) if

$$0 < f(0) < \infty. \tag{2.5}$$

Notice that

$$f(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k).$$

Thus, the memory of a time series is essentially a measure of the dependence between all the variables in the series, considering the effect of all correlations simultaneously.

For many reasons, it is important to characterize the rate at which the spectral density tends to $f(0)$ as $\omega \to 0$. An alternative definition of the memory of a time series (see, for instance, Beran 1994b) entails the existence of $\alpha < 1$ such that

$$f(\omega) \sim c_1 |\omega|^{-\alpha} \qquad \text{as } \omega \to 0, \tag{2.6}$$

where $c_1 > 0$ and "$\sim$" indicates that the ratio of left- and right-hand sides tends to 1. Corresponding to the Definition 2.3.1, X is then said to have long memory if $0 < \alpha < 1$, short memory if $\alpha = 0$ and to be anti-persistent if $\alpha < 0$. Note that $\alpha \geq 1$ implies $f(\omega) \notin L_1(\Pi)$, and thus $f(\omega)$ cannot represent a spectral density of a stationary time series. On the other hand, if $\alpha \leq -1$, then the series is not invertible in a sense that it cannot be used to reconstruct a series of a white noise by passing X through a linear filter (see Brockwell and Davies 1991, Section 4.10).

One aspect of condition (2.6) is that for the case $\alpha < 1$, $\alpha \neq 0$, this condition is equivalent to a hyperbolic decay of the autocovariances (cf. Yong 1974 and Zygmund 2002, Chapter V.2)

$$|\gamma(k)| \sim c_2 k^{\alpha-1} \quad \text{as } k \to \infty, \tag{2.7}$$

with

$$c_2 = 2c_1\Gamma(1-\alpha)\sin\left(\frac{\alpha}{2}\pi\right).$$

If $\alpha > 0$, then all the correlations of the series are positive, and they decay so slowly to zero that (2.3) holds. This is in contrast to other known dependent time series models such as ARMA processes (see Section 2.4), in which the asymptotic decay of the correlations is exponential, so that

$$|\gamma(k)| < ab^k,$$

where $0 < a < \infty$ and $0 < b < 1$. Because the absolute value of b is less than 1, (2.5) holds in this case, and the such processes have a short memory.

Note that equation (2.7) determines only the asymptotic decay of the correlations, and it does not imply the values of the correlations in some specific lags. Moreover, it determines only the rate of convergence and not the absolute size of the correlations. The effect of long memory on statistical inference can be extreme even for small sample sizes. Most estimates and test statistics have a slower rate of convergence so that assuming short memory leads to underrating uncertainty (measured by the size of the confidence interval) by a factor that tends to infinity as the sample size tends to infinity (see Beran 1994b, Section 2.1). For example, consider the variation in the sample mean $\overline{X}_N = N^{-1}\sum_{t=1}^{N} X_t$. If X is a short memory time series and $f(\omega)$ is continuous at $\omega = 0$, then according to Féjèr's theorem we get

$$Var\left(\overline{X}_N\right) = \frac{1}{N^2}\sum_{i,j=1}^{N}\gamma\left(|i-j|\right) = \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|j|}{N}\right)\gamma\left(k\right) \sim \frac{2\pi f\left(0\right)}{N}, \quad as\ N\to\infty, \tag{2.8}$$

Equation (2.8) can be viewed as a generalization of the σ^2/N rule frequently used to evaluate $Var\left(\overline{X}_N\right)$ if the correlation of the series are assumed to be negligible. However (2.8) shows that if the autocorrelations decay at a hyperbolic rate as in (2.7), then the variance of $\overline{X}_N$ differs from σ^2/N not just by a constant factor but by the speed at which it converges to zero. The behaviour of the sample mean under such circumstances was discussed by Adenstedt (1974). He proved that if (2.6) holds for any $-1 < \alpha < 1$, then

$$Var\left(\overline{X}_N\right) \sim c_3 N^{\alpha-1}, \tag{2.9}$$

where $c_3 > 0$.

There is not many evidence of time series showing anti-persistency. That is because equality (2.4) can only hold if the positive variance $\gamma(0)$ is precisely balanced by predominantly negative autocovariances $\gamma(j)\,,\ j \neq 0$. In practice, this condition is very unstable since any arbitrarily small disturbance added to the series destroys property (2.4). However, condition (2.4) may be a result of differencing, for instance, if we observe the differenced series $\nabla X \equiv \{X_{t+1} - X_t\}_{t\in\mathbb{Z}}$ when X is a nonstationary ARFIMA(0,d,0) series with $d \in \left(\frac{1}{2}, 1\right)$ then ∇X is an anti-persistent and stationary ARFIMA(0,d,0) series with $d \in \left(-\frac{1}{2}, 0\right)$ (see Section 2.5.2).

On the other hand, there is ample historical evidence that long memory processes occur in fields as diverse as hydrology, astronomy, biology, computer networks, chemistry, agriculture, geophysics and economics. In some of these fields long memory is in fact recognized to be the rule rather than the exception (cf. Beran 1992). Newcomb (1886) discussed the phenomenon of long memory in astronomical data sets and called it "semi-systematic" errors. Pearson (1902) observed slowly decaying correlations in simulated astronomical observations. Student (1927) also found long memory behaviour in the context of chemical measurements. Perhaps the most well-known example of long memory is the so-called Hurst

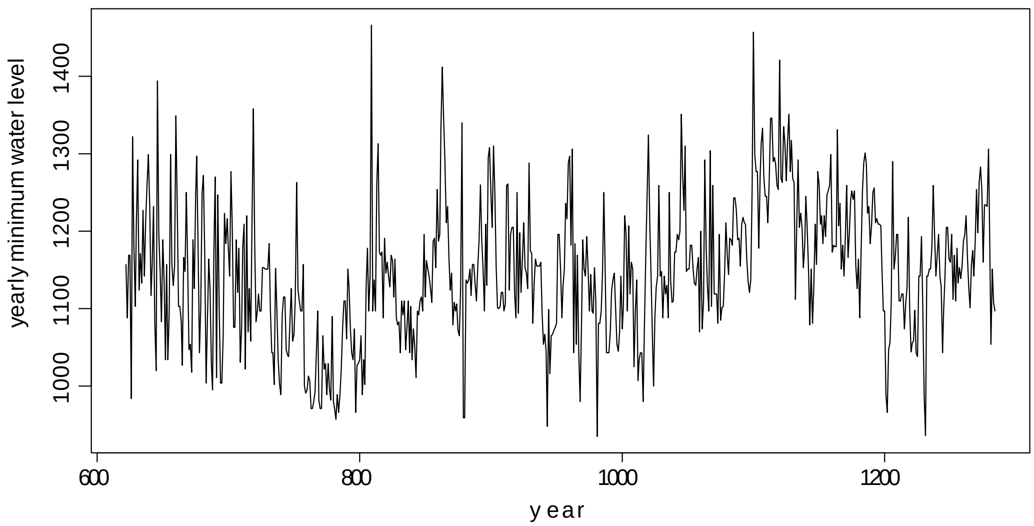


Figure 2.2: Yearly minimum water levels of the Nile River at the Roda Gauge (622-1284 A.D.)

effect in hydrology (see Section 3.4.1). Motivated to design the capacity of water reservoirs, Hurst (1951, 1956) analyzed runoff time series from the river Nile and other hydrological records. Hurst found deviations from the expected short range dependence behaviour of the time series. Figure 2.2 presents one of the time series that led to the discovery of the Hurst effect. This figure displays the yearly minimal water level of the Nile River for the years 622-1284, measured at the Roda Gauge near Cairo (The data of the Nile River is due to Beran (1994b, pp. 237-239). Also available at StatLib archive: http://lib.stat.cmu.edu/S/beran). This Figure reflects some of the typical characteristics of time series with long memory behaviour (Beran 1994b): There are long periods where the observations tend to stay at a high level, and, on the other hand, there are long periods with low levels. Looking at short time periods, there seem to be cycles or local trend. However, looking at the whole series, there is no apparent persisting cycle. In fact, the series seems to be homoscedastic and to fluctuate around a constant mean, perhaps consistent with stationarity. The Correlogram

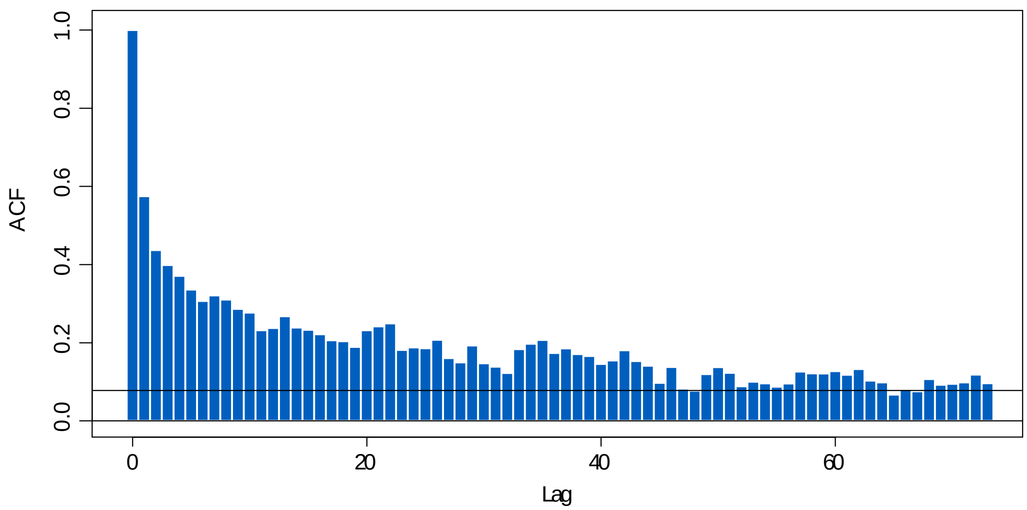


Figure 2.3: Nile River minima: sample autocorrelations

of this series in Figure 2.3 shows that the sample autocorrelations exhibit a clear pattern of slow decay and persistence. Recall that the Nile river is the site of the Biblical story of Joseph, the son of Jacob. Mandelbrot and Wallis (1968) suggested the term "Joseph Effect" for describing these characteristics of long memory, since these figures may be viewed as an reminiscent of the seven fat and seven lean years foreseen by Joseph in the Biblical story. The presence of long memory in the Nile River behaviour may provide an explanation why corresponding averages of the flow of the Nile River differ greatly from each other, and therefore from their common expectation, over successive intervals of several years.

There is also substantial evidence that long memory processes describe rather well economic and financial data such as forward premiums, interest rate differentials, and inflation rates. Apart from Mandelbrot's pioneering work on self-similar processes and their diverse applications (see Mandelbrot 1997), the importance of long memory in economic data was recognized by Granger (1966). Granger used different kinds of estimates of the spectral density for economic time series after known business cycles and trends are removed. He observed that the typical shape of the spectral density is that it has a pole at the origin and it decays monotonically with the absolute value of the frequency. The first characteristic corresponds to long memory behavior, while the second one can be modeled by many long memory models such as the Fractional Gaussian Noise and Fractional ARIMA models discussed below.

Perhaps the most dramatic empirical success of long memory processes has been in work on modeling the volatility of asset prices. Asset returns frequently exhibit little autocorrelations consistent with the efficient market hypothesis, whereas their squares are noticeably correlated. While commonly used models in finance like ARCH and GARCH try to represent this phenomena (see Engle 1982 and Bollerslev 1986), they imply that the autocorrelations of the squared asset returns decay exponentially. However, empirical evidence (see, for example, Whistler 1990, Ding, Granger and Engle 1993, and Dacorogna, Muller, Nagler, Olsen and Pictet 1993) rather suggests a slow decay of the correlations, consistent with long memory behaviour. These findings have led to formulation of a long memory conditional heteroscedastic model of time series, called the FIGARCH (Baillie, Bollerslev, and Mikkelsen 1996), a model that has since been widely applied in finance and may offer potentially important insights on market behavior.

A possible explanation for how long memory behaviour might arise has been provided by Robinson (1978) and Granger (1980). In some statistical applications, the observed time series can be regarded as aggregates of many individual time series. For instance, Macroeconomic series can be regarded as aggregates across many micro-units. Granger (1980) considered

$$X_t^N = \sum_{i=1}^{N} X_t\left(\theta_i\right),$$

which is the aggregate of N components of independent processes, $X_t\left(\theta_i\right)$, such that for $i = 1, 2, ..., N$,

$$X_t\left(\theta_i\right) = \phi\left(\theta_i\right) X_{t-1}\left(\theta_i\right) + \epsilon_t\left(\theta_i\right),$$

where $\epsilon_t\left(\theta_i\right)$ are independent zero mean and homoscedastic random variables, and $\phi\left(\theta_i\right)$ are drawn from the Beta $(c, 2-\alpha)$ distribution with support $(0,1)$, and $c > 0$, $0 < \alpha < 1$. Conditionally on the $\theta_i's$, $X_t\left(\theta_i\right)$ are stationary short memory AR(1) processes. Granger

(1980) showed that in the limit as $N \to \infty$, the unconditional autocovariance function of X_t^N decays like $k^{\alpha-1}$, as in (2.7). Thus, although each individual time series, $X_t(\theta_i)$, is a simple AR(1) process, the aggregated series X_t^N approaches a limiting time series that has a long memory. Previously, Mandelbrot (1971) suggested a similar idea in the context of Monte Carlo simulation of ARFIMA(0,d,0) model (to be discussed below).

Further references on long memory behavior, evidence of long memory and possible physical explanations for the behaviour that is typical of long memory time series, may be found in many survey type resources on long memory such as Taqqu (1986), Hampel (1987), Beran (1992, 1994b), Baillie (1996) and Robinson (1994, 2003).

2.4 *Stationary Increments of Self-Similar Processes*

2.4.1 Self-Similar Processes

The theory of self similar processes was developed by Kolmogorov (1940) and Lamperti (1962). The importance of such processes was recognized by Mandelbrot and co-workers who introduced them to into statistics. In general, self-similarity means that the phenomenon in interest seems to have a similar structure across a wide range of scale (cf. Mandelbrot 1971, 1983 and Taqqu 1986). In the context of stochastic processes, self-similarity is defined in terms of the distribution of the process, as in the following definition.

Definition 2.4.1 (Self-Similar process) *Let* $Y = \{Y_t\}_{t\in\mathbb{R}}$ *be a real valued stochastic process. Y is called self-similar with self-similarity parameter H (abbreviated as H-ss), if for any* $c > 0$*,*

$$Y_{ct} =_d c^H Y_t, \tag{2.10}$$

where $=_d$ *is equality in distribution.*

Equation (2.10) is also equivalent to

$$(Y_{ct_1}, ..., Y_{ct_k}) =_d c^H (Y_{t_1}, ..., Y_{t_k})$$

for any finite sequence of time point $t_1, ..., t_k$.

A sample path of an H-ss can be very complex, and we are not guaranteed in general that the process is stochastically continuous, nor also measurable. The following Proposition is due to Lamperti (1962, Theorem 2. See also Vervaat 1985).

Proposition 2.4.1 *A stochastically continuous (continuous in probability) process Y is an H-ss process if and only if there exists a process S_t and a positive function $a : \mathbb{R}_+ \to \mathbb{R}_+$ with $a(c) \to \infty$ as $c \to \infty$, such that*

$$S_{ct}/a(c) \to_d Y_t \quad \text{as } c \to \infty. \tag{2.11}$$

An important class of self similar processes is the class of self similar processes with stationary increments, defined as follow.

Definition 2.4.2 (Stationary Increments) *A stochastic process Y has stationary increments (or, in short, we say that Y is an si process) if for every $c \in \mathbb{R}$*

$$Y_{t+c} - Y_c =_d Y_t - Y_0. \tag{2.12}$$

Although H-ss with $H \leq 0$ is a relevant possibility for other classes of processes, there are only trivial (or pathological) $H \leq 0$-ss si processes, as the next theorem shows (Vervaat 1985, Theorem 1.3).

Proposition 2.4.2 *Suppose that Y is a self-similar process with self-similarity parameter H and stationary increments.*

(i) *If $H < 0$, then $Y_t = 0$ a.s. for each real t.*

(ii) *If $H = 0$ and Y_t is measurable, then $Y_t = Y_0$ a.s. for each real t..*

Remark 2.4.1 *There are nontrivial nonmeasurable 0-ss si processes. For example, Y iid and nondegenerate.*

Thus, in the following we consider only H-ss si processes, Y, with $H > 0$. We also want to exclude the trivial event $Y = 0$ a.s., and we assume that this event has zero probability. Then, If $H > 0$, The properties

$$Y_0 =_d a^H Y_0 \quad \text{for every } a > 0,$$

and

$$Y_t =_d t^H Y_1 \quad \text{for every } t > 0,$$

imply that $Y_0 =_d 0$ and $Y_t \to_d \infty$ as t tends to infinity. In particular, we may conclude that a non-degenerate H-ss si process cannot be stationary. However, we are guaranteed that the sample path of any H-ss si with $H > 0$ is stochastically continuous, since all such processes arise in (2.11) directly from (2.10) with $S = Y$, $a(c) = c^H$ and $=_d$ instead of $\to_d$. Furthermore, It is easy to see that Y in (2.11) is si if S_t is discretely si, i.e., satisfies (2.12) only for $t \in \mathbb{Z}$. So stochastic continuous ss si processes may arise as limits in (2.11) with discretely si S_t. This fundamental relation between stochastically continuous H-ss si processes and limits in distribution of discretely si processes was derived by Lamperti (1962, Theorem 2). The following proposition summarizes this idea (Beran 1994b).

Proposition 2.4.3 *Suppose that Y is a stochastic process such that $Y_1 \neq 0$ with positive probability, and Y_t is the limit in distribution of the sequence of normalized partial sums*

$$a_n^{-1} S_{nt} = a_n^{-1} \sum_{i-1}^{[nt]} X_t, \qquad n = 1, 2, ., , ,$$

Here $[nt]$ denotes the integer part of nt, $X_1, X_2, \ldots$ is a stationary sequence of random variables, and $a_1, a_2, \ldots$ is a sequence of positive normalizing constants such that $\log a_n \to \infty$. Then there exists an $H > 0$ such that $Y(t)$ is a stochastic continuous H-ss si process. Conversely, all stochastic continuous H-ss si processes with $H > 0$ can be obtained as the limit of such partial sums.

The form of the covariance function $Cov(Y_t, Y_s)$ of an ss si process follows directly from the definitions. To simplify notation, assume here $E(Y) = 0$. The process $X = \{Y_t - Y_{t-1}\}_{t \in \mathbb{Z}}$ is called the (stationary) increments process of Y, and we denote its variance by σ^2. Since

$$\sigma^2 = E\left[(Y_t - Y_{t-1})^2\right] = E\left[(Y_1 - Y_0)^2\right] = E\left[Y_1^2\right],$$

we then have

$$E\left[(Y_t - Y_s)^2\right] = E\left[Y_{t-s}^2\right] = \sigma^2 (t-s)^{2H}.$$

On the other hand,

$$E\left[(Y_t - Y_s)^2\right] = E\left[Y_t^2\right] + E\left[Y_s^2\right] - 2E\left[Y_t Y_s\right] = \sigma^2 t^{2H} + \sigma^2 s^{2H} - 2\gamma_y(t,s).$$

Hence,

$$\gamma_y(t,s) = \frac{1}{2}\sigma^2\left[t^{2H} - (t-s)^{2H} + s^{2H}\right]. \qquad (2.13)$$

It is possible to obtain the autocovariance function of the increments process $X_t = Y_t - Y_{t-1}$.

$$\begin{aligned}\gamma(k) &= Cov(X_t, X_{t+k}) = Cov(X_1, X_{k+1}) \\ &= Cov(Y_1, Y_{k+1} - Y_k) = Cov(Y_1, Y_{k+1}) - Cov(Y_1, Y_k).\end{aligned}$$

Using (2.13),

$$\gamma(k) = \frac{1}{2}\sigma^2\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \qquad (2.14)$$

for $k \geq 0$ and $\gamma(k) = \gamma(-k)$ for $k < 0$.

The asymptotic behaviour of $\gamma(k)$ follows by Taylor expansion, from which it can be shown that if $0 < H < 1$, $H \neq \frac{1}{2}$, then as k tends to infinity

$$\gamma(k) \sim H(2H-1)k^{2H-2}.$$

For $\frac{1}{2} < H < 1$, the correlations decay to zero so slowly that

$$\sum_{k=-\infty}^{\infty} \gamma(k) = \infty, \qquad (2.15)$$

Thus, the increments process X has long memory. For $H = \frac{1}{2}$, all the correlations at non-zero lags are zero, i.e., the observations are uncorrelated and the process has short memory. For $0 < H < \frac{1}{2}$, it can be shown that

$$\sum_{k=-\infty}^{\infty} \gamma(k) = 0, \qquad (2.16)$$

and therefore the process is anti-persistent.

For $H = 1$ (2.14) implies $\rho(k) \equiv 1$. This case is hardly of any practical importance in the stationary setup. For $H > 1$, $\gamma(k)$ diverges to infinity. This contradicts the fact that

$\rho(k)$ must be between -1 and 1. We conclude that if the second moments are finite and $\lim_{k\to\infty}\rho(k)=0$, then

$$0 < H < 1.$$

Under these assumptions, the spectral density of the increment process X is given by (Sinai 1976)

$$f(\omega) = F\left(\sigma^2, H\right)(1-\cos\omega)\sum_{k=-\infty}^{\infty}|\omega+2\pi k|^{-1-2H}, \qquad \omega\in\Pi,$$

where $\sigma^2 = Var(X)$ and $F\left(\sigma^2,H\right)$ is a normalizing factor designated to ensure $\int\limits_{-\pi}^{\pi} f(\omega)\,d\omega = \sigma^2$,

$$\begin{aligned} F\left(\sigma^2,H\right) &= \sigma^2\left\{\int_{-\pi}^{\pi}(1-\cos\omega)\sum_{k=-\infty}^{\infty}|\omega+2\pi k|^{-1-2H}\,d\omega\right\}^{-1} \\ &= \sigma^2\left\{\int_{-\infty}^{\infty}(1-\cos\omega)\,|\omega|^{-1-2H}\,d\omega\right\}^{-1}. \end{aligned}$$

The behaviour of $f(\omega)$ near the origin follows by Taylor expansion of $(1-\cos\omega)$ at zero, and it can be shown (see Theorem 4.2.1) that

$$f(\omega) \sim c_1\left(\sigma^2,H\right)|\omega|^{1-2H} \qquad \text{as } \omega\to 0.$$

Thus, the form of the spectral density $f(\omega)$ is uniquely determined by only two parameters σ^2 and H. Particularly, the behaviour of $f(\omega)$ near $\omega=0$ is compatible with property (2.6) of long memory with $\alpha = 2H-1$.

Figure 2.4 shows the theoretical autocorrelation function of several fractional Gaussian noise (FGN) series, a definition of which is given below. Figure 2.5 illustrates some typical sample paths of FGN series with $H = 0.9, 0.7, 0.3, 0.1$, produced by applying an FGN filter on the white noise series used in Figure 2.1. The characteristics of long memory are prominent for $H = 0.9$ but can also be seen for $H = 0.7$. In the cases of $H = 0.1$ and $H = 0.3$ the series has a more ragged path than that of a white noise, which is typical for anti-persistent series. Some more distinctive characteristics of long memory and anti-persistent time series are illustrated in Section 2.5.2.

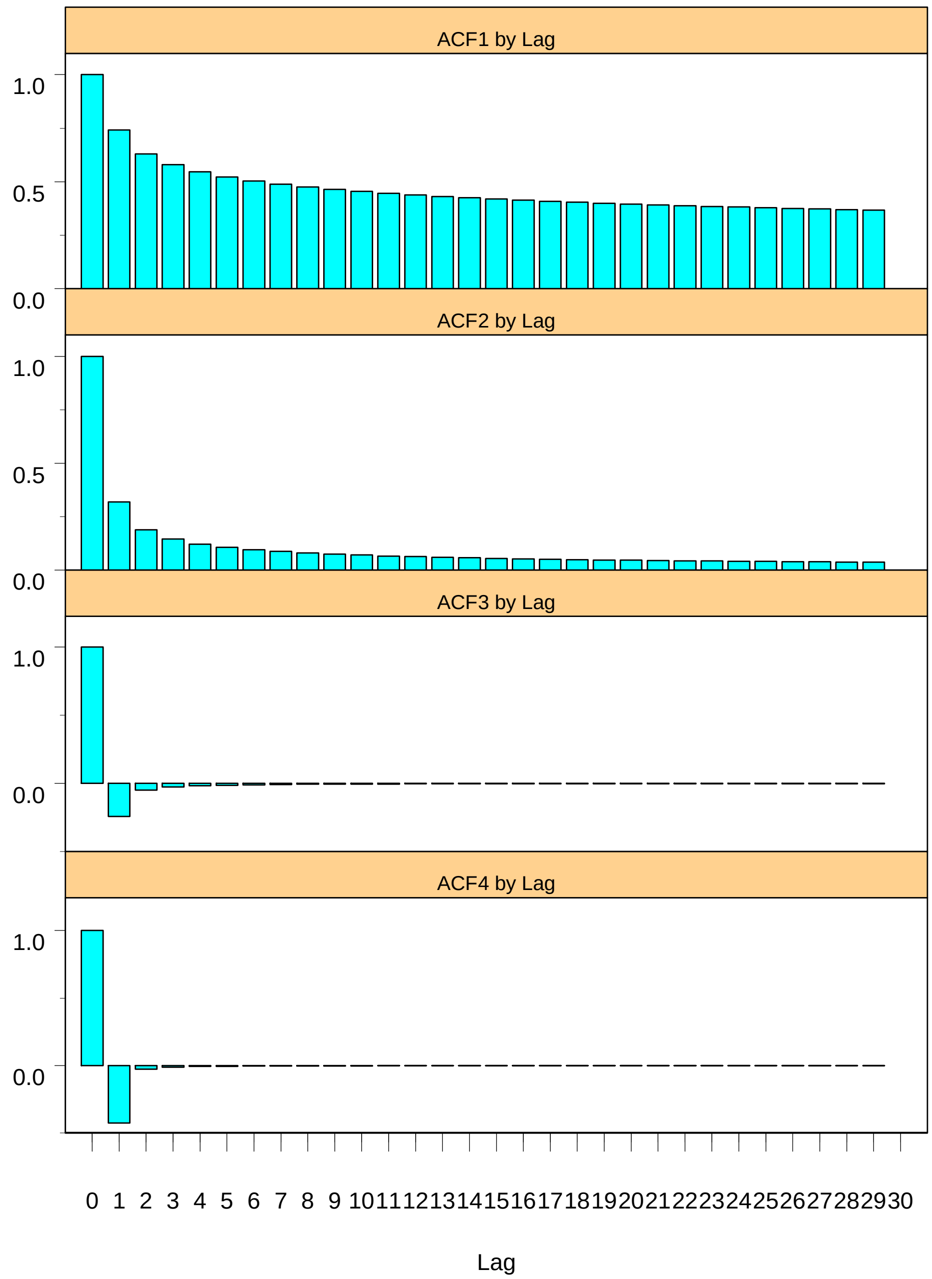


Figure 2.4: Autocorrelation function of FGN(H) series with $H = 0.9$ (ACF1), $H = 0.7$ (ACF2), $H = 0.3$ (ACF3), and $H = 0.1$ (ACF4).

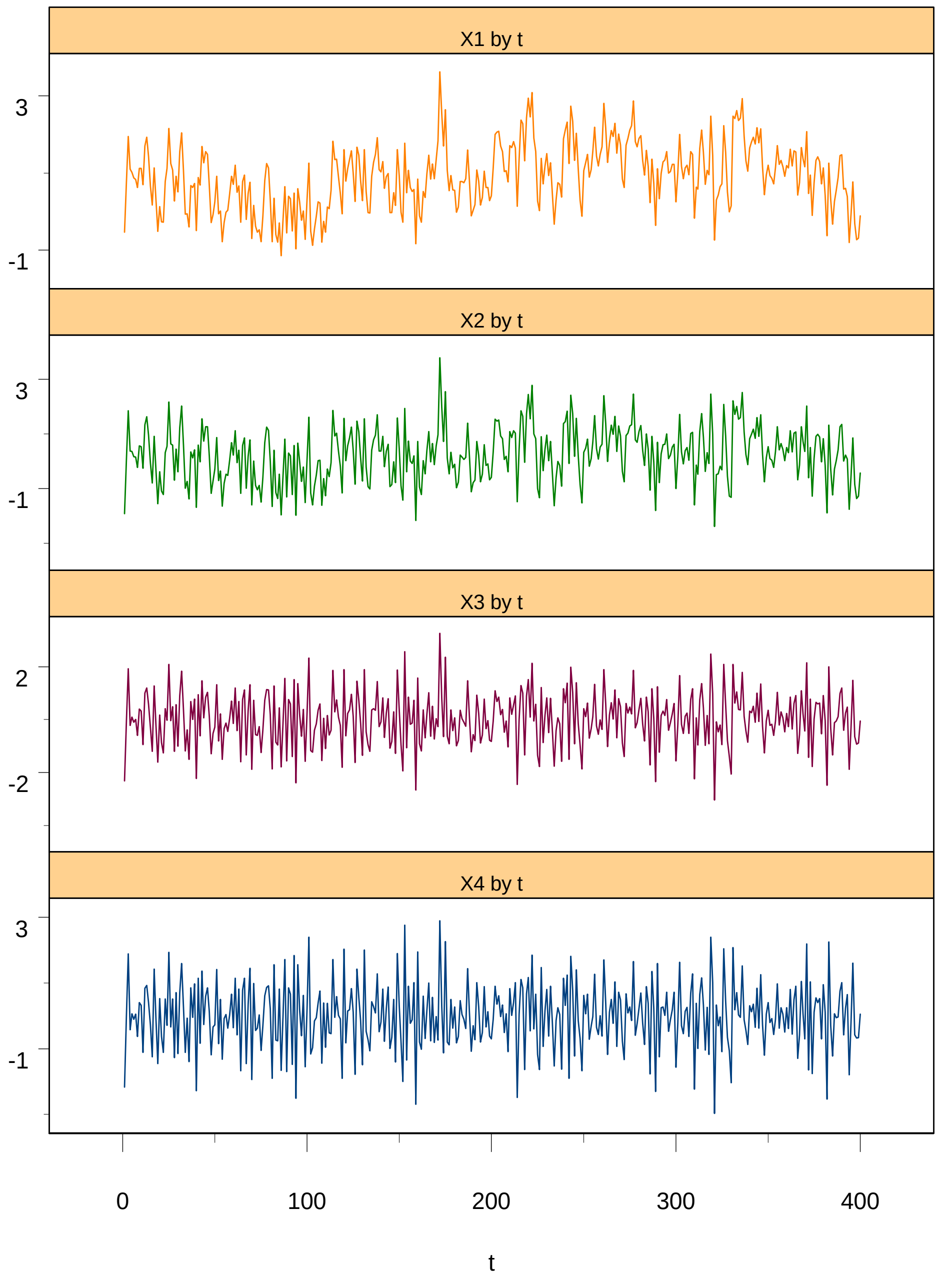


Figure 2.5: Simulated series of FGN(H) with $H = 0.9$ (X1), $H = 0.7$ (X2), $H = 0.3$ (X3), and $H = 0.1$ (X4).

2.4.2 Fractional Brownian Motion and Gaussian Noise

This section introduces a Gaussian model obtained as the increments of a Gaussian H-ss process with independent increments. The model was developed by Mandelbrot and Van Ness (1968, see also Mandelbrot and Walliss 1968, 1969(a,b)), although being implicitly considered earlier by Kolmogorov (1940), Hunt (1951) and Lamperti (1962). Results on non-Gaussian ss processes and their increments processes can be found, for instance, in Samorodnitsky and Taqqu (1994).

Consider an H-ss si, Gaussian, zero mean process Y with $0 < H < 1$. The distribution of Y is fully specified by the covariances of the process given by equation (2.13). Therefore, any $0 < H < 1$ uniquely defines a Gaussian stationary increment process, $X = \{Y_t - Y_{t-1}\}_{t\in\mathbb{Z}}$. The self-similar process Y is called a fractional Brownian motion, denoted by B^H, and the corresponding increments process is called fractional Gaussian noise. A more constructive definition of fractional Brownian motion is given below.

For $H = \frac{1}{2}$, the corresponding self-similar process $B^{\frac{1}{2}}$ turns out to be the standard Brownian motion.

Definition 2.4.3 (Brownian Motion) *A real-valued random process* $B = \{B_t\}_{t\in[0,\infty)}$ *with a.s. continuous sample path and starting point* $B_0 = 0$ *a.s. is said to be a standard Brownian motion (henceforth BM) if*

1. *It is an independent increments process, i.e., if* $0 \leq s \leq t \leq u \leq v$ *then* $B_t - B_s$ *and* $B_v - B_u$ *are independent random variables, and*
2. $B_t - B_s \sim N\left(0,\ \sigma^2 |t-s|\right)$ *for every* $t, s \in [0, \infty)$.

It is straightforward to verify from definition that BM is a stochastic continuous si process. We now show that standard BM, B, is also an H-ss with $H = \frac{1}{2}$. Because B is Gaussian, it is sufficient to look only at the expectation and covariances of B. We have for each t

$$E\left[B_t\right] = E\left[B_t - B_0\right] = 0.$$

Particularly, for any positive factor c,

$$E\left[B_{ct}\right] = c^{\frac{1}{2}} B_t. \tag{2.17}$$

Consider now the covariances $Cov\left(B_t, B_s\right)$ and assume wlg that $t \geq s$. Because the increments are independent,

$$\begin{aligned} Cov\left(B_t, B_s\right) &= Cov\left(\left[B_t - B_s\right] + \left[B_s - B_0\right], B_s\right) \\ &= Var\left(B_s - B_0\right) = \sigma^2 s = \sigma^2 \min\left(t, s\right). \end{aligned}$$

Therefore, for any $c > 0$

$$Cov\left(B_{ct}, B_{cs}\right) = c\sigma^2 \min\left(t, s\right) = Cov\left(c^{\frac{1}{2}} B_t, c^{\frac{1}{2}} B_s\right). \tag{2.18}$$

Thus, according to (2.17) and (2.18), B is $\frac{1}{2}$-ss si process.

Fractional BM can be constructed as a weighted average of a standard BM over the infinite past. A mathematically stringent definition along this line can be given in terms of a stochastic integral with respect to the Brownian motion (cf. Ash and Gardner 1975) of a kernel function whose form is determined by the self similarity parameter H. Fractional BM is then defined as follows (Beran 1994b).

Definition 2.4.4 (Fractional Brownian Motion, Fractional Gaussian Noise) *Let B be a standard BM, $0 < H < 1$, $\sigma > 0$ and*

$$w_H\left(t, u\right) = \begin{cases} 0 & \text{if } t \leq u, \\ (t-u)^{H-\frac{1}{2}} & \text{if } 0 \leq u < t, \\ (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} & \text{if } u < 0. \end{cases}$$

Also let $B^H = \left\{B_t^H\right\}_{t \in [0,\infty)}$ be defined by the stochastic integral

$$B_t^H = \sigma \int w_H\left(t, u\right) dB_u, \tag{2.19}$$

where the convergence of the integral is to be understood in the $L^2\left(\mathcal{L}\right)$-norm where $\mathcal{L}$ denotes the Lebesgue measure on the real numbers. Then B^H is said to be a fractional BM (henceforth FBM) with self-similarity parameter H. The corresponding increments process of FBM is called fractional Gaussian noise (henceforth FGN).

H-ss of B^H follows directly from the $\frac{1}{2}$-ss of B_t. Note that

$$w_H\left(ct, u\right) = c^{H-\frac{1}{2}} w_H\left(t, uc^{-1}\right).$$

Therefore,

$$B^H_{ct} = \sigma \int w_H\left(ct, u\right) dB\left(u\right) = \sigma c^{H-\frac{1}{2}} \int w_H\left(t, uc^{-1}\right) dB_u.$$

Defining $v = uc^{-1}$, we obtain

$$\sigma c^{H-\frac{1}{2}} \int w_H\left(t, v\right) dB_{cu}.$$

By self-similarity of B_t, this is equal in distribution to

$$\sigma c^{H-\frac{1}{2}} c^{\frac{1}{2}} \int w_H\left(t, v\right) dB_u = c^H B^H_t.$$

Thus, B^H_t defined by (2.19) is an H-ss si process.

It is informative to take a closer look at the weight function $w_H\left(t, u\right)$. Figure 2.6 shows $w_H\left(t, u\right)$ as a function of u for several values of H. For $H = \frac{1}{2}$

$$w_H\left(t, u\right) = \begin{cases} 1 & \text{if } 0 \leq u < t, \\ 0 & \text{otherwise.} \end{cases}$$

This imply that the increments of $B^{\frac{1}{2}}_t$ are iid as expected. If H is in the intervals $\left(0, \frac{1}{2}\right)$ or $\left(\frac{1}{2}, 1\right)$, then the weight function is proportional to $|u|^{H-\frac{3}{2}}$ as $u \to -\infty$. For $H > \frac{1}{2}$, $|u|^{H-\frac{3}{2}}$ tends so slowly to zero that

$$\int_{-\infty}^{t} w_H\left(t, v\right) du = \infty$$

for all $t \in \mathbb{R}$. For $H < \frac{1}{2}$, $|u|^{H-\frac{3}{2}}$ dies off very quickly and

$$\int_{-\infty}^{t} w_H\left(t, v\right) du = 0.$$

These properties are reflected in the corresponding properties of the correlations $\rho\left(k\right)$ of the increments process, (2.15) and (2.16).

2.5 Autoregressive Moving Average Processes

2.5.1 ARMA and ARIMA models

ARMA and ARIMA models were proposed by Box and Jenkins (1970). Because of their simplicity and flexibility, they became very popular among time series practitioners. The

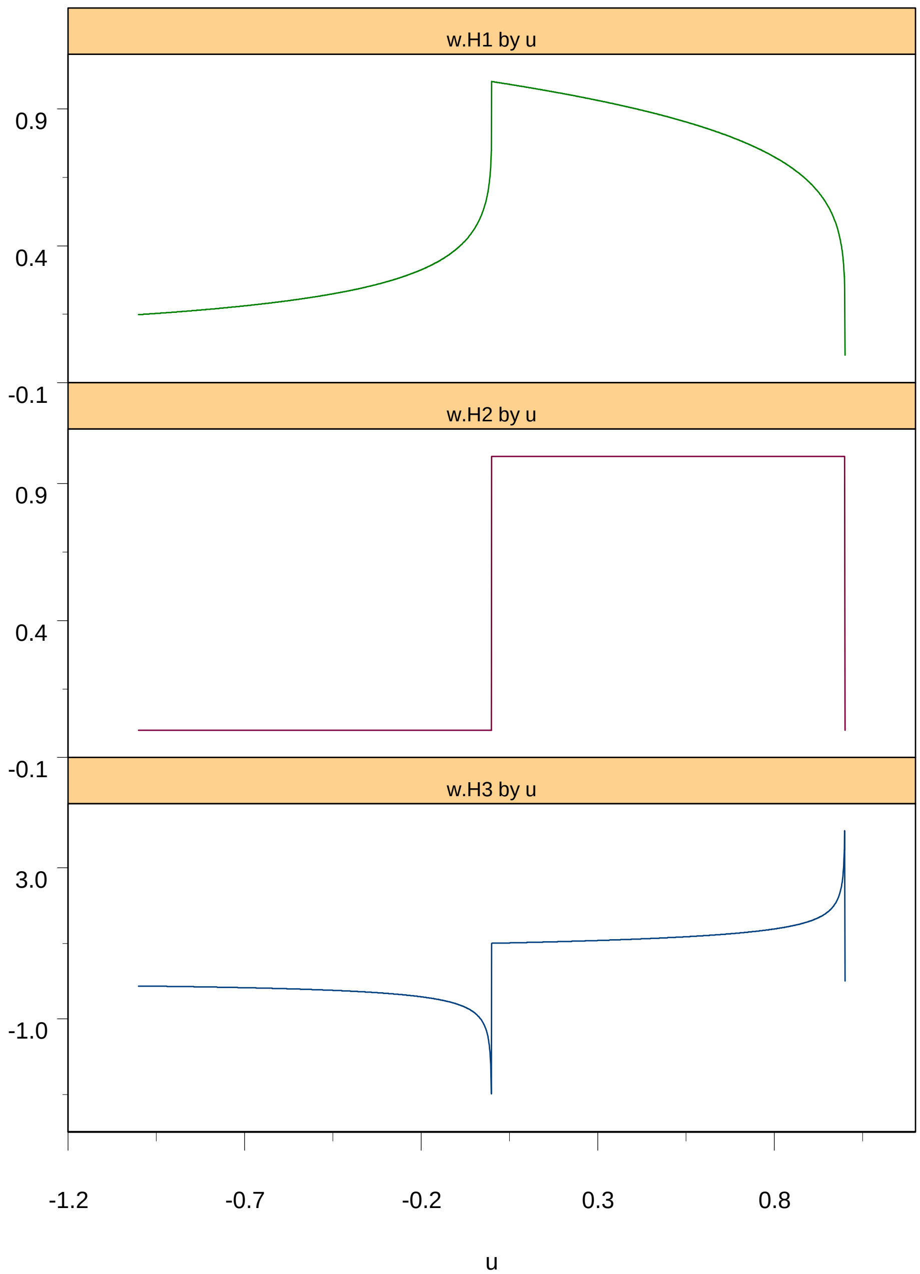


Figure 2.6: Weight function $w_H(t,u)$ for fractional Brownian motion with $H = 0.7$ (w.H1), $h = 0.5$ (w.H2), and $H = 0.3$ (w.H3).

importance of these models arises from the fact that for any autocovariance function $\gamma(\cdot)$ such that $\lim_{h\to\infty}\gamma(h) = 0$ it is possible to find an ARMA process with autocovariance function that identify with $\gamma(\cdot)$ up to an arbitrarily large finite lag, K, even though it is possible that the number of the parameters required for the ARMA process will tend to infinity as $K \to \infty$.

Let us briefly recall the definition of ARMA time series and some of its main properties. To simplify notations, assume $\mu = E(X) = 0$. Otherwise, X_t needs to be replaced by $X_t - \mu$ in all formulas. Denote by B the backshift operator, such that for a time series X,

$$B^j X_t = X_{t-j.} \tag{2.20a}$$

Let p and q be integers and define the polynomials on $\mathbb{C}$

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p,$$

and

$$\xi(z) = 1 + \xi_1 z + \ldots + \xi_q z^q.$$

Definition 2.5.1 (ARMA Process) *The process X is said to be an ARMA(p,q) process if for every t*

$$\phi(B) X_t = \xi(B)\epsilon_t, \tag{2.21}$$

where $\epsilon_t \sim WN\left(0, \sigma^2\right)$. The polynomials $\phi(z)$ and $\xi(z)$ are referred to as the autoregressive and moving average polynomials, respectively, of the ARMA process.

A fundamental result for ARMA models formulates an intimate relation between the properties of the ARMA process and the roots of the polynomials $\phi(z)$ and $\xi(z)$. We shall focus attention on autoregressive and moving average polynomials, $\phi(z)$ and $\xi(z)$ respectively, with no common zeros, and such that all solutions of $\phi(z)\xi(z) = 0$ are outside the unit circle $|z| \le 1$. In this case (cf. Box and Jenkins 1970, Chapter 3, and Brockwell and Davies 1991, Chapter 3), the asymptotic decay of the autocovariance function is exponential in the sense that there is an upper bound

$$|\gamma(k)| < ab^k$$

where $0 < a < \infty$ and $0 < b < 1$.

We now turn to describe ARFIMA models.

Definition 2.5.2 *Let d be a nonnegative integer such that the d'th difference $(1-B)^d X_t$ is an ARMA(p,q). In this case X is said to be an integrated ARMA(p,d,q) or ARIMA(p,d,q) process, satisfying a difference equation of the form*

$$\phi^*(B) X_t = \phi(B)(1-B)^d X_t = \xi(B)\epsilon_t, \qquad \epsilon_t \sim WN\left(0,\sigma^2\right). \tag{2.22}$$

Thus, for ARIMA(p,d,q) processes the autoregressive polynomial $\phi^*(z)$ of equation (2.22) has a zero of order d at $z = 1$. ARIMA(p,d,q) process X is stationary if and only if $d = 0$, in which case it reduces to an ARMA(p,q) process. For such process it would be natural to apply the operator $\nabla = 1 - B$ repeatedly d times, as $\nabla^d X = \left\{\nabla^d X_t\right\}_{t\in\mathbb{Z}}$ will be a stationary process with rapidly decaying sample autocorrelation function, compatible with that of an ARMA process with no zeros of the autoregressive polynomial on the unit circle.

The spectral density of an ARMA process is given (cf. Priestly 1981) by

$$f_X(\omega) = \frac{\sigma^2}{2\pi}\frac{\left|\xi\left(e^{-i\omega}\right)\right|^2}{\left|\phi\left(e^{-i\omega}\right)\right|^2}. \tag{2.23}$$

Since $f_X(\omega)$ is a rational function of $e^{-i\omega}$, it is uniquely determined by σ^2 and the autoregressive and moving average polynomials of the ARMA process.

2.5.2 Fractional ARIMA Models

Fractional autoregressive integrated moving average model of order p,d,q, abbreviated as ARFIMA(p,d,q), was proposed independently by Granger and Joyeux (1980) and Hosking (1981) as a natural extension of the classic ARIMA(p,d,q) model. In contrast to the ARIMA models, in ARFIMA models the differencing parameter, d, may take any real value. ARFIMA models have become very popular since they offer much efficacy in modeling both the long and short-run behaviour of a time series. While any stationary process can always be approximated by a simple ARMA(p,q) process, the orders p and q required to achieve a reasonably good approximation may be so large as to become unwieldy for estimation. In cases of long memory or anti-persistent time series, ARIMA(p,d,q) models offer a convenient and effective formulation of the dependence structure of the series.

The extension of ARIMA models to fractionally differenced models is achieved by the following way. For any real number, d, we define the difference operator $\nabla^d = (1-B)^d$ by means of the binomial expansion,

$$\nabla^d = (1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j,$$

where B is the backshift operator (2.20a), and

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} = \prod_{0<k\leq j} \frac{k-1-d}{k}, \qquad j = 0, 1, 2, \ldots. \tag{2.24}$$

Here, Γ is the Gamma function

$$\Gamma(x) = \begin{cases} \int_0^{\infty} t^{x-1} e^{-t} dt, & x > 0, \\ \infty, & x = 0, \\ x^{-1}\Gamma(1+x), & x < 0. \end{cases}$$

Definition 2.5.3 (ARFIMA process) *Let d be a real number such that the d'th difference $\nabla^d X$ is an ARMA(p,q). In this case X is said to be an fractionally integrated ARMA(p,d,q) or ARFIMA(p,d,q) process, satisfying a difference equation of the form*

$$\phi(B)\nabla^d X_t = \xi(B)\epsilon_t, \qquad \epsilon_t \sim WN\left(0, \sigma^2\right), \tag{2.25}$$

where $\phi(z)$ and $\xi(z)$ are the autoregressive and moving average polynomials of degree p and q, respectively.

As in Section 2.5.1, $\phi(z)$ and $\xi(z)$ are assumed hereinafter to have no common zeros, and to have no zeros in the complex unit circle $|z| \leq 1$.

For $d < \frac{1}{2}$ the ARFIMA process can be shown to be stationary (see Brockwell and Davies 1991, Section 13.2) and to have a spectral density, $f(\omega)$, given by

$$\begin{aligned} f(\omega) &= f_{ARMA}(\omega)\left|1-e^{-i\omega}\right|^{-2d} \qquad (2.26) \\ &= \frac{\sigma^2}{2\pi}\frac{\left|\xi\left(e^{-i\omega}\right)\right|^2}{\left|\phi\left(e^{-i\omega}\right)\right|^2}\left|1-e^{-i\omega}\right|^{-2d}, \qquad \omega \in \Pi. \end{aligned}$$

Similarly to the spectral density of an ARMA process (2.23), it is uniquely determined by $\xi(z), \phi(z), \sigma^2$ and by d.

Since $e^{-i\omega} \sim 1 - \omega$ as $\omega \to 0$, we see from (2.26) that the behaviour of the spectral density very close to the origin is given by

$$f(\omega) \sim \frac{\sigma^2}{2\pi} \frac{|\xi(1)|^2}{|\phi(1)|^2} \left|1 - e^{-i\omega}\right|^{-2d} \sim f_{ARMA}(0) |\omega|^{-2d}. \tag{2.27}$$

For $0 < d < \frac{1}{2}$, the spectral density has a pole at zero and the time series, hence, has long memory. For $-\frac{1}{2} < d < 0$, on the the hand, we have $f(0) = 0$, and the time series is, thus, anti-persistent. The rate at which the spectral density tends to $f(0)$ is compatible with property (2.6) of long memory where the memory parameter α corresponds to $2d$.

The range of the memory parameter corresponding to both causality and invertibility of the time series is $|d| < \frac{1}{2}$ (see Theorem 2.5.1 below). The cases $d > \frac{1}{2}$ or $d \le -\frac{1}{2}$ can be reduced to the case $-\frac{1}{2} < d \le \frac{1}{2}$ by taking $\left\lfloor |d| + \frac{1}{2} \right\rfloor$ differences or partial sums, respectively, where $\lfloor x \rfloor$ denotes the integer part of d. For instance, if X is an ARFIMA process with $d = 0.8$, then the differenced process $\nabla X = \{X_t - X_{t-1}\}_{t \in \mathbb{Z}}$ is a stationary ARFIMA process with $d = -0.2$. On the other hand, if X is an ARFIMA process with $d = -0.8$, then the partial sums $S_t = \sum_{j=1}^{t} X_j$ are an ARFIMA process with $d = 0.2$.

For ARFIMA(0,d,0) with $|d| < \frac{1}{2}$ the unique solution of (2.25) is given by

$$X_t = \nabla^{-d} \epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

Hosking (1981) gave the explicit expression for the coefficients ψ_j of the causality representation of the ARFIMA(0,d,0) process with $|d| < \frac{1}{2}$ as

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} = \prod_{0 < k \le j} \frac{k-1+d}{k}, \qquad j = 0, 1, 2, ... \tag{2.28}$$

Also, in this case, the coefficients π_j of the invertibility representation of the process are given by (2.24). The form of the covariances of ARFIMA(0,d,0) processes follows from a formula in Gradshteyn and Ryzhik (1965, p. 372),

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} = \sigma^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}.$$

Then, the correlations are equal to

$$\rho(k) = \frac{(-1)^k \Gamma(1-d)^2}{\Gamma(k-d+1)\Gamma(1-k-d)} \tag{2.29}$$
$$= \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} = \prod_{0<k\leq h} \frac{k-1+d}{k-d}, \qquad h = 1, 2, ...$$

Applying the Stirling's approximation $\Gamma(x) = \sqrt{2\pi/x}\left(\frac{x}{e}\right)^x\left(1+O\left(x^{-1}\right)\right)$ to (2.24), (2.28) and (2.29), we obtain

$$\pi_j \sim c_1 \frac{1}{\Gamma(-d)}\, j^{-d-1} \quad as \;\; j \to \infty, \tag{2.30}$$

$$\psi_j \sim c_2 \frac{1}{\Gamma(d)} j^{d-1} \quad as \;\; j \to \infty,$$

and

$$\rho(k) \sim c_3 \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1} \quad as \;\; j \to \infty,$$

where c_1, c_2, c_3 are positive constants. Figure 2.7 shows the autocorrelation function of several ARFIMA(0,d,0) series. This Figure demonstrates the effect of different values of d on the form of the correlations.

For general ARFIMA(p,d,q), the explicit formulas for the coefficients ψ_j, π_j and the autocovariances $\gamma(k)$ are quite complex (cf. Sowell 1992). However, similar limiting results are obtained as in the ARFIMA(0,d,0) case. The following proposition summarizes some of the properties of general ARFIMA(p,d,q) processes (cf., Brockwell and Davies 1991, pp. 524-525 and Beran 1994, pp. 63-65).

Proposition 2.5.1 *Suppose that $d \in (-0.5, 0.5)$ and let $\phi(z)$ and $\xi(z)$ be the autoregressive and moving average polynomials of order p and q respectively. Suppose also that $\phi(z)$ and $\xi(z)$ have no common zeros, and $\phi(z)\xi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Then*

(i) The ARFIMA equation (2.25) has a unique stationary solution, X, satisfying

$$X_t = \nabla^{-d}\phi^{-1}(B)\xi(B)\epsilon_t = \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}, \tag{2.31}$$

Thus, ψ_j is the corresponding coefficient of z^j, determined by the expansion of

$$(1-z)^{-d}\phi^{-1}(z)\xi(z)$$

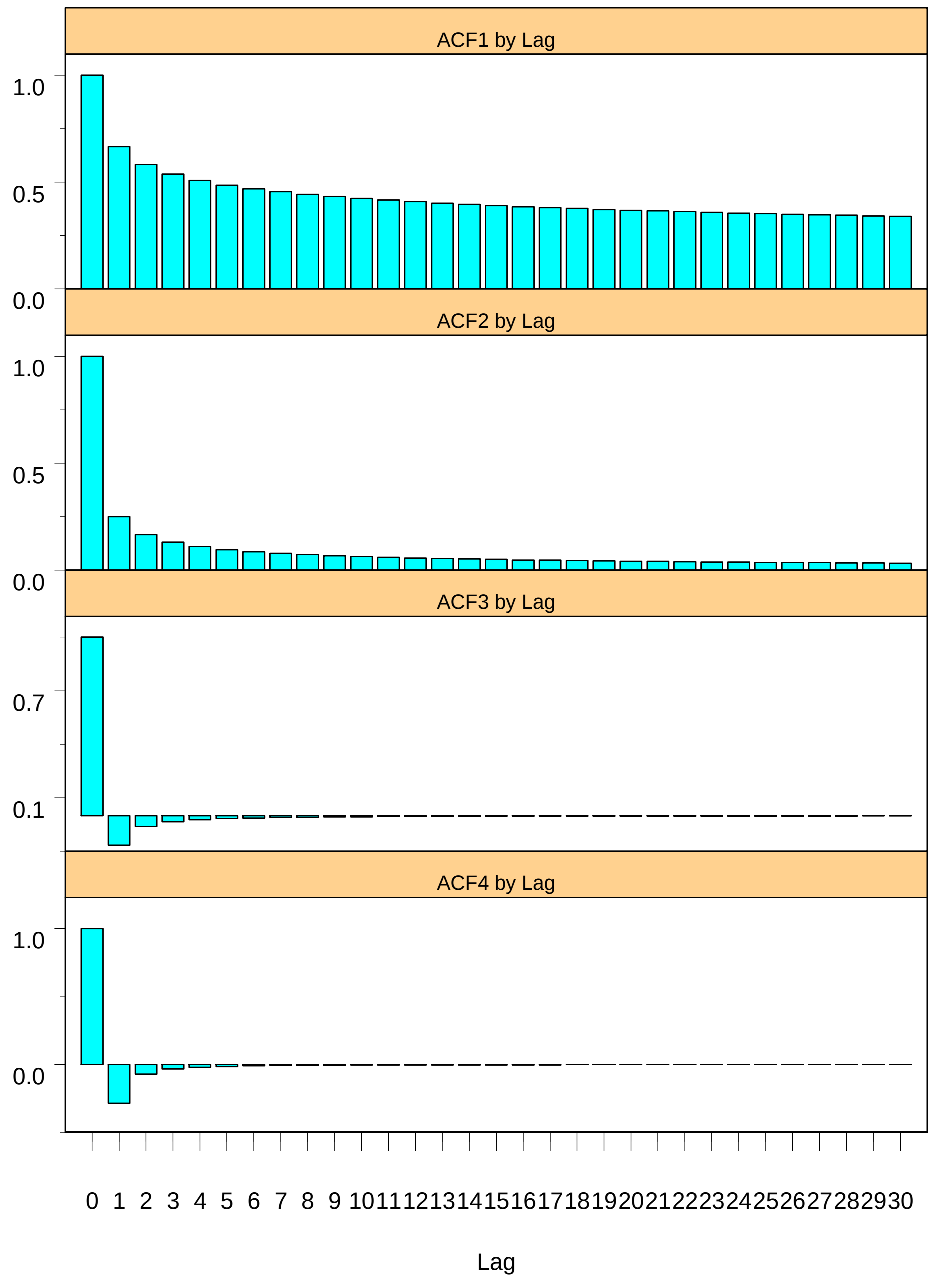


Figure 2.7: Autocorrelation function of ARFIMA(0,d,0) time series with $d = 0.4$ (ACF1), $d = 0.2$ (ACF2), $d = -0.2$ (ACF3), and $d = -0.4$ (ACF4).

as power series in z *for* $|z| \leq 1$*. Furthermore,*

$$\left|\psi_j\right| \sim c_1 j^{d-1} \quad as \quad j \rightarrow \infty,$$

for some $c_1 > 0$*.*

(ii) X *is causal, where the coefficients* ψ_j *of the causality representation of* X *are determined by (2.31).*

(iii) X *is invertible, where the coefficients* π_j *of the invertibility representation of* X *are determined by the relation*

$$\epsilon_t = \xi^{-1}(B)\,\phi(B)\,\nabla^d X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \tag{2.32}$$

Thus, π_j *is the corresponding coefficient of* z^j *, determined by the expansion of*

$$\xi^{-1}(z)\,\phi(z)\,(1-z)^d$$

as power series in z *for* $|z| \leq 1$*. Furthermore,*

$$|\pi_j| \sim c_2\ j^{-d-1} \quad as \quad j \rightarrow \infty,$$

for some $c_2 > 0$*.*

(iv) *The autocovariance function of* X *satisfy, for* $d \neq 0$

$$|\gamma(K)| \sim c_3 k^{2d-1} \quad as\ k \rightarrow \infty,$$

where $c_3 > 0$*.*

Figures 2.8 and 2.9 show sample paths of several ARFIMA(0,d,0) time series with $d = 0.4, 0.2, -0.2, -0.4$ and of AR(1) time series (AR(1) is an ARMA(1,0) process) with the same lag-1 correlation, respectively, as in Figure 2.8. Both figures were produced after applying ARFIMA and AR filters on the white noise series used for Figure 2.1. Again, we can see the prominent characteristics of the long and negative memory behaviour in Figure 2.8, while in Figure 2.9 it can be seen that although the series are certainly correlated, the correlations fade away rapidly. This can be seen, for example, by looking at time points where the series seems to have a sudden large "jump". Large jumps can be the result of an

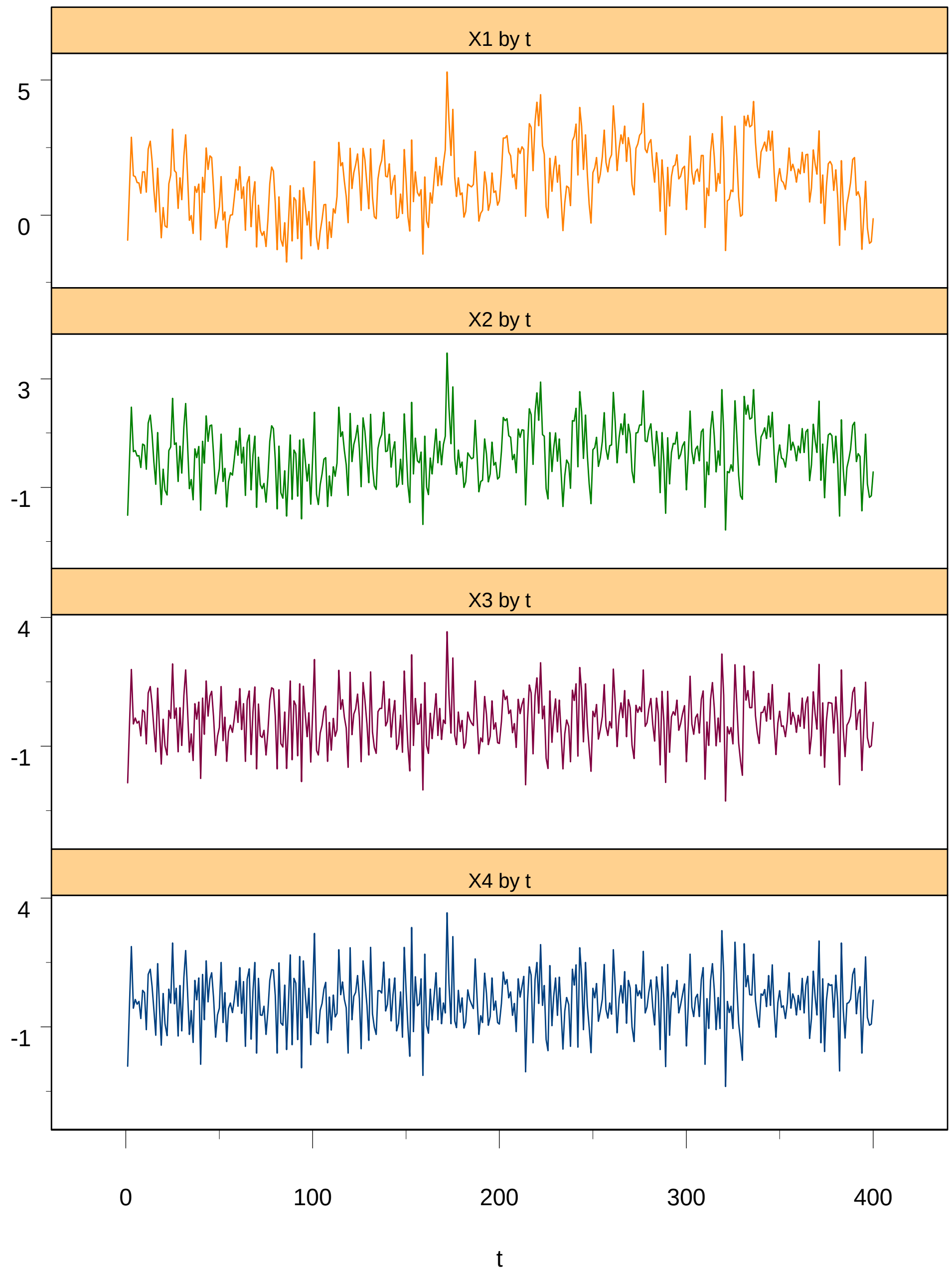


Figure 2.8: Simulated series of ARFIMA(0,d,0) with $d = 0.4$ (X1), $d = 0.2$ (X2), $d = -0.2$ (X3) and $d = -0.4$ (X4).

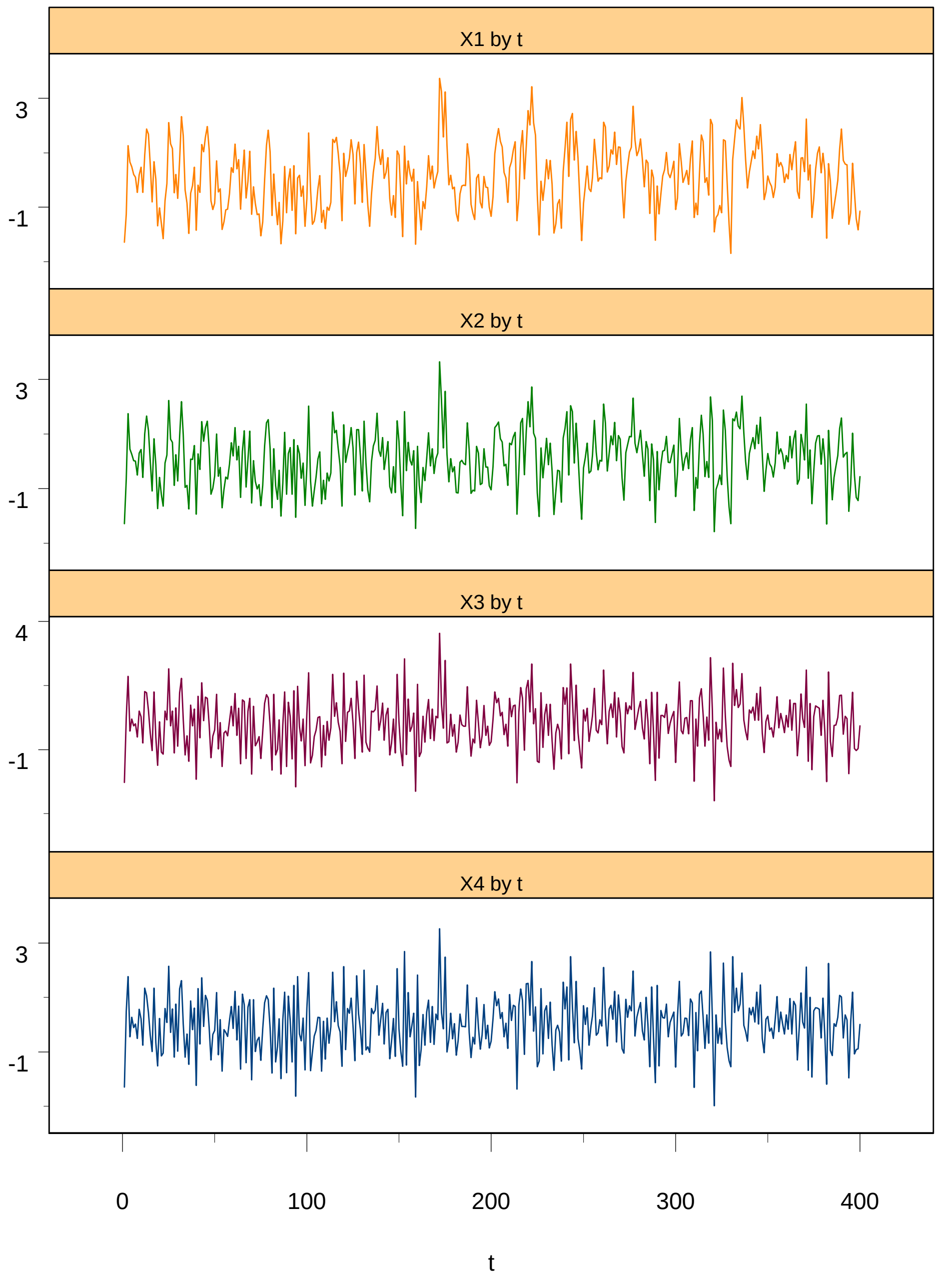


Figure 2.9: Simulated series of AR(1) with the same lag-1 autocorrelation as the process used for Figure 2.8.

instantaneous large noise term, say $|\epsilon_{t_1}| \gtrapprox 2\sigma$, that was added to the series at time t_1. If the time series has long memory or anti-persistent time series, then any such jump in the series is more probable to have an effect on the behaviour of the consequent observations, $X_t,\ t > t_0$, in relatively long times after t_0. While for time series with long memory this effect is results in relatively long period where the consequent observations tend to stay with similar level to that of the jump, for anti-persistent time series it results in relatively long period (though shorter than in the long memory case) where the consequent observations tend to stay in the opposite level to that of the jump. On the other hand, if the series is correlated with short-memory, then any such large jump would have a possibly sharp but short-lasting effect on the behaviour of the observations in consequent times after t_0.

Chapter 3

LITERATURE REVIEW

3.1 Introduction

This chapter provides a summary of some prominent methods used for estimation of long memory time series. The basic idea that stands behind each of the techniques of estimation is discussed, as well as the advantages and disadvantages of these methods. Some references to the literature are provided as a complement to the topics not covered here in detail. Since we could not possibly cover here all estimation methods used for long memory time series, we had to leave out some methods, and sometimes mention others only briefly. Some other comprehensive surveys on estimation of long range dependence, which consist also of other methods that are not included here, are Beran (1994b), Taqqu, Teverovski and Willinger (1995), Baillie (1996), Giraitis and Robinson (2003), Moulines and Soulier (2003), Chan and Palma (2006).

For ease of exposition, we use the memory parameter α related to the form of the spectral density near the origin, as in (2.6). Note that $\alpha = 2H - 1$ if H is the self-similarity parameter of an FGN, and $\alpha = 2d$ if d is the fractional-differencing parameter of an ARFIMA process (Sections 2.4-2.5).

The rest of Chapter 3 is organized as follows. Section 3.2 considers direct estimation of the mean and the variance of a time series with long-range dependence. In Section 3.3 we present some parametric methods for estimation. While the issue of order-determination is not discussed here, it should be noted that standard model choice procedures developed for short-memory models (see, e.g., Akaika 1973, Schwarz 1978, Parzen 1974, Hannan 1980, Shibata 1980) may also be applied for long memory time series (Crato and Ray 1996, Beran, Bhansali and Ocker 1998). Since parametric estimation may be a demanding procedure in terms of the required CPU time, we also discuss the computational aspects of the parametric methods and we provide their order of complexity, as given by Chan and Palma (2006). In Section 3.4 the reader is presented to semiparametric methods of estimation of the memory parameter. These methods are generally less efficient than the parametric ones, in case where the model is correct. This is true for all the semiparametric methods of estimation, which have a slower rate of convergence of the estimator to the true estimator. However, the semiparametric methods are naturally more robust, essentially having much weaker assumptions on the process, and therefore they are more reliable in cases that we do not have much knowledge about the underlying process.

3.2 Point Estimation

3.2.1 Estimation of Location

A widely used class of estimators of the mean, μ, of a time series is the class of linear unbiased estimators. An estimate of this class is given by a weighted average of $X_1, ..., X_n$,

$$\widehat{\mu}_{\mathbf{c}} = \sum_{j=1}^{N} c_j X_j = \mathbf{c}'\mathbf{X},$$

such that

$$\sum_{j=1}^{N} c_j = \mathbf{c}'\mathbf{1} = 1. \tag{3.1}$$

Here we used $\mathbf{c} = (c_1, ..., c_N)'$, $\mathbf{X} = (X_1, ..., X_N)'$ and $\mathbf{1} = (1, ..., 1)'$. The sample mean, for example, is obtained by $c_j = \frac{1}{N}$ for all $j = 1, ..., N$. The variance of a general linear unbiased estimator $\widehat{\mu}_c$ is equal to

$$Var\left(\widehat{\mu}_c\right) = \sum_{j,l=1}^{N} c_j c_l \gamma\left(j-l\right) = \mathbf{c}'\Sigma_N \mathbf{c}. \tag{3.2}$$

where $\Sigma_N = \left[\gamma\left(j-l\right)\right]_{j,l=1,...,N}$ is the covariance matrix of $\mathbf{X}$. Minimizing (3.2) with respect to $\mathbf{c}$, under the constraint (3.1), yields

$$\mathbf{c} = \Sigma_N^{-1}\mathbf{1}\left[\mathbf{1}'\Sigma_N^{-1}\mathbf{1}\right]^{-1}, \tag{3.3}$$

and thus the best linear unbiased estimator (BLUE) is given by

$$\widehat{\mu}_{BLUE} = \mathbf{c}'\mathbf{X} = \left[\mathbf{1}'\Sigma_N^{-1}\mathbf{1}\right]^{-1}\mathbf{1}'\Sigma_N^{-1}\mathbf{X}. \tag{3.4}$$

Its variance is equal to

$$Var\left(\widehat{\mu}_{BLUE}\right) = \left[\mathbf{1}'\Sigma_N^{-1}\mathbf{1}\right]^{-1}.$$

$\widehat{\mu}_{BLUE}$ is therefore optimal in the sense that it has the smallest possible variance among all unbiased estimators. Adenstedt (1974) discussed the form and behaviour of the BLUE in the case where the spectral density of the process $f(\omega)$ behaves, in accordance to property (2.6), like $c_1 \cdot |\omega|^{-\alpha}$ at the origin, where $c_1 > 0$ and $\alpha < 1$. Adenstedt (1974) found that, for a large class of spectral densities, the asymptotic form of $Var\left(\widehat{\mu}_{BLUE}\right)$ is determined solely by the behaviour of $f(\omega)$ near $\omega = 0$, and he proved that $Var\left(\widehat{\mu}\right) = O\left(N^{\alpha-1}\right)$ as $N \to \infty$ (see Lemma 4.4.1 of this thesis).

Although (3.3) and (3.4) are simple, one needs to know all covariances $\gamma(0), ..., \gamma(n-1)$ in order to calculate $\widehat{\mu}_{BLUE}$ and $Var\left(\widehat{\mu}_{BLUE}\right)$. Usually, the covariances are unknown and have to be substituted by their estimated values. This makes the estimation of μ rather complicated, and the distribution of $\widehat{\mu}$ becomes complex as well. A simple alternative estimate of μ, which can be calculated without knowing Σ_N, is the sample mean. Samarov and Taqqu (1988) investigated the relative asymptotic efficiency of $\overline{X}_N$ compared to the BLUE in long memory time series. They found that under mild regularity conditions the ratio of (3.4) divided by $Var\left(\overline{X}_N\right)$ is asymptotically equal to

$$eff\left(\overline{X}_N, BLUE\right) = \begin{cases} (\alpha+1)\,\frac{\Gamma\left(1+\frac{\alpha}{2}\right)\Gamma(2-\alpha)}{\Gamma(1-\alpha)} & : -1 < \alpha < 1 \\ 0 & : \alpha < -1 \end{cases}. \tag{3.5}$$

Figure 3.1 displays $eff\left(\overline{X}_N, BLUE\right)$ as a function of $\alpha \in (-1,1)$. For $\alpha \geq 0$ (in the cases of short and long memory), the asymptotic efficiency is always above 98%. Therefore, in these cases it is sufficient to use the sample mean instead of the much more complicated BLUE. However, as the values of α become negative (in the case that corresponds to negative memory), the sample mean tends to be inefficient. Particularly, if the spectrum has a zero at the origin of order 1 or greater, the sample mean has a slower rate of convergence than the BLUE. Vitale (1973) had earlier discussed similar issues in this case.

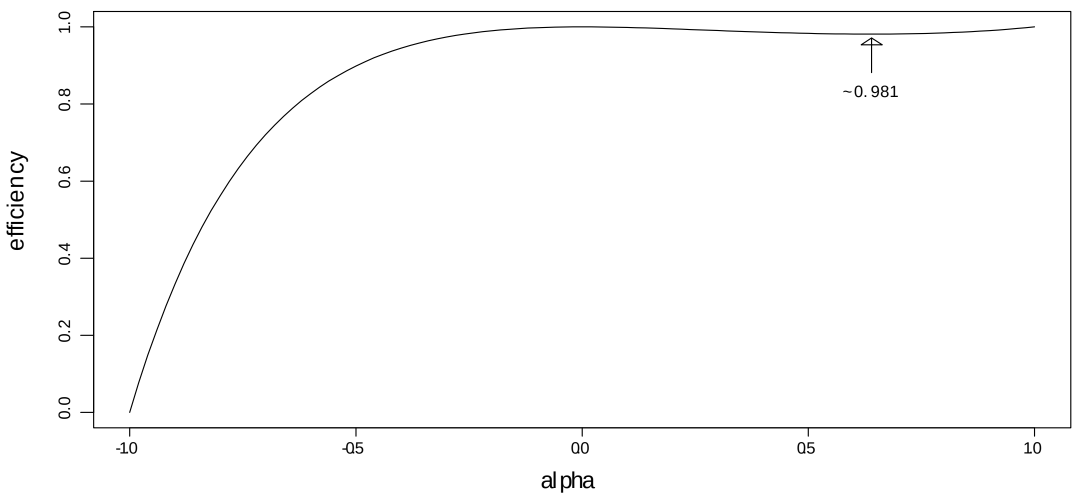


Figure 3.1: Asymptotic efficiency of the sample mean as a function of alpha.

If X is Gaussian, the BLUE is also the maximum likelihood estimator of μ. This implies also that for Gaussian long-memory processes, the sample mean is almost optimal. In practice, however, deviations from the normal distribution are expected. The BLUE and the sample mean are both very sensitive to outliers and other deviations from normality. A large number of useful location estimators that are less sensitive to deviations from the ideal distribution can be defined as or approximated by M-estimators (see e.g. Huber 1981, Hampel et al. 1986). An M-estimator $\widehat{\mu}$ of the location parameter μ of a distribution $F_{\mu,\sigma}(x) = F\left(\frac{x-\mu}{\sigma}\right)$ is defined by

$$\sum_{i=1}^{N} \Psi\left(\frac{X_i - \widehat{\mu}}{\sigma}\right) = 0, \tag{3.6}$$

where Ψ is a function such that

$$\int \Psi\left(\frac{x-\mu}{\sigma}\right) dF_{\mu,\sigma}(x) = 0.$$

If the scale is unknown, we have to replace σ by a suitable estimate. For example, the sample mean is defined by (3.6) with $\Psi(x) = x$. The median is obtained by setting Ψ

equal to the sign of x. Particularly applicable are bounded $\Psi-$functions, as the resulting estimates are not sensitive to deviations from the ideal distribution (see, e.g., Hampel et al. 1986). Beran (1991) investigated the asymptotic distribution of M-estimators of location for processes of the form

$$X_t = \mu + G\left(Z_t\right),$$

where$\{Z_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian process with zero mean, variance 1 and long-range dependence defined by (2.6), and where $G\left(\cdot\right)$ is a real measurable function. Beran (1991) proved that under some mild regularity conditions, the relative asymptotic efficiency of $\widehat{\mu}$ compared to the sample mean is equal to a positive constant, $0 < c \leq 1$, which depends on the function Ψ. In the special case of Gaussian observations with long memory, i.e. $G\left(x\right) \equiv x$, all M-estimators are asymptotically equivalent, that is $c = 1$, and no efficiency is lost by robustification. Beran (1991) also considered the special case of intermediate-memory processes where the sum of all correlations is zero, and proved that in this case, the asymptotic efficiency of all M-estimators with nonlinear Ψ-function is zero. The reason is that the variance of the sample mean converges to zero at a faster rate than N^{-1}. This follows from the fact that $Var\left(\overline{X}_N\right)$ is equal to the sum of all covariances $\gamma\left(i-j\right)$ $\left(i, j = 1, ..., N\right)$ and the assumption that the sum of all covariances is zero. This link is destroyed when a nonlinear Ψ-function issued. The variance of the resulting estimator converges to zero at the rate N^{-1}.

3.2.2 Estimation of Scale

Consider now direct estimation of the variance $\sigma^2 = Var(X_t)$ of a stationary dependent series. Assuming first that X has known zero mean, we can estimate σ^2 by

$$s^2 = \frac{1}{N}\sum_{t=1}^{N} X_t^2.$$

For a very wide range of dependent series under (2.6) with $-1 < \alpha < \frac{1}{2}$, s^2 is asymptotically normally distributed (Taqqu 1975, 1979). Particularly if X is Gaussian, we have

$$\sqrt{N}\left(s^2 - \sigma^2\right) \rightarrow N\left(0, 2\sum_{n=-\infty}^{\infty} \gamma\left(n\right)^2\right), \tag{3.7}$$

For $\alpha > \frac{1}{2}$, however, $\gamma(n)$ is not square summable, and the rate at which s^2 converges to σ^2 is different. Rosenblatt (1961) showed that $N^\alpha\left(s^2 - \sigma^2\right)$ has a nonnormal limiting distribution, which Taqqu (1975) termed the Rosenblatt distribution. If the mean of X is unknown, and instead we estimate σ^2 by

$$s^2 = \frac{1}{N-1}\sum_{t=1}^{N}\left(X_t - \overline{X}_N\right)^2$$

then, for $-1 < \alpha < \frac{1}{2}$, (3.7) still holds, but the limit distribution of $N^\alpha\left(s^2 - \sigma^2\right)$ for $\alpha > \frac{1}{2}$ contains an additional χ^2-distributed term, besides the Rosenblatt one. At any rate, for $\alpha > \frac{1}{2}$, the efficiency of the classical scale estimator s^2 is equal to zero, because its rate of convergence is slower than $\sqrt{N}$. This is in contrast to estimators obtained by maximum likelihood and related parametric methods discussed in the next section. They are $\sqrt{N}$ consistent for all $\alpha \in (-1, 1)$ and asymptotically normal.

3.3 Parametric Estimation

3.3.1 Exact Gaussian MLE

Suppose that X is a stationary Gaussian process with mean μ_0 and covariance matrix $\Sigma_N = [\gamma(j-l)]_{j,l=1,\dots,N}$. Let the spectral density be characterized by an unknown finite dimensional parameter vector, $\theta \in \Theta$. Thus, we assume that the spectral density belongs to a parametric family $f(\omega) = f_\theta(\omega)$ where $\theta \in \Theta \subseteq \mathbb{R}^p$.

The Gaussian likelihood function of X is equal to

$$f_N(\mathbf{X};\theta) = 2\pi^{-\frac{N}{2}}\left|\Sigma_N\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{X}-\mu\mathbf{1})'\Sigma_N(\theta)^{-1}(\mathbf{X}-\mu\mathbf{1})}, \tag{3.8}$$

where, as before, μ represents the mean of the process, $\mathbf{X} = (X_1, ..., X_N)'$ is the data vector and $\mathbf{1}$ is the N-length vector $(1, ..., 1)'$. The log-likelihood function is then given by

$$-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_N(\theta)\right| - \frac{1}{2}(\mathbf{X}-\mu\mathbf{1})'\Sigma_N(\theta)^{-1}(\mathbf{X}-\mu\mathbf{1}).$$

Since the log transformation is monotonic, the ML estimate is obtained by

$$\widehat{\theta}_N = \text{argmin}_\theta \mathcal{L}_N(x;\theta), \tag{3.9}$$

where $\mathcal{L}_N(\mathbf{X};\theta)$ is the normalized log-likelihood function

$$\begin{aligned}\mathcal{L}_N(\mathbf{X};\theta) &= -\frac{1}{N}\log f_N(\mathbf{X};\theta) - \frac{1}{2}\log(2\pi) \\ &= \frac{1}{2N}\log|\Sigma_N(\theta)| + \frac{1}{2N}(\mathbf{X}-\mu\mathbf{1})'\,\Sigma_N(\theta)^{-1}(\mathbf{X}-\mu\mathbf{1})\,.\end{aligned} \tag{3.10}$$

The literature on the Gaussian MLE for dependent observations developed first for short memory processes. Hannan (1973) established asymptotic normality and efficiency of the estimator in the case of the true mean μ_0 being known. Hannan also considered the Whittle's (1953) approach discussed in Section 3.3.3. For the long memory case, Yajima (1985) considered the Gaussian MLE of an ARFIMA(0,d,0) with $0 < d < \frac{1}{2}$, or equivalently $0 < \alpha < 1$. Dahlhaus (1989) generalized Yajima's (1985) result to general Gaussian long memory processes, $0 < \alpha < 1$, where the true mean of the process is possibly unknown. Dahlhaus (1989) proved that under some mild regularity conditions,

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \underset{d}{\rightarrow} N\left(0, \Gamma(\theta_0)^{-1}\right), \tag{3.11}$$

where θ_0 denotes the true parameter of the process and $\Gamma(\theta)$ is the Fisher information matrix given by

$$\Gamma(\theta) = -E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\mathcal{L}_N(\theta)\right)_{j,k=1,\ldots,p}.$$

Particularly, the asymptotic variance of $\left(\hat{\theta}_N - \theta_0\right)$ reaches the Cramér-Rao bound, and thus Dahlhaus proved efficiency of the Gaussian MLE in the sense of Fisher.

When μ_0 is unknown, Dahlhaus (1989) required to substitute the unknown mean μ, that appears in the Gaussian likelihood function, by any $N^{\{1-\alpha\}/2}$-consistent estimate of μ. The arithmetic mean is the most common estimate of μ, and it provides this required property (see Section 3.2.1). Thus, in the case of an unknown mean, a plug-in estimator is required, which usually replaces the unknown mean parameter μ with the sample mean, before estimating the other parameters with the exact or approximated MLE. For ease of exposition in this thesis, we frequently refer to Dahlhaus's Gaussian MLE as the exact Gaussian MLE (or just exact MLE) in both cases where the mean μ is assumed to be known or not. However, as mentioned by Lieberman (2005), the plug-in MLE can only be exact if the estimator of μ is its profile MLE (3.4).

In general, the solution of (3.9) does not have a closed form, and it needs to be assessed numerically by evaluating the normalized log-likelihood function, (3.10), for many trial values of $\theta \in \Theta$. Particularly, for each trial value of θ, this requires the calculation of the determinant and the inverse of $\Sigma_N(\theta)$. The computation of the determinant and the inverse of a nonsingular square $N \times N$ matrix has arithmetic complexity of order $O(N^3)$, i.e., by the Cholesky decomposition (cf. Press et al. 1992, p. 34). That makes the procedure very costly in terms of CPU, in particular if the dimension of θ is high or if the time series is very long. The number of trial values of θ may be reduced by using gradient descent procedure to search for the Likelihood maximum, usually with several starting values for the parameter θ. This was suggested by Sowell (1992), who discussed other computational aspects of the exact MLE procedure as well. Sowell (1992) derived an explicit numerical iterative procedures for obtaining the exact MLE of the Gaussian ARFIMA(p,d,q) process with known mean, and he analyzed the procedure performance for the whole range of $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, or equivalently, $\alpha \in (-1, 1)$. In order to carry out the evaluations of the Likelihood function faster, Sowell proposed an improvement for the Cholesky decomposition based on a recursive calculation of the determinant and the inverse of $\Sigma_N(\theta)$ with the Levinson's algorithm (Sowell 1989). Sowell's method reduces the arithmetic complexity of the procedure to $O(N^2)$. However, Sowell's method still suffer from computational drawback, since it requires an excessive memory in order to store the Cholesky factors (see Doornik and Ooms 2003). A different approach to handle the computational problem (cf. Beran 1994b, Chan and palma 2006) is to decompose the log-likelihood function as

$$f_N(X;\theta) = f_1(X_1;\theta) f_1(X_2;\theta|X_1) \cdots f_1(X_N;\theta|X_1, ..., X_{N-1}), \tag{3.12}$$

where $f_1(X_j;\theta|X_1, ..., X_{j-1})$ denotes a one-dimensional normal density function of X_j given $X_1, ..., X_{j-1}$. Since X is a stationary Gaussian process, then the conditional expectation $E(X_j|X_1, ..., X_{j-1})$ is equal to the best linear prediction of X_j given $X_1, ..., X_{j-1}$,

$$\mu_{X_j|X_1,...,X_{j-1}} = E(X_j;\theta|X_1, ..., X_{j-1}) = \sum_{s=1}^{j-1} \beta_{j-1,s}(X_{j-s} - \mu),$$

and the conditional variance is

$$\sigma^2_{X_j|X_1,...,X_{j-1}} = E\left[\left(X_j - \widehat{X}_{j+1}\right)^2; \theta|X_1, ..., X_{j-1}\right],$$

(see, e.g., Brockwell and Davis 1987, section 2.7). In fact $\left(X_t - \mu_{X_t|X_1,\ldots,\mathbf{x}_{t-1}}\right)_{t=1,\ldots,N}$ is a sequence of uncorrelated Gaussian heteroscedastic noise. The coefficients $\beta_{j,s}$ and the variance $\sigma^2_{X_j|X_1,\ldots,X_{j-1}}$ may be directly computed with the Durbin-Levinson algorithm (see Durbin 1960, see also, e.g., Brockwell and Davis 1987, pp. 169-170). The arithmetic complexity of this approach is $O(N^2)$, as in Sowell's method, while it avoids the computational drawback of Sowell's method. However, although the Durbin-Levinson algorithm enables a relatively fast evaluation of the likelihood function, the next sections will introduce some faster methods for finding the solution of (3.9). These methods are based on approximations of the exact Gaussian likelihood function.

Another sort of weakness of the Gaussian plug-in MLE was noted by Cheung and Diebold (1994). While it was shown by Sowell (1992) that when the mean of the process is known, the exact MLE of the differencing parameter d is substantially more efficient than the Whittle MLE (see Section 3.3.3) and the log-regression estimator (see Section 3.4.3) of d, Cheung and Diebold (1994) conducted a Monte Carlo simulation study and found out that when the mean of the process is unknown (as happens in most cases), the discrete Whittle estimator is much preferable to the exact MLE in terms of mean squared error (MSE). They showed that while the exact Gaussian MLE of d suffers from a weak negative bias, the sample mean plug-in MLE has a much higher negative bias than the exact MLE (see also Chapter 5). Lieberman (2005) derived asymptotic expansion for the exact and discrete Whittle likelihoods with either known mean or the plug-in versions with the sample mean replacing the unknown true mean. He showed that the plug-in Gaussian likelihood is contaminated by an additional second order negative bias term, which does not exist in the case of known mean. Lieberman (2005) proposed a bias correction for the plug-in Gaussian MLE as well as for the plug-in Whittle MLE, which seems to capture much of the difference between the cases of known mean and unknown mean. A different approach to handle this weakness of the plug-in MLE is to consider instead the differenced data series (Smith, Sowell and Zin 1997). Another suggested approach is to use a "modified" profile MLE, as was suggested by Cox and Reid (1987) and An and Bloomfield (1993). The idea of this estimator is to use a linear transformation of the parameters of interest and to make them orthogonal to nuisance parameters, and particularly in our case, to the mean μ. The asymptotic distribution of the modified profile MLE is unchanged compared to the exact MLE, and it eliminates some degree of the bias in the exact likelihood estimates (see An and Bloomfield 1993, Hauser 1999). Overall, however, empirical results show that both the exact MLE and the modified MLE are inferior to the Whittle MLE in most cases (Nielsen and Frederiksen 2005).

3.3.2 Autoregressive Approximation

The following method is based upon on the autoregressive representation of invertible time series. In the literature it is mostly been applied to ARFIMA(p,d,q) models, in which the inversion of the series may be translated into a simple Taylor expansion procedure (see Theorem 2.5.1). Thus, we will refer in this section particularly to ARFIMA(p,d,q) models and recall that the regular memory parameter α is just $2d$ in this case. An invertible ARFIMA(p,d,q) process X fulfills the relation

$$\epsilon_t = \sum\nolimits_{j=0}^{\infty} \pi_j X_{t-j}, \tag{3.13}$$

where $\epsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right)$, and the coefficients $\pi_j \equiv \pi_j\left(\theta\right)$ are given to us by the equation (2.32). Wlg we may fix $\pi_0 = 1$ by a proper normalization of the equation (3.13), and replace the sign of the rest of the coefficients, so the white noise series, ϵ_t, may be regarded as a series of error terms achieved by the $AR\left(\infty\right)$ representation of the process,

$$\epsilon_t = X_t - \sum\nolimits_{j=1}^{\infty} \pi_j X_{t-j}. \tag{3.14}$$

This suggests that instead of formulating the Likelihood function in terms of the observations X, as in (3.8), we may express it in terms of the white noise disturbances ϵ_t, which can be computed directly from the relation (3.14). The alternative log-likelihood function (normalized by $-N$) is then

$$\mathcal{L}_N\left(\mathbf{X};\theta\right) = \frac{N}{2}\log 2\pi + \frac{N}{2}\log \sigma_\varepsilon^2 + \frac{1}{2}\sum_{t=1}^{N}\left(\frac{\epsilon_t}{\sigma_\varepsilon}\right)^2. \tag{3.15}$$

In practice, however, only finite number of past observations, $X_t, X_{t-1}, ..., X_1$, are available to us, and therefore one needs to evaluate ϵ_t in a different manner, rather than by (3.14).

It is possible, however, to consider instead the truncated linear model

$$\widetilde{\epsilon}_t = X_t - \sum\nolimits_{j=1}^{t-1} \phi_j X_{t-j}, \tag{3.16}$$

where the coefficients in (3.16) are the coefficients of the best linear prediction of X_t given $X_{t-1}, X_{t-2}, ..., X_1$. In this approach, the resulted likelihood function is the exact likelihood (3.12), attained by multiplying the density functions of each of the disturbances conditional on the past observations. The coefficients, $\phi_1, \phi_2, ..., \phi_t$, as well as the variance of the prediction errors, $\widetilde{\epsilon}_t$, of the best linear prediction may be calculated, for example, with the Durbin-Levinson algorithm, which has arithmetic complexity of order $O(N^2)$.

Hasslett and Raftery (1989) dealt with the computational problem of the exact MLE procedure when they analyzed a very long time series in a spatial context. They modeled their data with an ARFIMA model, and they allowed the differencing parameter to have values in the range $0 \leq d \leq \frac{1}{2}$. Using some heuristic approximation arguments, they proposed an applicative approximate version of the likelihood (3.15), designed to accelerate the CPU time of the exact MLE procedure, and seems to perform very well. Their idea was to approximate the "tail" of the truncated series in the RHS of (3.16), $\sum\nolimits_{j=M}^{t-1} \phi_j X_{t-j}$, where M is some large integer smaller than $t-1$ (Hasslett and Raftery 1989 suggested the value M $= 100$, which gives good results over a wide range of values of d and N). Hasslett and Raftery's (1989) approximation is obtained by the formula

$$\sum\nolimits_{j=1}^{t-1} \phi_j X_{t-j} \approx \sum\nolimits_{j=0}^{M} \phi_j X_{t-j} - M\pi_M d^{-1} \left\{1 - (M/t)^d\right\} \overline{X}_{M+1,t+1-M},$$

where π_M is the M'th coefficient of the $AR(\infty)$ representation (3.14) and $\overline{X}_{M+1,t+1-M} = \frac{1}{t-1-2M} \sum\nolimits_{j=M+1}^{t+1-M} X_j$.The arithmetic complexity of Haslett and Raftery's method is of order $O(NM)$, which significantly improves the CPU time of the exact MLE procedure.

Beran (1994a) suggested to use a simpler Likelihood approximation, based on the truncated autoregressive approximation, for which he managed to show that similar first-order asymptotic results as those of the exact MLE holds, under the conditions that $\mu = 0$ is known and $0 < d < \frac{1}{2}$. Beran used the following approximation for the disturbances,

$$\epsilon_t = X_t - \sum\nolimits_{j=1}^{t-1} \pi_j X_{t-j}, \quad (t = 2, ..., N), \tag{3.17}$$

where π_j are the usual coefficient of the $AR(\infty)$ representation (3.14). Beran's approximation is therefore equivalent to assuming that all the observations before time $t \leq 0$ are equal to the mean 0. Beran's MLE is then obtained by minimizing

$$\frac{N}{2}\log 2\pi + \frac{N}{2}\log\sigma_\varepsilon^2 + \frac{1}{2}\sum_{t=2}^{N}\left(\frac{\epsilon_t}{\sigma_\varepsilon}\right)^2.$$

Beran (1994a) proposed a further refinement to his method. Let the first parameter θ_1 be σ_ε^2, and define

$$r_t(\theta) = \frac{\epsilon_t}{\sqrt{\theta_1}}.$$

Then Beran's MLE may be attained by minimizing

$$N\log\theta_1 + \sum_{t=2}^{N} r_t(\theta)^2$$

with respect to θ. Under mild regularity conditions, it is equivalent to the following system of p nonlinear equations,

$$\sum_{t=2}^{N} r_t(\theta)\frac{\partial}{\partial\theta_j}r_t(\theta) = 0, \qquad j = 2, ..., p$$

and

$$\sigma_\varepsilon^2 = \theta_1 = \frac{1}{N-1}\sum_{t=2}^{N} r_t^2(\theta).$$

Beran (1994a) proved consistency, asymptotic normality and efficiency of his approximated MLE, with the same properties as in (3.11).

A year later, Beran (1995) dealt with a more-general model, in which the m'th differenced process, $(1-B)^m X_{t-j} - \mu$, is an ARFIMA(p,δ,q), where μ is unknown and $\delta \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. While Beran's (1995) consistency proof seems to contain an invalid circular argument, as Velasco and Robinson (2000) point out, Beran's (1995) reported simulations support his conclusion that

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \underset{d}{\rightarrow} N\left(0, \Gamma(\theta^*)^{-1}\right),$$

where $\Gamma(\theta^*)$ is the Fisher information matrix evaluated at the parameter vector θ^* that equals to the true parameter θ_0, besides that $\delta \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ replaces the true differencing parameter $d = m + \delta$. Particularly note that $\theta^* = \theta_0$ if $m = 0$. Thus, for $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ the last result reduces to the same asymptotic property (3.11) that was proved for the exact MLE only for the cases of short memory, $d = 0$, or long memory with $d \in \left(0, \frac{1}{2}\right)$.

The arithmetic complexity of Beran's method is of order $O(N^2)$. Thus, from this respect, it is not preferable to the exact likelihood approach. Moreover, the conditional variance of the disturbances (3.17) depends on the number of given observations, t, while Beran's method represents the conditional disturbances as an homoscedastic process. This effect is asymptotically negligible, but for short time series it may yield relatively poor results (Beran 1995).

3.3.3 Whittle's Approximation

The Whittle MLE is named after Whittle (1951, 1953), who proposed a frequency-based approximation to the Gaussian Likelihood in the context of short memory time series. The method is based on two different approximations for the inverse and the determinant of the covariance matrix, $\Sigma_N(\theta)^{-1}$ and $|\Sigma_N(\theta)|$, respectively. These approximations are justified by the general theory (cf. Grenander and Szegö 1958) of symmetric Toeplitz matrices of the form

$$\Sigma_N(f) = \left[\int_{-\pi}^{\pi} e^{i(r-s)\omega} f(\omega)\, d\omega\right]_{r,s=1,\ldots,N}, \tag{3.18}$$

where $f(\omega)$ is a symmetric, nonnegative and integrable function defined on $\Pi = [-\pi, \pi]$. Thus, the r, s-entry of the Toeplitz matrix is assumed to be the $|r-s|$ Fourier coefficient of $f(\omega)$. Note that if $f(\omega)$ is the spectral density of the time series, $f_\theta(\omega)$, then $\Sigma_N(f_\theta)$ is the covariance matrix of the process (see Section 2.2.2). For clear exposition, we shall frequently denote the covariance matrix by $\Sigma_N(f_\theta)$, to indicate that it depends on θ through the spectral density $f_\theta(\omega)$.

There are few different versions of the Whittle's approximations, depending on the model and the norm of interest, see for example Hannan (1973, Lemma 4), Dahlhaus (1989, Lemma 5.2) and Theorem (4.3.2) in this thesis. The Whittle's approximations are given informally by

$$\Sigma_N(f_\theta)^{-1} \approx \frac{1}{4\pi^2}\Sigma_N\left(f_\theta^{-1}\right),$$

and

$$\frac{1}{N}\log\left|\Sigma_N\left(f_\theta\right)\right| \approx \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_\theta\left(\omega\right)d\omega,$$

where both of the approximations are justified asymptotically as $N\to\infty$.

Substituting these approximations into (3.10), we get that the Whittle estimate of θ_0, given by minimizing

$$\int_{-\pi}^{\pi}\log f_\theta\left(\omega\right)d\omega + \frac{1}{2\pi}\frac{\left(\mathbf{X}-\mu\mathbf{1}\right)'\Sigma_N\left(f_\theta^{-1}\right)\left(\mathbf{X}-\mu\mathbf{1}\right)}{N}. \tag{3.19}$$

This likelihood enables to avoid the heavy CPU time required for the computation of the inverse and the determinant of the covariance matrix. However though, the computation of (3.19) requires to compute the quadratic form $\left(\mathbf{X}-\mu\mathbf{1}\right)'\Sigma_N\left(f_\theta^{-1}\right)\left(\mathbf{X}-\mu\mathbf{1}\right)$ for every trial value of θ. In order to evaluate $\Sigma_N\left(f_\theta^{-1}\right)$ one has to calculate N integrals of the form

$$\left[\int_{-\pi}^{\pi}e^{ik\omega}f_\theta^{-1}d\omega\right]_{k=0,\ldots,N-1} \tag{3.20}$$

(see (3.18). Note that since f_θ^{-1} is a symmetric function, $\int_{-\pi}^{\pi}e^{ik\omega}f_\theta^{-1}d\omega = \int_{-\pi}^{\pi}e^{-ik\omega}f_\theta^{-1}d\omega$, we may consider here only $k\geq 0$). That can still be very costly in CPU time, in particular for large sample sizes and if θ has a high dimension.

A major improvement in the computation time may be achieved by representing the Whittle likelihood (3.19) by means of the periodogram of the process,

$$I\left(\omega\right) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N}\left(X_j-\overline{X}_N\mathbf{1}\right)e^{ij\omega}\right|^2 = \frac{1}{2\pi}\sum_{k=-(N-1)}^{N-1}\widehat{\gamma}\left(k\right)e^{-ik\omega}, \qquad \omega\in\Pi, \tag{3.21}$$

where $\widehat{\gamma}\left(k\right)_{k=1,2,\ldots N-1}$ are the sample autocovariances

$$\widehat{\gamma}\left(k\right) = \frac{1}{N}\sum_{t=1}^{N-|k|}\left(X_t-\overline{X}_N\mathbf{1}\right)\left(X_{t+|k|}-\overline{X}_N\mathbf{1}\right),$$

and $\overline{X}_N$ is the sample mean $\overline{X}_N = \frac{1}{N}\sum_{t=1}^{N}X_t$. A frequency-domain representation of the

Whittle's MLE is obtained by the following argument.

$$\begin{aligned}
& \frac{1}{2\pi}\frac{(\mathbf{X}-\mu\mathbf{1})'\,\Sigma_N\left(f_\theta^{-1}\right)(\mathbf{X}-\mu\mathbf{1})}{N} \\
= & \frac{1}{2\pi N}\sum_{i,j=1}^{N}(X_i-\mu\mathbf{1})(X_j-\mu\mathbf{1})\int_{-\pi}^{\pi}e^{i(i-j)\omega}f_\theta^{-1}(\omega)\,d\omega \\
= & \int_{-\pi}^{\pi}\left[\frac{1}{2\pi N}\sum_{i,j=1}^{N}(X_i-\mu\mathbf{1})(X_j-\mu\mathbf{1})\,e^{i(i-j)\omega}\right]f_\theta^{-1}(\omega)\,d\omega \\
= & \int_{-\pi}^{\pi}\left[\frac{1}{2\pi}\sum_{k=-(N-1)}^{N-1}\left(\frac{1}{N}\sum_{t=1}^{N-|k|}(X_t-\mu\mathbf{1})\left(X_{t+|k|}-\mu\mathbf{1}\right)\right)e^{-ik\omega}\right]f_\theta^{-1}d\omega \\
= & \int_{-\pi}^{\pi}\frac{\frac{1}{2\pi}\sum_{k=1}^{N}\widehat{\gamma}(k)\,e^{-ik\omega}}{f_\theta(\omega)}d\omega \\
= & \int_{-\pi}^{\pi}\frac{I(\omega)}{f_\theta(\omega)}d\omega.
\end{aligned}$$

Hence, the Whittle MLE may be obtained by computing the periodogram function $I(\omega)$ and minimizing

$$\int_{-\pi}^{\pi}\log f_\theta(\omega)\,d\omega+\int_{-\pi}^{\pi}\frac{I(\omega)}{f_\theta(\omega)}d\omega \tag{3.22}$$

with respect to θ. The evaluation of the Whittle Likelihood in this form may be done very fast by computing the periodogram via the Fast Fourier Transform with computational order of $O(N\log N)$ (Press et al., 1992, p. 498).

A further refinement of the method may be obtained by replacing the integrals in (3.22) by simple sums. In most applications, we usually assume that the spectral density of the process satisfies positivity a.e. and

$$f(\omega)=O\left(|\omega|^{-\alpha}\right)$$

for some $\alpha>-1$ (It basically corresponds to the invertibility condition). Thus $\frac{1}{f(\lambda;\theta)}$ is a Rieman-integrable function on Π and we may get the following discrete approximation of the Whittle Likelihood,

$$\frac{4\pi}{N}\sum_{j=0}^{[N/2]}\left[\log f\left(\omega_{j,N};\theta\right)+\frac{I\left(\omega_{j,N}\right)}{f\left(\omega_{j,N};\theta\right)}\right]. \tag{3.23}$$

where

$$\omega_{j,N}=\frac{2\pi j}{N}, \qquad j=1,...,N^{*},$$

and N^{*} is the integer part of $\frac{N-1}{2}$.

Hannan (1973) considered both the Whittle's likelihood (3.22) and the discrete version of it, (3.23), for linear time series with absolutely positive and twice continuously differentiable spectral densities. Hannan's (1973) result is thus applicable only for the short-memory case, for which he proved consistency and asymptotic normality of the Whittle's estimates. Fox and Taqqu (1986) dealt again with the Whittle MLE (3.22) for the long memory case, $\alpha\in(0,1)$, with μ known or zero, and they were able to show that the Whittle's estimator is being asymptotically normal with the same rate of convergence as in the short memory case. Dahlhaus (1989) extended Fox and Taqqu's (1986) result to the case where μ_0 is unknown. Dahlhaus (1989) also proved asymptotic efficiency of the Whittle estimate for long memory processes, and showed that the exact and the Whittle MLE obtain the same first order asymptotic properties as was shown in (3.11). Giraitis and Surgailis (1990) relaxed Gaussianity and established similar results with the Whittle MLE of general linear processes with long memory,

$$X_t=\sum_{k=-\infty}^{k=\infty}\psi_j(k)\epsilon_{t-j},$$

where ϵ_t are any zero mean and iid disturbances. Heyde and Gay (1993) and Hosoya (1997) considered multivariate non-Gaussian case. Velasco and Robinson (2000) considered the discrete Whittle likelihood (3.23) and proved asymptotic normality and efficiency of the Whittle MLE for any memory parameter in the range $\alpha>-1$, where α is permitted to exist in the nonstationary region $[1,\infty)$. However, in the case of nonstationarity the spectral density is not defined. In this case, it is common to define a generalized spectral density, and the periodogram is then viewed as an estimate of that generalized spectral density. In addition, in the case of nonstationarity, some smoothing operation on the periodogram, called tapering, is usually required to reduce the periodogram bias (see, e.g., Priestly 1981 p. 563).

Overall, the Whittle MLE is a very popular technique for estimation of long memory, and is considered to be a competitive rival for the exact MLE. It is a relatively fast and efficient method. As mentioned in Section 3.3.1, Cheung and Diebold (1994) and Nielsen and Frederiksen (2005) showed that the discrete Whittle MLE is preferable to the exact MLE (and also to the non-discrete Whittle MLE) in the case of unknown mean. Hauser (1999), however, showed by simulations that the discrete Whittle estimate has serious deficiencies for large parameter ranges. He recommends using the Whittle Likelihood with a tapered periodogram in all cases.

3.4 Semi-Parametric Estimation

3.4.1 R/S Statistics

Hurst (1951, 1956) discovered the characteristics of long memory while investigating the Nile river's flow stream as well as other records of hydrological time series (see Section 2.3). For a given time series, X, Hurst (1951, 1956) examined the behaviour of adjusted rescaled range (R/S) statistic, given at time point t and lag k by

$$R/S\,(t,k) = \frac{\max\limits_{1\le i\le k}\left[S_{t+i} - S_t - \frac{i}{k}\left(S_{t+k} - S_t\right)\right] - \min\limits_{1\le i\le k}\left[S_{t+i} - S_t - \frac{i}{k}\left(S_{t+k} - S_t\right)\right]}{\left\{k^{-1}\sum\limits_{i=t+1}^{t+k}\left(X_i - \overline{X}_{t,k}\right)^2\right\}^{1/2}}, \quad (3.24)$$

where $S_t = \sum\limits_{i=1}^{t} X_i$. He observed that for many of his records, $R/S\,(t,k)$ can be described to behave asymptotically like ck^H where c is some positive constant and $H > \frac{1}{2}$. This empirical finding was in contradiction to results for stationary process with short-memory usually considered at that time, from whom $R/S\,(t,k)$ should behave asymptotically like a constant times $k^{\frac{1}{2}}$. Hurst's finding that the $R/S\,(t,k)$ statistic behaves like k^H with $H > \frac{1}{2}$ is called the Hurst effect. Motivated by Hurst's empirical findings, Mandelbrot and co-workers later introduced FGN as a statistical model with long memory (see Section 2.4.2), where the memory parameter of the model is denoted by H for Hurst.

Mandelbrot (1975) showed that under mild regularity conditions $k^{-H}R/S\,(t,k)$ converges in distribution to a nondegenerate random variable in the limit where $k \to \infty$. This suggests that for large values of k, a plot of $\log R/S$ against $\log k$ should be randomly scattered around a straight line with slope H,

$$\log E\,[R/S] \approx a + H \log k.$$

For a given time series with length N, we may sample $N - k + 1$ different values of the $R/S\,(t,k)$ statistics that correspond to lag k in $1, .., N$, by taking all possible values of time points $t = 0, 1, 2, ..., N - k$. The Hurst coefficient H is then estimated by fitting a linear line to the plot and finding its slope for large values of k with a least squares regression or "by eye".

A desired property of the R/S-statistic proved by Mandelbrot (1975) is that this method is robust against long-tailed marginal distribution, in the sense that if X is iid with $E\left(X_t^2\right) < \infty$ and it is in the domain of attraction of stable distributions with index $0 < \alpha < 2$, the asymptotic slope in the R/S plot remains $\frac{1}{2}$. However, many other difficulties arise while using this method. The distribution of the R/S-statistic seems to be very complex, and it is neither normal, nor symmetric. The values of $R/S\,(t,k)$ for different time points t and lags k are not independent from each other, and as k tends to N we get less samplings of the R/S-statistic.

3.4.2 Correlogram-Based Estimation

The plot of the sample correlations $\widehat{\rho}\,(k)$ against the lag k (correlogram) is a standard tool for a preliminary analysis of a time series analysis. A common rule of thumb is to consider only correlations outside of the band of $\pm\frac{2}{\sqrt{N}}$ (see, e.g., Priestly 1981, p. 340). As we have seen in Chapter 2, one of the characteristics of long-memory processes is that the autocorrelations $\rho\,(k)$ decay at a hyperbolic rate as $k \to \infty$ (see (2.7)). This asymptotic rule suggests estimating the memory parameter α by evaluating the asymptotic rate of decay of the sample autocorrelations.

For example, a suitable plot can be obtained by taking the plot of $\log|\rho\,(k)|$ against $\log k$.

If the asymptotic decay of the correlations is hyperbolic, then for large lags, the points in the plot should be scattered around a straight line with negative slope approximately equal to $\alpha - 1$. In contrast, for short-memory processes, the log-log correlogram should show divergence to minus infinity at a rate that is at least exponential. A closed-form estimate based on this idea was proposed by Robinson (1994) and was studied by Hall, Koul and Turlach (1997). A different proposal of Robinson (1994) is to consider the estimator obtained by nonlinear regression that minimizes the sum of square $\sum_{k=M}^{N} \left(\widehat{\rho}\left(k\right) - c_2 k^{\alpha-1}\right)^2$, where M increases with N but significantly slower than N. Hall, Koul and Turlach (1997) also discuss this suggestion.

Essentially, similar difficulties, as with the R/S plot method, apply here. The distribution of the regression coefficients is complex, and it is not clear how to build confidence intervals for the estimators. Because of these reasons, the semiparametric time domain methods are usually considered only as heuristic methods. This is in contrast to semiparametric methods of estimation in the frequency domain, discussed in the next sections.

We finally mention briefly that Tieslau, Schmidt and Baillie (1995) proposed a parametric version of the minimum distance estimator, which is similar to Robinson's (1994) semiparametric nonlinear regression estimator, but where the autocorrelations are taken to be those of an ARFIMA(p,d,q) process. Tieslau, Schmidt, and Baillie (1995) and Chung and Schmidt (1995) managed to achieve with this method the $\sqrt{N}$ rate of convergence of the parameter estimates as in the other parametric methods.

3.4.3 Log-Periodogram Regression

The log-periodogram regression method is a frequency domain semiparametric estimator for the memory of the series, α. It does not require the specification of a parametric model for the data, but only relies on the shape of the spectral density near the origin (2.6). This behaviour of the spectral density can be formulated by a simple linear relation in α,

$$\log f\left(\omega\right) \simeq \log c_1 - \alpha \log\left|\omega\right| \quad \text{as} \ \ \left|\omega\right| \to 0, \tag{3.25}$$

suggesting that a linear regression on very low frequencies can be applied for estimation of α.

Geweke and Porter-Hudak (1983) suggested a semiparametric estimator (also referred to as the GPH estimator) achieved by regressing the periodogram on the low frequencies in a log-log scale. With analogy to (3.25), the Geweke and Porter-Hudak's (1983) proposed model is

$$\log I\left(\omega_{j,N}\right) \simeq \log c - \alpha \log \omega_{j,N} + \varepsilon_j. \tag{3.26}$$

Here $\omega_{j,N}$ are taken to be only the low frequencies

$$\omega_{j,N} = \frac{2\pi j}{N}, \qquad j = 1, ..., M,$$

where M increases with N but significantly slower than N, so that particularly $\omega_{M,N} \to 0$ as $N \to \infty$. The estimate $\widehat{\alpha}$ of the true memory parameter α is then obtained by calculating the regression slope. Relying on some heuristic arguments, Geweke and Porter-Hudak (1983) claimed that under this setting

$$\sqrt{M}\left(\widehat{\alpha} - \alpha\right) \underset{d}{\to} N\left(0, \frac{\pi^2}{6}\right). \tag{3.27}$$

In fact, in the case of a long memory process where the periodogram is calculated at increasing number of Fourier frequencies, and in particular if $\omega_{j,N} \to 0$ for fixed j, it was shown by Künsch (1986), Hurvich, and Beltrao (1993) and Robinson (1995a) that the ε_j in (3.26) may have a complex distribution, not taken into account by Geweke and Porter-Hudak (1983). Particularly, the ε_j are not homoscedastic and they may be seriously biased and correlated, even asymptotically. Nevertheless, Robinson (1995a) proved (3.27) to hold for Gaussian, zero mean, stationary processes with memory parameter in the range $-1 < \alpha < 1$. Robinson (1995a) also showed that by pooling the periodogram values at J adjacent frequencies

$$\omega_{j,N}, \omega_{j-1,N}, ..., \omega_{j-J,N}, \qquad j = J+1, ..., M,$$

(J does not depend on N, so particularly it is assumed that $J << N$), the asymptotic variance in (3.27) can be reduced. Particularly, it converges to 1 from above as $J \to \infty$.

Hurvich and Ray (1995) extended Robinson's (1995a) result for all the range $\alpha < 3$. Thus, they provided a uniform theory that covers the whole range of noninvertibility, and also the possibility of some degree of nonstationarity, in which they used a tapered periodogram. Velasco (1999a) further developed the use of tapering for this method and established a general tapered version of the method for every $\alpha > 1$. Velasco (2000) also relaxed Gaussianity and provided a version of (3.27) for general Gaussian processes. Hurvich, Deo and Brodsky (1998) dealt with the issue of choosing a bandwidth, M, and proved that $M \sim N^{4/5}$ implies an optimal estimator of α in the Mean Square Error sense. However, they also showed that the multiplying constant of the optimal M depends on the unknown parameters. Hurvich and Deo (1999) proposed a data-dependent consistent estimate of this constant.

Log-periodogram regression has become a popular method for estimation or for first diagnostic of the memory of a time series. It is simple to apply and in contrast to the parametric methods discussed above, almost no assumptions are made on the underlying process, besides on the shape of the spectral density near the origin. On the other hand, the rate of convergence of the memory parameter is slower than the $\sqrt{N}$ rate achieved by the parametric methods. Moreover, small sample studies of the the log-periodogram regression have indicated a serious bias of the estimator $\widehat{\alpha}$, which is a consequence of the non-ideal distribution of the ε_j, addressed above (see Hurvich and Beltrao 1993).

3.4.4 Local Whittle

The local Whittle estimate was proposed by Künsch (1987) and was later developed by Robinson (1995b). They consider a narrow band version of the discrete Whittle MLE (3.23) that, in a similar way to the log-periodogram Regression, relies only on the shape of the spectral density near the origin (2.6).

The local Whittle estimation is obtained by minimizing the objective function, given by

$$\frac{1}{M}\sum_{j=1}^{M}\left\{\log\left(c_1\omega_{j,N}^{-\alpha}\right)+\frac{I\left(\omega_{j,N}\right)}{c_1\omega_{j,N}^{-\alpha}}\right\} \tag{3.28}$$

with respect to parameters $c_1 > 0$ and $\alpha \in \Theta$. Here, $\omega_{j,N}$, M plays the same role as in the log-periodogram regression, that is to say, M increases with N but significantly slower than N such that $\omega_{M,N} \to 0$ as $N \to \infty$.

Note also that we may plug-in the (local) profile estimator of the parameter

$$\widehat{c}_1 = \frac{1}{M} \sum_{j=1}^{M} \frac{I(\omega_{j,N})}{\omega_{j,N}^{-\alpha}}$$

into (3.28) instead of the unknown true c_1 to get an objective function that depends on the memory parameter only,

$$Q(H) = \log \left\{ \frac{1}{M} \sum_{j=1}^{M} \frac{I(\omega_{j,N})}{\omega_{j,N}^{-\alpha}} \right\} - \alpha \frac{1}{M} \sum_{j=1}^{M} \log \omega_{j,N}. \tag{3.29}$$

As in the parametric methods discussed above, the local Whittle estimation requires a numerical procedure for finding the maximum of (3.28) or (3.29). However, the parameter dimension is low and the estimator is usually easy to locate.

Robinson (1995b) established for the local Whittle estimator $\widehat{\alpha}$ that under $|\alpha| < 1$ and zero mean,

$$\sqrt{M} (\widehat{\alpha} - \alpha) \underset{d}{\rightarrow} N(0, 1). \tag{3.30}$$

Thus, the local Whittle estimation performs asymptotically better then the log-periodogram method (see (3.27)). In particular, the limit result (3.30) corresponds to the pooled log-periodogram estimator of Robinson (1995a) when the degree of pooling, J, tends to infinity.

Lobato (1999) proposed a multivariate extension for the local Whittle estimation. Velasco (1999b) extended Robinson's (1995b) results for the whole range of $\alpha > -1$. As with all the previous estimators in the frequency domain, a tapering procedure is required for the nonstationary region of the memory parameter. Particularly for the local Whittle estimator, tapering is required to ensure consistency of the estimator for $\alpha > 2$ and asymptotic normality for $\alpha \geq 1\frac{1}{2}$.

A main problem with tapering is that tapering may strongly inflate the variance of the estimator α (see Velasco 1999(a,b)). Hurvich and Chen (2000) proposed a taper that improves the efficiency of the local whittle estimator for differenced data with memory parameter $\alpha < 3$, but for $\alpha \geq 1\frac{1}{2}$ it still does not reach the asymptotic variance achieved in the range $\alpha < 1\frac{1}{2}$. Moreover, Phillips (1999), Phillips and Shimotsu (2003), Kim and Phillips (2006) showed that when $\alpha > 2$ the local whittle, as well as the untaperred log-periodogram regression estimators, are inconsistent, and when $\alpha \geq 1\frac{1}{2}$ they exhibit a nonstandard limit distribution. Shimotsu and Phillips (2004) proposed a new semiparametric estimation, called the exact local Whittle, which is similar to the local Whittle estimation in that it is based on a narrow band of the discrete Whittle MLE, but when the periodogram $I(\omega_{j,N})$ is replaced by a more efficient data-dependent approximation of the spectral density near the origin, basically relying on the relation (Phillips 1999, Theorem 2.2)

$$I(\omega)\,\omega_{j,N}^{\alpha} \approx I_{\nabla^{\alpha/2}X}(\omega)\,, \text{ as } \omega \to 0, \tag{3.31}$$

where $I_{\nabla^{\alpha/2}X}(\omega_{j,N})$ is the periodogram of the (short memory) fractionally differenced series $\nabla^{\alpha/2}X$. The exact local Whittle estimator is then obtained by minimizing

$$Q(G;H) = \frac{1}{M}\sum_{j=1}^{M}\left\{\log\left(c_1\omega_{j,N}^{-\alpha}\right) + \frac{I_{\nabla^{\alpha/2}X}(\omega)}{c_1}\right\} \tag{3.32}$$

with respect to parameters $c_1 > 0$ and $\alpha \in \Theta$. Shimotsu and Phillips (2004) proved that the exact local Whittle estimator is consistent and achieves the asymptotic result (3.30) for each value of $\alpha \in \mathbb{R}$ if the true mean of the series is known. Shimotsu (2006) showed that the same result holds in the feasible case of unknown mean and $\alpha \in (-1, 4)$.

Chapter 4

MAIN RESULTS

4.1 Introduction

Throughout this chapter we consider a stationary Gaussian time series X with mean μ_0 and spectral density $f_{\theta_0}\ (\omega)\,,\ \omega \in \Pi \equiv [-\pi, \pi]$, where μ_0 and θ_0 are unknown parameters of the process which have to be estimated. We are interested in sequences with spectral densities $f_{\theta_0}(\omega)$ that belongs to the parametric family $\{f_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$ such that for all $\theta \in \Theta$

$$f_\theta(\omega) \sim |\omega|^{-\alpha(\theta)} L_\theta(\omega) \qquad \text{as } \omega \to 0,$$

where $-1 < \alpha(\theta) < 1$ and $L_\theta(\omega)$ positive and varies slowly in the sense that

$$L_\theta(\omega) = O\left(|\omega|^{-\delta}\right) \qquad \text{for each } \delta > 0.$$

Note that according to the definition described in Section 2.3, such a series is said to have long memory if $0 < \alpha(\theta_0) < 1$, short memory if $\alpha(\theta_0) = 0$ and intermediate memory if $-1 < \alpha(\theta_0) < 0$. Under this setting, the Gaussian maximum likelihood estimates (MLE) might be expected to have optimal asymptotic statistical properties.

As seen in Section 3.3.1, Hannan (1973) proved consistency and asymptotic normality of the Gaussian MLE for the case of possibly dependent observations, but with short memory. The corresponding proof for the case of long memory, $\alpha(\theta) \in (0, 1)$, where μ_0 is possibly unknown, is due to Dahlhaus (1989, 2005). He considered the estimator obtained by minimizing the plug-in log-likelihood function (normalized by $-N$),

$$\frac{1}{2N} \log\det \Sigma_N(f_\theta) + \frac{1}{2N}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right)' \Sigma_N(f_\theta)^{-1}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right),$$

with respect to Θ, where $\Sigma_N(f_\theta) = \left[\int_{-\pi}^{\pi} e^{i(r-s)x} f_\theta(x)\, dx\right]_{r,s=1,\ldots,N}$ is the covariance matrix of the process, $\mathbf{1} = (1, \ldots, 1)'$ and $\hat{\mu}_N$ is a consistent estimator of μ_0 (e.g., the arithmetic mean. see Section 3.2.1) that is plugged into the likelihood functions instead of the unknown mean. Dahlhaus (1989, 2005) proved

$$\sqrt{N}\left(\hat{\theta}_N-\theta_0\right)\underset{d}{\rightarrow}N\left(0,\Gamma\left(\theta_0\right)^{-1}\right),$$

where $\hat{\theta}_N$ denotes the maximum likelihood estimate and $\Gamma(\theta)$ is the Fisher information matrix given by

$$\Gamma(\theta)=-E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\mathcal{L}_N\left(\theta\right)\right)_{j,k=1,\ldots,p}.$$

Thus, Dahlhaus proved the efficiency of plug-in MLE.

The main object of this chapter is to extend Dahlhaus's (1989, 2005) results to include possibly short-memory or anti-persistent time series, without a priori knowledge of the memory of the time series. We adapt the proof technique of Dahlhaus essentially based on the asymptotic behaviour of Toeplitz matrices. Many of Dahlhaus's arguments are extended and simplified by a development of a uniform limit for the plug-in log-likelihood function that is valid on any compact parameter subspace Θ_1 of Θ in which $\alpha\left(\theta\right)>\alpha\left(\theta_0\right)-1$. Not surprisingly, this limit is identical to the limit of the Whittle log-likelihood in the same parameter subspace (cf. Fox and Taqqu 1986 and Velasco and Robinson 2000). In order to derive this limit result, we use a uniform version of Dahlhaus's Theorem 5.1 due to Lieberman, Rousseau and Zucker (2003) and Lieberman and Phillips (2004). We also establish a uniform version of Grenander and Szegö's (1958) Theorem on the limit of determinants of Toeplitz matrices.

The rest of Chapter 4 is organized as follows. Section 4.2 states the assumptions for the rest of the chapter. These assumptions are proved to hold for both FGN with memory parameter $0<H<1$ and for ARFIMA time series with memory parameter $-\frac{1}{2}<d<\frac{1}{2}$. Section 4.3 introduces some limit theorems for the traces and log-determinants of Toeplitz and inverse-Toeplitz matrices. These results assist us in Section 4.4 to derive the limit distribution of quadratic forms that involve inverse Toeplitz matrices. Section 4.5 contains a proof of the consistency of the exact MLE. Section 4.6 provides a proof of asymptotic normality and efficiency of the estimator.

4.2 Model Assumptions

The results are proved under the following assumptions:

(A.0) (a) $X = \{X_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian sequence with mean μ_0 and spectral density $f_{\theta_0}(\omega)$, $\omega \in \Pi \equiv [-\pi,\pi]$, where μ_0 and $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ are unknown parameters. If θ and θ' are distinct elements of Θ, we assume that the set $\{\omega | f_\theta(\omega) \neq f_{\theta'}(\omega)\}$ has a positive Lebesgue measure.

(b) We suppose that θ_0 lies in the interior of Θ, and that Θ is compact.

In addition, we require the following assumptions on $f_\theta(\omega)$. There exists $\alpha : \Theta \to (-1,1)$ such that for each $\delta > 0$

(A.1) $f_\theta(\omega)$, $f_\theta^{-1}(\omega)$, $\partial/\partial\omega f_\theta(\omega)$ are continuous at all (ω,θ), $\omega \neq 0$, and

$$\begin{aligned} f_\theta(\omega) &= O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \\ f_\theta^{-1}(\omega) &= O\left(|\omega|^{\alpha(\theta)-\delta}\right), \\ \frac{\partial}{\partial\omega} f_\theta(\omega) &= O\left(|\omega|^{-\alpha(\theta)-1-\delta}\right). \end{aligned}$$

(A.2) $\partial/\partial\theta_j f_\theta(\omega)$ and $\partial^2/\partial\theta_j\theta_k f_\theta(\omega)$ are continuous at all (ω,θ), $\omega \neq 0$, and

$$\begin{aligned} \frac{\partial}{\partial\theta_j} f_\theta(\omega) &= O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \quad 1 \le j \le p, \\ \frac{\partial^2}{\partial\theta_j\partial\theta_k} f_\theta(\omega) &= O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \quad 1 \le j,k \le p. \end{aligned}$$

(A.3) The function $\alpha(\theta)$ is continuous, and the constants appearing in the $O(\cdot)$ above can be chosen independently of θ (not of δ).

If assumptions (A.1)-(A.3) hold for $f_\theta(\omega)$ on some parameter subspace, say $\Theta^* \subseteq \Theta$, and not necessarily on Θ, we say that $f_\theta(\omega)$ satisfies assumption (A.1)-(A.3) on Θ^* with exponent $\alpha^* : \Theta^* \to (-1,1)$.

Assumptions (A.0)-(A.3) are modifications of Dahlhaus's (1989) assumptions (A0), (A2), (A3) and (A7)-(A9) to our case. The most important aspect of the assumptions is that $\alpha(\theta)$ may have values in the interval $(-1,1)$. It is, as aforesaid, extending Dahlhaus (1989) who limited $\alpha(\theta)$ to the interval $(0,1)$. As a result, f_θ^{-1} and its derivatives in θ are not assumed to be continuous at all (ω,θ), as in Dahlhaus. It should be also noted that besides the assumption $f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right)$, the rest of Dahlhaus's assumptions are presented in terms of f_θ^{-1} and its derivatives instead of f_θ. However, because $f_\theta(\omega)$ and $f_\theta^{-1}(\omega)$ have lower bounds (see (4.1) and the related discussion below), it is easily seen that the two presentations are equivalent. For example,

$$\left|\frac{\partial}{\partial\theta_j} f_\theta^{-1}(\omega)\right| = \left|\frac{\partial}{\partial\theta_j} f_\theta(\omega)\right| \Big/ f_\theta^2(\omega).$$

Assumption (A.0) is a background assumption of the model, and it corresponds to assumption (A0) of Dahlhaus (1989). We have divided this assumption into two parts. While part (a) indicates the general features of the parametric representation of the process, part (b) is in fact a technical condition on the form of the boundaries of the parameter space. This latter condition is required to ensure uniform consistency of the MLE. It is always possible, however, in our situation to "extend" the boundaries of the parameter space and to make (A.0)(b) hold, without changing the validity of the rest of the assumptions. Assumption (A.1) and (A.2) are needed to derive the limit of the plug-in log-likelihood function and its derivatives in θ. Assumption (A.1) corresponds to assumptions (A2) and (A7) (for $k = 0, 1$) of Dahlhaus (1989), and assumption (A.2) corresponds to assumption (A3) (without the third order differentiability) of Dahlhaus (1989). Dahlhaus also assumed that $\frac{\partial^2}{\partial\omega^2} f_\theta(\omega)$ and $\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f_\theta(\omega)$ are continuous at all (ω, θ), $\omega \neq 0$, and $\frac{\partial^2}{\partial\omega^2} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-2-\delta}\right)$, $\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right)$ for every $\delta > 0$. He used the second order differentiability in ω in order to bound $f_\theta(\omega)$ uniformly above zero, and the third order differentiability in θ to establish asymptotic equicontinuity of quadratic forms that involve inverse Toeplitz matrices. However, it turned up that both these conditions are not necessary for the establishment of the results. Assumption (A.3) corresponds to Dahlhaus's (1989, 2005) assumptions (A8) and (A9). This assumption is required for the uniform approximation of the log-likelihood function, as in Lieberman, Rousseau and Zucker (2003).

Note that the upper bounds for $f_\theta(\omega)$ and $f_\theta^{-1}(\omega)$ that appear in assumption (A.1) imply that for each $\delta > 0$ there are positive constants C_1, C_2 such that

$$C_1 |\omega|^{-\alpha(\theta)+\delta} \leq f_\theta(\omega) \leq C_2 |\omega|^{-\alpha(\theta)-\delta}. \tag{4.1}$$

Condition (4.1) appears explicitly, e.g., in Fox and Taqqu (1986, Section 4) for the case of long memory. Dahlhaus (1989) also relies on this implication in his proof of Lemma 5.3 (see also Lemma 4.3.1 of this thesis). We will make a direct use of the lower bound of f_θ in the proof of the consistency of the plug-in MLE.

In addition to the foregoing assumptions, an additional assumption is required on the plug-in estimate of μ_0. This assumption will be formally presented in Section 5.4 as assumption (A.4).

We show now that assumptions (A.0)(a) and (A.1)-(A.3) hold if $\{X_t - \mu\}_{t\in\mathbb{Z}}$ is FGN or Gaussian ARFIMA process (Condition (A.0)(b) is assumed to be satisfied anyway by a proper choice of Θ's boundaries). Recall from Chapter 2 that the spectral density $f_H(\omega)$ of an FGN is given by

$$f_{\sigma^2,H}(\omega) = F\left(\sigma^2, H\right)(1-\cos\omega)\sum_{k=-\infty}^{\infty}|\omega+2\pi k|^{-1-2H}, \quad \omega\in\Pi,$$

with

$$F\left(\sigma^2, H\right) = \sigma^2\left\{\int_{-\pi}^{\pi}(1-\cos\omega)\sum_{k=-\infty}^{\infty}|\omega+2\pi k|^{-1-2H}d\omega\right\}^{-1} \tag{4.2}$$

and $\sigma^2 = Var(\mathbf{X}_i)$. ARFIMA(p,d,q), on the other hand, has a more-easy to handle spectral density $f_{d,\xi,\phi}(\omega)$ given by

$$f_{\sigma^2,d,\xi,\phi}(\omega) = \frac{\sigma^2}{2\pi}\frac{\left|\xi\left(e^{-i\omega}\right)\right|^2}{\left|\phi\left(e^{-i\omega}\right)\right|^2}\left|1-e^{i\omega}\right|^{-2d}, \quad \omega\in\Pi,$$

with $\phi(z) = 1+\sum_{j=1}^{p}\phi_j z^j$ and $\xi(z) = 1+\sum_{j=1}^{q}\xi_j z^j$. The polynomials $\phi(z)$ and $\xi(z)$ are assumed to have have no common zeros and to have all their zeros outside the unit circle. This implies that both $\frac{\xi(z)}{\phi(z)}$ and $\frac{\phi(z)}{\xi(z)}$ have a power series expansion for all $z\in\mathbb{C}$, $|z|\le 1$ (see Box and Jenkins 1970).

Theorem 4.2.1 *Suppose that the parameter space* Θ *is compact. Then conditions (A.0)(a) and (A.1)-(A.3) are fulfilled if* $\{X_t-\mu\}_{t\in\mathbb{Z}}$ *is an FGN where* $\theta = \left(\sigma^2, H\right)$ *and* $0 < H < 1$ *or a Gaussian ARFIMA(p,d,q) where* $\theta = \left(\sigma^2, d, \xi_1, ..., \xi_p, \phi_1, ..., \phi_p\right)$ *and* $|d| < \frac{1}{2}$.

Proof. Note that (A0)(a) is satisfied for both FGN and ARFIMA(p,d,q) by the Gaussianity of the processes and by the fact that $f_{\sigma^2,H}(\omega)$ and $f_{\sigma^2,d,\xi,\phi}(\omega)$ are determined uniquely by

$\left(\sigma^2, H\right)$ and $\left(\sigma^2, d, \xi_1, ..., \xi_p, \phi_1, ..., \phi_p\right)$, respectively (see Sections 2.4-2.5). We start by proving that (A.1)-(A.3) hold for an FGN with $\alpha(\theta) = 2H - 1$. We write $f_{\sigma^2,H}(\omega)$ as

$$f_{\sigma^2,H}(\omega) = \sigma^2 \left\{ \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} f_k(\omega, H)\, d\omega \right\}^{-1} \sum_{k=-\infty}^{\infty} f_k(\omega, H).$$

where

$$f_k(\omega, H) = (1 - \cos\omega)\, |\omega + 2\pi k|^{-1-2H}.$$

$f_0(\omega, H) = (1 - \cos\omega)\, |\omega|^{-1-2H}$ is twice continuously differentiable and positive at all (ω, H), $\omega \neq 0$, and

$$\begin{aligned}
\frac{\partial}{\partial\omega} f_0(\omega, H) &= \sin\omega\, |\omega|^{-1-2H} + (1 - \cos\omega)\, |\omega|^{-2-2H}, \\
\frac{\partial}{\partial H} f_0(\omega, H) &= -2(1 - \cos\omega)\, |\omega|^{-1-2H}\, log(|\omega|), \quad 1 \leq j \leq p, \\
\frac{\partial^2}{\partial H^2} f_0(\omega, H) &= 4(1 - \cos\omega)\, |\omega|^{-1-2H}\, log^2(|\omega|), \quad 1 \leq j, k \leq p.
\end{aligned}$$

Since $1 - \cos\omega \sim |\omega|^2/2$ and $\sin\omega \sim |\omega|$ as $\omega \to 0$, we get that

$$\begin{aligned}
&f_0(\omega, H)/\,|\omega|^{1-2H}, \quad f_0^{-1}(\omega)/\,|\omega|^{-1+2H}, \quad \frac{\partial}{\partial\omega} f_0(\omega, H) \Big/ |\omega|^{-2H}, \\
&\frac{\partial}{\partial H} f_0(\omega, H) \Big/ \left[|\omega|^{1-2H}\, log(|\omega|)\right] \text{ and } \frac{\partial^2}{\partial H^2} f_0(\omega, H) \Big/ \left[|\omega|^{-1-2H}\, log^2(|\omega|)\right]
\end{aligned}$$

are continuous at all (ω, H), $\omega \neq 0$, and note the discontinuity in $\omega \neq 0$ is removable. These imply that $f_0(\omega, H)$ satisfies assumption (A.1)-(A.3), where the uniform bounding assumption in (A.3) is fulfilled because of the compactness of Θ.

Consider now $\sum_{k\neq 0} f_k(\omega, H)$. For each $k \neq 0$, $f_k(\omega, H)$ is nonnegative, twice continuously differentiable and bounded from above by $g_k = 4\pi\, |k|^{-1-2H_m}$ at all (ω, H), where

$$H_m = \min_{\theta\in\Theta} H.$$

Since $H_m > 0$, we have $\sum_{k\neq 0} g_k < \theta$, and by Weierstrass's M-test the series $\sum_{k\neq 0} f_k(\omega, H)$ converges uniformly. As a result, $\sum_{k\neq 0} f_k(\omega, H)$ is continuous, and hence also bounded, at all (ω, H). Similar arguments show that the derivatives of $\sum_{k\neq 0} f_k(\omega, H)$ are given as the infinite sum over $k \neq 0$ of the corresponding derivatives of the summands $f_k(\omega, H)$, and that they are also continuous and bounded at all (ω, H). As a results from what we have proved so far, it follows that $\sum_{k=-\infty}^{\infty} f_k(\omega, H)$ satisfies assumption (A.1)-(A.3).

We also get now that the normalizing factor $\int_{-2\pi}^{2\pi} \sum_{k=-\infty}^{\infty} f_k(\omega, H)\, d\omega$ can be bounded from below and above by some positive constants that are independent of H. Hence, the assertion will follow if $\int_{-2\pi}^{2\pi} \sum_{k=-\infty}^{\infty} f_k(\omega, H)\, d\omega$ is twice continuously differentiable at all (ω, H). This can be shown with the dominated convergence theorem. For example, the continuity of $\int_{-2\pi}^{2\pi} \sum_{k=-\infty}^{\infty} f_k(\omega, H)\, d\omega$ follows from the fact that $\sum_{k=-\infty}^{\infty} f_k(\omega, H) \leq K|\omega|^{-1-2H}$, where K is independent of H, and therefore

$$\lim_{\omega^* \to \omega} \int_{-2\pi}^{2\pi} \sum_{k=-\infty}^{\infty} f_k(\omega^*, H)\, d\omega = \int_{-2\pi}^{2\pi} \lim_{\omega^* \to \omega} \sum_{k=-\infty}^{\infty} f_k(\omega^*, H)\, d\omega.$$

For Gaussian ARFIMA(p,d,q) process it is much simpler to verify conditions (A.1)-(A.3) with $\alpha(\theta) = 2d$. The polynomials $\xi(z)$ and $\phi(z)$ are assumed to have all their roots outside the unit circle, and therefore $\frac{|\xi(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}$ is a positive and twice continuously differentiable at all $(\omega, \xi_1, ..., \xi_p, \phi_1, ..., \phi_p)$. Since by Taylor expansion $|1 - e^{i\omega}| \sim |\omega|$ as $\omega \to 0$, assumptions (A.1) and (A.2) are satisfied by $f_{\sigma^2,d,\xi,\phi}$. The uniform bounding assumption in (A.3) is also fulfilled, as above, by the fact that Θ is compact. ■

4.3 Limit Theorems for Toeplitz Matrices

Toeplitz matrices arise quite naturally in the study of stationary processes, since the $N \times N$ covariance matrices of such processes are Toeplitz. To be precise, in the general case of stationary processes we are concerned with $N \times N$ matrices of the form

$$\Sigma_N(f) = \left[\int_{-\pi}^{\pi} e^{i(r-s)\omega} f(\omega)\, d\omega\right]_{r,s=1,\ldots,N} \tag{4.3}$$

where $f(\omega)$ is the spectral density of the process, which is nonnegative, integrable, real and symmetric on $\Pi \equiv [-\pi, \pi]$ (see Section 2.2). The most common and complete reference describing the Theory of Toeplitz matrices is Grenander and Szegö (1958). However, some essential theory of Grenander and Szegö is applicable for Toeplitz matrices corresponding to spectral density functions which are positive and uniformly bounded from beneath and from above. For long memory or anti-persistency cases a more general setting is required.

Some more recent theory of Fox and Taqqu (1987), Avram (1988), Dahlhaus (1989, 2005), Lieberman, Rousseau and Zucker (2003) as well as Lieberman and Phillips (2004) sheds light on some crucial asymptotic properties of Toeplitz matrices of the form (4.3) that may be applied in exploring the likelihood function of long or anti-persistent Gaussian process. The goal of this section is to present some of these asymptotic results in a uniform version required for our later arguments.

Suppose A is an $N \times N$ matrix and denote by $\|A\|$ the spectral norm of A, that is,

$$\|A\| = \sup_{x \in \mathbb{C}^n} \left(\frac{x^* A^* A x}{x^* x} \right)^{1/2},$$

where A^* is the conjugate transpose of A. The following well known relations between matrix norms are used in the rest of the chapter without further reference (see, for instance, Golub and Van Loan 1996, Section 2.2).

(i) $\qquad \|AB\| \leq \|A\| \cdot \|B\|,$

(ii) $\qquad \|A + B\| \leq \|A\| + \|B\|.$

If A is Hermitian and positive-definite, then

(iii) $\qquad x^* A x \leq x^* x \|A\| \quad \text{for } x \in \mathbb{C}^N.$

The following Lemma is a direct extension of Lemma 5.3 of Dahlhaus (1989), established in a similar way to the proof of Dahlhaus (1989, Lemma 5.3). Nevertheless, we provide here an expository proof.

Lemma 4.3.1 *Let Θ^* be a compact subset of Θ. Suppose that $f_\theta(\omega)$ and $g_\theta(\omega)$ are symmetric nonnegative functions, $\omega \in \Pi$, $\theta \in \Theta^*$, such that there exist continuous functions $\alpha(\theta), \beta(\theta) < 1$ and positive constants C_1, C_2 that are not dependent of θ, with*

$$\begin{aligned} f_\theta^{-1}(\omega) &\leq C_1 |\omega|^{\alpha(\theta)}, \\ g_\theta(\omega) &\leq C_2 |\omega|^{-\beta(\theta)}. \end{aligned}$$

Then we obtain with a positive constant K independent of θ and N

$$\left\| \Sigma_N(f_\theta)^{-1/2} \Sigma_N(g_\theta)^{1/2} \right\| = \left\| \Sigma_N(g_\theta)^{1/2} \Sigma_N(f_\theta)^{-1/2} \right\| \leq K N^{\max\{(\beta(\theta)-\alpha(\theta))/2,\ 0\}}.$$

Proof. Referring to Lemma 5.3 of Dahlhaus (1989), Dahlhaus stated his Lemma for strictly positive $f_\theta(\omega)$ and $g_\theta(\omega)$ with $0 < \alpha(\theta), \beta(\theta) < 1$ and where C_1, C_2, and hence also K, may depend on θ and N. In a similar way to Dahlhaus (1989, Lemma 5.3) we achieve the bound

$$\begin{aligned} & \left\| \Sigma_N (f_\theta)^{-1/2} \Sigma_N (g_\theta)^{1/2} \right\|^2 \\ = & \sup_{|x|=1} \frac{\int_{-\pi}^{\pi} g_\theta(x) \left| \sum_{n=1}^{N} x_n \exp(-i\gamma n) \right|^2 d\gamma}{\int_{-\pi}^{\pi} f_\theta(x) \left| \sum_{n=1}^{N} x_n \exp(-i\gamma n) \right|^2 d\gamma} \\ \le & K_1 \sup_{h \in P_N} \frac{\int_{-\pi}^{\pi} |\gamma|^{-\beta(\theta)} h(\gamma) d\gamma}{\int_{-\pi}^{\pi} |\gamma|^{-\alpha(\theta)} h(\gamma) d\gamma}, \end{aligned}$$

where $P_N = \left\{ h(\gamma) : \ h(\gamma) \le N \text{ and } \int_{-\pi}^{\pi} h(\gamma) \ = 1 \right\}$. Here, K_1 is a positive constant that can be chosen independently of θ and N because C_1 and C_2 are independent of θ. For each $\theta \in \Theta^*$ such that $\beta(\theta) \le \alpha(\theta)$, the above expression is bounded. For each $\theta \in \Theta^*$ such that $\beta(\theta) > \alpha(\theta)$, the sup is attained by $h(\gamma) = N\chi_{\{|\gamma|<1/2N\}}$ and

$$\left\| \Sigma_N (f_\theta)^{-1/2} \Sigma_N (g_\theta)^{1/2} \right\|^2 \le K_2 N^{\beta(\theta)-\alpha(\theta)},$$

where K_2 can be chosen independently of of θ since $\alpha(\theta), \beta(\theta) \le 1-\epsilon$ for some $\epsilon(\Theta^*) > 0$.
■

The next Theorem is a generalized version of Theorem 5.1 of Dahlhaus (1989) and it deals with asymptotic approximation of traces of products of Toeplitz and inverse-Toeplitz matrices. Part (a) of the Theorem is a direct adaptation of Theorem 7 of Lieberman and Phillips (2004), which extended Dahlhaus's Theorem 5.1 to a wider range of Toeplitz matrices and also established error orders for the limits approximation. The uniformity in θ is obtained by similar arguments to those of Lieberman, Rousseau and Zucker (2003), which developed a uniform version of Dahlhaus's Theorem 5.1. Part (b) of the Theorem obtains similar result to those of Theorem 1(b) of Fox and Taqqu (1987) and Theorem 1(b) of Avram (1988), but when also inverse Toeplitz matrices are allowed in the products within the trace term. The proof of part (b) follows directly with the same lines of arguments as in Theorem 5.1 of Dahlhaus (1989) and Theorem 7 of Lieberman and Phillips (2004), but when part (b) of Theorem 1 of Fox and Taqqu (1987) is used instead of part (a).

Theorem 4.3.1 *Let Θ^* be a compact subset of Θ. Let $p \in \mathbb{N} \cup \{0\}$ and let $f_{\theta,j}(\omega)$ and $g_{\theta,j}(\omega)$, $j = 1, ..., p+1$, be symmetric, real-valued functions on Π. Suppose that for each $j = 1, ..., p+1$, $f_{\theta,j}(\omega)$ satisfies assumptions (A.1)-(A.3) on Θ^* with exponent $\alpha(\theta)$ that does not depend on j. Suppose also that for each $\delta > 0$*

$$|g_{\theta,j}(\omega)| \leq K(\delta)|\omega|^{-\beta(\theta)-\delta}, \qquad j = 1, ..., p,$$

and

$$|g_{\theta,p+1}(\omega)| \leq K(\delta)|\omega|^{-\beta^*(\theta)-\delta}.$$

where $\beta(\theta), \beta^*(\theta) < 1$, continuous on Θ^* and do not depend on j. Then

(a) *If $[p\max(\beta(\theta) - \alpha(\theta), 0) + \max(\beta^*(\theta) - \alpha(\theta), 0)] < 1$ at all $\theta \in \Theta^*$,*

$$\lim_{N\to\infty} \frac{1}{N} tr\left[\prod_{j=1}^{p+1}\left\{\Sigma_N(f_{\theta,j})^{-1}\Sigma_N(g_{\theta,j})\right\}\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\prod_{j=1}^{p+1}\frac{g_{\theta,j}(\omega)}{f_{\theta,j}(\omega)}\right\}d\omega$$

uniformly in $\theta \in \Theta^*$.

(b) *Let $\eta > 1$. If $[p\max(\beta(\theta) - \alpha(\theta), 0) + \max(\beta^*(\theta) - \alpha(\theta), 0)] < \eta$ at all $\theta \in \Theta^*$,*

$$\lim_{N\to\infty} \frac{1}{N^{\eta}} tr\left[\prod_{j=1}^{p+1}\left\{\Sigma_N(f_{\theta,j})^{-1}\Sigma_N(g_{\theta,j})\right\}\right] = 0$$

uniformly in $\theta \in \Theta^*$.

Remark 4.3.1 *The result of Theorem 4.3.1 also holds for more general $f_{\theta,j}(\omega)$ and $g_{\theta,j}(\omega)$ that satisfy similar conditions with $\alpha_j(\theta)$, $\beta_j(\theta)$ that may depend on j. Note however that the proof of Theorem 5.1 of Dahlhaus (1989) uses a chaining of matrices square roots of the form $\Sigma_N(g_{j-1})^{1/2}\Sigma_N(f_j)^{-1}\Sigma_N(g_j)^{1/2}$ and $\Sigma_N(g_{k-1})^{1/2}\Sigma_N\left\{(4\pi^2 f_k)^{-1}\right\}\Sigma_N(g_k)^{1/2}$. Therefore, in the general case, the conditions on the value of*

$$[p\max(\beta(\theta) - \alpha(\theta), 0) + \max(\beta^*(\theta) - \alpha(\theta), 0)],$$

need to be replaced by equivalent conditions on the value of

$$\sum_{j=1}^{p+1}\left[\max\left\{\left(\beta_j(\theta) - \alpha_j(\theta)\right)/2, 0\right\} + \max\left\{\left(\beta_j(\theta) - \alpha_{j+1}(\theta)\right)/2, 0\right\}\right]$$

where $\alpha_{p+2} = \alpha_1$.

The second Theorem presented in this section is a uniform version of Grenander and Szegö's (1958) Theorem, which represents the first part of the Whittle approximation, i.e. of $\frac{1}{N}logdet\Sigma_N(f)$ by the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log 2\pi f(\omega)d\omega$. The next Lemma is the original result of Grenander and Szegö, and it is followed by its uniform version.

Lemma 4.3.2 (Grenander and Szegö 1958, p. 65(12)) *Let $f(\omega)$ be a real-valued, non-negative and integrable function on Π such that*

$$\int_{-\pi}^{\pi}\log f(\omega)\,d\omega > -\infty.$$

Then

$$\lim_{N\to\infty}\frac{1}{N}logdet\Sigma_N(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(2\pi f(\omega))\,d\omega. \tag{4.4}$$

Remark 4.3.2 *Result (4.4) of Grenander and Szegö (1958) is in fact formulated for a function $f^*(\omega)$ with positive and finite lower and upper bounds, where $f^*(\omega)$ represents $\frac{1}{2\pi}f(\omega)$ in Lemma 4.3.2. The condition on the bounds of $f^*(\omega)$ is then used to obtain a weaker version of Theorem 4.3.1. Grenander and Szegö's proof (p. 66(d)) of (4.4) is based on the fact that the minimum, μ_N, of the integral*

$$T_n = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|1+u_1e^{i\omega}+...+u_Ne^{i\omega N}\right| f^*(\omega)\,d\omega$$

where $u_1,...,u_N$ are complex variables, is given by

$$\mu_N = \frac{det\Sigma_N(f^*)}{det\Sigma_{N-1}(f^*)}, \tag{4.5}$$

while on the other hand they show that

$$\lim_{N\to\infty}\mu_N = exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f^*(\omega)\,d\omega\right\}. \tag{4.6}$$

Hence, by using (4.5) and (4.6),

$$\lim_{N\to\infty}\frac{det\Sigma_N(f^*)}{det\Sigma_{N-1}(f^*)} = exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f^*(\omega)\,d\omega\right\}.$$

This implies (4.4). Both (4.5) and (4.6), however, hold for any $f^*(\omega)$ that obeys the weaker conditions of Lemma 4.3.2 (see Grenander and Szegö 1958, 2.2(a) and 3.1(a)).

The following Theorem states a uniform version for Grenander and Szegö's result.

Theorem 4.3.2 *Let Θ^* be a compact subset of Θ. Suppose (A.1)-(A.3) hold. Then*

$$\lim_{N\to\infty}\frac{1}{N}logdet\Sigma_N\left(f_\theta\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_\theta\left(\omega\right)\right)d\omega$$

uniformly in $\theta\in\Theta^*$.

Proof. Set $\varepsilon>0$. We prove first that for each $\theta'\in\Theta$ there exists a $\delta>0$ and an integer M_δ such that for all $N\geq M_\delta$

$$\sup_{|\theta-\theta'|<\delta}\left|\frac{1}{N}logdet\Sigma_N\left(f_\theta\right)-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_\theta\left(\omega\right)\right)d\omega\right|<\varepsilon, \tag{4.7}$$

that is, $\frac{1}{N}logdet\Sigma_N\left(f_\theta\right)$ converges uniformly to $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_\theta\left(\omega\right)\right)d\omega$ on the δ-ball with center at θ', $B\left(\theta',\delta\right)$. Since $\alpha\left(\theta\right)$ is continuous, we may suppose that δ is always chosen such that $\alpha\left(\theta\right)\leq\alpha^*<1$ on $B\left(\theta',\delta\right)$. Note that the LHS of (4.7) is smaller than

$$\begin{aligned}&\sup_{|\theta-\theta'|<\delta}\left|\frac{1}{N}logdet\Sigma_N\left(f_\theta\right)-\frac{1}{N}logdet\Sigma_N\left(f_{\theta'}\right)\right| \qquad (4.8)\\&+\left|\frac{1}{N}logdet\Sigma_N\left(f_{\theta'}\right)-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_{\theta'}\left(\omega\right)\right)d\omega\right|\\&+\sup_{|\theta-\theta'|<\delta}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_{\theta'}\left(\omega\right)\right)d\omega-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(2\pi f_\theta\left(\omega\right)\right)d\omega\right|.\end{aligned}$$

Consider now the first term in (4.8). We obtain with a mean value θ^*

$$\frac{1}{N}logdet\Sigma_N\left(f_\theta\right)-\frac{1}{N}logdet\Sigma_N\left(f_{\theta'}\right)\leq\sum_{j=1}^{p}\left(\theta_j-\theta'_j\right)\frac{1}{N}tr\left[\Sigma_N\left(\frac{\partial}{\partial\theta_j}f_{\theta^*}\right)\Sigma_N^{-1}\left(f_{\theta^*}\right)\right],$$

which converges uniformly to $\sum_{j=1}^{p}\left(\theta_j-\theta'_j\right)\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta_j}f_{\theta^*}\left(\omega\right)f_{\theta^*}^{-1}\left(\omega\right)\right\}d\omega$ by Theorem 4.3.1. Since

$$\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta_j}f_{\theta^*}\left(\omega\right)f_{\theta^*}^{-1}\left(\omega\right)\right\}d\omega\right|\leq\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\left|\frac{\partial}{\partial\theta_j}f_{\theta^*}\left(\omega\right)\right|\left|f_{\theta^*}^{-1}\left(\omega\right)\right|\right\}d\omega,$$

assumptions (A.1)-(A.3) imply that $\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta_j}f_{\theta^*}(\omega)\,f_{\theta^*}^{-1}(\omega)\right\}d\omega\right|$ is bounded by some positive constant, K, that is independent of θ^* and j. Therefore we have

$$\sup_{|\theta-\theta'|<\delta}\left|\frac{1}{N}logdet\Sigma_N(f_\theta)-\frac{1}{N}logdet\Sigma_N(f_{\theta'})\right|\leq\delta pK+|\varepsilon_N^*|$$

with $\varepsilon_N^*\to 0$. The last expression can be made as small as desired by appropriate choices of δ and N (or M_δ), and particularly it can be made smaller than $\frac{\varepsilon}{3}$. The second term in (4.8) can also be made smaller than $\frac{\varepsilon}{3}$ for large enough M_δ due to the fact that $\frac{1}{N}logdet\Sigma_N(f_{\theta'})$ converges to $\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log(2\pi f_{\theta'}(\omega))\,d\omega$ by Lemma 4.3.2. Since $\log(2\pi f_{\theta'}(\omega))$ is continuous and integrable, $\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log(2\pi f_{\theta'}(\omega))\,d\omega$ is continuous as well by the dominated convergence theorem, and the third term in (4.8) can also be made smaller than $\frac{\varepsilon}{3}$ for sufficiently small δ. Thus, we proved (4.7). Now, since Θ^* is compact, it is possible to construct a finite open covering of Θ^* by δ_j-balls with centers at θ_j', $j=1,...,K$, such that (4.7) holds for each θ_j'. As a result, we get that $\frac{1}{N}logdet\Sigma_N(f_\theta)$ converges uniformly to $\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\log(2\pi f_\theta(\omega))\,d\omega$ on Θ^*. ■

4.4 Distribution of Quadratic Forms

It is convenient to adopt Dahlhaus's (1989) notation, so ∇g_θ and $\nabla^2 g_\theta$ are the gradient vector and Hessian matrix of g_θ with respect to θ, that is,

$$\nabla g_\theta=\left(\frac{\partial}{\partial\theta_j}g_\theta\right)_{j=1,\dots,p}\qquad and\qquad \nabla^2 g_\theta=\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}g_\theta\right)_{j,k=1,\dots,p}.$$

Denote $\Sigma_N(\nabla f_\theta)=\nabla\Sigma_N(f_\theta)$ as the p-length gradient vector with j'th component equals to the $N\times N$ matrix $\Sigma_N\left(\frac{\partial}{\partial\theta_j}f_\theta\right)$. Alternatively, $\Sigma_N(\nabla f_\theta)$ can be viewed as a three-dimensional matrix (called cubix) in $\mathbb{R}^{p\times N\times N}$, where the ijk entry of $\Sigma_N(\nabla f_\theta)$ represents the partial derivative of the jk'th entry of $\Sigma_N(f_\theta)$ with respect to θ_j. Similarly, $\Sigma_N\left(\nabla^2 f_\theta\right)=\nabla^2\Sigma_N(f_\theta)$ is the $p\times p$ Hessian matrix, with jk'th component equal to the $N\times N$ matrix $\Sigma_N\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}f_\theta\right)$, or alternatively, a four-dimensional matrix (called quartix) in $\mathbb{R}^{p\times p\times N\times N}$. (cf. Dattorro 2005, Appendix D). Note that as an example,

$$|||\Sigma_N(\nabla f_\theta)||| = \left(\sum_{j=1}^{p}\left\|\Sigma_N\left(\frac{\partial}{\partial\theta_j}f_\theta\right)\right\|^2\right)^{1/2}.$$

In many cases we use Σ_θ and $\Sigma_{\nabla,\theta}$ as $\Sigma_N(f_\theta)$ and $\Sigma_N(\nabla f_\theta)$, respectively. The following notations are used as well:

$$A_\theta^{(0)} = \Sigma_N(f_\theta)^{-1},$$

$$A_\theta^{(1)} = \Sigma_N(f_\theta)^{-1}\Sigma_N(\nabla f_\theta)\Sigma_N(f_\theta)^{-1},$$

$$A_\theta^{(2)} = \Sigma_N(f_\theta)^{-1}\Sigma_N\left(\nabla^2 f_\theta\right)\Sigma_N(f_\theta)^{-1}$$

and

$$A_\theta^{(3)} = \Sigma_N(f_\theta)^{-1}\Sigma_N(\nabla f_\theta)\Sigma_N(f_\theta)^{-1}\Sigma_N(\nabla f_\theta)'\Sigma_N(f_\theta)^{-1}.$$

In the present section we aim to derive the asymptotic distribution of the quadratic forms

$$(\mathbf{X}-\hat{\mu}_N\mathbf{1})' A_\theta^{(i)}(\mathbf{X}-\hat{\mu}_N\mathbf{1}), \qquad i = 0,1,2,3,$$

where $\hat{\mu}_N$ is a consistent estimate of μ_0. The following Proposition states a well-known expression for the joint cumulants for quadratic forms of stationary Gaussian time series where the true mean of the series is known (see, e.g., Dahlhaus 1989, p. 1757).

Proposition 4.4.1 *For each $j = 1,...,l$, suppose that $R_{N,j}$ is an $N\times N$ nonnegative definite matrix, and let*

$$Q_{N,j} = (\mathbf{X}-\mu_0\mathbf{1})' R_{N,i}(\mathbf{X}-\mu_0\mathbf{1}), \qquad j = 1,...,l,$$

be the corresponding quadratic form with matrix $R_{N,j}$. Then, the l'th order joint cumulant of $(Q_{N,1},...,Q_{N,l})$ is given by

$$k_l(Q_{N,1},...,Q_{N,l}) = 2^{l-1}\frac{1}{l}\sum tr\left[\prod_{k=1}^{l}\left(\Sigma_N(f_{\theta_0})R_{N,i_k}\right)\right], \tag{4.9}$$

where the summation is over all permutations $(i_1, ..., i_l)$ *of* $(1, ..., l)$*. Particularly, if* $Q_{N,i} = Q_N$, $i = 1, ..., l$*, the ordinary* l*'th order cumulant of* Q_N *is given by*

$$k_l(Q_N) = 2^{l-1}(l-1)!tr\left[(\Sigma_N(f_{\theta_0})R_N)^l\right]. \tag{4.10}$$

Because of the nonuniform behaviour of $tr\left[\Sigma_N(f_\theta)^{-1}\Sigma_N(f_{\theta_0})\right]$ around $\alpha(\theta_0) - \alpha(\theta) = 1$, implied by Theorem 4.3.1, we need to consider separately the case where θ lies in some compact subset of Θ in which $\max_\theta(\alpha(\theta_0) - \alpha(\theta)) < 1$, and the case where it is possible that $\alpha(\theta_0) - \alpha(\theta) \geq 1$. A similar distinction between the two cases was taken by Fox and Taqqu (1987) and Terrin and Taqqu (1990), who considered the distribution of the quadratic form $Q_N = (\mathbf{X} - \mu_0\mathbf{1})'\Sigma_N(f_\theta)(\mathbf{X} - \mu_0\mathbf{1})$. In the context of estimation, a similar treatment appears in Robinson's (1995b) consideration of semiparametric estimation for nonstationary and invertible time series and Velasco and Robinson's (2000) discrete-frequency version of Whittle Likelihood estimation. We therefore define

$$\Theta_1 = \Theta \cap \{\theta \in \mathbb{R}^p : \alpha(\theta) \geq \alpha(\theta_0) - 1 + \varepsilon\} \tag{4.11}$$

for some $\varepsilon \in (0, 1)$. ε can be taken as small as desired. Particularly, if $\alpha(\theta_0) \leq 0$ we consider $\Theta_1 = \Theta$.

Theorem 4.4.1 *Suppose (A.0)-(A.3) hold, and let* $Q_{N,\theta}^{(i)} = (\mathbf{X} - \mu_0\mathbf{1})' A_\theta^{(i)}(\mathbf{X} - \mu_0\mathbf{1})$*. Then*

$$\begin{aligned} \frac{1}{N} \cdot Q_N^{(0)} &\to_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)}{f_\theta(\omega)}d\omega, \\ \frac{1}{N} \cdot Q_N^{(1)} &\to_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\nabla f_\theta(\omega)}{f_\theta(\omega)^2}d\omega, \\ \frac{1}{N} \cdot Q_N^{(2)} &\to_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\nabla^2 f_\theta}{f_\theta(\omega)^2}d\omega, \end{aligned}$$

and

$$\frac{1}{N} \cdot Q_N^{(3)} \to_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)(\nabla f_\theta(\omega))^2}{f_\theta(\omega)^3}d\omega$$

uniformly in $\theta \in \Theta_1$*.*

Proof. We prove the result for $Q_N^{(3)}$. The results for $Q_N^{(i)}$ where $i = 0, 1, 2$ are obtained similarly. Using the cumulants formula (4.10) we have

$$E\left(Q_N^{(3)}\right) = \frac{1}{N}tr\left[A_\theta^{(3)}\Sigma_{\theta_0}\right],$$

which converges to $\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\nabla f_\theta(\omega)^2}{f_\theta(\omega)^3}d\omega$ uniformly on Θ_1 by Theorem 4.3.1. Similarly,

$$Cov\left(Q_N^{(3)}\right)=\frac{2}{N^2}tr\left[\left(A_\theta^{(3)}\Sigma_{\theta_0}\right)^2\right],$$

which decays to 0 uniformly on Θ_1 by Theorem 4.3.1. The result is then obtained with Markov's inequality. ■

The next two Lemmas derive asymptotic order for $\mathbf{1}'A_\theta^{(i)}\mathbf{1}$, $i=0,1,2,3$. The first Lemma is for $i=0$, and it is due to Adenstedt (1974, Theorem 5.2). The second Lemma is of Dahlhaus (1989, Lemma 5.4d), who proved a more general result for $A_\theta^{(i)}$ with $i=0,1,2,3$.

Lemma 4.4.1 (Adenstedt 1974, Theorem 5.2) *Suppose* $f(\omega)=\frac{1}{2\pi}\left|1-e^{i\omega}\right|^{-\alpha}\cdot g(\omega)$ *where* $g(\omega)$ *is symmetric, real valued, has positive upper and lower bounds and continuous at* $\omega=0$. *Then as* $N\to\infty$,

$$\left|\mathbf{1}'\Sigma_\theta^{-1}\mathbf{1}\right|\sim\frac{B\left(1-\frac{\alpha}{2},1-\frac{\alpha}{2}\right)}{\Gamma(1-\alpha)\,g(0)}N^{1-\alpha},$$

where $B(p,q)=\Gamma(p)\Gamma(q)/\Gamma(p+q)$ *is the Beta function and* $\Gamma(p)$ *is the gamma function.*

Lemma 4.4.2 (Dahlhaus 1989, Lemma 5.4d) *Let* Θ^* *be a compact subset of* Θ. *Suppose (A.0)-(A.3) hold on* Θ^*. *Then for each* $\theta\in\Theta^*$, $\delta>0$ *and* $i=0,1,2,3$,

$$\left|\mathbf{1}'A_\theta^{(i)}\mathbf{1}\right|\le KN^{1-\alpha(\theta)+\delta},$$

where K *is positive and independent of* θ *and* N.

In order to obtain an analogous result to Theorem 4.4.1 with $\hat\mu_N$ substituted for μ_0 in the quadratic form $Q_N^{(i)}$, we assume that $\hat\mu_N$ fulfills the following condition.

(A.4) With $\alpha(\theta)$ as in assumptions (A.1)-(A.3), the following holds for each $\delta>0$

(a) If $\hat\mu_N$ is independent of θ then

$$N^{\{1-\alpha(\theta_0)\}/2-\delta}\left|\hat\mu_N-\mu_0\right|=o_p(1).$$

(b) If $\hat\mu_N=\hat\mu_N(\theta)$ then

$$\sup_{\theta\in\Theta}\left[N^{\{1-\max(\alpha(\theta_0),\alpha(\theta))\}/2-\delta}\left|\hat\mu_N-\mu_0\right|\right]=o_p(1).$$

Part (a) of assumption (A.4) corresponds to the assumption on $\hat{\mu}_N$ in Theorem 3.2 of Dahlhaus (1989). This condition is fulfilled, for example, by the arithmetic mean and linear M-estimates (see Section 2.4). The case where $\hat{\mu}_N$ is dependent of θ is also of particular interest, for example, for the profile MLE of μ_0 (3.4), which depends on θ through the covariance matrix $\Sigma_N(f_\theta)$. In the latter case we expect a different rate of convergence of $\hat{\mu}_N$ to μ_0 for different θ's (see, for example, Sections 7-8 of Adenstedt 1974). Consequently, some additional considerations are required to derive a uniform convergence of the plug-in log-likelihood and its derivatives. The next two Theorems show that we may get the desired extension for any estimate of μ_0 that fulfills assumption (A.4).

The following Theorem summarizes and generalizes some of the ideas of Dahlhaus (1989, pp. 1757-1758), and it states an upper bound for the asymptotic order of the expression $\left|\widehat{Q}_N^{(i)} - Q_N^{(i)}\right|$ where

$$\widehat{Q}_N^{(i)} = (\mathbf{X} - \hat{\mu}_N \mathbf{1})' A_\theta^{(i)} (\mathbf{X} - \hat{\mu}_N \mathbf{1})$$

and

$$Q_N^{(i)} = (\mathbf{X} - \mu_0 \mathbf{1})' A_\theta^{(i)} (\mathbf{X} - \mu_0 \mathbf{1}).$$

Theorem 4.4.2 *Suppose (A.0)-(A.4) hold. Then for every $\delta > 0$, $i = 0, 1, 2, 3$,*

(a) If $\alpha(\theta_0) \geq \alpha(\theta)$,

$$\left|\widehat{Q}_N^{(i)} - Q_N^{(i)}\right| \leq K N^{\alpha(\theta_0) - \alpha(\theta) + \delta},$$

with K is positive and independent of θ and N.

(b) if $\alpha(\theta_0) < \alpha(\theta)$,

$$\left|\widehat{Q}_N^{(i)} - Q_N^{(i)}\right| \leq K N^{\delta},$$

with K is positive and independent of θ and N.

Proof. Statements (a) and (b) are proved together for $i = 3$. The results for $i = 0, 1, 2$ are obtained similarly. We have

$$\begin{aligned} \left|\widehat{Q}_N^{(i)} - Q_N^{(i)}\right| &= \left|2(\mu_0 - \hat{\mu}_N) \mathbf{1}' A_\theta^{(3)} (\mathbf{X} - \mu_0 \mathbf{1}) + (\mu_0 - \hat{\mu}_N)^2 \mathbf{1}' A_\theta^{(3)} \mathbf{1}\right| \\ &\leq 2|\mu_0 - \hat{\mu}_N| \left|\mathbf{1}' A_\theta^{(3)} (\mathbf{X} - \mu_0 \mathbf{1})\right| + |\mu_0 - \hat{\mu}_N|^2 \left|\mathbf{1}' A_\theta^{(3)} \mathbf{1}\right|. \end{aligned}$$

The second term is smaller than $KN^{\max(\alpha(\theta_0),\alpha(\theta))-\alpha(\theta)+3\delta} = KN^{\max(\alpha(\theta_0)-\alpha(\theta),0)+3\delta}$ with Lemma 4.4.2. Concentrate now on the first term. We prove that

$$2\left|\mu_0-\hat{\mu}_N\right|\left|\mathbf{1}'A_\theta^{(3)}\left(\mathbf{X}-\mu_0\mathbf{1}\right)\right| \leq KN^{\max(\alpha(\theta_0)-\alpha(\theta),0)+4\delta}. \tag{4.12}$$

We obtain

$$\begin{aligned}
& E_{\theta_0}\left\{\left|\mathbf{1}'A_\theta^{(3)}\left(\mathbf{X}-\mu_0\mathbf{1}\right)\right|^2\right\} \qquad (4.13)\\
\leq\ & \left|\mathbf{1}'A_\theta^{(3)}\Sigma_{\theta_0}A_\theta^{(3)}\mathbf{1}\right| \\
\leq\ & \left\|\Sigma_\theta^{-1/2}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\Sigma_{\theta_0}\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1/2}\right\|\left|\mathbf{1}'A_\theta^{(3)}\mathbf{1}\right| \\
\leq\ & \left\|\Sigma_\theta^{-1/2}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1/2}\right\|^2\left\|\Sigma_\theta^{-1/2}\Sigma_{\theta_0}\Sigma_\theta^{-1/2}\right\|\left|\mathbf{1}'A_\theta^{(3)}\mathbf{1}\right|
\end{aligned}$$

Let $\nabla f_\theta = g_\theta^+ - g_\theta^-$ with $g_\theta^+, g_\theta^- \geq 0$.

$$\left\|\Sigma_\theta^{-1/2}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1/2}\right\| \leq \left\|\Sigma_\theta^{-1/2}\Sigma_N\left(g_\theta^+\right)\Sigma_\theta^{-1/2}\right\| + \left\|\Sigma_\theta^{-1/2}\Sigma_N\left(g_\theta^-\right)\Sigma_\theta^{-1/2}\right\|.$$

We may therefore assume here wlg that ∇f_θ is nonnegative, and by Lemmas 4.3.1 and 4.4.2 we get that (4.13) is smaller than

$$KN^{\max(\alpha(\theta_0)-\alpha(\theta),0)+2\delta}N^{1-\alpha(\theta)+\delta} = KN^{\max(1+\alpha(\theta_0)-2\alpha(\theta),1-\alpha(\theta))+3\delta},$$

for each $\delta > 0$, with K independent of θ and N. An application of the Markov's inequality then yields (4.12). ■

As a direct corollary from Theorems 4.4.1 and 4.4.2, the following Theorem is obtained.

Theorem 4.4.3 *Suppose (A.0)-(A.4) hold, and let* $Q_N^{(i)} = \left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)' A_\theta^{(i)}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)$. *Then*

$$\begin{aligned}
\frac{1}{N}\cdot Q_N^{(0)} &\rightarrow_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)}{f_\theta(\omega)}d\omega, \\
\frac{1}{N}\cdot Q_N^{(1)} &\rightarrow_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\nabla f_\theta(\omega)}{f_\theta(\omega)^2}d\omega, \\
\frac{1}{N}\cdot Q_N^{(2)} &\rightarrow_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\nabla^2 f_\theta}{f_\theta(\omega)^2}d\omega,
\end{aligned}$$

and

$$\frac{1}{N}\cdot Q_N^{(3)} \rightarrow_p \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}(\omega)\left(\nabla f_\theta(\omega)\right)^2}{f_\theta(\omega)^3}d\omega$$

uniformly in $\theta \in \Theta_1$.

We now provide a short proof that the condition in assumption (A.4)(b) is satisfied by Adenstedt's (1974) profile MLE of μ_0, (3.4), denoted here as $\hat{\mu}_N^A$. This Theorem is not needed for the rest of chapter, and it may be safely skipped.

Lemma 4.4.3 *Suppose (A.1) and (A.3) hold on a compact set* Θ. *Then (A.4)(b) holds for* $\hat{\mu}_N^A(\theta)$.

Proof. Adenstedt's estimator is given by

$$\hat{\mu}_N^A = \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{X}_N$$

It is easily seen that

$$E\left(\hat{\mu}_N^A\right) = \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\mu_0 = \mu_0.$$

Furthermore,

$$\begin{aligned}
Var\left(\hat{\mu}_N^A\right) &= \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-2}\cdot tr\left(\Sigma_N(f_{\theta_0})\Sigma_N(f_\theta)^{-1}\mathbf{1}\mathbf{1}'\Sigma_N(f_\theta)^{-1}\right) \\
&= \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-2}\cdot\left|\mathbf{1}'\Sigma_N(f_\theta)^{-1}\Sigma_N(f_{\theta_0})\Sigma_N(f_\theta)^{-1}\mathbf{1}\right| \\
&\leq \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-2}\cdot\left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)\cdot\left\|\Sigma_N(f_\theta)^{-1/2}\Sigma_N(f_{\theta_0})\Sigma_N(f_\theta)^{-1/2}\right\| \\
&= \left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{-1}\left\|\Sigma_N(f_\theta)^{-1/2}\Sigma_N(f_{\theta_0})\Sigma_N(f_\theta)^{-1/2}\right\|.
\end{aligned}$$

By Lemmas 4.3.1 and 4.4.1 this term is equal to $N^{\{\max(\alpha(\theta_0),\alpha(\theta))-1\}-\delta}$, which yields the result. ■

4.5 Consistency of the MLE

Let $\hat{\theta}_N$ be the maximum likelihood estimator, obtained by minimizing the normalized (by $-N$) Gaussian plug-in log-Likelihood function

$$\mathcal{L}_N(\theta) = \frac{1}{2N}\log\det\Sigma_N(f_\theta) + \frac{1}{2N}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)'\Sigma_N(f_\theta)^{-1}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right) \tag{4.14}$$

with respect to Θ, where $\Sigma_N(f_\theta)$ is the covariance matrix of X given by (4.3) and $\hat{\mu}_N$ is an estimate of μ_0.

In this section the consistency of $\hat{\theta}_N$ is established. The proof makes use of the results established in the former sections in order to derive an asymptotic limit for $\mathcal{L}_N(\theta)$ if $\theta \in \Theta_1$. However, since the desirable parameter space includes all $\alpha \in (-1,1)$, an additional theory is required in order to handle the possibility that $\theta \in \Theta_1^{-1}$, in which our knowledge about the limit distribution of the quadratic form $Q_N = (\mathbf{X} - \hat{\mu}_N \mathbf{1})' \Sigma_N (f_\theta)^{-1} (\mathbf{X} - \hat{\mu}_N \mathbf{1})$ is relatively poor. To handle this case, we have adapted the idea of Velasco and Robinson (2000, Theorem 1), who considered a parametric estimation based on a generalization of the discrete-frequency of Whittle estimation, and proved that in the region of $\theta' s$ where $\alpha(\theta_0) - \alpha(\theta) \geq 1$, the Whittle log-Likelihood (normalized by $-N$) diverges to $+\infty$ a.s. as $N \to \infty$. Our proof shows a similar property for the plug-in Likelihood, and therefore $\hat{\theta}_N$ cannot be found in that region for N large enough.

Theorem 4.5.1 *Suppose (A.0)-(A.4) hold. Then*

$$\hat{\theta}_N \to_p \theta_0. \tag{4.15}$$

Proof. *Suppose that $\hat{\theta}_N$ a the minimizer of $\mathcal{L}_N(\theta)$ in Θ such that for all $\theta' \in \Theta$ we have $\mathcal{L}_N\left(\hat{\theta}_N\right) \leq \mathcal{L}_N(\theta')$. For any $\delta \in (0,1)$ with $U_\delta(\theta_0) = \{\theta \in \Theta : |\theta - \theta_0| < \delta\}$*

$$\begin{aligned}
\lim_{N\to\infty} P_{\theta_0}\left(\left|\hat{\theta}_N - \theta_0\right| > \delta\right) &\leq \lim_{N\to\infty} P_{\theta_0}\left(\exists \theta^* \in \operatorname*{arginf}_{\theta\in\Theta} \mathcal{L}_N(\theta),\ \theta^* \in \{\Theta \setminus U_\delta(\theta_0)\}\right) \\
&= \lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\{\Theta\setminus U_\delta(\theta_0)\}} \mathcal{L}_N(\theta) = \inf_{\theta\in\Theta} \mathcal{L}_N(\theta)\right) \\
&= \lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\{\Theta\setminus U_\delta(\theta_0)\}} \mathcal{L}_N(\theta) \leq \inf_{\theta\in U_\delta(\theta_0)} \mathcal{L}_N(\theta)\right).
\end{aligned}$$

Consider first subspace $\Theta_1 \subseteq \Theta$ defined as in (4.11), and assume that ε in the definition is small enough such that $U_\delta(\theta_0) \subsetneq \Theta_1$. We prove that

$$\lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\{\Theta_1\setminus U_\delta(\theta_0)\}} \mathcal{L}_N(\theta) > \mathcal{L}_N(\theta_0)\right) = 1. \tag{4.16}$$

With Theorems 4.3.2 and 4.4.3 we obtain for all $\theta \in \Theta_1$

$$\sup_{\theta\in\Theta_1} |\mathcal{L}_N(\theta) - \mathcal{L}(\theta)| \to_p 0, \tag{4.17}$$

where

$$\mathcal{L}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log(2\pi f_\theta(\omega))\, d\omega + \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{f_{\theta_0}(\omega)}{f_\theta(\omega)} d\omega.$$

It is therefore suffice to show that

$$\begin{aligned} & \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\left(2\pi f_{\theta}(\omega)\right) d\omega + \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{f_{\theta_0}(\omega)}{f_{\theta}(\omega)} d\omega \\ > \ & \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\left(2\pi f_{\theta_0}(\omega)\right) d\omega + \frac{1}{2} \quad \textit{for all } \theta \in \Theta_1,\ \theta \neq \theta_0, \end{aligned}$$

or, equivalently, that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_{\theta_0}(\omega)}{f_{\theta}(\omega)} d\omega > 1 + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{f_{\theta_0}(\omega)}{f_{\theta}(\omega)} d\omega \quad \textit{for all } \theta \in \Theta_1,\ \theta \neq \theta_0. \tag{4.18}$$

By (4.1), $\frac{f_{\theta_0}(\omega)}{f_{\theta}(\omega)}$ *is positive and finite almost everywhere on* Π*. Moreover, by (A.0) the set* $\left\{\omega : \frac{f_{\theta_0}(\omega)}{f_{\theta}(\omega)} \neq 1\right\}$ *has a positive Lebesgue measure. Therefore, the inequality*

$$x \geq 1 + \log x, \quad \textit{for all } x > 0,$$

where the inequality is strict for all $x \neq 1$*, immediately yields (4.18), and (4.16) is established.*

Suppose now that $\theta \notin \Theta_1$*. With the same* ε *as in the definition of* Θ_1*, let*

$$\Theta_2 = \Theta_1^{-1} = \Theta \cap \left\{\theta \in \mathbb{R}^p : \alpha(\theta) < \alpha(\theta_0) - 1 + \varepsilon\right\}.$$

In this case Θ *is just* $\Theta_1 \cup \Theta_2$ *and for any* $\delta \in (0,1)$

$$\begin{aligned} & \lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\{\Theta\setminus U_\delta(\theta_0)\}} \mathcal{L}_N(\theta) \leq \inf_{\theta\in U_\delta(\theta_0)} \mathcal{L}_N(\theta)\right) \\ \leq \ & \lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\{\Theta_1\setminus U_\delta(\theta_0)\}} \mathcal{L}_N(\theta) \leq \inf_{\theta\in U_\delta(\theta_0)} \mathcal{L}_N(\theta)\right) \\ & + \lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta\in\Theta_2} \mathcal{L}_N(\theta) \leq \inf_{\theta\in U_\delta(\theta_0)} \mathcal{L}_N(\theta)\right), \end{aligned} \tag{4.19}$$

where the first probability tends to 0 as $N \to \infty$ *by (4.16). To show that the second probability is negligible as well, note first that for any two parameter vectors* $\theta_1, \theta_2 \in \Theta$ *such that* $\alpha(\theta_1) \geq \alpha(\theta_2)$,

$$\begin{aligned} \inf_{x\neq 0} \frac{x'\Sigma_N(f_{\theta_2})^{-1} x}{x'\Sigma_N(f_{\theta_1})^{-1} x} &= \inf_{x\neq 0} \frac{x'x}{x'\Sigma_N(f_{\theta_2})^{1/2}\Sigma_N(f_{\theta_1})^{-1}\Sigma_N(f_{\theta_2})^{1/2} x} \\ &= \left[\sup_{x\neq 0}\left(\frac{x'\Sigma_N(f_{\theta_2})^{1/2}\Sigma_N(f_{\theta_1})^{-1}\Sigma_N(f_{\theta_2})^{1/2} x}{x'x}\right)\right]^{-1} \\ &= \left\|\Sigma_N(f_{\theta_1})^{-1/2}\Sigma_N(f_{\theta_2})^{1/2}\right\|^{-2}, \end{aligned} \tag{4.20}$$

which, by Lemma 4.3.1, is greater than some positive constant K independent of θ_1, θ_2 and N. Consider now any parameter vectors $\theta_2 \in \Theta_2$ and θ_1 on the boundary of Θ_1. We have

$$\alpha(\theta_2) < \alpha(\theta_1) = \alpha(\theta_0) - 1 + \varepsilon.$$

Substituting $x = \mathbf{X} - \hat{\mu}_N \mathbf{1}$ into (4.20) yields

$$\frac{1}{2N}(\mathbf{X} - \hat{\mu}_N \mathbf{1})' \Sigma_N(f_{\theta_2})^{-1}(\mathbf{X} - \hat{\mu}_N \mathbf{1}) \geq K \frac{1}{2N}(\mathbf{X} - \hat{\mu}_N \mathbf{1})' \Sigma_N(f_{\theta_1})^{-1}(\mathbf{X} - \hat{\mu}_N \mathbf{1}). \tag{4.21}$$

If $\hat{\mu}_N = \hat{\mu}_N(\theta)$, then $\hat{\mu}_N$ in both sides of (4.21) should be computed at the same θ, say θ_2. In this case, replace $\hat{\mu}_N(\theta_2)$ in the RHS of (4.21) by $\hat{\mu}^*_N(\theta_1)$, where

$$\hat{\mu}^*_N(\theta) = \begin{cases} \hat{\mu}_N(\theta_2) & \theta = \theta_1 \\ \hat{\mu}_N(\theta) & \textit{otherwise} \end{cases}.$$

Note also that $\hat{\mu}^*_N$ satisfies (A.4)(b). This result, together with Theorems 4.3.2 and 4.4.3, yields for some positive $K_1,\ K_2, c_1, c_2 \in \mathbb{R}$, independent of θ_1, θ_2 and N,

$$\begin{aligned} \lim_{N\to\infty} \mathcal{L}_N(\theta_2) &\geq \frac{1}{4\pi}\int_{-\pi}^{\pi} \log(2\pi f_{\theta_2}(\omega))\,d\omega + \frac{K_1}{4\pi}\int_{-\pi}^{\pi} \frac{f_{\theta_0}(\omega)}{f_{\theta_1}(\omega)}d\omega \\ &\geq \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\left(2\pi c_1 |\omega|^{-\alpha(\theta_2)+\delta}\right) d\omega + \frac{K_1}{4\pi}\int_{-\pi}^{\pi} c_2 |\omega|^{\alpha(\theta_1)-\alpha(\theta_0)+2\delta}\, d\omega \\ &\geq -K_2 + \frac{K_1 c_2}{4\pi}\int_{-\pi}^{\pi} |\omega|^{-1+\varepsilon+2\delta}\, d\omega = \frac{K_1 c_2}{2\pi}\frac{\pi^{\varepsilon+2\delta}}{\varepsilon+2\delta} - K_2 = C(\varepsilon,\delta). \end{aligned}$$

for each $\delta > 0$. The last term can be made as large as desired for any θ_0 and $\theta_2 \in \Theta_2$ by taking sufficiently small $\varepsilon, \delta > 0$. Particularly, δ and ε in the definitions of Θ_1 and Θ_2 may be chosen such that $C(\varepsilon) > \mathcal{L}_N(\theta_0)$. Therefore, the second probability in (4.19) tends to 0 as $N \to \infty$, and thus $\hat{\theta}_N \to_p \theta_0$. ■

4.6 Central Limit Theorem

Theorem 4.6.1 *Suppose (A.0)-(A.4) hold. Then $\sqrt{N}\left(\hat{\theta}_N - \theta_0\right)$ tends in distribution to a normal random vector with mean 0 and covariance matrix $\Gamma(\theta_0)^{-1}$ where*

$$\Gamma(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} (\nabla \log f_\theta(\omega))(\nabla \log f_\theta(\omega))'\, d\omega.$$

Proof. An application of the mean value theorem yields

$$\nabla \mathcal{L}_N \left(\hat{\theta}_N, \hat{\mu}_N\right) - \nabla \mathcal{L}_N \left(\theta_0, \hat{\mu}_N\right) = \nabla^2 \mathcal{L}_N \left(\overline{\theta}_N, \hat{\mu}_N\right) \left(\hat{\theta}_N - \theta_0\right).$$

with $\left|\overline{\theta}_N - \theta_0\right| \leq \left|\widehat{\theta}_N - \theta_0\right|$. The assertion will follow if we prove

(i) $\sqrt{N}\nabla \mathcal{L}_N \left(\hat{\theta}_N, \hat{\mu}_N\right) \rightarrow_p 0.$

(ii) $\sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \hat{\mu}_N\right) \rightarrow_D N\left(0, \Gamma\left(\theta_0\right)\right),$

(iii) $\nabla^2 \mathcal{L}_N \left(\overline{\theta}_N, \hat{\mu}_N\right) \rightarrow_p \Gamma\left(\theta_0\right).$

Start with (i). $\mathcal{L}_N\left(\theta_N, \hat{\mu}_N\right)$ is minimized by $\widehat{\theta}_N$ and therefore $\nabla \mathcal{L}_N \left(\hat{\theta}_N, \hat{\mu}_N\right) = 0$ if $\widehat{\theta}_N$ is in the interior of Θ. Since θ_0 lies in the interior of Θ, we get for all positive ε

$$\begin{aligned} & \lim_{N\rightarrow\infty} P_{\theta_0}\left(\sqrt{N}\nabla \mathcal{L}_N \left(\hat{\theta}_N, \hat{\mu}_N\right) > \varepsilon\right) \\ = & \lim_{N\rightarrow\infty} P_{\theta_0}\left(\sqrt{N}\nabla \mathcal{L}_N \left(\hat{\theta}_N, \hat{\mu}_N\right) > \varepsilon \text{ and } \widehat{\theta}_N \text{ lies on the boundary of } \Theta\right) \\ \leq & \lim_{N\rightarrow\infty} P_{\theta_0}\left(\widehat{\theta}_N \text{ lies on the boundary of } \Theta\right) = 0 \end{aligned}$$

by Theorem 4.5.1.

For part (ii), we have

$$\sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \hat{\mu}_N\right) = \frac{1}{2\sqrt{N}} tr\left\{\Sigma_{\theta_0}^{-1}\Sigma_{\nabla,\theta_0}\right\} - \frac{1}{2\sqrt{N}}\left(\mathbf{X} - \hat{\mu}_N\mathbf{1}\right)' A_{\theta_0}^{(1)}\left(\mathbf{X} - \hat{\mu}_N\mathbf{1}\right).$$

According to Theorem 4.4.2

$$\sup_{\theta\in\Theta}\sqrt{N}\left|\nabla \mathcal{L}_N \left(\theta_0, \hat{\mu}_N\right) - \nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)\right| \rightarrow_p 0,$$

and hence it is sufficient to prove the assertion for $\sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)$. We denote by $Q_{N,j}$, $j \in \{1, ..., p\}$, the j'th entry of the p-length vector $\left(\mathbf{X} - \mu_0\mathbf{1}\right)' A_{\theta_0}^{(1)}\left(\mathbf{X} - \mu_0\mathbf{1}\right)$. By Proposition 4.4.1 the l^{th} joint cumulant of $\left(Q_{N,j_1}, ..., Q_{N,j_l}\right)$ is given by

$$k_l\left(Q_{N,j_1}, ..., Q_{N,j_l}\right) = 2^{l-1}\frac{1}{l}\sum tr\left[\prod_{k=1}^{l}\left(\Sigma_N\left(f_{\theta_0}\right)\left(A_{\theta_0}^{(1)}\right)_{i_k}\right)\right],$$

where the summation is over all permutations $\left(i_1, ..., i_l\right)$ of $\left(j_1, ..., j_l\right)$. Thus, the l^{th} joint cumulant of $\left(\sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)_{j_1}, ..., \sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)_{j_l}\right)$ is

$$\begin{aligned} & k_l\left(\sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)_{j_1}, ..., \sqrt{N}\nabla \mathcal{L}_N \left(\theta_0, \mu_0\right)_{j_l}\right) \\ = & \left\{\begin{array}{ll} 0 & if\ l = 1 \\ \frac{1}{2}N^{-l/2}\left(-1\right)^l\frac{1}{l}\sum tr\left[\prod_{k=1}^{l}\left\{\Sigma_{\theta_0}^{-1}\Sigma\left(\frac{\partial}{\partial\theta_j}f_{\theta_0}\right)\right\}\right] & if\ l \geq 2 \end{array}\right. . \end{aligned}$$

An application of Theorem 4.3.1 yields

$$\lim_{N\to\infty} k_l\left(\sqrt{N}\nabla\mathcal{L}_N\left(\theta_0,\mu_0\right)\right)=\begin{cases}0 & if\ l\neq 2\\ \Gamma\left(\theta_0\right) & if\ l=2\end{cases}$$

and therefore (ii) is obtained.

For part (iii), we have

$$\begin{aligned}\nabla^2\mathcal{L}_N\left(\theta,\hat{\mu}_N\right) &= \frac{1}{2N}tr\left\{\Sigma_\theta^{-1}\Sigma_N\left(\nabla^2 f_\theta\right)\right\}-\frac{1}{2N}tr\left\{\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\right\}\\ &\quad -\frac{1}{2N}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)' A_\theta^{(2)}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)+\frac{1}{N}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)' A_\theta^{(3)}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right).\end{aligned}$$

If $\theta\in\Theta_1$ then $\nabla^2\mathcal{L}_N\left(\theta,\hat{\mu}_N\right)$ converges in probability by Theorems 4.3.1 and 4.4.3 to

$$\begin{aligned}&\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\nabla^2 f_\theta\left(\omega\right)}{f_\theta\left(\omega\right)}d\omega-\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\left(\nabla f_\theta\left(\omega\right)\right)^2}{f_\theta\left(\omega\right)^2}d\omega\\ &-\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}\left(\omega\right)\nabla^2 f_\theta\left(\omega\right)}{f_\theta\left(\omega\right)^2}d\omega+\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f_{\theta_0}\left(\omega\right)\left(\nabla f_\theta\left(\omega\right)\right)^2}{f_\theta\left(\omega\right)^3}d\omega\end{aligned}$$

Moreover, since $\left|\overline{\theta}_N-\theta_0\right|\leq\left|\widehat{\theta}_N-\theta_0\right|\to_p 0$, we obtain by the smoothness conditions (A.1)-(A.2) that

$$\nabla^2\mathcal{L}_N\left(\overline{\theta}_N,\hat{\mu}_N\right)\to_p\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\left(\nabla f_{\theta_0}\left(\omega\right)\right)^2}{f_{\theta_0}\left(\omega\right)^2}d\omega=\Gamma\left(\theta_0\right)$$

which establishes (iii). ■

To conclude the chapter, the asymptotic efficiency of the estimator $\hat{\theta}_N$ is now proved. Since X is assumed to be Gaussian, it is sufficient to show that the achieved asymptotic variance of $\hat{\theta}_N$ is equal to the Cramér-Rao bound. This property may be established in a direct way, as in Dahlhaus (1989, Theorem 4.1).

Theorem 4.6.2 *Suppose (A.0)-(A.4) hold. Then* $\hat{\theta}_N$ *is an efficient estimate of* θ_0 *in the sense of Fisher.*

Proof. Denote the Fisher information matrix by $\Gamma_N\left(\theta_0\right)$, and note that because of the smoothness conditions (A.1) and (A.2) we have

$$\Gamma_N\left(\theta_0\right)=NE_{\theta_0}\nabla^2\mathcal{L}_N\left(\theta_0,\mu_0\right).$$

As in the proof of part (iii) in Theorem 4.6.1, we obtain

$$N^{-1}\Gamma_N(\theta_0) \to \Gamma(\theta_0).$$

Therefore, according to the Cramér-Rao theorem, $\Gamma(\theta_0)^{-1}$ is the lower bound on the asymptotic variance of $\sqrt{N}\left(\widetilde{\theta}_N - \theta_0\right)$, where $\widetilde{\theta}_N$ is any unbiased estimator of θ. Hence, Theorem 4.6.2 is proved. ■

Chapter 5

MONTE CARLO STUDY

5.1 Introduction

In this chapter we analyze the finite sample properties of the exact and plug-in Gaussian MLE of several Gaussian ARFIMA(0,d,0), ARFIMA(1,d,0) and ARFIMA(0,d,1) time series.

Similar studies of the BIAS and MSE of the Exact and plug-in Gaussian MLE were conducted, for example, by Sowell (1992), Cheung and Diebold (1994), Hauser (1999) and Nielsen and Frederiksen (2005). They showed that the exact Gaussian MLE of the differencing parameter d suffers from a systematic negative bias that increases with d, and that the plug-in MLE with the sample mean has a generally higher negative bias than the exact MLE. They also showed that the bias in the estimates decreases in absolute value with the series length, N, so asymptotically we get an unbiased estimate in accordance with Theorem (4.6.1). While similar aspects are studied here, we also compare between the performance of the plug-in Gaussian MLE of ARFIMA(0,d,0) series with either the standard sample mean or the profile mean estimate of Adenstedt (1974, see Section 3.2.1). Some aspects of the MLE's finite sample distribution are examined as well.

5.2 Summary of results

We generated several Gaussian ARFIMA(0,d,0), ARFIMA(1,d,0) and ARFIMA(0,d,1) time series with zero-mean, unit variance and different values of the memory parameter and the AR or MA parameters, accordingly to the simulated model. For the ARFIMA(0,d,0) time series, we considered $\theta = d$ where $d \in [-0.49, 0.49]$. The true parameter values were $d_0 = 0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4$. For the simulated ARFIMA(1,d,0) and ARFIMA(0,d,1) time series, the AR and MA parameters ϕ and ξ, respectively, were estimated as well. In these cases we had $\theta = (d, \phi) \in \Theta_{d,\phi}$ and $\theta = (d, \xi) \in \Theta_{d,\xi}$, respectively to the ARFIMA(1,d,0) or ARFIMA(0,d,1) process. Here again, we considered $d \in [-0.49, 0.49]$, while both ϕ and ξ were limited to $[-0.99, 0.99]$. The true parameter values in these cases were $d_0 = 0, \pm 0.3$, and both ϕ_0 and ξ_0 took the values ± 0.6. For each of the models, 1000 replications of lengths $N = 100, 200$, and also $N = 500$ for the ARFIMA(0,d,0) models, were simulated for each value of θ_0. The likelihood was numerically maximized on a grid of length 0.01 with respect to the parameter's components.

The simulations were made with R 2.6.1 program. In the program we used Davies and Harte's (1987) method to simulate a stationary Gaussian time series based on the Fourier transform of the autocovariances (see also Beran 1994, pp. 216-217). Sowell's (1992) procedure was used to compute the autocovariances of an ARFIMA(p,d,q) model in terms of sums of hypergeometric functions. We implemented the Durbin-Levinson algorithm in order to reduce the time required for direct computation of the determinants and the inverses of the covariance matrices (see Section 3.3.1).

We now turn to the simulations results. Tables 5.1, 5.2 and 5.3 summarize the sample bias, the sample standard deviation and the square root of the sample MSE that were obtained for the parameter estimates in each of the generated processes. Figures 5.1-5.5 present some kernel density plots of the marginal sample probability density of the estimates. The plots were made by using Gaussian kernel density estimators. The kernel's bandwidths were chosen by Scott's (1992) variation of Silverman's (1986) "rule of thumb", both methods aimed to minimize the mean integrated squared error. In cases where some of the estimates were lying on the parameter space boundary (for instance, $\widehat{d} = -0.49$), the total area below the density plot is equal to 1 minus the estimated mass of probability on the boundary. Figures 5.6-5.10 display normal Q-Q plots to asses whether the marginal distributions of the MLE are close to a normal distribution.

We see that the estimates of d are negatively biased, with larger bias in absolute value for the plug-in estimates. In addition, for fixed N, the negative bias in the estimates of d generally increases as d_0 becomes larger. On the other hand, for a fixed d_0, the bias decreases in absolute value with N. The short memory parameters ϕ and ξ in the ARFIMA(1,d,0) and ARFIMA(0,d,1) models suffer in most of the cases from a positive bias. Relatively highly biased results were obtained in the cases of ARFIMA(1,d,0) series with $d = 0.3$, $\phi = 0.6$, or

ARFIMA(0,d,1) with $d = 0.3$, $\xi = -0.6$. Overall, our results compare with previous studies (e.g., Cheung and Diebold 1994, Hauser 1999, Nielsen and Frederiksen 2005).

A comparison between the exact MLE and the two plug-in estimates for ARFIMA(0,d,0) series shows that the plug-in MLE with the profile mean does not perform better than the plug-in MLE with the sample mean for all values of the memory parameter. In fact, in all cases where d_0 is negative, the plug-in MLE with the sample mean has a lower MSE than its counterpart with the profile mean. This result is somewhat surprising in view of the fact that the sample mean is asymptotically less efficient than the profile mean where the memory parameter is negative (see Section 3.2.1). However, it may stem from the fact that the profile mean plugged in to the likelihood function depends on the parameter values in which the likelihood is evaluated. This may cause a larger dependency between the estimated mean and the other estimated parameters, which may adversely affect the estimation of the parameters of interest. In general, we conclude that the benefit gained by plugging in the profile mean to the likelihood function is limited, while in addition it is a less convenient estimate relatively to the sample mean.

Figures 5.1-5.5 demonstrate the tendency of the MLE's finite sample probability density toward a narrow and symmetric distribution around the true parameter values as N increases, while Figures 5.1-5.5 presents the Q-Q-plots corresponding to these distributions. In some cases, and in particular in the ARFIMA(1,0.3,0) series with $\phi = 0.6$ and the ARFIMA(0,0.3,1) series with $\xi = -0.6$, the distribution of the plug-in MLE is relatively skewed and dispersed, even when the sample size is $N = 200$. In most other cases, the distribution of the MLE is close to a Normal when the sample size is $N = 200$, with some exceptions where d_0 is relatively low and the distribution of the MLE of d has a "chopped left tail" as a result of an accumulation of estimates on the parameter space boundary $d = -0.49$.

d_0	μ known BIAS	STD	$\sqrt{\text{MSE}}$	$\widehat{\mu}$ sample mean BIAS	STD	$\sqrt{\text{MSE}}$	$\widehat{\mu}$ profile BIAS	STD	$\sqrt{\text{MSE}}$
				N=100					
-0.4	-0.0004	0.0694	0.0694	-0.0124	0.0707	0.0718	-0.0235	0.0670	0.0710
-0.3	-0.0094	0.0827	0.0833	-0.0303	0.0861	0.0913	-0.0402	0.0865	0.0954
-0.2	-0.0108	0.0844	0.0851	-0.0372	0.0898	0.0972	-0.0438	0.0914	0.1014
-0.1	-0.0113	0.0832	0.0839	-0.0416	0.0912	0.1002	-0.0451	0.0914	0.1019
0	-0.0117	0.0815	0.0824	-0.0444	0.0913	0.1016	-0.0465	0.0904	0.1017
0.1	-0.0122	0.0794	0.0803	-0.0472	0.0909	0.1024	-0.0483	0.0894	0.1016
0.2	-0.0136	0.0767	0.0779	-0.0506	0.0894	0.1027	-0.0513	0.0878	0.1017
0.3	-0.0155	0.0722	0.0738	-0.0559	0.0858	0.1023	-0.0565	0.0843	0.1015
0.4	-0.0207	0.0615	0.0649	-0.0672	0.0777	0.1027	-0.0677	0.0768	0.1024
				N=200					
-0.4	-0.0037	0.0538	0.0539	-0.0102	0.0540	0.0549	-0.0160	0.0536	0.0559
-0.3	-0.0058	0.0577	0.0580	-0.0174	0.0596	0.0621	-0.0211	0.0613	0.0649
-0.2	-0.0058	0.0570	0.0573	-0.0203	0.0600	0.0634	-0.0216	0.0615	0.0652
-0.1	-0.0059	0.0565	0.0568	-0.0219	0.0602	0.0641	-0.0222	0.0612	0.0651
0	-0.0060	0.0558	0.0561	-0.0233	0.0600	0.0644	-0.0231	0.0606	0.0648
0.1	-0.0063	0.0551	0.0554	-0.0247	0.0597	0.0647	-0.0242	0.0601	0.0648
0.2	-0.0067	0.0539	0.0543	-0.0260	0.0590	0.0644	-0.0257	0.0591	0.0645
0.3	-0.0073	0.0524	0.0529	-0.0288	0.0573	0.0642	-0.0285	0.0577	0.0643
0.4	-0.0105	0.0469	0.0481	-0.0357	0.0522	0.0632	-0.0358	0.0532	0.0641
				N=500					
-0.4	-0.0020	0.0365	0.0366	-0.0063	0.0354	0.0360	-0.0080	0.0362	0.0371
-0.3	-0.0021	0.0370	0.0371	-0.0088	0.0361	0.0372	-0.0085	0.0371	0.0380
-0.2	-0.0020	0.0368	0.0369	-0.0101	0.0361	0.0375	-0.0086	0.0369	0.0379
-0.1	-0.0019	0.0368	0.0369	-0.0109	0.0362	0.0378	-0.0087	0.0368	0.0379
0	-0.0018	0.0365	0.0365	-0.0113	0.0362	0.0379	-0.0089	0.0365	0.0376
0.1	-0.0019	0.0361	0.0361	-0.0118	0.0360	0.0379	-0.0094	0.0361	0.0373
0.2	-0.0018	0.0358	0.0359	-0.0125	0.0359	0.0380	-0.0100	0.0357	0.0371
0.3	-0.0019	0.0352	0.0353	-0.0136	0.0354	0.0379	-0.0113	0.0354	0.0371
0.4	-0.0031	0.0330	0.0331	-0.0164	0.0337	0.0375	-0.0143	0.0334	0.0363

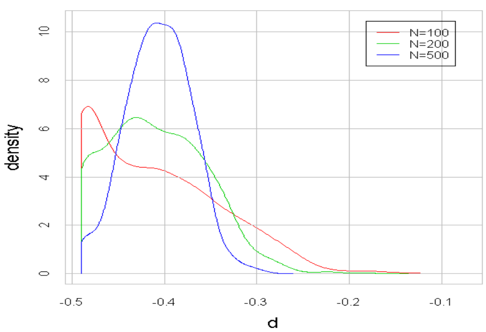
Table 5.1: Estimated Bias, standard deviation (STD) and square root of MSE obtained by 1000 replications of the Gaussian MLE of ARFIMA(0,d,0) time series with length N=100, 200, 500.

ϕ_0	$\widehat{\theta}_i$	$d_0 = -0.3$			$d_0 = 0$			$d_0 = 0.3$		
		BIAS	STD	$\sqrt{\text{MSE}}$	BIAS	STD	$\sqrt{\text{MSE}}$	BIAS	STD	$\sqrt{\text{MSE}}$
					N=100					
-0.6	d:	-0.0336	0.0959	0.1017	-0.0578	0.1128	0.1268	-0.0822	0.1073	0.1352
	ϕ :	0.0340	0.1009	0.1064	0.0390	0.1024	0.1096	0.0494	0.1061	0.1170
0.6	d:	-0.0323	0.1741	0.1771	-0.1553	0.1630	0.2251	-0.2441	0.1393	0.2810
	ϕ :	-0.0031	0.1705	0.1706	0.0845	0.1426	0.1658	0.1525	0.1031	0.1841
					N=200					
-0.6	d:	-0.0218	0.0689	0.0723	-0.0283	0.0709	0.0764	-0.0435	0.0697	0.0822
	ϕ :	0.0191	0.0672	0.0699	0.0204	0.0694	0.0723	0.0261	0.0698	0.0745
0.6	d:	-0.0478	0.1467	0.1543	-0.1420	0.1544	0.2098	-0.1845	0.1416	0.2326
	ϕ :	0.0225	0.1473	0.1490	0.0949	0.1324	0.1629	0.1263	0.1141	0.1702

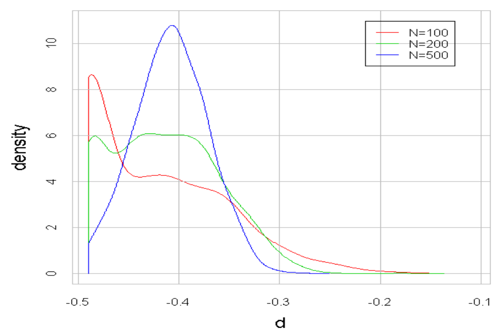
Table 5.2: Estimated Bias, standard deviation (STD) and square root of MSE obtained by 1000 replications of the Gaussian MLE of ARFIMA(1,d,0) time series with length N=100, 200.

ξ_0	$\widehat{\theta}_i$	$d_0 = -0.3$			$d_0 = 0$			$d_0 = 0.3$		
		BIAS	STD	$\sqrt{\text{MSE}}$	BIAS	STD	$\sqrt{\text{MSE}}$	BIAS	STD	$\sqrt{\text{MSE}}$
					N=100					
-0.6	d:	0.0777	0.2243	0.2374	-0.1672	0.2268	0.2818	-0.3084	0.1974	0.3662
	ξ :	-0.0696	0.2064	0.2178	0.1482	0.2482	0.2891	0.2902	0.2089	0.3576
0.6	d:	-0.0305	0.0983	0.1029	-0.0535	0.1027	0.1158	-0.0772	0.0982	0.1249
	ξ :	0.0179	0.0961	0.0978	0.0217	0.0976	0.0999	0.0349	0.0916	0.0980
					N=200					
-0.6	d:	0.0177	0.1898	0.1906	-0.1321	0.1846	0.2270	-0.1916	0.1542	0.2460
	ξ :	-0.0121	0.1809	0.1813	0.1286	0.1890	0.2286	0.1831	0.1646	0.2462
0.6	d:	-0.0179	0.0694	0.0716	-0.0277	0.0702	0.0755	-0.0431	0.0675	0.0801
	ξ :	0.0085	0.0672	0.0677	0.0118	0.0691	0.0701	0.0174	0.0653	0.0675

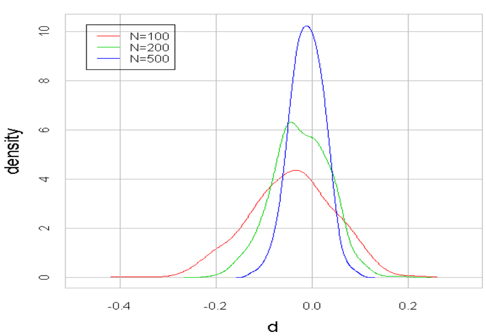
Table 5.3: Estimated Bias, standard deviation (STD) and square root of MSE obtained by 1000 replications of the Gaussian MLE of ARFIMA(1,d,0) time series with length N=100, 200.



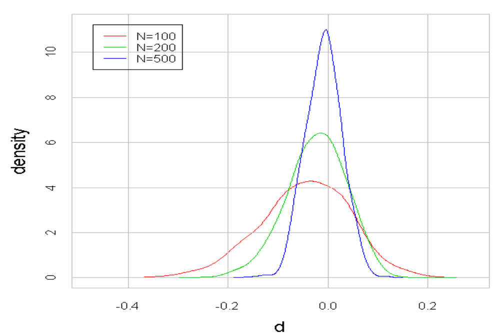
$\widehat{\mu}$ sample mean, $d_0 = -0.4$



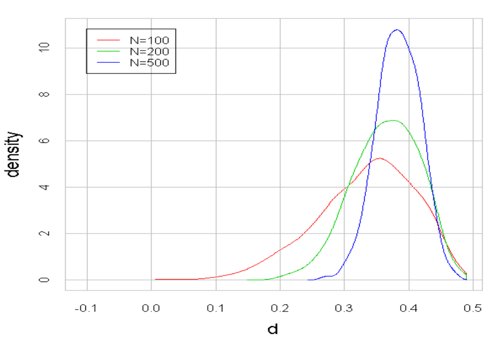
$\widehat{\mu}$ profile, $d_0 = -0.4$



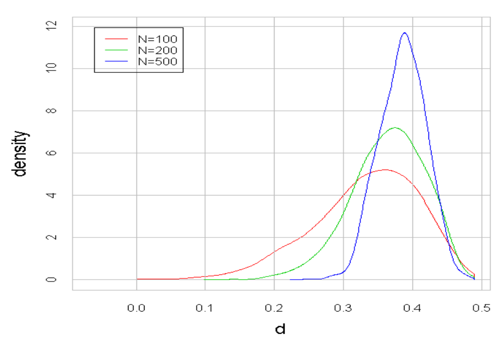
$\widehat{\mu}$ sample mean, $d_0 = 0$



$\widehat{\mu}$ profile, $d_0 = 0$

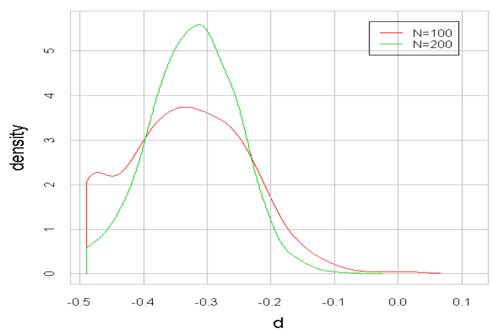


$\widehat{\mu}$ sample mean, $d_0 = 0.4$

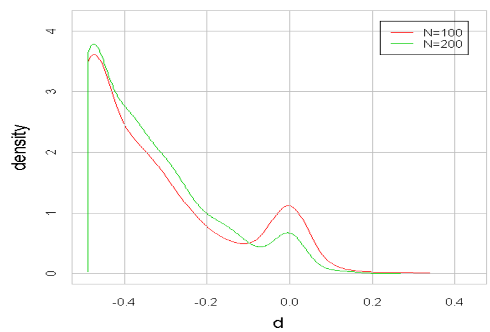


$\widehat{\mu}$ profile, $d_0 = 0.4$

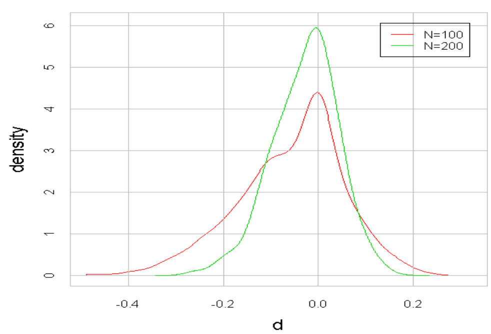
Figure 5.1: Kernel density plots of the sample mean and the profile mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,0) series with length N=100, 200, 500.



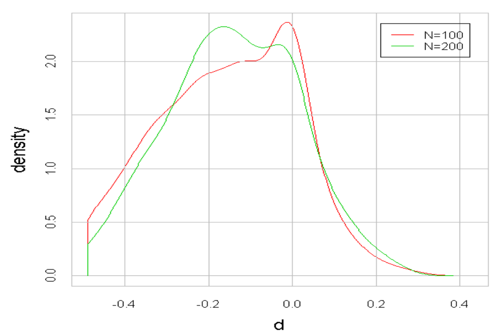
$\phi_0 = -0.6,\ d_0 = -0.3$



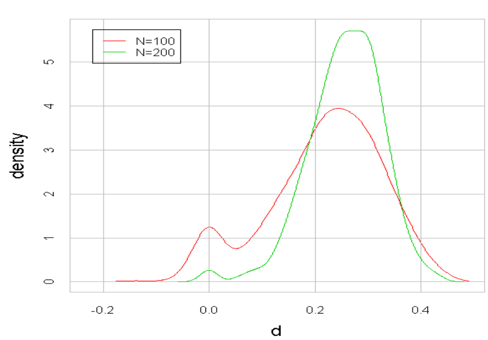
$\phi_0 = 0.6,\ d_0 = -0.3$



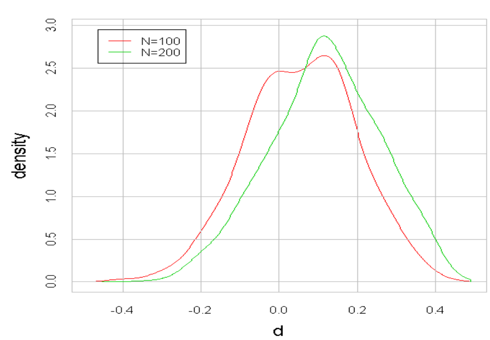
$\phi_0 = -0.6,\ d_0 = 0$



$\phi_0 = 0.6,\ d_0 = 0$

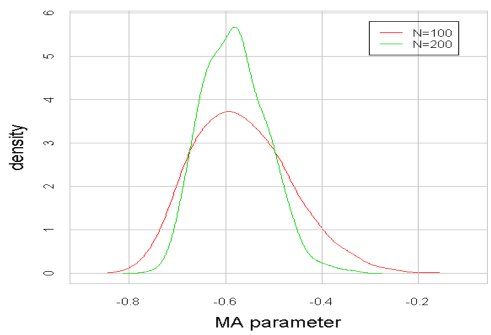


$\phi_0 = -0.6,\ d_0 = 0.3$

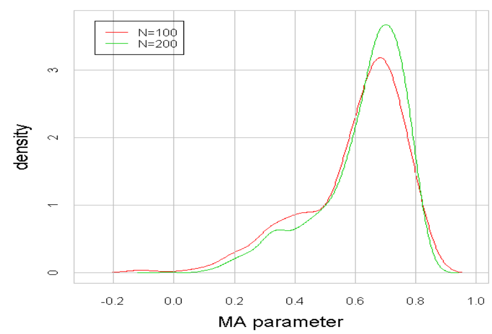


$\phi_0 = 0.6,\ d_0 = 0.3$

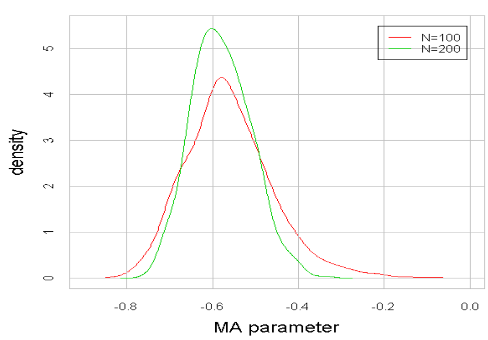
Figure 5.2: Kernel density plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.



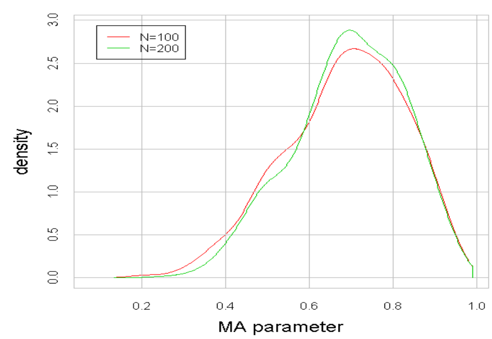
$\phi_0 = -0.6,\ d_0 = -0.3$



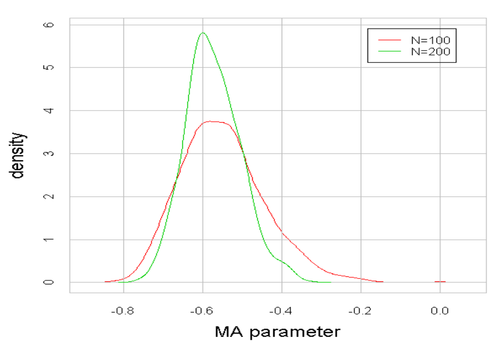
$\phi_0 = 0.6,\ d_0 = -0.3$



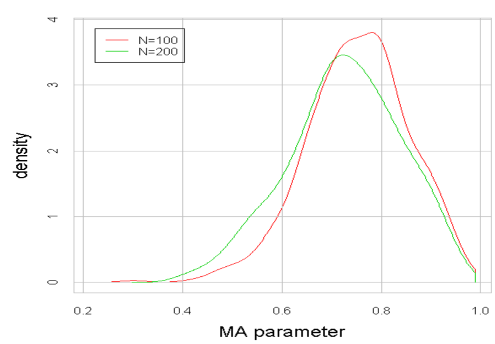
$\phi_0 = -0.6,\ d_0 = 0$



$\phi_0 = 0.6,\ d_0 = 0$



$\phi_0 = -0.6,\ d_0 = 0.3$



$\phi_0 = 0.6,\ d_0 = 0.3$

Figure 5.3: Kernel density plots of the sample mean plug-in Gaussian MLE of the AR parameter obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.

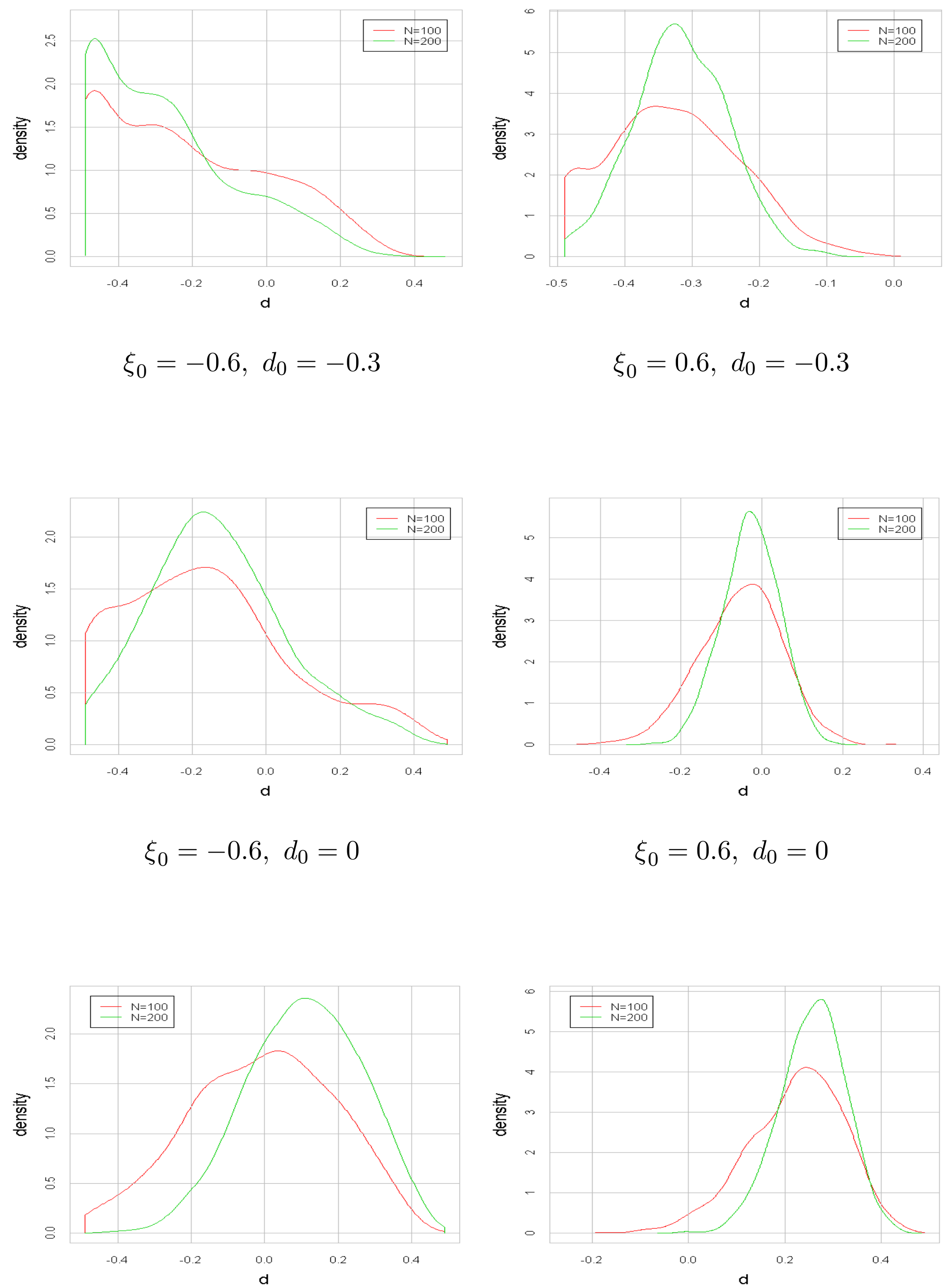
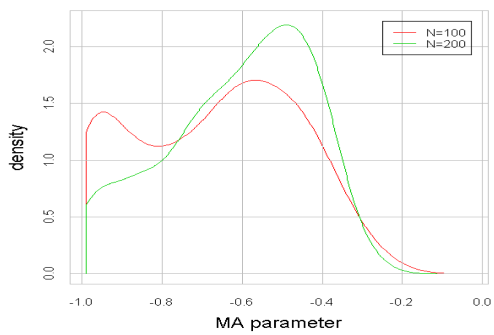
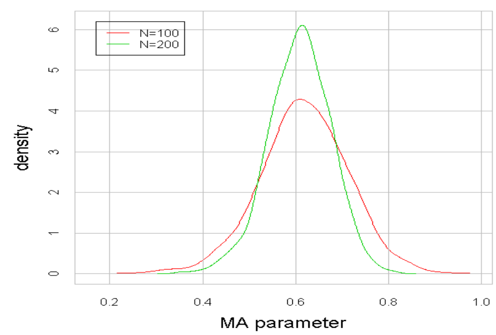


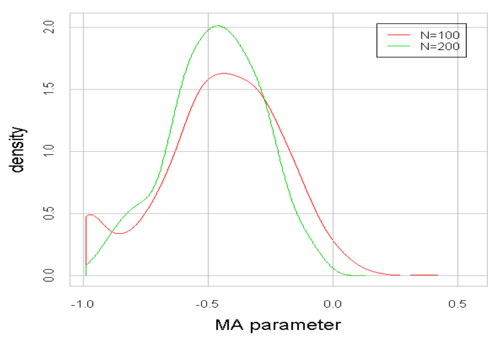
Figure 5.4: Kernel density plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.



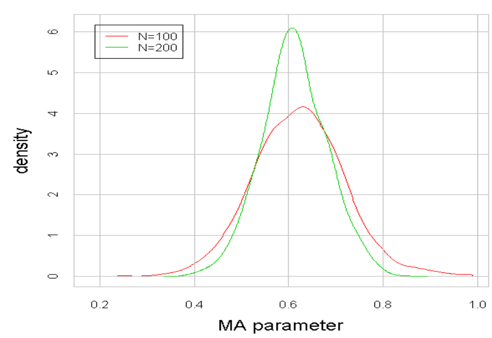
$\xi_0 = -0.6,\ d_0 = -0.3$



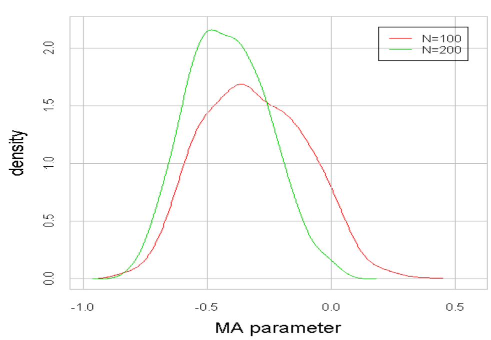
$\xi_0 = 0.6,\ d_0 = -0.3$



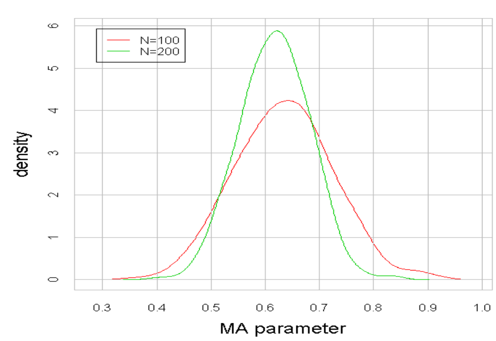
$\xi_0 = -0.6,\ d_0 = 0$



$\xi_0 = 0.6,\ d_0 = 0$

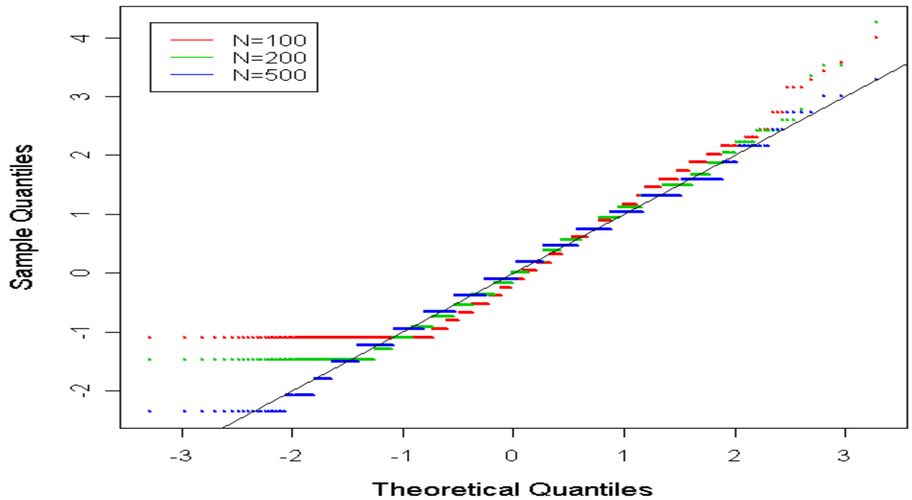


$\xi_0 = -0.6,\ d_0 = 0.3$

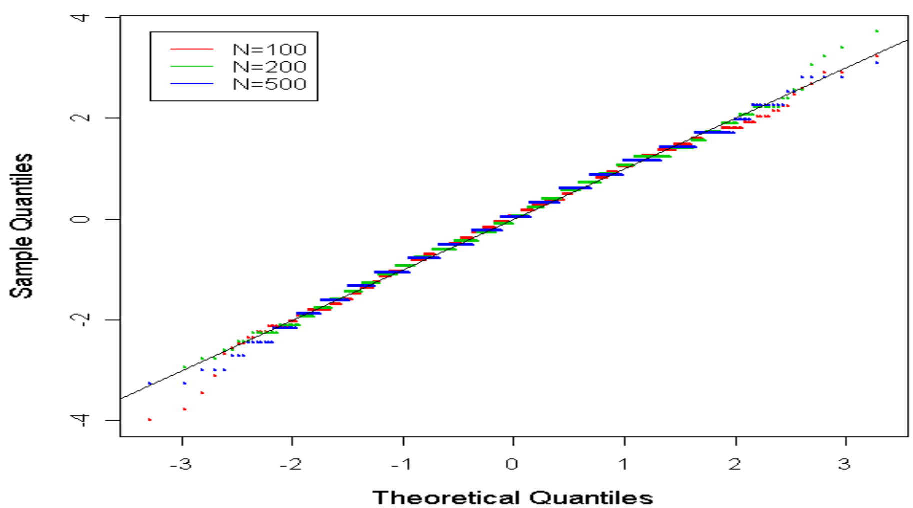


$\xi_0 = 0.6,\ d_0 = 0.3$

Figure 5.5: Kernel density plots of the sample mean plug-in Gaussian MLE of the MA parameter obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.

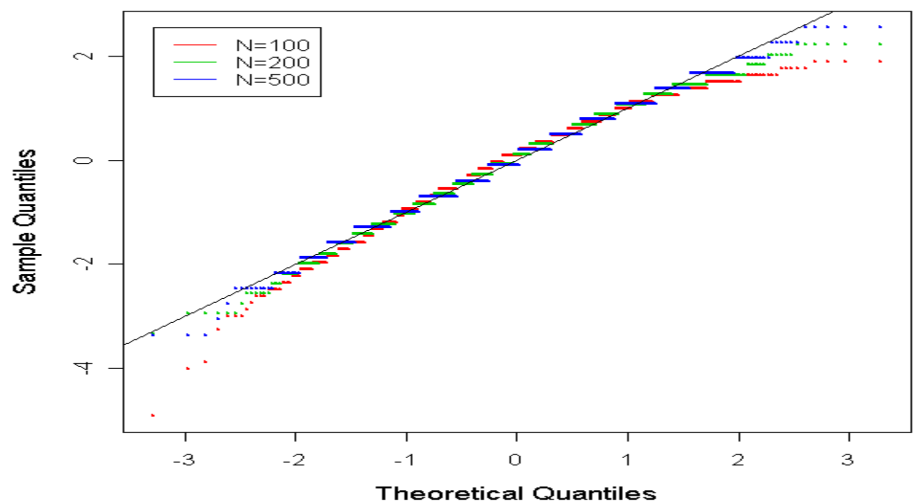


$\widehat{\mu}$ sample mean, $d_0 = -0.4$

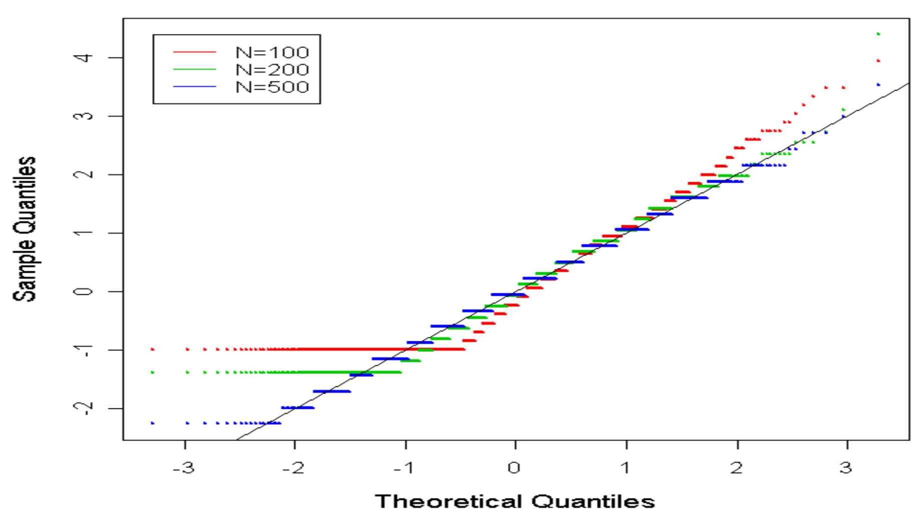


$\widehat{\mu}$ sample mean, $d_0 = 0$

Figure 5.6: Normal Q-Q plots of the sample mean and the profile mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,0) series with length N=100, 200, 500.

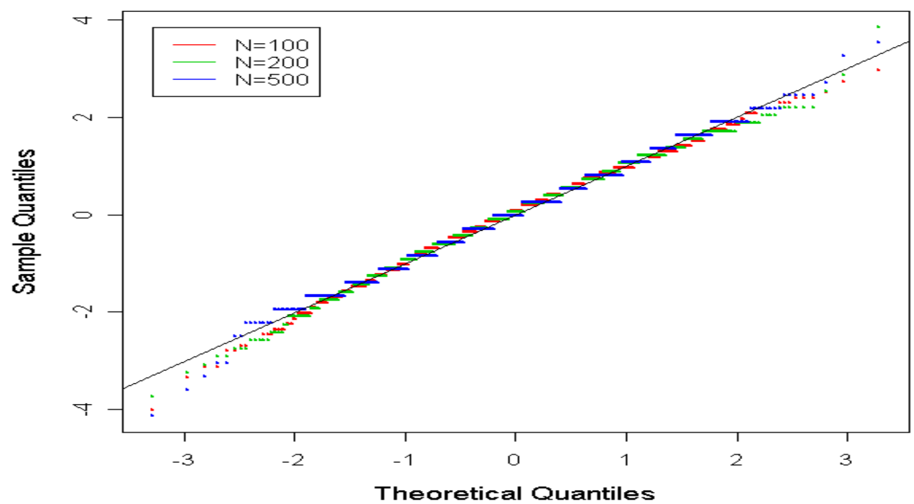


$\widehat{\mu}$ sample mean, $d_0 = 0.4$

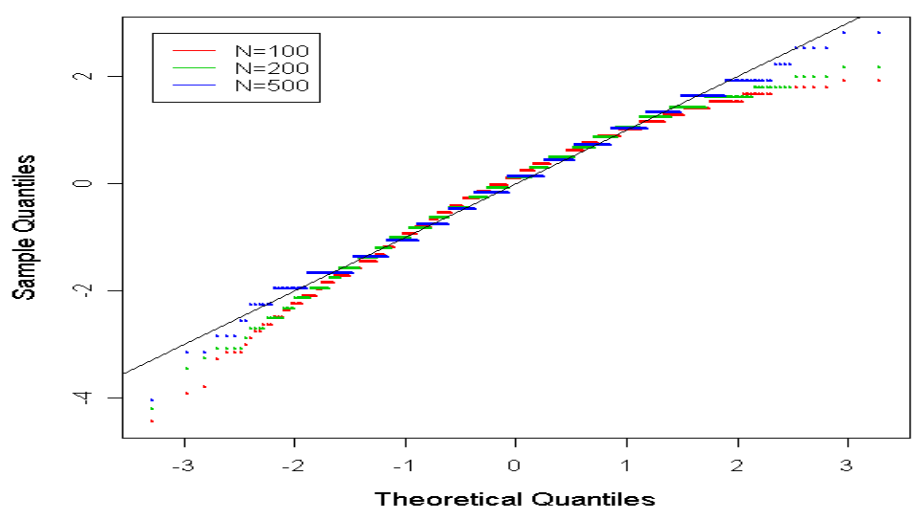


$\widehat{\mu}$ profile, $d_0 = -0.4$

Figure 5.6 (continued): Normal Q-Q plots of the sample mean and the profile mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,0) series with length N=100, 200, 500.

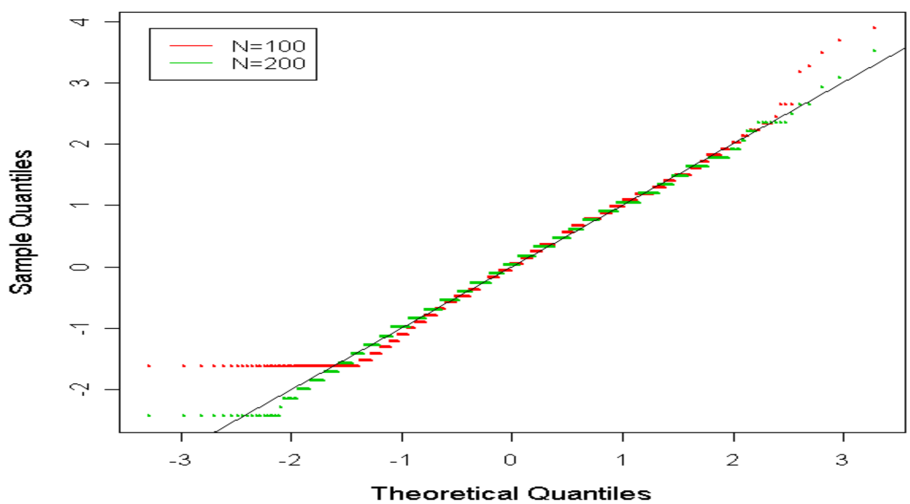


$\widehat{\mu}$ profile, $d_0 = 0$

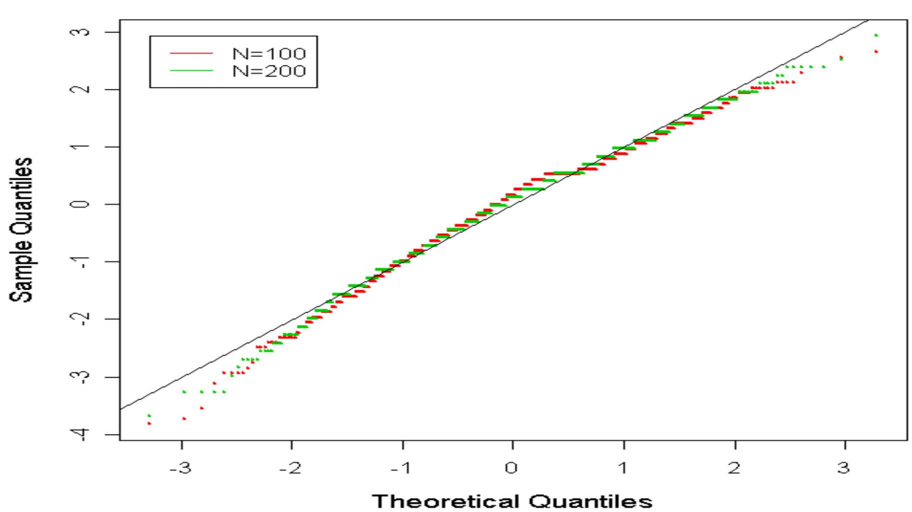


$\widehat{\mu}$ profile, $d_0 = 0.4$

Figure 5.6 (continued): Normal Q-Q plots of the sample mean and the profile mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,0) series with length N=100, 200, 500.

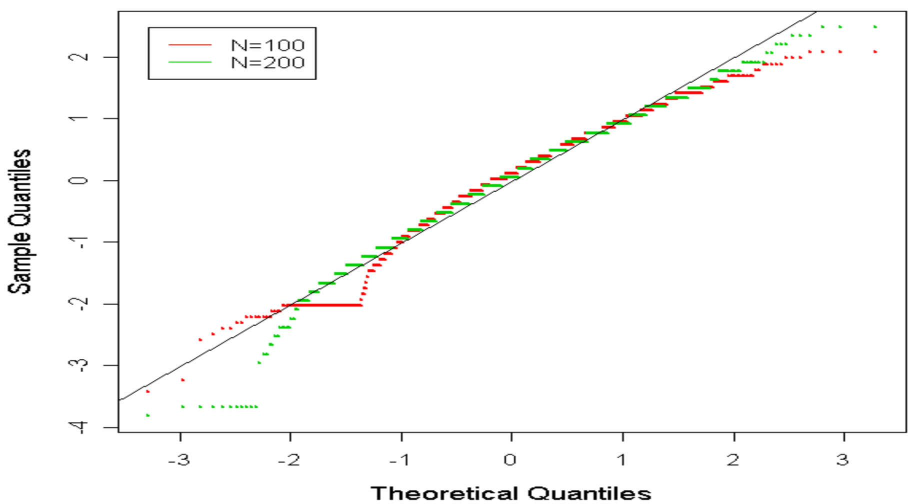


$\phi_0 = -0.6,\ d_0 = -0.3$

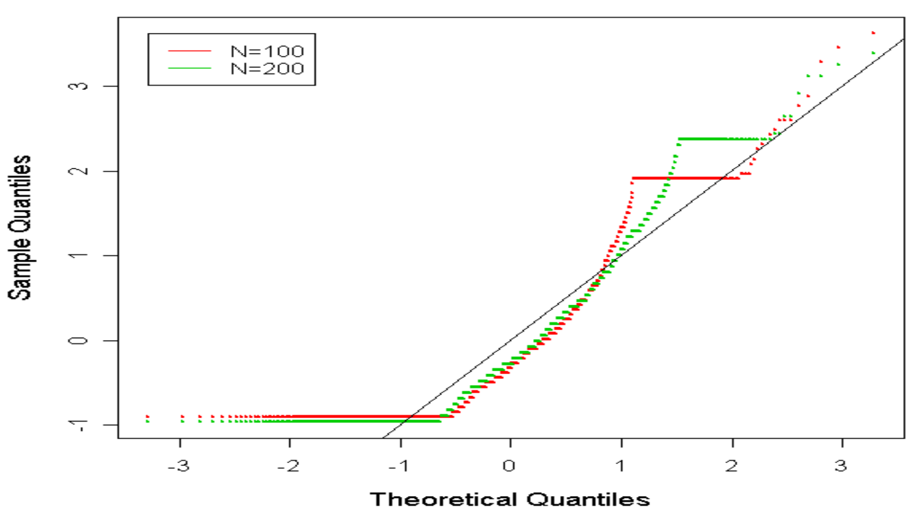


$\phi_0 = -0.6,\ d_0 = 0$

Figure 5.7: Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.

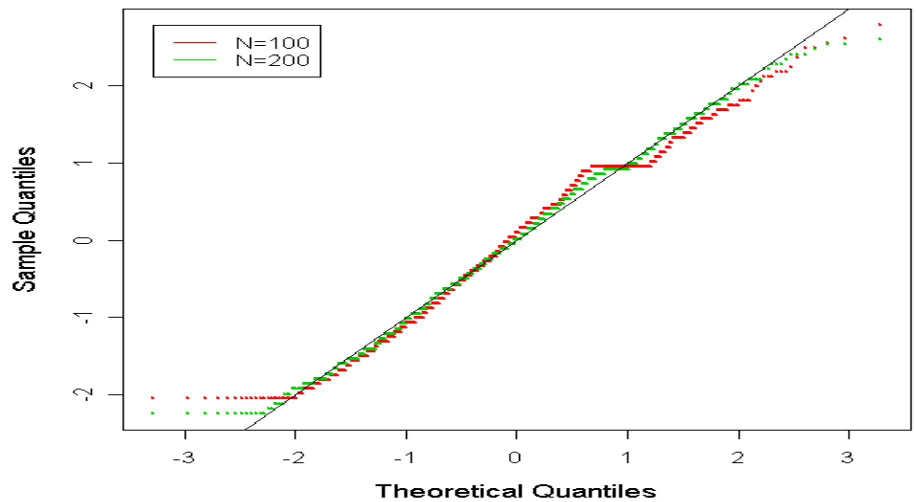


$\phi_0 = -0.6, \ d_0 = 0.3$

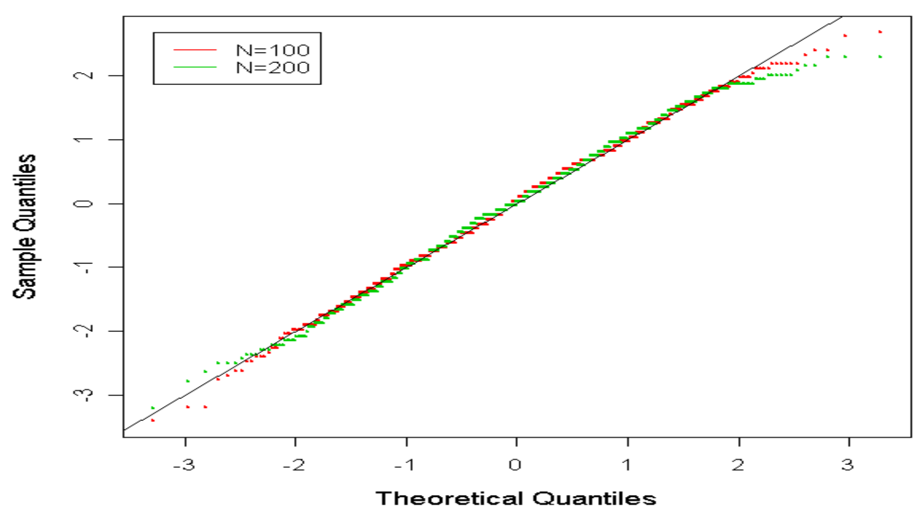


$\phi_0 = 0.6, \ d_0 = -0.3$

Figure 5.7 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.

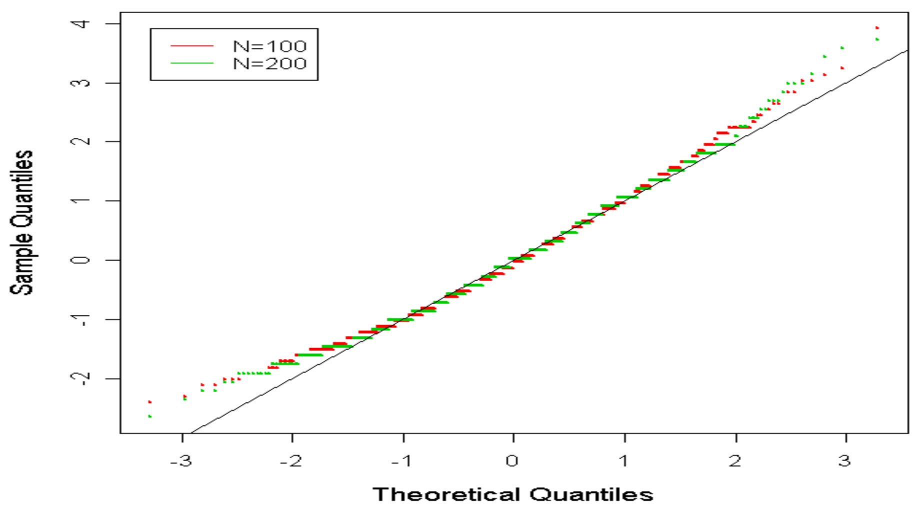


$\phi_0 = 0.6, \; d_0 = 0$

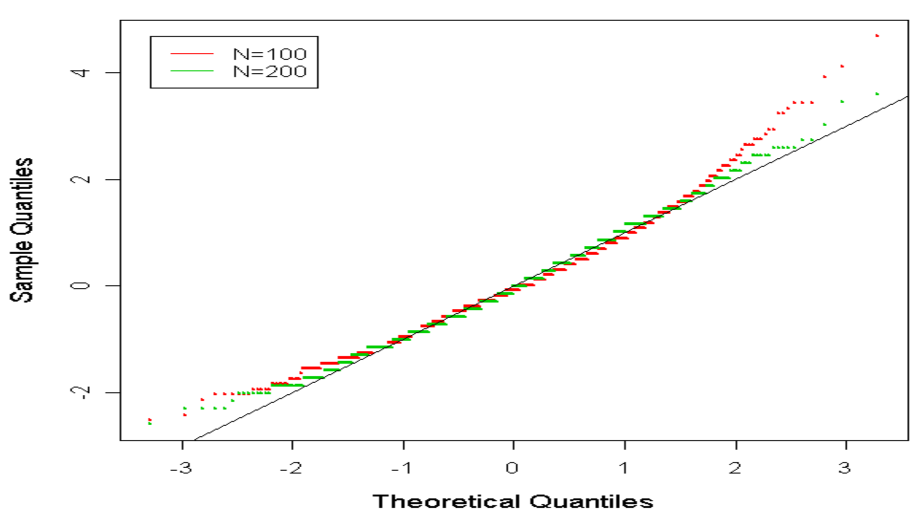


$\phi_0 = 0.6, \; d_0 = 0.3$

Figure 5.7 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.

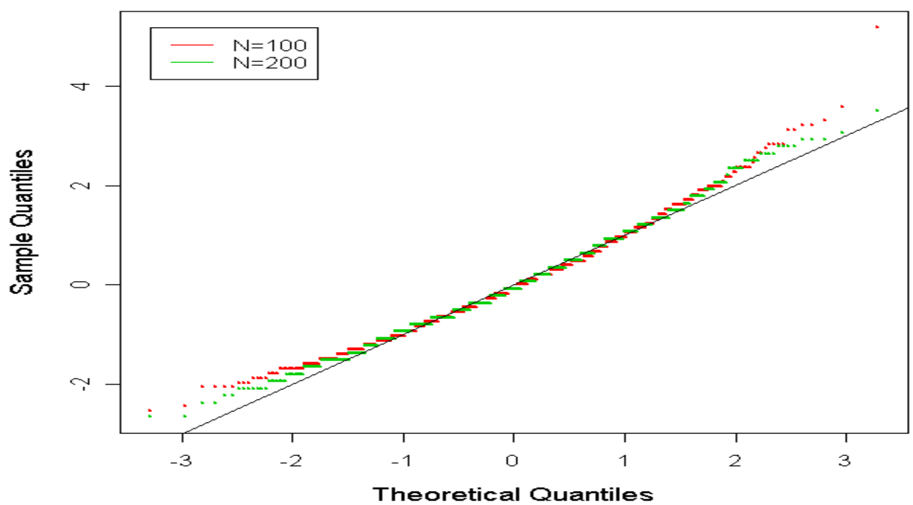


$\phi_0 = -0.6,\ d_0 = -0.3$

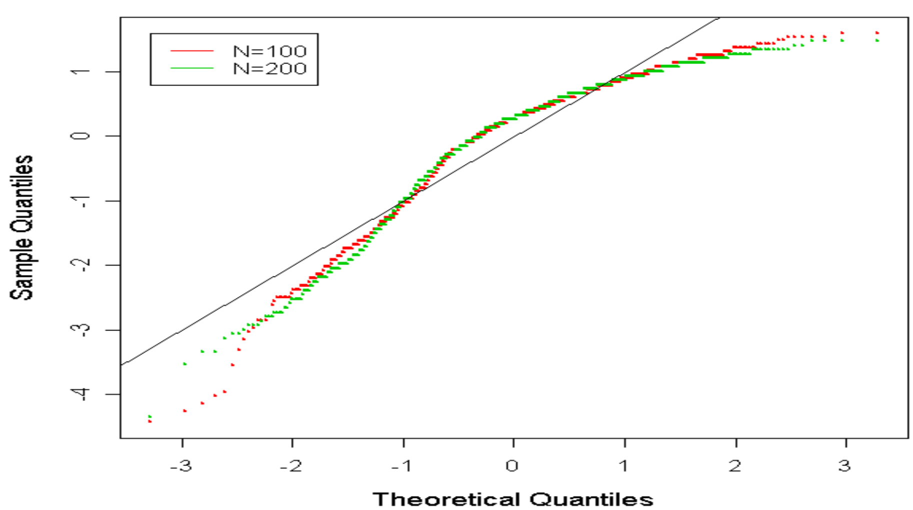


$\phi_0 = -0.6,\ d_0 = 0$

Figure 5.8: Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the AR parameter obtained by 1000 replications of ARFIMA(1,d,0) series with length N=100, 200.

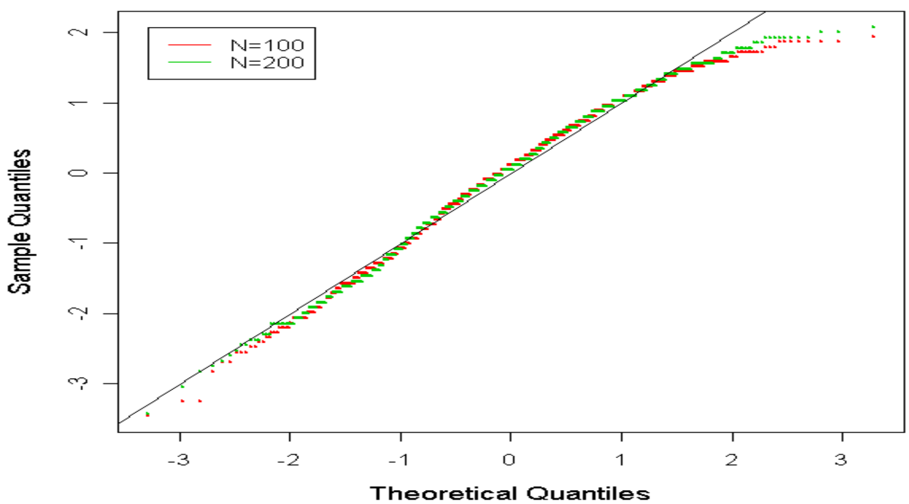


$\phi_0 = -0.6,\ d_0 = 0.3$

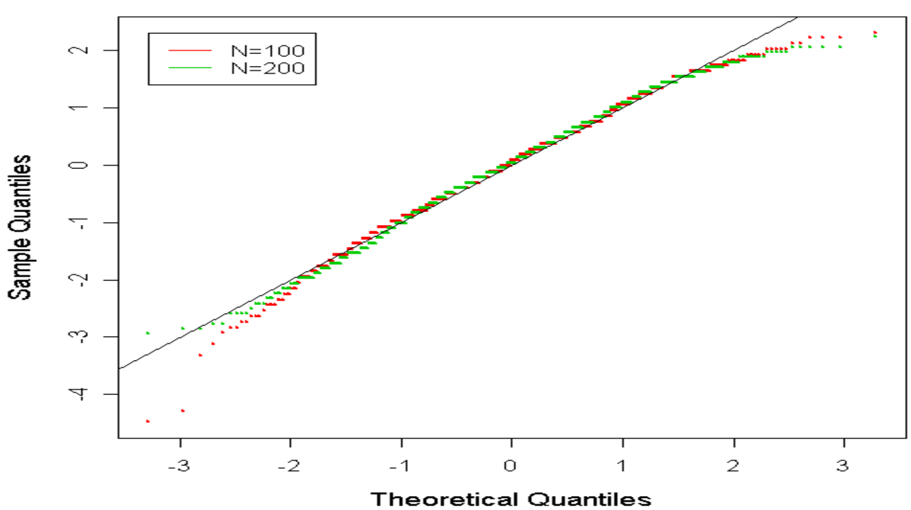


$\phi_0 = 0.6,\ d_0 = -0.3$

Figure 5.8 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the AR parameter obtained by 1000 replications of ARFIMA(1,d,0) series with length N= 100, 200.

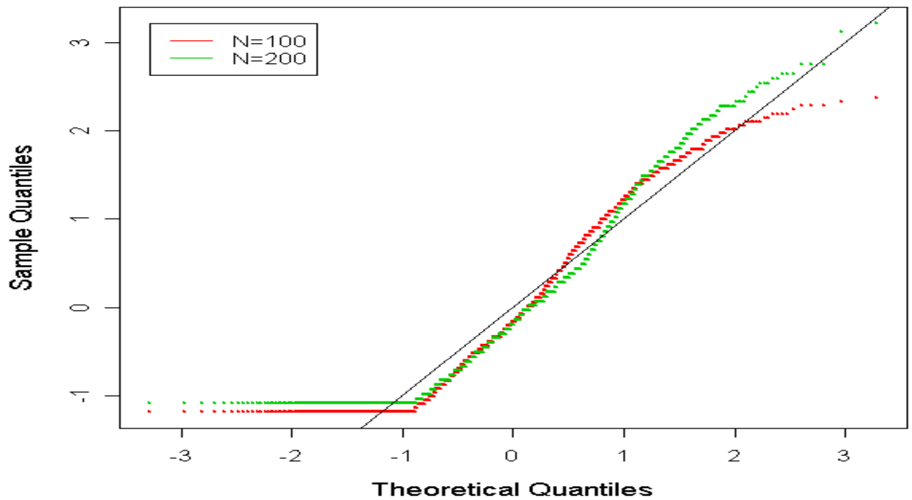


$\phi_0 = 0.6,\ d_0 = 0$

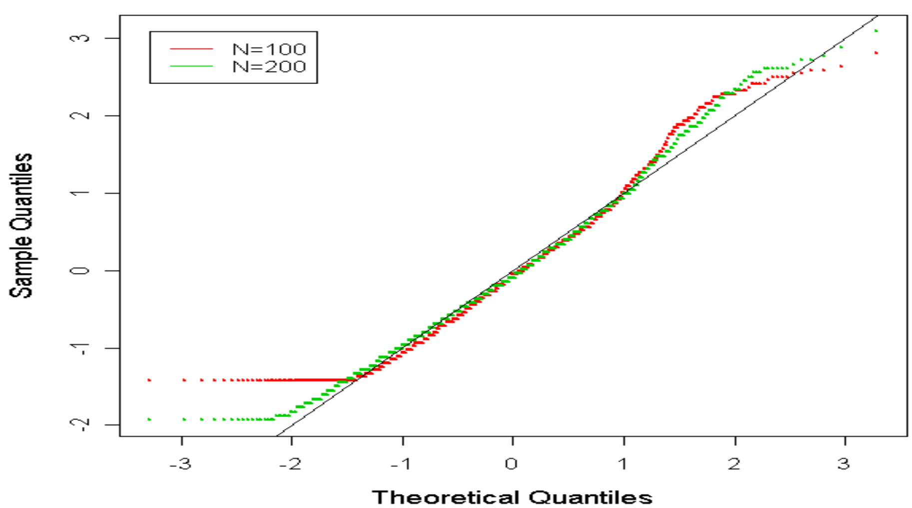


$\phi_0 = 0.6,\ d_0 = 0.3$

Figure 5.8 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the AR parameter obtained by 1000 replications of ARFIMA(1,d,0) series with length N= 100, 200.

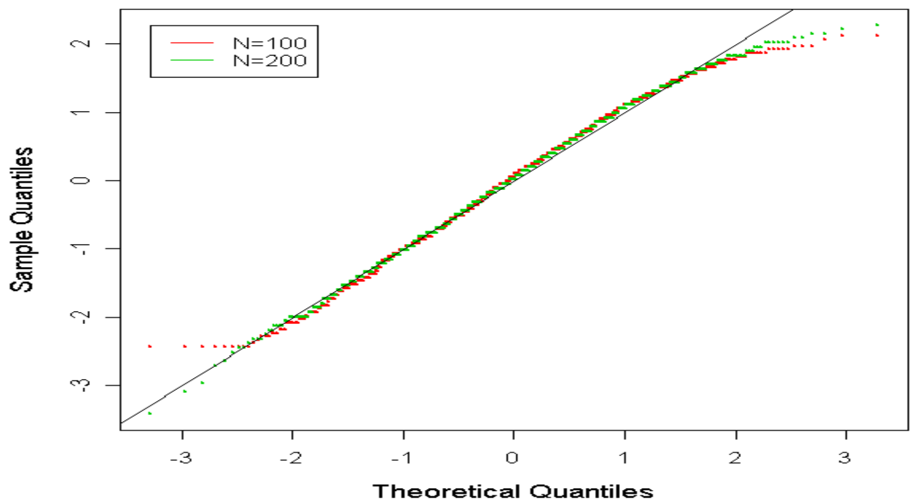


$\xi_0 = -0.6,\ d_0 = -0.3$

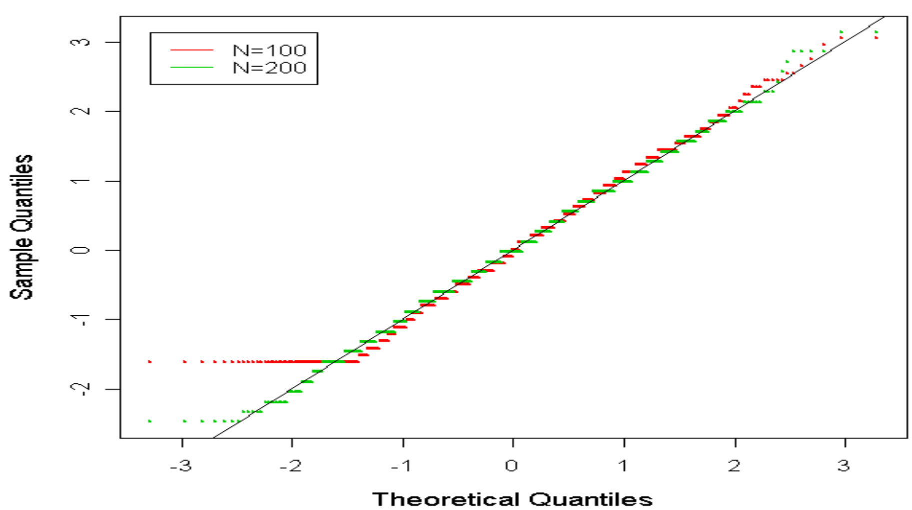


$\xi_0 = -0.6,\ d_0 = 0$

Figure 5.9: Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.

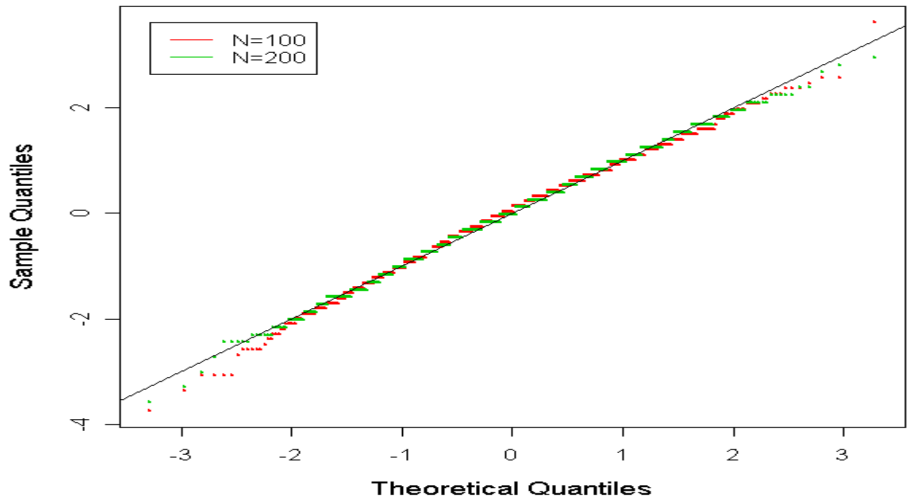


$\xi_0 = -0.6, \ d_0 = 0.3$

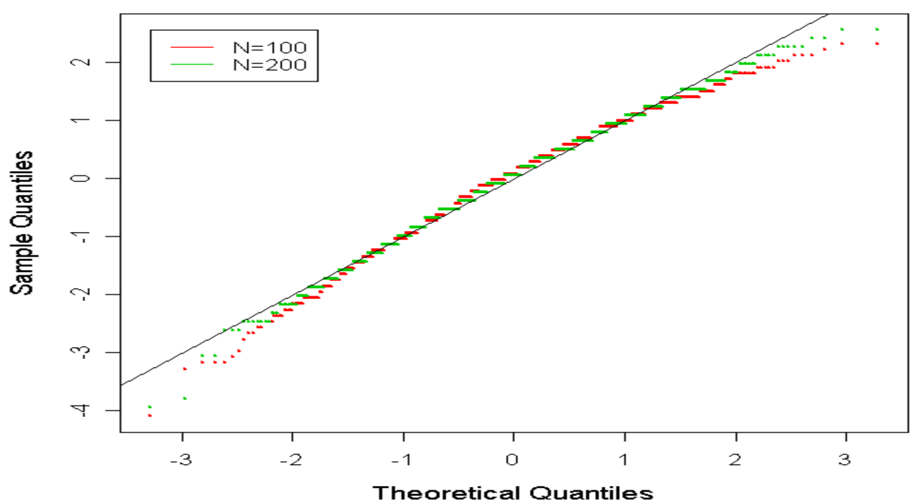


$\xi_0 = 0.6, \ d_0 = -0.3$

Figure 5.9 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.

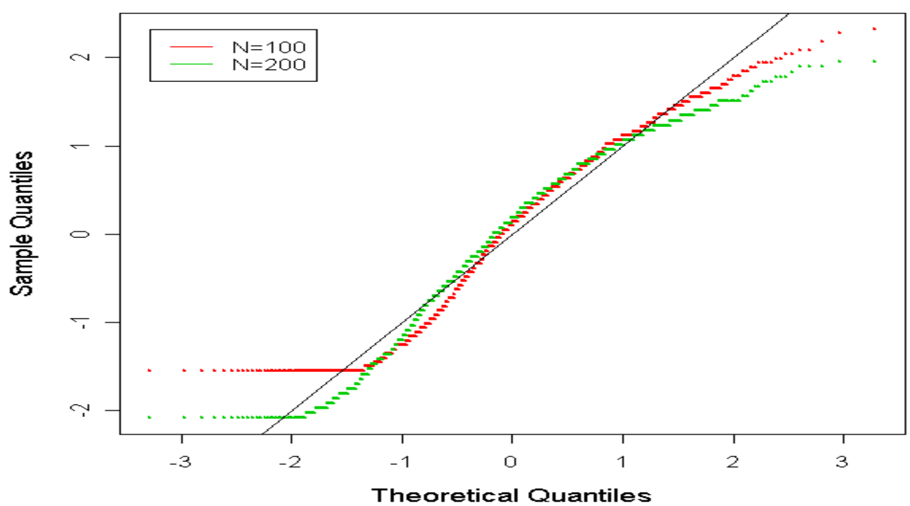


$\xi_0 = 0.6, \; d_0 = 0$



$\xi_0 = 0.6, \; d_0 = 0.3$

Figure 5.9 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of d obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.



$\xi_0 = -0.6, \; d_0 = -0.3$

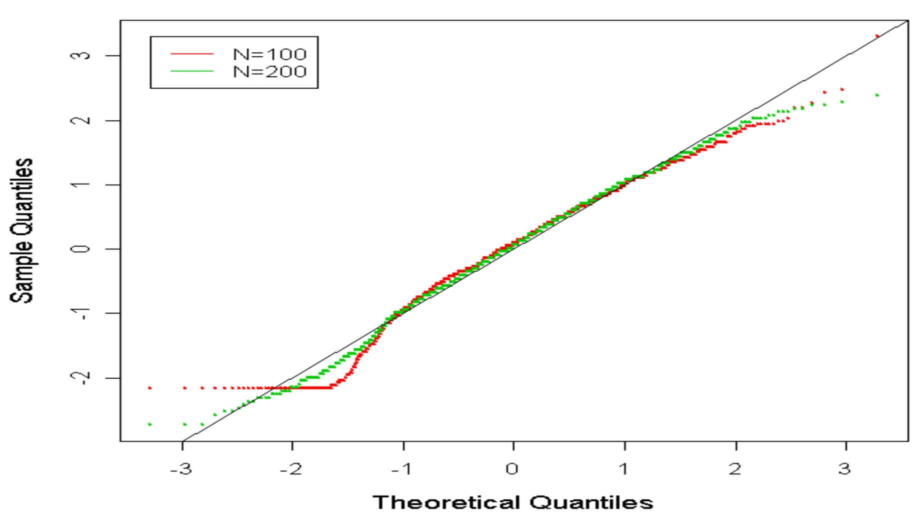
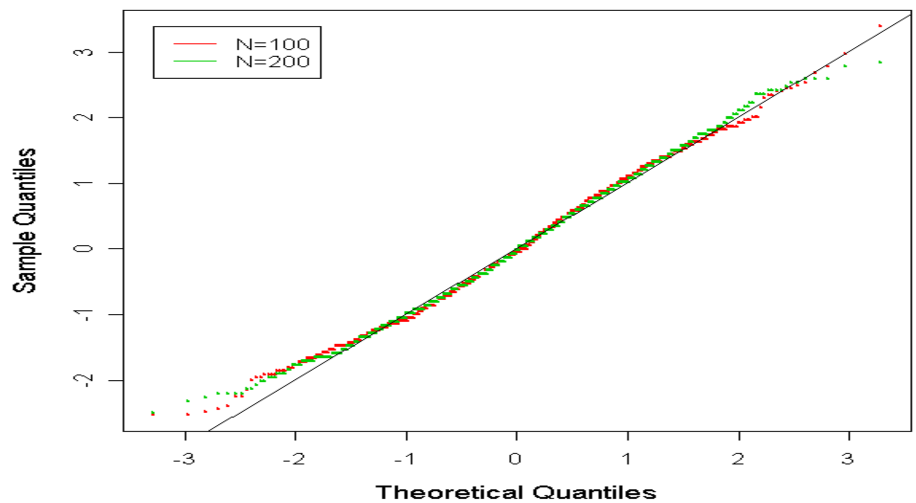
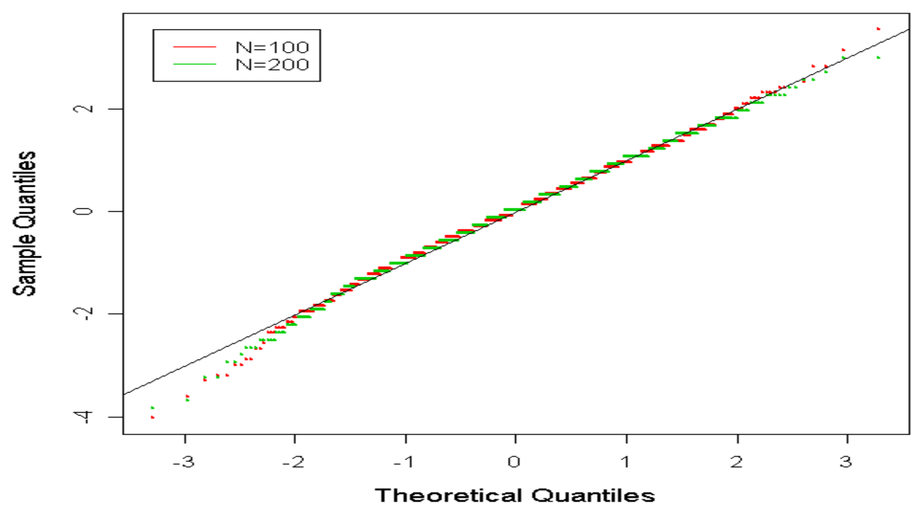


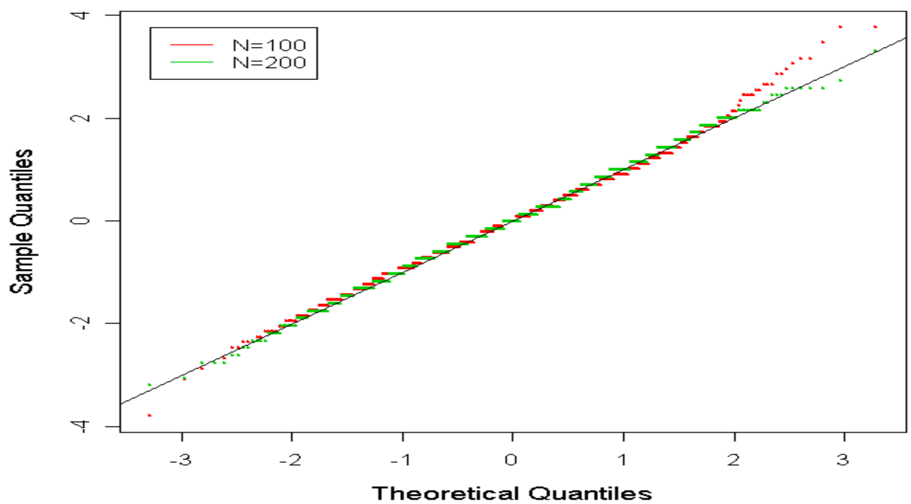
Figure 5.10: Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the MA parameter obtained by 1000 replications of ARFIMA(0,d,1) series with length N=100, 200.



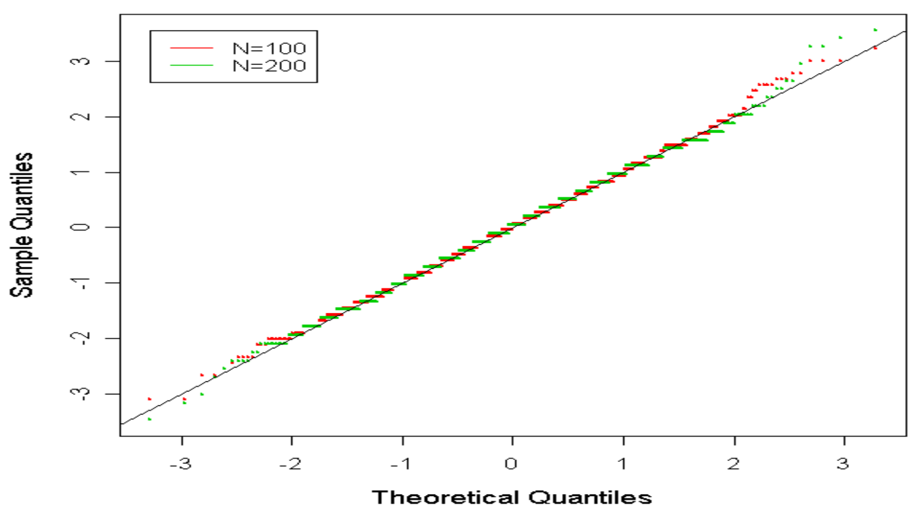


$\xi_0 = 0.6, \; d_0 = -0.3$

Figure 5.10 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the MA parameter obtained by 1000 replications of ARFIMA(0,d,1) series with length N= 100, 200.



$\xi_0 = 0.6,\ d_0 = 0$



$\xi_0 = 0.6,\ d_0 = 0.3$

Figure 5.10 (continued): Normal Q-Q plots of the sample mean plug-in Gaussian MLE of the MA parameter obtained by 1000 replications of ARFIMA(0,d,1) series with length N= 100, 200.

Chapter 6

SUMMARY

In this work the exact and the plug-in Gaussian MLE of a stationary and invertible time series were proved to be consistent, asymptotically normal and efficient. We showed that the same first-order asymptotic properties of the Gaussian MLE that were established for time series with short-memory (Hannan 1973) or with long-memory (Dahlhaus 1989, 2005), still hold when the parameter space is expanded so that the true memory parameter may take values in the range that includes both cases of long-memory, short-memory and anti-persistent time series.

The contribution of this work was enabled mainly through an extension of Dahlhaus's (1989) proof of consistentcy along the following lines. Using an extended uniform version of Dahlhaus's (1989) Theorem 5.1 on the asymptotic behaviour of products of Toeplitz and inverse-Toeplitz matrices due to Lieberman, Rousseau and Zucker (2003) and Lieberman and Phillips (2004), the normalized Gaussian log-likelihood was shown to converge uniformly to a known, finite limit function that is uniquely maximized at the true parameter θ_0 in any compact subset $\overline{\Theta}$ of the parameter space Θ that contains θ_0, as long as $\overline{\Theta}$ is such that for all $\theta \in \overline{\Theta}$ the memory parameter, $\alpha(\theta)$, satisfies $\alpha(\theta) \geq \alpha(\theta_0) - 1$, where $\alpha(\theta_0)$ is the true memory parameter of the series. On the other hand, when $\alpha(\theta) < \alpha(\theta_0) - 1$, we used a similar argument to that of Velasco and Robinson (2000, Theorem 1) in the context of a discrete Whittle MLE and showed that the Gaussian likelihood function converges to zero w.p. 1 (in fact, it was shown that the normalized minus log-likelihood is asymptotically larger than any constant w.p. 1).

A simulation study was conducted to test the Gaussian MLE performance in finite samples, whose results seems to conform with the asymptotic theory.

There is much room for future research on the Gaussian MLE of long memory time series. We point out some directions left open for further research in this area:

- Our results on the exact Gaussian MLE are restricted to time series with memory parameter $-1 < \alpha < 1$. For instance, when dealing with a nonstationary Gaussian time series with a memory parameter $\alpha \geq 1$, it would be necessary to difference the series until getting stationarity of the differenced series. This may be done by exploratory way or by more formal tests (see, for instance, Dickey and Fuller 1979). Recall that α represents $2d$ where d is a differencing order required to get a short memory series as in the fractional ARIMA models (see Section 2.5.2), so differencing the series n-times would result in a series with memory parameter $\alpha - 2n$. Thus, unless the memory parameter is exactly an odd integer, differencing the series a proper number of times would yield a series with memory parameter within the open interval $(-1, 1)$, for which we may apply our exact Gaussian MLE. A similar approach may be used in the case of noninvertibility where the memory parameter satisfies $\alpha \leq -1$. Here, we would use partial sums of the original series instead of differences. However, so far there is no theory that seems to handle the nonstationarity or noninvertibility cases by means of the Gaussian MLE. A possible approach to extend our result to nonstationary time series may be to use the Autoregressive-based exact Gaussian MLE along the lines of Beran's (1995) Autoregressive-based pseudo MLE (see Section 3.3.2).

- While our results apply only for univariate Gaussian time series, it may be worthwhile to consider relaxing the Gaussianity assumption as well as extending the results to the multivariate case. While both generalizations seem rather complicated, a possible starting point toward achieving them, might be considering similar generalizations that already exist in the literature for the Whittle approximate MLE (see Section 3.3.3) such as Giraitis and Surgailis (1990), Heyde and Gay (1993) and Hosoya (1997).

- It is required is to develop a better understanding of the behaviour of the exact and plug-in Gaussian MLE in small samples sizes and different parameter regions. Such theories may assist in improving the Gaussian MLE accuracy, and particularly reducing its relatively large bias. Some recent works on higher order expansions such as Lieberman, Rousseau and Zucker (2003), Lieberman (2005) and Lieberman and

Phillips (2005) made a progress toward this goal. Bearing in mind that the exact Gaussian MLE is a relatively efficient method in the case where the true mean of the process is known, developing a rigorous procedure that reduce the bias of the exact Gaussian MLE even when the mean of the process is unknown may yield some more practicable and competitive MLE-based procedures of estimation.

- As explained above, whenever $\alpha(\theta) \geq \alpha(\theta_0) - 1$, we may establish a uniform limit of the normalized log-likelihood function, while in the cases in which

$$\alpha(\theta) < \alpha(\theta_0) - 1, \tag{6.1}$$

our knowledge about the behaviour of the log-likelihood function is much more limited. The main effort in establishing a limit theory for the (exact or Whittle) log-likelihood function is put into developing the limit distribution of the quadratic forms

$$\mathbf{X}'\Sigma_N(f_\theta)\mathbf{X} \tag{6.2}$$

and

$$\mathbf{X}'\Sigma_N(f_\theta)^{-1}\mathbf{X}, \tag{6.3}$$

where $\mathbf{X}$ here represents a zero-mean Gaussian stationary time series with true parameter value θ_0, and $\Sigma_N(f_\theta)$ is the covariance matrix evaluated at parameter θ. While in this work, we heavily relied on some limit theorems for traces of products Toeplitz and inverse Toeplitz matrices, and particularly on Theorem (4.3.1), there are also other approaches to establish asymptotic properties of the quadratic forms (see, for instance, Avram and Taqqu 2005). Unfortunately, so far none of these methods seem to provide a full description of the limit distribution of the quadratic forms (6.2) or (6.3) in the case of (6.1). A significant progress in this direction was made by Terrin and Taqqu (1990) who handle the quadratic form (6.2) under (6.1). They showed that in this case, under proper normalization and by subtracting its mean, (6.2) converges weakly to a non-Gaussian self similar process that can be viewed as a generalization of the Rosenblatt process. However still, the asymptotic mean of (6.2), as well as the limit distribution of (6.3), remained unclear.

Finally, and in a broader view, the general research goals in this area are to provide researchers with efficient and simple tools to test, estimate and make predictions in long memory time series. As seen in Chapter 3, many different approaches were developed and proposed, but it seems that there is still much more to be done in filling some left-open gaps, refining existing methods and gaining a better understanding of the pros and cons of the different approaches.

BIBLIOGRAPHY

[1] Adenstedt, R.K. (1974). On large-sample estimation for the mean of a stationary random sequence, Ann. of Math. Stat., 2, 1095-1107.

[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd Int. Symp. Inf. Theory (B. N. Petrov and F. Czáki, eds), Budapest: Akad. Kiadó, 267-281.

[3] An, S. and Bloomfield, P. (1993). Cox and Reid's modification in regression models with correlated errors, Technical Report, North Carolina State University.

[4] Avram, F. (1988). On bilinear forms in Gaussian random variables and Toeplitz matrices, Probab. Theory Related Fields, 79, 37–45.

[5] Avram, F. and Taqqu, M. (2005). Applications of a generalized Brascamp-Lieb-Barthe inequality and a Generalized Szego theorem to the asymptotic theory of sums, integrals and quadratic forms, within "Dependence in Probability and Statistics", edited by Bertail, P., Doukhan, P. and Soulier, P., Springer, New York.

[6] Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics, J. Econometrics, Vol. 73, 5-59.

[7] Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, J. Econometrics, 74, 3-30.

[8] Beran, J. (1991). M-estimator of location for data with slowly decaying serial correlations, J. Amer. Statist. Assoc. 86, 704-708.

[9] Beran, J. (1992). Statistical methods for data with long-range dependence, Statistical Science, 7, 404–427.

[10] Beran, J. (1994a). On a class of M-estimators for long-memory Gaussian models, Biometrica., 81, 755-766.

[11] Beran, J. (1994b). Statistics for long-memeory processes, Chapman & Hall, New-York.

[12] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARIMA models, JRSS, Vol. B, 57, 659-672.

[13] Beran, J., Bhansali, R.J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, Biometrika, 85, 921-934.

[14] Bollerslev, T. (1986). Generalized autoregressive conditional beteroskedasticity, J. Econometrics, 31, 307-327.

[15] Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis, Forecasting and Control, Holden-Day, San Fransisco.

[16] Brockwell, P.J. and Davis, R.A. (1991). Time Series: Theory and Methods (2nd ed.), New York: Springer-Verlag.

[17] Chan, N.H. and Palma, W. (2006). Estimation of long-memory time series models: A survey of likelihood-based methods, Advances in Econometrics, 20, 89-121.

[18] Cheung, Y.W. and Diebold, F.X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean, J. Econometrics, 62, 301-316.

[19] Chung, C.F. and Schmidt, P. (1995). The minimum distance estimator for fractionally integrated ARMA processes, Econometrics and Economic Theory Papers, 9408, Michigan State University.

[20] Cox, D.R. (1977). Contribution to discussion of paper by A.J. Lawrance and N.T. Kottegoda, J. the Royal Stat. Socirty, Ser. A, 140, 34.

[21] Cox, D R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, J. the Royal Stat. Society, Ser. B, 49, 1-39.

[22] Crato, N. and Ray, B.K. (1996). Model selection and forecasting for long-range dependent processes, J. Forecasting, 15, 107-125.

[23] Dacorogna M.M., Muller, U.A., Nagler R.J., Olsen R.B. and Pictet O.V., (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market, J. Intern. Money and Finance, 12, 413-438.

[24] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes, Ann. Statist, 17, 1749-1766.

[25] Dahlhaus, R. (2005). Extended correction note for "Efficient parameter estimation for self-similar processes", Universität Heidelberg, Available at math.uni-heidelberg.de/stat/people/dahlhaus/Publications.html.

[26] Dattorro J. (2005). Convex Optimization & Euclidean Distance Geometry, Meboo Publishing, USA.

[27] Davies, R.B. (1973). Asymptotic inference in stationary Gaussian time series, Adv. in Appl. Probab, 5, 469-497.

[28] Davies, R.B. and Harte, D.S. (1987). Tests for Hurst effect, Biometrika 74, 95–101.

[29] Ding, Z., Granger, C.W.J. and Engle R.F. (1993). A long memory property of stock market returns and a new model, J. Empirical Finance, 1, 83–106.

[30] Doornik, J.A. and Ooms M. (1999). A package for estimating, forecasting and simulating ARFIMA models: Arfima package 1.0 for Ox. Discussion paper, Nuffield College, Oxford.

[31] Doornik, J.A. and Ooms M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models, Comput. Stat. and Data Anal., 42, 333–348.

[32] Durbin, J. (1960). The fitting of time series models, Int. Statist. Rev., 28, 233-244.

[33] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica, 50, 987-1008.

[34] Fox, R. and Taqqu, M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, Ann. Statist, 14, 157-532.

[35] Fox, R. and Taqqu, M.S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence, Probab. Theory Related Fields, 74, 213-240.

[36] Geweke, J. and Porter-Hudak S. (1983). The estimation and application of long memory time series models, J. Time Ser. Anal., 4, 221-238.

[37] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, Prob. Theory Related Fields, 86, 87-104.

[38] Giraitis, L. and Robinson, P.M. (2003). Parametric Estimation Under Long Range Dependence, Theory and Applications of Long Range Dependence, Birkhauser, 229-249.

[39] Golub, G.H. and Van Loan, C.F. (1996). Matrix Computations, 3rd ed. Baltimore, MD: Johns Hopkins.

[40] Gradshteyn, I.S. and Ryzhik, I.M. (1965). Tables of Integrals, Series and Products, Academic Press, New York.

[41] Granger, C.W.J. (1966). The typical spectral shape of an economic variable, Econometrica, 34, 150-161.

[42] Granger, C.W.J. (1980). Long Memory relationships and the aggregation of dinamic models, J. Econometrics, 14, 227-223.

[43] Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing, J. Time Ser. Anal., 1, 15-31.

[44] Gray, R.M. (2002). Toeplitz and circulant matrices: A review, Available at http://ee.stanford.edu/~gray/toeplitz.pdf.

[45] Grenander, U. and Szegö, G. (1958). Toeplitz forms and their applications, University of Calif. Press, Berkeley and Los Angeles.

[46] Guégan, D. (2005). How can we define the concept of long memory? an econometric survey, Econometric Rev., 24, 113–149.

[47] Hall, P., Koul, H.L. and Turlach, B.A. (1997). Note on convergence rates of semiparametric estimators of dependence index, Ann. Stat., 25, 1725-1739.

[48] Hampel, F. (1987). Data analysis and self-similar processes, Bull. 46th Session of the ISI, Tokyo,. Vol. LII, Book 4, 235–254.

[49] Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.

[50] Hannan, E.J. (1970). Multiple Time Series, Wiley, New York.

[51] Hannan, E.J. (1973). The asymptotic theory of linear time series models, J. Appl. Probab., 10, 130-145.

[52] Hannan, E.J. (1980). The estimation of the order of an ARMA process, Ann. Statist., 8, 1071-1081.

[53] Haslett, J. and Raftery, A.E. (1989). Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource (with discussion), J. Appl. Stat., 38, 1-50.

[54] Hauser, M.A. (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study, J. Stat. Planning and Inference, 80, 229-255.

[55] Heyde C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence, Stochastic Process. Appl., 45, 169–182.

[56] Hosking, J.R.M. (1981). Fractional differencing, Biometrika, 1981, 68, 165-176.

[57] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models, Ann. Stat., 25, 105–137.

[58] Huber, P.J. (1981). Robust statistics, Wiley, New York.

[59] Hunt, G.A. (1951). Random Fourier transforms, Trans. Amer. Math. Soc., 71, 38-69.

[60] Hurst, H.E. (1951). Long-term storage capacity of reservoirs, Trans. Am. Soc. Civil Engineers, 116, 770-799.

[61] Hurst, H.E. (1956). Methods of using long term storage in reservoirs, Proc. the Inst. of Civil Eng., I, 519-543.

[62] Hurvich, C.M. and Beltrao, K.I. (1993). Asymptotics for the low frequency ordinates of the periodogram of a long memory time series, J. Time Ser. Anal., 14, 455-472.

[63] Hurvich, C.M. and Deo, R.S. (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series, J. Time Ser. Anal., 20, 331-341.

[64] Hurvich, C.M., Deo, R.S. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series, J. Time Ser. Anal., 19, 19-46.

[65] Hurvich, C.M. and Ray, B.K. (1995). Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes, J. Time Ser. Anal., 16, 17-41.

[66] Kim, C.S. and Phillips, P.C.B. (2006). Log-periodogram Regression: the Nonstationary Case, Cowles Foundation Discussion Papers, 1587, Cowles Foundation, Yale University.

[67] Kolmogorov, A.N. (1940). Wienersch Spiralen und einige andere interessante Kurven in Hlbertschen Raum, Comptes Rendus (Doklady) Académie des Sciences de l'URSS (N.S.), 26, 115-118.

[68] Künsch, H.R. (1986). Discrimination between long range dependence and monotonic trends, J. Applied Probability, 23, 1025–1030.

[69] Künsch, H.R. (1987). Statistical aspects of self-similar processes, In Proceedings of the First World Congress of the Bernoulli Society (Yu. Prokhorov and V. V. Sazanov, eds.), 1, 67-74, VNU Science Press, Utrecht.

[70] Lamperti, J. (1962). Semi-stable stochastic processes, Trans. Amer. Math. Soc., 104, 62-78.

[71] Lieberman, O. (2005). On plug-in estimation of long memory models, Econometric Theory, 21, 2, 431–454.

[72] Lieberman, O. and Phillips, P.C.B. (2004). Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra,. J. Time Ser. Anal., 25, 733-753.

[73] Lieberman, O. and Phillips, P.C.B. (2005). Expansions for approximate maximum likelihood estimators of the fractional difference parameter, The Econometrics J., 8, 367-379, 2005.

[74] Lieberman, O., Rousseau, J. and Zucker, D. (2003). Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. The Ann. Stat., 31, 586-612.

[75] Lo, A.W. (1991). Long term memory in stock market prices, Econometrica, 59, 1279–1313.

[76] Lobato, I. (1999). A semiparametric two-step estimator for a multivariate long memory process, J. Econometrics, 90, 129-153.

[77] Mandelbrot, B.B. (1971). A fast fractional Gaussian noise generator, Wat. Resour. Res. 7, 543-553.

[78] Mandelbrot, B.B. (1975). Limit theorems of the self-normalized range for weakly and strongly dependent processes, Z. Wahr. verw. Geb., 31, 271-285.

[79] Mandelbrot, B.B. (1983). The Fractal Geometry of Nature, W.H. Freeman, New York.

[80] Mandelbrot, B.B. (1997). Fractals and Scaling in Finance: Discontinuity, Concentration, Risk, Springer-Verlag, New York.

[81] Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications, SIAM Review, 10, 422-37.

[82] Mandelbrot, B.B. and Walliss, J.R. (1968). Noah, Joseph and operational hydrology, Water Resources Research, 4, 909-918.

[83] Mandelbrot, B.B. and Walliss, J.R. (1969a). Computer experiments with fractional Gaussian noises, Parts 1,2,3, Water Resources Research, 5, 228-267.

[84] Mandelbrot, B.B. and Walliss, J.R. (1969b). Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence, Water Resources Research, 5, 967-988.

[85] Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result, American J. Math, 8, 343-366.

[86] Nielsen, M.and Frederiksen, P.H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration, Econometric Reviews, 24, 405-443.

[87] Parzen, E. (1974). Some recent advances in time series modelling, IEEE Trans. Automatic Control, AC-19, 723-730.

[88] Pearson, K. (1902). On the mathematical theory of errors of judgement, with special reference to personal equation, Phil. Trans. R. Soc. Lond., Ser A, 198, 235- 299.

[89] Phillips, P.C.B. (1999). Discrete Fourier Transforms of Fractional Processes, Cowles Foundation Discussion Papers 1243, Cowles Foundation, Yale University.

[90] Phillips, P.C.B. and Shimotsu, K. (2001). Local Whittle estimation in nonstationary and unit root cases, The Ann. Stat., 34, 656-692.

[91] Press, S.A., Teukolsky, W.H., Vetterling, W.T. and B. P. Flannery (1992). Numerical recipes in FORTRAN, Cambridge University Press, Cambridge, 2nd ed.

[92] Priestly, M.B. (1981). Spectral Analysis and Time Series, Academic Press, New York.

[93] Robinson, P.M. (1978). Statistical inference for a random coefficient autoregressive model. Scandinavian J. stat., 5, 163-168.

[94] Robinson, P.M. (1994). Time series with strong dependence. In Advances in Econometrics: Sixth World Congress, C. A. Sims ed., 1, 47–95.

[95] Robinson, P.M. (1995a). Log-periodogram regression of time series with long range dependence, Ann. Stat.., 23, 1048-1072.

[96] Robinson, P.M. (1995b). Gaussian semiparametric estimation of long-range dependence, The Ann. Stat., 23, 1630–1661.

[97] Robinson, P.M. (2003). Long memory time series, within "Time series with long memory", edited by Robinson, P.M., Oxford University Press, Oxford.

[98] Rosenblatt, M. (1961). Independence and dependence, Proc. Fourth Berkeley Symp. Math. Statist. Probab, 431-443, Univ. California Press, Berkeley.

[99] Samarov, A. and Taqqu, M.S. (1988). On the efficiency of the sample mean in tong-memory noise, J. Time Ser. Anal., 9, 191-200.

[100] Samorodnitsky, G. and Taqqu, M.S. (1994). Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman-Hall, New York.

[101] Schwarz, G. (1978). Estimating the dimension of a model. The Ann. Stat., 6, 461–464.

[102] Scott, D. W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.

[103] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, Ann. Statist., 8, 147-164.

[104] Shimotsu, K. (2006). Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend, Working Papers 1061, Queen's University, Department of Economics.

[105] Shimotsu, K. and Phillips, P.C.B. (2004). Exact Local Whittle Estimation of Fractional Integration, The Ann. Stat., 33, 1890-1933.

[106] Silverman, B. W. (1986). Density Estimation, Chapman and Hall, London.

[107] Sinai, Y.G. (1976). Self-similar probability distributions, Theory of Probability and its applications, 21, 64-80.

[108] Smith, A.A.J., Sowell, F.B. and Zin, S.E. (1997). Fractional integration with drift: Estimation in small samples, Empirical Economics, 22, 103-116.

[109] Sowell, F.B. (1989). A decomposition of block Toeplitz matrices with applications to vector time series, manuscript, Carnegie Mellon University.

[110] Sowell, F.B. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models, J. Econometrics, 53, 165-188.

[111] Student, (1927). Errors of routine analysis. Biometrika, 19, 151-164.

[112] Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process, Z. Wahrsch., 31, 287-302.

[113] Taqqu, M.S. (1986). A bibliographical guide to self-similar processes and long-range dependence. E. Eberlein and M. S. Taqqu, eds. Dependence in Probability and Statistics, Boston: Birkhauser, 137–162.

[114] Taqqu, M.S., Teverovski, V. and Willinger, W. (1995). Estimators for long-range dependence: an empirical study, Fractals, 3 ,785-798.

[115] Terrin, N. and Taqqu, M.S. (1990). A noncentral limit theorem for quadratic forms of Gaussian stationary sequences, J. Theor. Prob., 3, 449-475.

[116] Tieslau, M., Schmidt, P. and Baillie R. (1996). A minimum distance estimator for long-memory processes, J. Econometrics, 71, 249-264.

[117] Velasco, C. (1999a). Non-stationary log-periodogram regression. J. Econometrics, Elsevier, 91, 325-371.

[118] Velasco, C. (1999b). Gaussian semiparametric estimation of nonstationary time series, J. Time Ser. Anal., 20, 87-127.

[119] Velasco, C. (2000). Non-Gaussian Log-Periodogram Regression, Econometric Theory, Cambridge University Press, 16, 44-79.

[120] Velasco, C. and Robinson P.M. (2000). Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series, J. the Amer. Stat. Assoc., 95, 1229-1243.

[121] Vervaat, W. (1985). Sample path properties of self-similar processes with stationary increments, Ann. Probab., 13, 1-27.

[122] Vitale, R.A. (1973). An asymptotically efficient estimate in time series analysis, Quart. Appl. Math., 30, 421-440.

[123] Whistler, D. (1990). Semiparametric Models of Daily and Intra-Daily Exchange Rate Volatility. Ph.D. thesis, University of London.

[124] Whittle, P. (1951). Hypothesis Testing in Time Series Analysis, Hafner, New York.

[125] Whittle, P. (1953). Estimation and information in stationary time series, Ark. Mat., 2, 423-434.

[126] Yajima, Y. (1985). On estimation of long-memory time series models, Aust. J. Statist., 27, 303-20.

[127] Yajima, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes, J. Time Ser. Anal., 10, 375-383.

[128] Yong, C. H. (1974). Asymptotic Behaviour of Trigonometric Series, Chinese Univ. Hong Kong, Hong Kong.

[129] Zygmund, A. (2002). Trigonometric series (3rd ed.), Cambridge Univ. Press, London.

תקציר

לסידרה עתית סטציונרית והפיכה עם פונקציה ספקטרלית המקיימת $f(\omega) \sim |\omega|^{-\alpha} L(\omega)$ כש- $\omega \to 0$, כאשר $|\alpha| < 1$ ו- $L(\omega)$ חיובית ומשתנה לאט ב- $\omega = 0$, אנו אומרים שיש זכרון ארוך אם $0 < \alpha < 1$, זכרון קצר אם $\alpha = 0$ וזכרון שלילי אם $-1 < \alpha < 0$. הוכח כי אומד ניראות מקסימלי של סידרה עתית גאוסיאנית הוא עקיב ובעל התפלגות אסימפטוטית נורמלית על ידי Hannan (1973) למקרה של זכרון קצר ועל ידי Dahlhaus (1989) למקרה של זכרון ארוך. מטרתה העיקרית של עבודה זו היא להכליל את התוצאות הללו למקרים של זכרון ארוך, קצר או שלילי, ללא ידע מוקדם על הזכרון של הסדרה העתית. אנחנו מאמצים את טכניקת ההוכחה של Dahlhaus (1989) המסתמכת בעיקרה על ההתנהגות האסימפטוטית של מטריצות Toeplitz, אולם רבים מהטיעונים של Dahlhaus מוכללים ומפושטים. ישימותן של התוצאות עבור סדרות fractional Gaussian noise ו-fractional ARMA מוכחת, וכן מודגמים הביצועים של האומדנים באמצעות סימולציות של סדרות fractional ARMA.

אוניברסיטת תל-אביב
הפקולטה למדעים מדויקים
ע״ש ריימונד ובברלי סאקלר

אמידת נראות מקסימלית לתהליכים גאוסיינים פרקציונלים סטציונרים והפיכים

חיבור זה הוגש כחלק מהדרישות לקבלת התואר
״מוסמך אוניברסיטה״ - M.Sc. באוניברסיטת תל-אביב

ביה״ס למדעי המתמטיקה
החוג לסטטיסטיקה והסתברות

על ידי
רועי רוזמרין

העבודה הוכנה בהדרכתם של
פרופסור עופר ליברמן
פרופסור יצחק מלכסון

אב תשס״ח