ABSTRACT

Title of dissertation: STATISTICAL ANALYSIS OF EYE GAZE DATA

Ritaja Sur, Doctor of Philosophy, 2010

Dissertation directed by: Professor Benjamin Kedem
Mathematical Statistics Program
Department of Mathematics

In this dissertation, we present analysis of eye gaze data in response to simple one-arm movements. We report the presence of long memory property in the eye gaze. Using this property, we model the eye gaze data. The best model involves present arm coordinates as well as lagged eye gaze and arm coordinates. Further analysis for classification of eye gaze under two conditions "Watch" and "Imitate" is presented. This analysis uses time domain and spectral domain methods. In the time domain, properties of higher order crossing (HOC) sequence is investigated for long memory processes. Further, we present its application to eye gaze data. The application reveals differences in the two conditions for some of the subjects. In the spectral domain, logarithmic ratio of spectral densities are assumed to follow either the exponential (EXP) or fractional exponential (FEXP) model form. By applying assumptions on the distribution of periodogram ordinates of long memory processes and using FEXP model, classification of long memory processes are studied using simulations. The results are compared when EXP models are also used. Both techniques are then applied for classification of eye gaze data.

STATISTICAL ANALYSIS OF EYE GAZE DATA

by

Ritaja Sur

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Dr. Benjamin Kedem, Chair
Dr. Paul Smith
Dr. Doron Levy
Dr. Myron Katzoff
Dr. Michel Wedel, Dean's Representative.

Acknowledgments

I would like to thank all whose support made this dissertation possible. I would like to first thank my advisor, Dr. Benjamin Kedem for his support. This dissertation would not have been possible without his constant guidance and encouragement. His advice and suggestions in all aspects of this dissertation have been invaluable to me. For the entire period of my stay in University of Maryland, he has not only helped me in understanding various research topics but also helped me become a better person.

I would like to thank Dr. Lior Noy and Dr. Tamar Flash for giving the eye gaze data and providing us with challenging problems. All our interactions have been a great help for me to understand this new topic. The comments provided and the questions raised in the methodology have always helped me think and write better.

I would also like to thank all my committee members. Special thanks to Dr. Smith for all your advice. From graduate courseworks, TA work to Job search, your advice has helped me in my endeavours year after year. I would like to thank Dr. Myron Katzoff, Dr. Michel Wedel and Dr. Doron Levy for your valuable comments and for agreeing to serve on my committee.

Special thanks to Dr. Konstantinos Fokianos for your help and for providing valuable papers, references and codes.

I would also like to thank my friends in my department. Thank you Anastasia, Carolina and Katya for making my graduate life a valuable experience. Thanks to

Kostas and Ziliang for being great friends and providing me guidance at all times. I would like to thank all my friends in US, UMD and also my friends in India. Special thanks to Shraddha Guzraty for your support all through the years.

I would like to thank my family specially my father Amitava Sur and my father-in-law, Suprakas Datta. I would like to thank all my family in India for their love and help. Special thanks to my husband Dr. Supratik Datta. You are a wonderful person, a caring and understanding husband. I have learnt a lot from you and have much more left. I could pass this phase of graduate life just because of you. You are my world. Thank you for being there for me all the time and giving me the support and encouragement to complete this dissertation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Studying human imitation is important for the understanding the mechanism behind how humans learn. One of the important questions that has been under investigation in this area is whether there is any difference in human behaviour when humans are asked to imitate given a stimulus versus when they just watched the stimulus without the requirement to imitate them. One of the possible paths to address this question, is through the study of eye gaze movements. Studies of eye movements during imitation have been performed with various hand/finger movement stimuli and it has been shown that given a movement stimulus, the pupil dilation is larger for a person who imitates it [37], [24]. In another study with arm movements as stimuli it has been observed that the average fixation duration of the eye was longer in imitation [26].

In this dissertation, we consider the same question of discriminating between eye movements under two conditions, "Watch" and "Imitate". We approach the problem from two directions, one based in time domain and other based in spectral domain. The time domain method is based on using counts of *visual* features referred to as (the number of) *higher order crossings* (HOC). The precise definition of higher order crossings and some of their properties will be addressed in Chapter 4. It is worth mentioning that these counts summarize significant spectral or oscil-

latory information which has been found useful in a number of applications where other methods proved somewhat problematic. Successful applications of higher order crossings include detection of brain injury in piglets using EEG signals [21], discrimination between various discontinuous breath sounds [16], prediction of an approaching seismic events [1] and nondestructive evaluation of ultrasonic signals obtained from metal and composite adhesion under different conditions [25]. This background and breadth of applications, pointing to HOC as a useful tool in time series discrimination, motivates us to apply it in the context of our problem. This dissertation discusses the first time application of HOC in the detection of differences in eye gaze data. In particular, for discrimination purposes distance from white noise process (reference series) were applied.

The other method is in the spectral domain. It is based on comparing the spectral densities of several processes. This method assumes a spectral density ratio model. The model assumes the log ratio of two spectral densities to have a certain linear parametric form. Such an analysis using model assumptions was first discussed by Coates and Diggle [10]. Their work was extended by Savvides to the case of multiple time series [31]. Few details of the analysis is given in Chapter 2. In the case of replicated time series, comparison of several spectral densities was studied by Han by using the method of smoothing spline ANOVA [17]. For more references on this topic, see [13], [31] and [17]. Finally, the parameters used in the spectral density ratio model can be used in classification/clustering of time series.

The goal of classification (unsupervised) problem is to look for a natural seperation among datasets. This seperation into groups is formed based on criteria where

the time series within a group are closer to each other in terms of distance compared to the distances measured between time series of different groups. For this purpose, there is a need to identify similarities or dissimilarities in the time series data using various distance metrics. There are many such measures already available in the time series literature which are based on parametric as well as non-parametric measures [9], [35]. In general, the performance of these measures depends on the type of the data, and is a subject of active research.

This dissertation applies the spectral density ratio based clustering technique to the eye gaze data. The idea is again to look for differences (if any) between the two conditions "Watch" and "Imitate". Investigating the eye gaze time series suggests a long memory type behaviour. Few details on long memory processes have been summarized in Chapter 2. An extensive overview of long memory processes can be found in [27], [4]. Since the goal is to detect differences in eye gaze, the two methods mentioned above have been studied specifically for long memory processes. Several simulation studies have been run to test the performance of the two methods for long memory processes. Further, clustering has been performed using a more generalized form of the spectral density ratio model that is specific to the case of long memory time series.

This dissertation is organized as follows: Chapter 2 discusses some concepts related to short memory and long memory processes. In Chapter 3, we give details of eye gaze data and suggest ways to model the eye behaviour. Chapter 4 gives some background on the HOC and its properties are studied under the long memory processes. Then it is applied on the eye gaze data. Chapter 5 extends

the methodology of comparing several spectral densities to the case of long memory processes. Further, the parameters of the more generalized model are applied for classification purposes in simulation. Finally, clustering methods are applied to eye gaze data. Chapter 6 presents some concluding remarks.

Chapter 2

Long Memory Processes

In this chapter, we look at some basic definitions and notation. We discuss the concept of stationarity, short and long memory processes and their spectral densities. We also define the periodogram which is an estimator of the spectral density and state its distributional properties for both long and short memory processes. In addition, we study the exponential (EXP) model introduced by Bloomfield (1973) and its generalized version referred to as the fractional exponential (FEXP) model that was introduced by Beran [5], [3]. We present some examples for both classes of models. We discuss the estimation procedure for comparing several spectral densities using the EXP model. We also look at various estimation techniques for estimating the fractional parameter in a long memory process.

2.1 Definitions and Notation

In this section, we consider discrete time processes such as $\{Y_t, t = 0, \pm 1, \pm 2, ...\}$. The second order moments of the joint distribution of $\{Y_t\}$ is quantified by the autocovariance function as described below.

Autocovariance: The autocovariance function of the process $\{Y_t; t = 0, \pm 1, \pm 2, ..., \}$ with $var(Y_t) < \infty$ for each $t \in \{0, \pm 1, \pm 2, ...\}$ is defined by,

$$\gamma_Y(t_1, t_2) = cov(Y_{t_1}, Y_{t_2}) = E(Y_{t_1} Y_{t_2}) - E(Y_{t_1})E(Y_{t_2}), \forall t_1, t_2 \in \{0, \pm 1, \pm 2, ...\} \quad (2.1)$$

The function $\gamma_Y(t_1, t_2)$ is a measure of the linear dependence of $\{Y_t\}$ between two time points t_1 and t_2 of the same time series. Specifically for $t_1 = t_2 = t$, it gives the variance $V(Y_t)$. Most time series literature make certain assumptions on the first two momemts of the time series in order to attain a certain regularity in the behaviour of the time series known as stationarity. In general, more assumptions on the higher order moments will lead to a more stable system approaching strict stationarity. In the second order case, one important assumption is to let the first order moment be independent of time t and the second moments depend on only the seperation between the two time points t_1 and t_2. This stationarity is referred to as weak stationarity, second order stationarity or just stationarity. The definition is as follows.

Stationarity: A time series process Y_t is said to be weakly stationary, if its second order properties follow the conditions,

1. $E(Y_t) = m$.
2. $E(Y_t^2) < \infty$, for all $t \in 0, \pm 1, \pm 2, ...$
3. $\gamma(t, t+h) = \gamma(t+r, t+h+r) \equiv$ function of h, for $h = 0, \pm 1, ...$

The periodocity of a time series can be studied in the frequency domain by taking the inverse fourier transform of the (absolutely summable) autocovariance function which is the spectral density. Hence, for any stationary process with (absolutely summable) autocovariance function, the autocovariance function and the spectral density form a Fourier transform pair that exists and is unique.

Spectral Density: Consider a stationary time series $\{Y_t, t = 0, \pm 1, \pm 2, ...\}$ with zero mean and autocovariance function $\gamma(h)$ such that $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. Then, spectral density of the process $\{Y_t\}$, $f(\omega)$, is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}, -\pi \leq \omega \leq \pi \tag{2.2}$$

The spectral density function $f(\omega)$ for a real valued processes Y_t has the following properties:

1. $f(\omega)$ is a real valued function
2. $f(\omega) \geq 0$
3. $f(\omega) = f(-\omega)$
4. $f(\omega + 2\pi) = f(\omega)$

A popular time series process is autoregressive moving average (ARMA) models discussed by Box, Jenkins and Reinsel in [7]. The estimation procedure of model coefficients, confidence intervals of the estimates and prediction intervals have been extensively studied in the time series literature and can be found in details in [6], [33].

ARMA Process: The process $\{Y_t\}$ is said to be an ARMA(p,q) process if $\{Y_t\}$ is stationary and for every t,

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \psi_j \epsilon_{t-j} \tag{2.3}$$

where $\{\epsilon_t\}$ is a white noise process with mean 0 and common variance σ^2. If the equations $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i = 0$ and $\psi(z) = 1 - \sum_{j=1}^{q} \psi_j z^j = 0$ have no common

zeroes and the roots are outside the unit circle, then the ARMA model is said to be causal and invertible, respectively. Under the above mentioned conditions, the autocovariance function of the process $\{Y_t\}$ decays exponentially and therefore summable (also absolutely summable). Such processes are also referred to as the short memory process. In contrast to an ARMA process, a process for which the autocovariance function decays slowly (hyperbolic rate) and hence is not absolutely summable is referred to as a long memory process. In the frequency domain, a stationary time series is a long memory process if there is a non-zero $d \in (-.5, .5)$ and a positive constant c such that the spectral density function $f(\lambda)$ satisfies the following condition.

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{c\,|\lambda|^{-2d}} = 1 \tag{2.4}$$

This condition implies that the spectral density has a pole at zero frequency when $d \in (0, 0.5)$. For alternate definitions of long memory processes see [4], [27]. For $d = 0$, the process is a short memory process. A well-known class of long memory models are the autoregressive fractionally integrated moving average models (ARFIMA).

ARFIMA Process: The process Y_t is said to be an ARFIMA(p,d,q) process with $d \in (-0.5, 0.5)$ if Y_t is stationary and satisfies the difference equations,

$$\phi(B)Y_t = \psi(B)(1 - B)^{-d}\epsilon_t \tag{2.5}$$

where ϵ_t is a white noise process with mean 0 and common variance σ^2, B is the

backshift operator such that,

$$B^j Y_t = Y_{t-j}, j = 1, 2, ... \tag{2.6}$$

and $\phi(.)$ and $\psi(.)$ are polynomials of degree p and q respectively that have no common roots. If the zeros of $\phi(.)$ and $\psi(.)$ lie outside the unit circle then there is a causal and invertible solution $\{Y_t\}$.

We note here that $d \in (-0.5, 0.5)$ in (2.5) yields a stationary solution. In particular, for $d = 0$, the ARFIMA process reduces to the ARMA case, that is short memory. For $0 < d < 0.5$, the process is said to have long memory and for $-0.5 < d < 0$, intermediate memory [8]. In this section and later, we will specifically look at results for the long memory case $(0 < d < 0.5)$ only.

The fractional differencing operator $(1 - B)^{-d}$ is defined by the following binomial expansion for $d \in (0, 0.5)$,

$$(1 - B)^{-d} = \eta(B) = \sum_{j=0}^{\infty} \eta_j B^j \tag{2.7}$$

where for $j = 0, 1, 2...$

$$\eta_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} \tag{2.8}$$

$$= \prod_{0<l\leq j} \frac{l - 1 - d}{l} \tag{2.9}$$

The spectral densities for the stationary ARMA and ARFIMA models discussed above are given as [8],

$$f_{ARMA}(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{1 - \sum_{r=1}^{q} \psi_r e^{ir\omega}}{1 - \sum_{r=1}^{p} \phi_r e^{ir\omega}} \right|^2 \tag{2.10}$$

$$f_{ARFIMA}(\omega) = \frac{\sigma^2}{2\pi} \left| 1 - e^{i\omega} \right|^{-2d} \left| \frac{1 - \sum_{r=1}^{q} \psi_r e^{ir\omega}}{1 - \sum_{r=1}^{p} \phi_r e^{ir\omega}} \right|^2 \tag{2.11}$$

where $-\pi < \omega \leq \pi$. Alternatively, the spectral density of an ARFIMA process can be written as

$$f_{ARFIMA}(\omega) = \left|1 - e^{i\omega}\right|^{-2d} f_{ARMA}(\omega) \tag{2.12}$$

The spectral density function $f_{ARFIMA}(\omega)$ is unbounded at zero frequency. As $|\omega| \to 0$, we have,

$$f_{ARFIMA}(\omega) \approx \frac{\sigma^2}{2\pi} |\omega|^{-2d} \frac{|\psi(1)|^2}{|\phi(1)|^2} \tag{2.13}$$

To estimate these spectral densities, one of the commonly used estimator is the smoothed periodogram. The periodogram is defined as follows.

Periodogram: Let Y_t be a time series of length N. The periodogram at frequency ω_i is defined by,

$$I_Y(\omega_i) = \frac{1}{2\pi N} \left| \sum_{t=1}^{n} Y_t \exp(-it\omega_i) \right|^2 \tag{2.14}$$

where $\omega_i = \frac{2\pi i}{N}$, $i = 1, 2, ..., [(N-1)/2]$.

The asymptotic properties of the periodogram vary for the short and long memory processes. We state a few results for the periodogram for both cases (See [8], [4]).

Proposition: If Y_t is stationary with mean μ and absolutely summable autocovariance function $\gamma(.)$, then

(i) $EI_Y(0) - \frac{N}{2\pi}\mu^2 \to f_Y(0)$

(ii) $EI_Y(\omega) \to f_Y(\omega)$, for $\omega \neq 0$.

This implies that the periodogram is asymptotically an unbiased estimator of the spectral density for all frequencies but zero. If the mean of the process $\mu = 0$, that

establishes the asymptotic unbiasedness for all frequencies.

The next result states that periodogram ordinates for an $IID(0, \sigma^2)$ process or more generally for a linear process (as described below) are asymptotically independent and exponentially distributed.

Let $\{Y_t\}$ be a general linear process such that

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \tag{2.15}$$

where the Z_t are $IID(0, \sigma^2)$, $\sigma^2 < \infty$, and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Theorem 1: Let $\{Y_t\}$ be the linear process as given above. Let $I_Y\{\omega\}$ denote the periodogram, and let $f_Y(\omega)$ be the spectral density of the process Y_t, respectively.

(i) If $f_Y(\omega) > 0$ for all $\omega \in [-\pi, \pi]$ and if $0 < \omega_1 < \omega_2 < .... < \omega_m < \pi$, then the vector $(I_Y(\omega_1), I_Y(\omega_2), ..., I_Y(\omega_m))'$ converges in distribution to a vector of independently and exponentially distributed random variables, the ith component of which has mean $f_Y(\omega_i), i = 1, 2, ...m$.

(ii) If $\sum_{j=-\infty}^{\infty} |\psi_j| \, |j|^{1/2} < \infty$, $EZ_1^4 = \eta\sigma^4 < \infty$, $\omega_j = 2\pi j/N \geq 0$, then

$$Cov(I_Y(\omega_j), I_Y(\omega_k)) = \begin{cases} 2f(\omega_j)^2 + O(N^{-1/2}), & \text{if } \omega_j = \omega_k = 0 \text{ or } \pi \\ f(\omega_j)^2 + O(N^{-1/2}), & \text{if } 0 < \omega_j = \omega_k < \pi \\ O(N^{-1}), & \text{if } \omega_j \neq \omega_k \end{cases} \tag{2.16}$$

where the terms $O(N^{-1/2})$ and $O(N^{-1})$ can be bounded uniformly in j and k by $c_1 N^{-1/2}$ and $c_2 N^{-1}$ respectively , for some constants c_1 and c_2.

The above asymptotic result of the periodogram ordinates holds for a stationary ARMA process which was described earlier. The same result of asymptotic

independence and exponential distribution of the periodogram ordinates holds for long memory processes in the case where m is finite and also under certain assumptions on the cumulants of Z_t, except that the asymptotic behaviour of the periodogram ordinates are quite different for frequencies close to zero. For fourier frequencies close to zero, the following result holds [4].

Theorem 2: Let $\{Y_t\}$ be a stationary process with long memory and a spectral density f given by

$$f(\omega) = f^*(\omega) \left|1 - \exp(i\omega)\right|^{-2d} \tag{2.17}$$

where $f^* : [-\pi, \pi] \to R_+$ is a continuous positive function and $d \in (0, 0.5)$. For a fixed integer j, let $\omega_j = 2\pi j/N$ be the jth Fourier frequency. Also, define the normalized periodogram

$$I^*(\omega_j) = \frac{I(\omega_j)}{f(\omega_j)} \tag{2.18}$$

Then the following results hold.

a) The asymptotic expectation of $I^*(\omega_j)$ is equal to

$$\lim_{N\to\infty} E[I^*(\omega_j)] = 2\Delta_j(d, -1) \tag{2.19}$$

where,

$$\Delta_j(d, u) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2 \frac{x}{2}}{(2\pi j - x)(2\pi j - ux)} \left|\frac{x}{2\pi j}\right|^{-2d} dx \tag{2.20}$$

b) The asympotic variance of $I(\omega_j)$ is of the order

$$var(I(\omega_j)) = O(N^{2d}) \tag{2.21}$$

c) For $j \neq j'$, the normalized periodograms $I^*(\omega_j)$ and $I^*(\omega_{j'})$ are asympotically correlated.

d) If Y_t is Gaussian, then $I^*(\omega_j) \rightarrow^d \eta_j$ where η_j is defined by

$$\eta_j = [\frac{1}{2}\Delta_j(d,-1) - \Delta_j(d,1)]Z_1^2 + [\frac{1}{2}\Delta_j(d,-1) + \Delta_j(d,1)]Z_2^2 \tag{2.22}$$

with independent standard normal random variables Z_1, Z_2 independent of j.

This theorem for long memory processes shows that the periodogram ordinates for frequencies near zero are not asympotically independent, and that their normalized counterparts do not possess the same distribution as in the case of short memory processes. In the next section, we look at the parametric forms of the spectral density functions for both short and long memory processes.

2.2 Spectral density models

The exponential model of an estimated spectrum of a stationary time series was introduced by Bloomfield [5]. It is based on the observation that the logarithm of the spectral density is a fairly well-behaved function and can be approximated by a truncated Fourier series of order say p. The form of the exponential model denoted as EXP(p) is given by

$$f(\omega) = \frac{\sigma^2}{2\pi}\exp(2\sum_{r=1}^{p} a_r \cos(r\omega)), 0 < \omega < \pi \tag{2.23}$$

with σ^2 and $a_r; r = 1, 2, ..p$ as parameters. Clearly $f(\omega) \geq 0$ and since this Fourier expansion is in terms of the cosine basis we have the eveness property of the spectral

density function satisfied. Alternatively, the logarithm of EXP(p) model can be written as

$$\log f(\omega) = a_0 + a_1 \cdot 2 \cdot \cos(\omega) + a_2 \cdot 2 \cdot \cos(2\omega) + ... + a_p \cdot 2 \cdot \cos(p\omega) \tag{2.24}$$

or,

$$\log f(\omega) = a^T Z(\omega) \tag{2.25}$$

where $a^T = (a_0, a_1, ..., a_p)^T$ and $Z(\omega) = (1, 2\cos(\omega), 2\cos(2\omega), ..., 2\cos(p\omega))^T$.

An extension to the EXP(p) models as described above was given by Beran for long memory processes [4]. The form of the model denoted as FEXP(p) is given by,

Let $g : [-\pi, \pi] \to R_+$ be a positive function such that

$$\overset{lim}{\omega \to 0} \frac{g(\omega)}{\omega} = 1 \tag{2.26}$$

and $g(\omega) = g(-\omega)$. Define $f_0 \equiv 1$, and let $f_1, f_2, ..., f_p$ be smooth even functions on $[-\pi, \pi]$. Assume that the matrix $H = (h_{ij})$ with $h_{ij} = f_j(2\pi i/N)$, $j = 0, 1, ..., p$ and $i = 1, 2..., N^*$ is nonsingular. Furthermore, let $\theta = (\theta_0, d, \theta_1, ..., \theta_p)$ is a real vector with $-0.5 < d < 0.5$. Then Y_t is an FEXP(p) process with short memory components $f_1, f_2, ..., f_p$ and a long memory component g if its spectral density is given by

$$f(\omega) = g(\omega)^{-2d} \exp(\sum_{j=0}^{p} \theta_j f_j(\omega)) \tag{2.27}$$

2.2.1 Example of a EXP model

Consider an ARMA(p,q) process given by

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \psi_j \epsilon_{t-j} \tag{2.28}$$

where ϵ_t is a white noise process with mean 0 and common variance σ^2. If the equations $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i = 0$ and $\psi(z) = 1 - \sum_{j=1}^{q} \psi_j z^j = 0$ have no common zeroes and the roots are outside the unit circle, the spectral density function is given by

$$f_y(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{1 - \sum_{r=1}^{q} \psi_r e^{ir\omega}}{1 - \sum_{r=1}^{p} \phi_r e^{ir\omega}} \right|^2 \tag{2.29}$$

Based on some calculations, it can be shown that the logarithm of spectral density (2.29) can be expressed as

$$\log f_y(\omega) = log(\sigma^2) - log(2\pi) + 2 \sum_{r=1}^{\infty} (\sum_{k=1}^{p} b_k^r - \sum_{l=1}^{q} c_l^r) \frac{\cos(r\omega)}{r} \tag{2.30}$$

where c_l and b_k are the reciprocals of the roots of polynomials $\psi(z)$ and $\phi(z)$ respectively. Notice here that since the roots lie outside the unit circle, the reciprocal roots c_l^r and b_k^r tend to zero as the value of r increases. Comparing Eqns. (2.30) and (2.25), we have

$$a_0 = log(\sigma^2) - log(2\pi) \tag{2.31}$$

$$a_r = \frac{2}{r} (\sum_{k=1}^{p} b_k^r - \sum_{l=1}^{q} c_l^r) \tag{2.32}$$

2.2.2 Example of a FEXP model

Consider an ARFIMA(p,d,q) process given by

$$\phi(B)(1-B)^d Y_t = \psi(B)\epsilon_t \tag{2.33}$$

The spectral density function is given by

$$f_y(\omega) = \frac{\sigma^2}{2\pi}\left|1-e^{i\omega}\right|^{-2d}\left|\frac{1-\sum_{r=1}^{q}\psi_r e^{ir\omega}}{1-\sum_{r=1}^{p}\phi_r e^{ir\omega}}\right|^2 \tag{2.34}$$

$$f_y(\omega) = \left|1-e^{i\omega}\right|^{-2d} f_{ARMA}(\omega) \tag{2.35}$$

Taking logarithm on both sides,

$$\log f_y(\omega) = \log\left|1-e^{i\omega}\right|^{-2d} + \log(f_{ARMA}(\omega)) \tag{2.36}$$

Using the EXP model form for the $\log(f_{ARMA}(\omega))$, we have

$$\log f_y(\omega) = \log\left|1-e^{i\omega}\right|^{-2d} + a^T Z(\omega) \tag{2.37}$$

Or more generally,

$$\log f_y(\omega) = -d\log\left|g(\omega)\right|^2 + a^T Z(\omega) \tag{2.38}$$

Comparing with FEXP(p) model (2.27), $\{f_i(\omega), i = 0,1,...,p\} = Z(\omega)$ and $\{\theta_i, i = 0,1,...,p\} = \mathbf{a}$.

We next look at examples of ARFIMA(p,d,q) models. Using known values of parameters in these models, we evaluate the coefficient vector in the FEXP model.

Example 1 Consider the stationary ARFIMA(p,d,q) process with $p = 1, q = 0$.

$$X_t = \phi X_{t-1} + (1-B)^{-d}\epsilon_t \tag{2.39}$$

Table 2.1: True parameter values in the FEXP(p) model (Example 1)

Process	θ_0	d	θ_1	θ_2	θ_3	θ_4	θ_5
$\phi = -0.3, d = 0.3$	0	0.3	-0.3	0.045	-0.0090	0.002025	-4.86E-4
$\phi = -0.1, d = 0.3$	0	0.3	-0.1	0.005	-0.0003	2.5E-5	-2E-6
$\phi = 0, d = 0.3$	0	0.3	0.0	0.000	0.0000	0.0000	0.0000
$\phi = 0.1, d = 0.3$	0	0.3	0.1	0.005	0.0003	2.5E-5	2E-6
$\phi = 0.3, d = 0.3$	0	0.3	0.3	0.045	0.0090	0.002025	4.86E-4

Table 2.2: True parameter values in the FEXP(p) model (Example 2)

Process	θ_0	d	θ_1	θ_2	θ_3	θ_4	θ_5
$\phi = -0.3, d = 0.3, \psi = -0.2$	0	0.3	-0.1	0.025	-0.0063	0.001625	-0.000422
$\phi = -0.1, d = 0.3, \psi = -0.4$	0	0.3	0.3	-0.075	0.021	-0.006375	0.002046
$\phi = 0, d = 0.3, \psi = 0$	0	0.3	0.0	0.000	0.0000	0.0000	0.0000
$\phi = 0.1, d = 0.3, \psi = 0.2$	0	0.3	-0.1	-0.015	-0.0023	-0.000375	-6.2E-05
$\phi = 0.3, d = 0.3, \psi = 0.4$	0	0.3	-0.1	-0.07	-0.0123	-0.004375	-0.001562

For different values of d and ϕ, Table (2.1) shows the true values of parameters $\{\theta_0, d, ..., \theta_5\}$ in the FEXP(p) models. Notice that as j increases, θ_j becomes negligible. In this example, for $j > 5$, $\theta_j \approx 0$.

Example 2 Consider the ARFIMA(p,d,q) process with $p = 1, q = 1$.

$$X_t = \phi X_{t-1} + (1 - B)^{-d}(\epsilon_t - \psi\epsilon_{t-1}) \tag{2.40}$$

For different values of d, ϕ and ψ, Table (2.2) shows the true parameter values in the FEXP(p) models. Again, for $j > 5$, the values of parameters θ_j are negligible and are not reported in the table.

2.2.3 Parameter estimation in FEXP(p) models using GLM

Beran discusses the estimation of the parameters in the generalized linear model (GLM) setting [3]. In a GLM, the mean of the response variable is dependent on the explanatory variables via a link function. If y is the response variable

with mean μ and distribution function F and $u_1, u_2, ..., u_K$ are the K explanatory variables, then we have the following

$$\nu(\mu) = \beta_0 + \beta_1 u_1 + ... + \beta_K u_K \tag{2.41}$$

where $\nu(.)$ is the link function.

Let us consider the response variable y to be the peridogram ordinate. Therefore, for $j = 1, 2, ..., [(N-1)/2]$,

$$y_j = I(\omega_j) \tag{2.42}$$

By assuming independence of periodogram ordinates (as in the case of the short memory processes), we have

$$y_j = I(\omega_j) \approx f(\omega_j) Z_j \tag{2.43}$$

where Z_j are IID exponential distributed random variables with mean 1.
Hence the response variable y_j has an exponential distribution with mean $f(\omega_j)$. The spectral density of long memory processes given by Eqn. (2.27) can be rewritten by taking the natural logarithm as

$$\log f(\omega_j) = -d \log(g(\omega_j)^2) + \sum_{i=0}^{p} \theta_i f_i(\omega_j) \tag{2.44}$$

Therefore, the estimation of the parameters $\{\theta_0, d, \theta_1, ..., \theta_p\}$ can be carried out using GLM methods with explanatory variables $g^2(\omega), f_1(\omega), f_2(\omega), ..., f_p(\omega)$, response variable y with exponential distribution and logarithmic link function.

The choice of the functions $\{g(\omega), f_i(\omega), i = 1, 2, ..., p\}$ give rise to different classes of FEXP (p) models. Specifically consider two classes,

Case 1 (Polynomial Model) : $g(\omega) = \sin^2(\omega/2), f_i(\omega) = |\omega|^i, i = 1, 2, ..., p$

Case 2 (Cosine Model): $g(\omega) = \sin^2(\omega/2), f_i(\omega) = \cos(i\omega), i = 1, 2, ..., p.$

The estimation of the parameters $\{\theta_0, d, \theta_1, ..., \theta_p\}$ for the given choice of functions mentioned in the two cases can be carried out using R software for both cases. Below are examples considering the case with cosine functions only. The estimation was implemented in R software using the longmemo package.

Example 1 Consider the ARFIMA process of length N with $p = q = 0$ and $d = 0.3$, and $\epsilon_t \sim N(0, 1)$.

The logarithm of the spectral density is

$$\log(f(\omega) = -\log(2\pi) - d\log|1 - \exp(i\omega)|^2 \tag{2.45}$$

Notice that true value of θ_0 is $-\log(2\pi)$ and d is 0.3. Further $\epsilon_t \sim N(0, 1)$ is used only for simulating ARFIMA$(0, d = 0.3, 0)$ process. The GLM estimation procedure is as discussed above using Eqns. (2.43) and (2.44). The estimates of the parameters $\{\theta_0, d\}$ is calculated for different values of p in Table (2.3). Since the process has no short memory component, the parameters θ_i are all zero for $i \geq 1$. The Table (2.3) lists the parameter estimates for the true model i.e. with $p = 0$ and also the misspecified model with $p = 1$. As N increases, the parameter estimates are closer to the true values under the true model. For all values of N, the test for $\theta_1 = 0$ is accepted at 5% significance level.

Example 2 Consider the ARFIMA process of length N with $\phi = 0.4$ and $d = 0.2$, and $\epsilon_t \sim N(0, 1)$

Table 2.3: Parameter estimates in the FEXP(p) model using Cosine model

N	p	Parameter	True	Estimate	Std Error	t-stat	pvalue
50	0	θ_0	-1.8379	-1.4039	0.1594	-8.8091	1.1502E-08
		d	0.3	0.125	0.2222	-1.1288	0.2711
	1	θ_0	-1.8379	-1.3933	0.1716	-8.1183	6.5033E-08
		d	0.3	0.2610	0.4622	-1.1283	0.2719
		θ_1	0	-0.3323	0.4769	-0.6968	0.4935
100	0	θ_0	-1.8379	-1.8955	0.1283	-14.7683	$<$ 2E-16
		d	0.3	0.175	0.1641	-2.1314	0.0383
	1	θ_0	-1.8379	-1.894	0.1305	-14.5137	$<$ 2E-16
		d	0.3	0.195	0.3021	-1.293	0.2024
		θ_1	0	-0.0567	0.3371	-0.1668	0.8683
500	0	θ_0	-1.8379	-1.9179	0.0571	-33.5978	$<$ 2.22E-16
		d	0.3	0.197	0.0661	-5.9581	8.7419E-09
	1	θ_0	-1.8379	-1.9190	0.0570	-33.6892	$<$ 2.22E-16
		d	0.3	0.235	0.1101	-4.2713	2.7790E-05
		θ_1	0	-0.1381	0.1348	-1.0245	0.3066

The logarithm of the spectral density can be written as

$$\log(f(\omega) = -\log(2\pi) - d\log|1-\exp(i\omega)|^2 + 2\sum_{r=1}^{\infty}\frac{\phi^r}{r}cos(r\omega) \tag{2.46}$$

In this example, a short memory component is present. There are infinite number of parameters $\theta_r = \frac{2\phi^r}{r}$, for $r = 1, 2, ...$ Since $0 < \phi < 1$, ϕ^r tends to zero as r increases and hence the sum is Eqn. (2.46) can be truncated to p terms. We consider the values of p to be $0, 1, 2$. The true values of the parameters $\{\theta_0, d\,\theta_1, \theta_2\}$ are $\{-\log(2\pi), 0.2, 0.8, 0.16\}$. The corresponding estimates of parameters $\{\theta_0, d\}$, $\{\theta_0, d\,\theta_1\}$ and $\{\theta_0, d\,\theta_1, \theta_2\}$ are calculated and reported in Table (2.4). From the table it is evident that better results are obtained as N and p increases.

In the next section, we discuss a method of comparing several spectral densities using EXP(p) model. This is from the work of Fokianos and Savvides [14], [31]. *This is important since this method compares the spectral densities of several processes*

Table 2.4: Parameter estimates in the FEXP(p) model using Cosine model

N	p	Parameter	True	Estimate	Std Error	t-stat	pvalue
50	0	θ_0	-1.8379	-1.4395	0.2197	-6.5524	1.3683E-06
		d	0.2	0.3590	0.1530	-2.3454	0.02843
	1	θ_0	-1.8379	-2.2606	0.2031	-11.1325	2.8679E-10
		d	0.2	0.4160	0.2730	-1.5205	0.1433
		θ_1	0.8	0.2410	0.5643	0.4270	0.6737
	2	θ_0	-1.8379	-2.2675	0.2099	-10.8015	8.5211E-10
		d	0.2	0.3430	0.4280	-0.7998	0.4332
		θ_1	0.8	0.3738	0.8154	0.4584	0.6516
		θ_2	0.16	0.0879	0.4528	0.1941	0.8480
100	0	θ_0	-1.8379	-1.8685	0.1439	-12.9884	$<$ 2.22E-16
		d	0.2	0.5260	0.0920	-5.7217	7.0813E-07
	1	θ_0	-1.8379	-1.8824	0.1426	-13.1951	$<$ 2E-16
		d	0.2	0.4120	0.1650	-2.4979	0.0161
		θ_1	0.8	0.2622	0.3684	0.7116	0.4803
	2	θ_0	-1.8379	-1.9102	0.1464	-13.0443	$<$ 2E-16
		d	0.2	0.2	0.2410	-0.8319	0.4104
		θ_1	0.8	0.6309	0.4950	1.2747	0.2090
		θ_2	0.16	0.3765	0.2966	1.2696	0.2108
500	0	θ_0	-1.8379	-1.8387	0.0671	-27.4034	$<$ 2.22E-16
		d	0.2	0.4910	0.0390	-12.62824	¡ 2.22E-16
	1	θ_0	-1.8379	-1.8724	0.0652	-28.6971	$<$ 2.22E-16
		d	0.2	0.2580	0.0630	-4.0923	5.7943E-05
		θ_1	0.8	0.6142	0.1544	3.9779	9.1477E-05
	2	θ_0	-1.8379	-1.8748	0.0648	-28.9390	$<$ 2.22E-16
		d	0.2	0.2030	0.0830	-2.4467	0.0151
		θ_1	0.8	0.7195	0.1870	3.8475	0.0002
		θ_2	0.16	0.1054	0.1218	0.8654	0.3876

with respect to a reference spectral density. This fact will be taken into account in the later chapters.

2.3 Comparing several spectral densities using EXP(p) models

Consider G stationary independent time series $\{Y_{jt}, j = 1, 2, ..., G; t = 1, 2, ..., n\}$ and let $\{f_j(\omega), j = 1, 2, ..., G\}$ be the corresponding continuous spectral densities. Based on the EXP(p) model, an useful expansion of the ratio of two spectral densities $f_j(\omega)$, $f_G(\omega)$ is given by

$$\log(\frac{f_j(\omega)}{f_G(\omega)}) = a_i^T Z(\omega), i = 1, 2, ...G-1, -\pi < \omega < \pi \tag{2.47}$$

where $Z(\omega) = (1, 2\cos(\omega), 2\cos(2\omega), ..., 2\cos(p\omega))^T$, and

$a_i = (a_{0i}, a_{1i}, ..., a_{pi})^T$ is a $p+1$ vector that can be estimated using the method of the maximum likelihood as described below.

The framework expressed by (2.47) is useful in the following hypothesis testing problems.

1. Test $f_j(\omega) = f_G(\omega), \forall i = 1, 2, ..., (G-1)$ by

   $H_0 : a_i = 0, \forall i = 1, 2, ..., (G-1)$

2. Test $f_j(\omega) \propto f_G(\omega), \forall i = 1, 2, ..., (G-1)$ by

   $H_0 : a_{1i} = a_{2i} = ... = a_{pi} = 0, \forall i = 1, 2, ..., (G-1)$

3. $H_0 : \mathbf{Aa} = \mathbf{0}$

   where $\mathbf{A}$ is a matrix of dimension $s \times (G-1)(p+1)$, $s \leq (G-1)(p+1)$ and is assumed full rank.

2.3.1 Estimation procedure using EXP(p) model

Consider a set of G independent stationary processes $\{Y_{jt}, t = 1, 2, ..., N\}$ and assume that each of the G time series possesses a spectral density function $f_j(\omega)$. Let the Gth time series Y_{Gt} be the reference time series, and assume the spectral densities are unknown and follow the EXP(p) model. Hence to estimate the coefficient vector $\mathbf{a}$ via the method of maximum likelihood estimation (MLE) the results on periodogram ordinates mentioned in the previous section are applied. The steps to get the log likelihood function and therefore the MLE estimators $\hat{\mathbf{a}}$ can be found in Savvides [31]. The procedure is summarized as follows.

1. Using Theorem 1, the periodogram ordinates $I_j(\omega_i)$ for the jth time series at ith fourier frequency $\omega_i = \frac{2\pi i}{N}, i = 1, 2, ...[(N-1)/2]$ are asymptotoically exponentially distributed with mean $f_j(\omega_i)$. Note the π is not included.

2. The joint distribution of log of ratio of periodogram ordinates $T_{ji} = \log \frac{I_j(\omega_i)}{I_G(\omega_i)}$ for $j = 1, ..., (G-1)$ can be obtained using the independence of G periodogram ordinates $I_j(\omega_i), j = 1, ..., G$ and then considering suitable transformations. For each fourier frequency ω_i, the asympotic joint distribution of $(T_{1i}, T_{2i}, ..., T_{(G-1)i})$ is given as:

$$h(t_{1i}, t_{2i}, .., t_{(G-1)i}) = \frac{(G-1)! \exp(\sum_{j=1}^{G-1} t_{ji} - \log \frac{f_j(\omega_i)}{f_G(\omega_i)})}{(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \log \frac{f_j(\omega_i)}{f_G(\omega_i)}))^G} \tag{2.48}$$

3. Now since the periodogram ordinates $I_j(\omega_i)$ are asympotically independent at fourier frequencies, the joint distribution of the vectors $\mathbf{T}_i = (T_{1i}, T_{2i}, ..., T_{(G-1)i})^T, i = 1, 2, ..., [(N-1)/2]$ is given by

$$H(\mathbf{T}_1, \mathbf{T}_2, ..., \mathbf{T}_{(G-1)}) = \prod_{i=1}^{[\frac{N-1}{2}]} h(t_{1i}, t_{2i}, .., t_{(G-1)i}) \tag{2.49}$$

4. Using the above results the log likelihood function up to a constant is given by,

$$l(\mathbf{a}) = \sum_{i=1}^{[(N-1)/2]} \sum_{j=1}^{(G-1)} (t_{ji} - a_j^T Z_i) - G \sum_{i=1}^{[(N-1)/2]} \log(1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - a_j^T Z_i)) \tag{2.50}$$

where $\mathbf{a} = \{a_1^T, ..., a_{G-1}^T\}$ and $Z_i = Z(\omega_i)$.

The maximum likelihood estimates of the coefficient vector $\mathbf{a}$ can then be obtained by setting the system of equations $S(\mathbf{a}) = 0$, where $S(\mathbf{a})$ is the score function corresponding to the likelihood function $l(\mathbf{a})$.

$$S(\mathbf{a}) = \frac{\partial l(\mathbf{a})}{\partial \mathbf{a}} \tag{2.51}$$

For testing the hypothesis for equality of G spectral densities, the likelihood ratio test statistic is given by,

$$\begin{aligned} LRT &= 2(l(\mathbf{a}) - l(0)) && (2.52) \\ &= -2 \sum_{i=1}^{[(N-1)/2]} \sum_{j=1}^{(G-1)} a_j^T Z_i + 2G \sum_{i=1}^{[(N-1)/2]} \log\left(\frac{1 + \sum_{j=1}^{(G-1)} \exp(T_{ji})}{1 + \sum_{j=1}^{(G-1)} \exp(T_{ji} - {a_j}^T Z_i)}\right) && (2.53) \end{aligned}$$

The distribution of the LRT statistic is chi-square variable with $(G-1)(p+1)$ degrees of freedom.

2.4 Estimation of paramater d

There are techniques available to estimate the long memory parameter or the fractional parameter d. A method proposed by Geweke and Porter-Hudak is based

on spectral density regression around zero frequency [18]. Consider $\{y_t\}$ to be a stationary long memory time series of length N. Assuming the spectral density of $\{y_t\}$ has the following form,

$$f(\lambda) = f_0(\lambda)[2\sin(\lambda/2)]^{-2d} \tag{2.54}$$

and based on some additional calculations [18], the spectral regression around zero frequency is given by

$$\ln(I(\lambda_j)) = \ln(f_0(0)) - d\ln(4\sin^2(\lambda_j/2)) + \eta_j, \text{j} = 1,2,...,\text{n} \tag{2.55}$$

where $I(\lambda_j)$ is the periodogram of y_t at the Fourier frequency $\lambda_j = 2\pi j/N$ ($j = 0, ..., (N-1)/2$) and $n = g(N) = N^{\alpha}(<< N; 0 < \alpha < 1)$ is the number of observations $I(\lambda_j)$ included in the spectral regression. The estimate of d is the slope of the regression line above and is obtained by least square estimation. As an extension to this method, Reisen replaced the discrete periodogram ordinates by smoothed periodogram using Parzen lag window. In this method, the function $g(N)$ is chosen as before and the truncation point in the Parzen lag window estimate is $m = N^{\beta}; 0 < \beta < 1$. For details, see [29].

Another method for estimation of memory parameter d for an ARFIMA process is based on the variance of the sample mean of the process. For a long memory process of m observations where m is large and c is a positive constant, we have

$$var(\bar{y_m}) \approx cm^{2d-1} \tag{2.56}$$

Therefore, a process with n observations ($n > m$) can be divided into k blocks each of size m. For the jth block ($j = 1, 2, ...k$) we can then write the following

using Equation(2.56)

$$\log(var(\bar{y_j})) \approx c + (2d-1)\log(j) \tag{2.57}$$

Here $\bar{y_j}$ denotes the mean in the jth block. Again, an estimator of d can be obtained using least square estimation procedure. It can be noted here that $d = 0$ implies the slope of the regression equation being -1 which refers to the short memory process.

The parameter d can also be estimated using maximum likelihood methods. For an ARFIMA(p, d, q) process the estimation procedure involves estimating parameter $\theta = (\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ...\theta_q, d)'$. It is based on maximizing the following likelihood under the assumption that Y_t is a zero mean stationary Gaussian process.

$$L(\theta) = \frac{-1}{2}\log det(\Gamma_\theta) - \frac{-1}{2} y'\Gamma_\theta^{-1} y, \text{where } \Gamma_\theta = Var(y) \tag{2.58}$$

The asymptotic properties of MLE estimators are given in Dahlhaus [11] and are as follows:

Theorem 3: Let $\hat{\theta_n}$ be the value that maximises the exact log-likelihood and θ_0 be the true parameter. under some regularity conditions,

(i) Consistency: $\hat{\theta_n} \to \theta_0$ in probabilty as $n \to \infty$.

(ii) Central limit theorem: As $n \to \infty$,

$\sqrt{n}(\hat{\theta_n} - \theta_0) \to N(0, \Gamma^{-1}(\theta_0))$

where

$\Gamma_{ij}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}[\frac{\partial \log f_\theta(\lambda)}{\partial \theta_i}][\frac{\partial \log f_\theta(\lambda)}{\partial \theta_j}]d\lambda,$

and where f_θ is the spectral density of the process Y_t.

2.4.1 Whittle estimation

Another well-known methodology for estimation is based on the approximation of the Gaussian log-likelihood function (Eqn. 2.58) using the Whittle's approximation [36]. The procedure is as follows. The log-likelihood function (Eqn. 2.58) divided by the sample size n is given as

$$L(\theta) = \frac{-1}{2n} \log det(\Gamma_\theta) - \frac{-1}{2n} y' \Gamma_\theta^{-1} y \tag{2.59}$$

where

$$(\Gamma_\theta)_{ij} = \gamma_\theta(i-j) \tag{2.60}$$

$$\gamma_\theta(k) = \int_{-\pi}^{\pi} f_\theta(\lambda) \exp(i\lambda k) d\lambda \tag{2.61}$$

The two terms in the log-likelihood function (Eqn. 2.59) are approximated in order to obtain the Whittle estimates. The approximations are given below and further details can be found in [Palma(2007)].

$$\frac{-1}{2n} \log |\Gamma_\theta| \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(2\pi f_\theta(\lambda)) d\lambda \tag{2.62}$$

$$\frac{-1}{2n} y' \Gamma_\theta^{-1} y \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda \tag{2.63}$$

Hence applying the approximations and substituting the integral by Riemann sums, the discrete log-likelihood function is

$$L_1(\theta) = -\frac{1}{2n} [\sum_{j=1}^{n} \log(f_\theta(\lambda_j)) + \sum_{j=1}^{n} \frac{I(\lambda_j)}{f_\theta(\lambda_j)}] \tag{2.64}$$

The estimates thus obtained by maximizing L_1 share similar asymptotic properties as that of the exact MLE [34], [11]. Also, if the assumption of normality is dropped

then under a few conditions, the Whittle estimates are consistent and asympotically normal distributed [15]. This technique however is dependent on the specification of the spectral density function $f_\theta(\lambda)$. In this regard, there are other semiparametric methods that rely on the shape of the spectral density instead of its parametric form [30]. In addition to the methods discussed above, there are other techniques such as rescaled range statistic (R/S), wavelet-based approach and detrended fluctuation analysis that provide estimates for the memory parameter d.

2.4.2 Estimation using autoregressive expansion

As mentioned in the previous section, the estimates of the ARFIMA coefficients θ can be obtained by maximizing (2.58). This can be computed using Durbin-Levinson or the state space approach with numerical complexity $O(n^2)$. So to speed up the computational speed, we could consider a autorgressive approximation of the process [27]. Since the ARFIMA(p,d,q) process has an infinite AR expansion, we can consider the following truncated version for estimation purposes.

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + ... + \pi_m y_{t-m} + \tilde{\epsilon_t}$$

Hence the approximate MLE estimates can be obtained using a ordinary least squares procedure, i.e. minimising $L_1(\theta)$

$$L_1 = \sum_{t=m+1}^{n} [y_t - \pi_1 y_{t-1} - \pi_2 y_{t-2} + ... - \pi_m y_{t-m}]^2$$

Chapter 3

Modeling long memory eye gaze data

3.1 Introduction

In this chapter, we describe the eye gaze data. We consider modeling eye gaze data of subjects in response to arm movements. We investigate the presence of long memory in the eye gaze data and apply procedures to find estimates of the parameter d. On the basis of long memory properties, we develop prediction models that are dependent on the arm component of movements as well as the past eye gaze data.

3.2 Eye Movement Data

The eye movement data were obtained from the work of Noy [26]. The experiment involved seven individuals referred to as subjects. In total, 10 distinct hand movements were viewed by the subjects under two conditions. The first condition was just to watch the movements and answer a few questions regarding the movement. In the second condition the subjects were asked to watch and then imitate the hand movements as closely as possible. The first was referred to as the "Watch" condition and second as the "Imitate" condition. On average the duration of each movement was 10.13 seconds. The x-y coordinates of the wrist, elbow and shoulder

of the arm of the movement performer and those of the eye of the subjects were recorded. An eye gaze tracker was used in order to get the data for each of the 7 subjects corresponding to the 10 movements. The eye gaze data were sampled at 120Hz (every 8th millisecond). The arm data were taken from a video recorded at 15Hz. In this dataset, the time series length ranges from 801 (6.68s) to 1609 (13.41s) data points.

For application purposes, we consider the (x, y) coordinates as well as the polar coordinates $(r,\ \theta)$ of the eye movement data which were obtained using standard transformations from the x and y components.

The standard transformations are given as follows:

$$r_t = (x_t^2 + y_t^2)^{1/2} \tag{3.1}$$

$$\theta_t = \begin{cases} 0, & \text{if } y_t \geq 0 \\ \arctan(\frac{y_t}{x_t}) + 2\pi, & \text{if } x_t \geq 0 \ \&\ y_t < 0 \\ \arctan(\frac{y_t}{x_t}) + \pi, & \text{if } x_t < 0 \\ \frac{\pi}{2}, & \text{if } x_t = 0 \ \&\ y_t > 0 \\ \frac{3\pi}{2}, & \text{if } x_t = 0 \ \&\ y_t < 0 \\ 0 & \text{if } x_t = 0 \ \&\ y_t = 0 \end{cases} \tag{3.2}$$

For purpose of analysis, we consider x_t, y_t, r_t and θ_t as univariate time signals.

3.3 Eye data modeling

In this section, we consider modeling the x- coordinate of eye gaze data. Let this time series data be denoted as x_t. For exploratory purposes, we first consider a simple model (Eqn. 3.8) which chooses as regressors the x- coordinates of wrist (w_t^x), elbow(e_t^x) and shoulder (sh_t^x) only,

$$\text{(M1)} \quad x_t = \beta_0 + \beta_1 w_t^x + \beta_2 e_t^x + \beta_3 sh_t^x + \epsilon_t. \tag{3.3}$$

For regression purposes, we truncated the first 50 time points to eliminate transient effects at the beginning of each signal. The estimates of coefficient in model M1 are obtained using least square fitting. Figure (3.1) shows the estimated coefficients $\hat{\beta}_1$ for one of the movements for both the conditions.

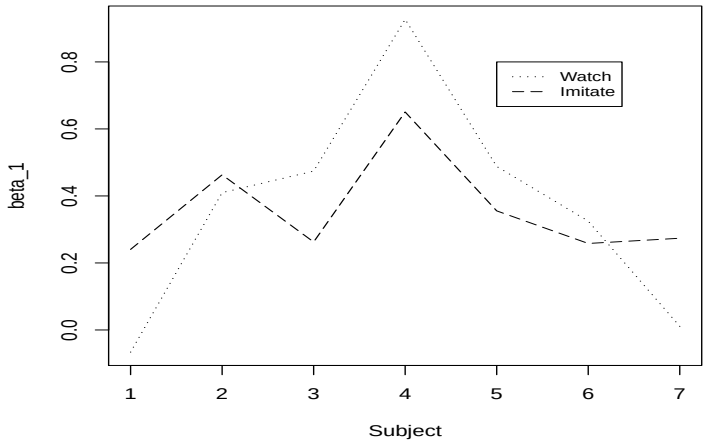

Figure 3.1: Estimated wrist coefficients $\hat{\beta}_1$ for seven subjects for "Watch" and "Imitate" conditions

Now let us consider a particular case (say Subject 4 movement 1 under the

"Watch" condition) which is a representative case. We perform the least square fit and obtained goodness-of-fit statistics and residuals. The sample autocorrelation function of the residuals of the linear regression fit is given in Figure (3.2). The plot clearly does not show an exponential decay and also shows a significant dependence even at lag greater than 30. In addition, the plot of the residuals variances for 50 blocks based on Equation (2.56) indicate presence of long range dependence (slope = -0.14($\neq$ -1))[Figure (3.3)]. This implies that the above regression model is not adequate and that the long memory structure has not been taken into account. Also, a stationarity check was performed on the regression residuals using pp.test and kpss.test in R software. For this case, the alternative hypothesis of stationarity was accepted in the case of pp.test and the null hypothesis of stationarity was rejected in kpss.test. It was observed that such conflicting results did hold for most of the cases which indicates that neither of $d = 0$ ($I(0)$) nor $d = 1$ ($I(1)$) processes can be used to model the residual series [2], [22]. Hence, a fractional value for parameter d such that $0 < d < 1$ may provide a better description of the data. The value of adjusted R^2 was 0.64. Figure (3.4) shows the fit of the model. This fit can be improved further. Figure (3.5) gives the distribution of R^2 values across all cases when simple linear regression (3.8) was performed.

3.3.1 Long memory in eye data

From the previous section, we have that the residuals obtained from the simple linear regression on the arm components indicate long memory behaviour. In this

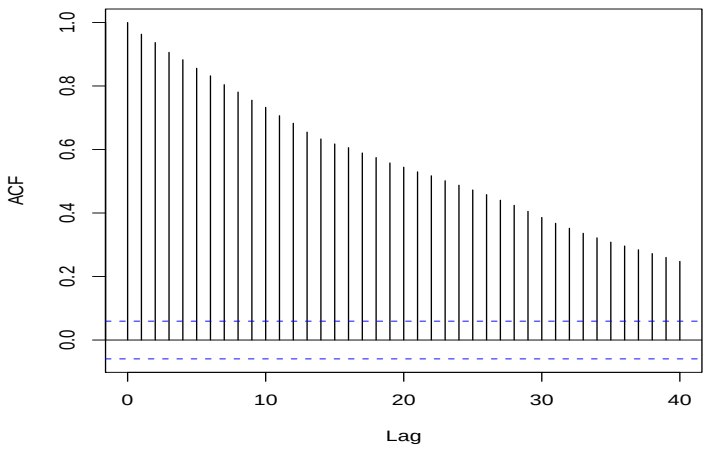

Figure 3.2: Autocorrelation of S4M1W residuals

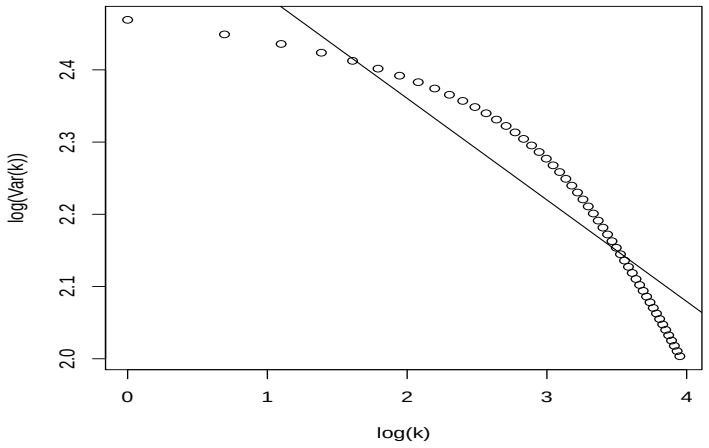

Figure 3.3: Slope = (-0.05), S4M1W residuals

section, we study in detail the long memory characteristics of residuals for all the cases under the two conditions. We note here that the linear regression residuals are not Gaussian distributed. This is supported by the histogram plot, qqplot and various tests for normaility. Figure (3.6) shows the histogram for a particular case.

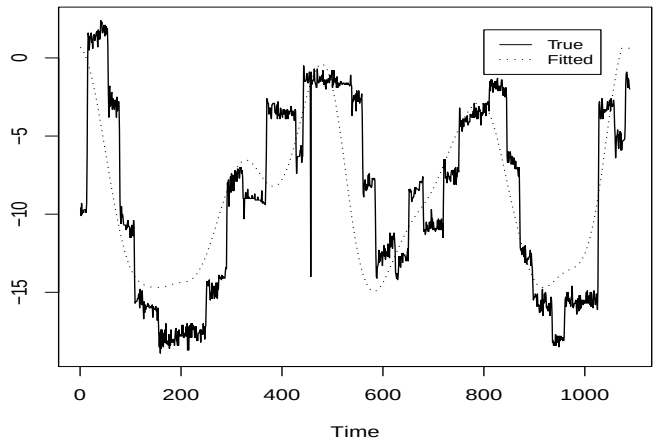

Figure 3.4: True and fitted values for Subject 4, Movement 1 "Watch" condition.

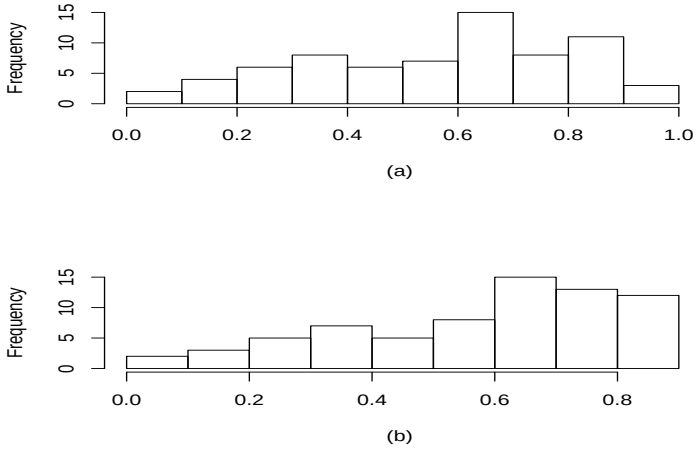

Figure 3.5: Distribution of Adjusted R-square values for (a) "Watch" and (b) "Imitate conditions". Each of the two conditions have 70 cases.

Specifically, Pearson Chi-square test rejects the null hypothesis of Gaussianity at 1% siginicance level for all the 140 cases.

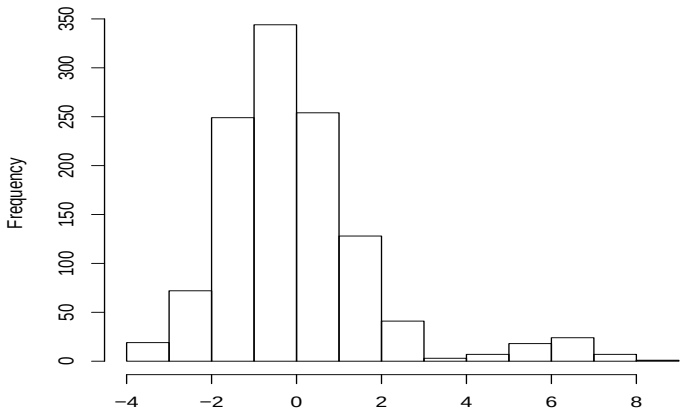

Figure 3.6: Histogram of linear regression residuals of Subject 3 Movement 4

Now, to estimate the memory parameter d we employ the methods mentioned in the previous chapter. Tables (3.3.1) and (3.3.1) summarizes the spectral regression estimates (GPH) of d for various values of α. In addition, we also report the estimates obtained by the Whittle's method and variance regression method (VarTest). For variance regression, block sizes of $m = 20, 50$ were considered. The Whittle estimates were computed using the FDWhittle in R. The estimates in the tables are the average for each subject across 10 movements. The standard deviation are also reported in parenthesis. Both tables suggests the presence of long memory with values of the estimate d being > 0. It is interesting to note here that though the mean parameter values differ across the six estimation methods, they are broadly consistent across the two conditions "Watch" and "Imitate".

Table 3.1: Average $\hat{d}$ and their corresponding standard deviations in parenthethis from Spectral regression, Variance regression and Whittle method under "Watch" condition

Subject	GPH($\alpha = 0.4$)	GPH($\alpha = 0.5$)	GPH($\alpha = 0.6$)	VarTest(20))	VarTest(50))	Whittle
1	0.23	0.52	0.75	0.40	0.45	0.29
	(0.24)	(0.16)	(.08)	(.03)	(.02)	(0.05)
2	0.71	0.85	0.97	0.45	0.48	0.34
	(0.34)	(0.17)	(0.10)	(0.02)	(0.01)	(0.16)
3	0.47	0.71	0.82	0.41	0.46	0.32
	(0.23)	(0.11)	(0.05)	(0.04)	(0.02)	(0.11)
4	0.49	0.66	0.79	0.42	0.46	0.26
	(0.27)	(0.25)	(0.12)	(0.03)	(0.02)	(0.10)
5	0.52	0.76	0.87	0.43	0.47	0.47
	(0.32)	(0.18)	(0.12)	(0.03)	(0.01)	(0.08)
6	0.43	0.74	0.86	0.43	0.47	0.17
	(0.22)	(0.12)	(0.05)	(0.03)	(0.01)	(0.09)
7	0.34	0.72	0.84	0.43	0.47	0.44
	(0.31)	(0.15)	(0.09)	(0.02)	(0.01)	(0.09)

3.3.2 Modeling long memory

To account for the long memory errors in the regression equation, we consider fitting an ARFIMA (p, d, q) model to the residuals. The estimates of the fractional parameter d, autoregressive order p and moving average parameter q can be computed using the maximum likelihood methods. As mentioned in previous section, we can use an alternative estimation method based on the representation of a long memory process by an infinite autoregressive expansion under certain conditions. Since the length of each time series in the dataset is finite, we consider fitting a AR process of finite order $(\tilde{p})$ to the regression residual process $\hat{\epsilon}_t$. The residuals obtained from fitting the truncated AR($\tilde{p}$) model did not show any long range dependence. The autocorrelation plot, the cumulative periodogram plot along with

Table 3.2: Average $\hat{d}$ and their corresponding standard deviations in parenthethis- from Spectral regression, Variance regression and Whittle method under "Imitate" condition

Subject	GPH($\alpha = 0.4$)	GPH($\alpha = 0.5$)	GPH($\alpha = 0.6$)	VarTest(20))	VarTest(50))	Whittle
1	0.22	0.53	0.72	0.39	0.45	0.28
	(0.27)	(0.20)	(0.13)	(0.04)	(0.02)	(0.09)
2	0.54	0.74	0.83	0.44	0.47	0.36
	(0.23)	(0.16)	(0.12)	(0.02)	(0.01)	(0.08)
3	0.50	0.70	0.78	0.42	0.46	0.35
	(0.26)	(0.15)	(0.07)	(0.03)	(0.01)	(0.10)
4	0.43	0.76	0.80	0.42	0.46	0.31
	(0.28)	(0.17)	(0.10)	(0.03)	(0.01)	(0.07)
5	0.77	0.83	0.92	0.45	0.48	0.46
	(0.28)	(0.16)	(0.09)	(0.02)	(0.01)	(0.09)
6	0.33	0.56	0.85	0.40	0.46	0.29
	(0.20)	(0.21)	(0.10)	(0.04)	(0.02)	(0.10)
7	0.42	0.77	0.90	0.43	0.47	0.56
	(0.24)	(0.15)	(0.16)	(0.03)	(0.01)	(0.13)

Portmanteau tests confirmed the whiteness of these residuals for all cases. For one particular case the plots are given in Figures 3.7 and 3.8.

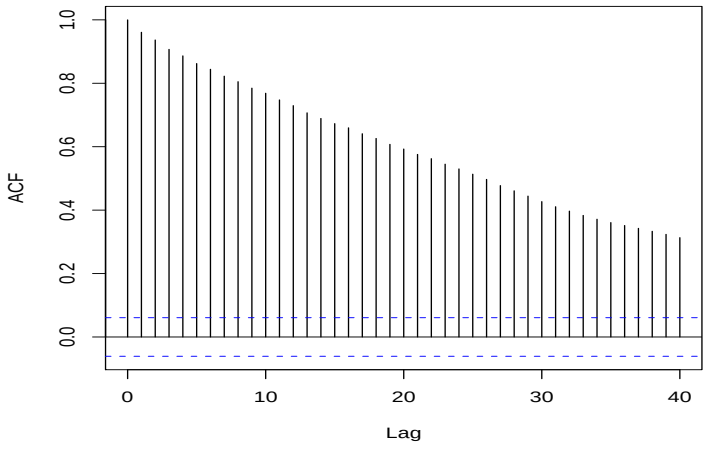

Figure 3.7: Autocorrelation plot for truncated AR process residuals

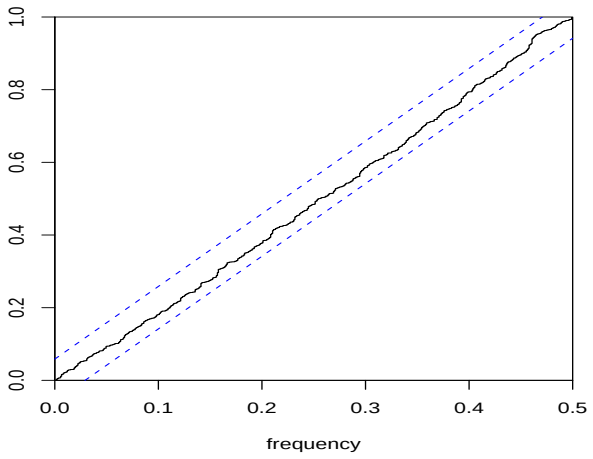

Figure 3.8: Cumulative periodogram plot for truncated AR process residuals

Hence we have a two step procedure, first estimating the coefficients of the covariates β_i and then estimating coefficients α_p of truncated autoregressive model for the residuals.

$$x_t = \beta_0 + \beta_1 w_t^x + \beta_2 e_t^x + \beta_3 sh_t^x + \epsilon_t \tag{3.4}$$

$$\epsilon_t = \sum_{j=1}^{\tilde{p}} \alpha_j \epsilon_{t-j} + \eta_t \tag{3.5}$$

Now to further simiplify the computations for fitting and prediction purposes, we could also consider the following model instead of the two step procedure as mentioned above.

$$x_t = c_0 + \sum_{i=0}^{p1} c_{1i} w_{t-i}^x + \sum_{i=0}^{p2} c_{2i} e_{t-i}^x + \sum_{i=0}^{p3} c_{3i} sh_{t-i}^x + \sum_{i=1}^{p} \alpha_i x_{t-i} + \eta_t \tag{3.6}$$

Solving for the error terms ϵ_t in and replacing ϵ_t, ϵ_{t-1}.. consecutively in the error model 3.5, we obtain the above model 3.6 with $p1 = p2 = p3 = p$. This

model form now involves the lagged covariates $w^x_{t-i}, e^x_{t-i}, sh^x_{t-i}$. The order of the lags $(p, p1, p2, p3)$ is large to account for the long memory behaviour. The estimates of the coefficients can be obtained from least square fitting procedure. The orders $(p, p1, p2, p3)$ used in regressing the eye data where chosen to be $(12, 4, 4, 4)$. The choice of the values of the lag orders will be discussed in the next section. A stepwise linear regression was performed in R to get the optimal model. *The approximations used in this modeling approach provided remarkable fits.* As an example Figure (3.9) depicts the fit for a particular case. The R^2 value for this case was 0.977. Across all cases, the distribution of the adjusted R^2 values were very high as shown in Fig (3.10). In order to investigate the performance of the optimal model so obtained, we compare the optimal model with other model forms in the next section.

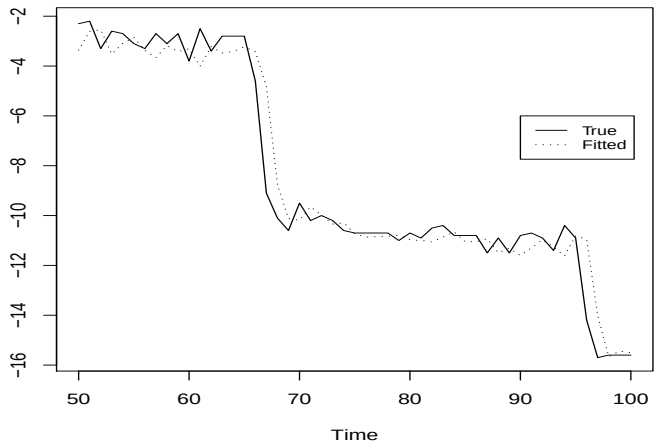

Figure 3.9: True vs fitted values for Subject 4 Movement 1 ("Watch").

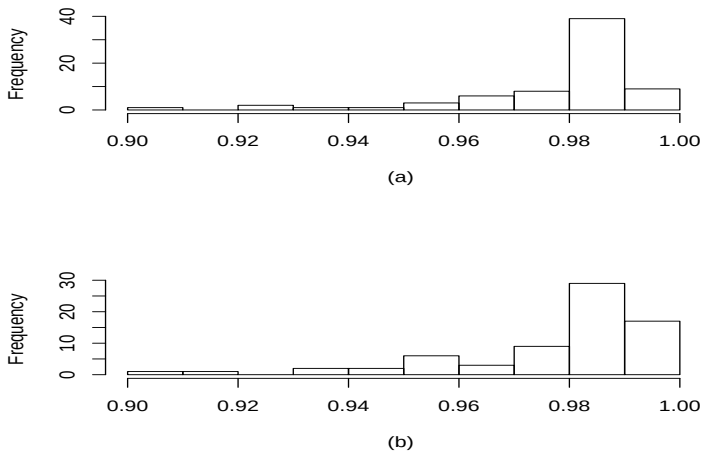

Figure 3.10: Distribution of Adjusted R-square values for (a) "Watch" and (b) "Imitate" conditions. Each condition has 70 cases.

3.3.3 Comparison of models

In this section, we specifically look at model forms which have been mentioned in the previous sections and compare them with other models using goodness-of-fit and prediction statistics. The model forms that will be considered here are given below.

$$Model1 : x_t = \beta_0 + \beta_1 w_t^x + \beta_2 e_t^x + \beta_3 sh_t^x + \epsilon_t \tag{3.7}$$

$$Model2 : x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} \tag{3.8}$$

$$Model3 : x_t = c_0 + \sum_{i=0}^{p1} c_{1i} w_{t-i}^x + \sum_{i=0}^{p2} c_{2i} e_{t-i}^x + \sum_{i=0}^{p3} c_{3i} sh_{t-i}^x + \sum_{i=1}^{p} \alpha_i x_{t-i} + \eta_t \tag{3.9}$$

Model 1 is the simple linear regression model which only depends on the arm input. Model 2 considers long memory behaviour of the process in terms of its au-

toregressive expansion but does not include the covariates wrist, elbow and shoulder. Model 3 combines the effect of the covariates and the long memory characteristic of the data. The estimation of coefficients in each case is obtained through least squares. Along with the three models, we also consider one other form that is derived from Model 3. Let Model 4 be the optimal model obtained as a result of stepwise regression in Model 3 using Akaike information criterion (AIC). Stepwise regression identifies the most significant covariates and therefore the number of covariates in Model 4 is less than or equal to the number of covariates in Model 3. The lag parameter p for Model 2 was also selected by minimizing the AIC criterion. The lag orders $(p, p1, p2, p3)$ used in Model 3 where chosen to be $(12, 4, 4, 4)$. This is due to the fact this particular order sequence showed consistently good results across all the 140 cases and that any orders higher than $(12, 4, 4, 4)$ did not result in a significantly better fit or prediction performance improvement. In general, Model 3 can be compared with various other models by increasing the value of p, p1, p2 or p3 which can result in infinite number of possibilities. As an example, comparison is presented with Model 5 which has a lag order (13, 4, 4, 4). All these models are compared using the AIC, R^2 and the residual standard error (Resid SE). The prediction performance is compared using the averaged absolute prediction error statistic (AAPE) and the averaged square prediction error (ASPE). The computation formulas for prediction errors are shown in Eqn. (3.10) and (3.11)

$$AAPE(n) = \frac{\sum_{t=1}^{n} |x_t - \hat{x_t}|}{n} \tag{3.10}$$

$$ASPE(n) = \frac{\sum_{t=1}^{n} (x_t - \hat{x_t})^2}{n} \tag{3.11}$$

where x_t is the true value and $\hat{x}_t$ is the predicted value at time t. The results of comaparison of these 5 models using the fit and prediction statistics is illustrated through the examples below.

In our first example to compare the models, we again look at modeling the x-coordinate of Subject 4, movement 1 under the "Watch" condition. The length of this time series data is 1089. The data points from 51 to 1000 were considered for modeling purposes. The last 89 data points were used in prediction comparisons. Table (3.3) shows the results from fitting models 1 to 5 to the data. As indicated from the Table (3.3), Model 1 clearly doesnot describe the data very well. Lower R^2, and higher AIC, residual standard error was observed in this case. Now in Model 2, considering only the long memory characteristic of the data gave a very good fit in terms of a high R^2 value but failed in terms of prediction as the error AAPE value was quite high. *This indicates that both the arm covariate and the long memory structure need to be taken into account.* This is supported by the results for Models 3,4 and 5. Clearly, with large R^2 values and small prediction errors, any one of the models could serve as the final model. Since Model 4 is obtained as a result of stepwise regression, it is the optimal model and is for this case given by,

$$x_t = x_{t-1}+x_{t-2}+x_{t-3}+x_{t-4}+x_{t-7}+x_{t-8}+w^x_{t-2}+w^x_{t-3}+w^x_{t-4}+e^x_{t-2}+e^x_{t-3}+\eta_t \quad (3.12)$$

The values of AAPE for this model is comparable with that of Models 3 and 5. It is also important to note here that the results of Model 5 were not significantly different than that of Model 3. Hence, we do not consider lag orders greater than $(p, p1, p2, p3)$ for fitting and prediction purposes.

Table 3.3: Comparison between models for Subject 4 Movement 1 ("Watch" Condition)

Model	AIC	R^2	Resid SE	AAPE	ASPE
Model 1	5001.7010	0.6561	3.3550	4.0500	21.1550
Model 2	2352.1490	0.9772	0.8529	7.0550	69.6380
Model 3	2574.0260	0.9784	0.8770	3.3810	22.2200
Model 4	2552.2320	0.9786	0.8743	3.4908	24.7260
Model 5	2569.7850	0.9785	0.8757	3.3838	21.8930

The residuals of the optimal model reaffirm whiteness (see Figure (3.11)). Also, a closer look at the histogram (Figure (3.12)) and qqplot (Figure (3.13)) suggests that the residuals are not normally distributed. In addition, the first two lagged terms x_{t-1}, x_{t-2} in the optimal model Eqn. (3.12) play an significant role, exclusion of which leads to poor fit and predictions. The residuals of Model 4 excluding x_{t-1}, x_{t-2} terms does not behave as a white noise process. Figure (3.14) shows the cumulative periodogram of the residuals of Model 4 excluding x_{t-1}, x_{t-2} terms.

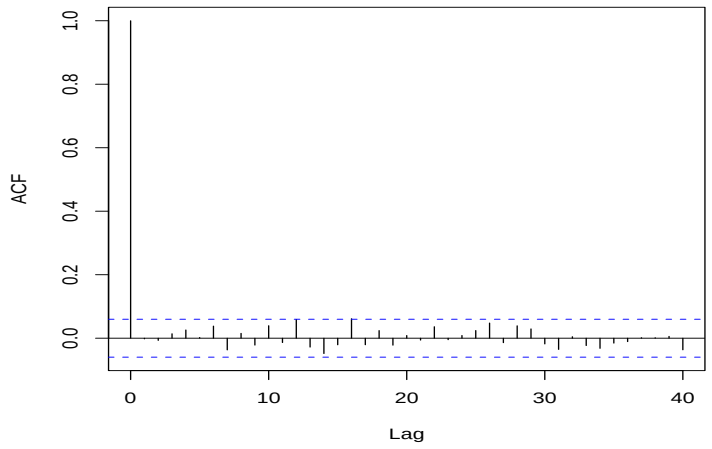

Figure 3.11: S4W1: Autocorrelation plot for Model 4 (optimal) residuals

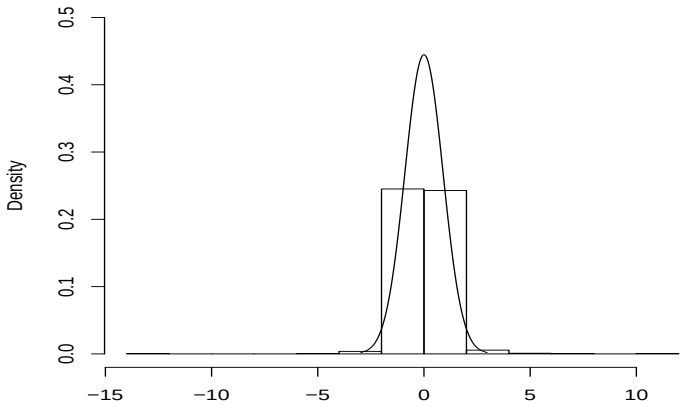

Figure 3.12: Histogram plot for Model 4 (optimal) residuals vs N(0,0.897)

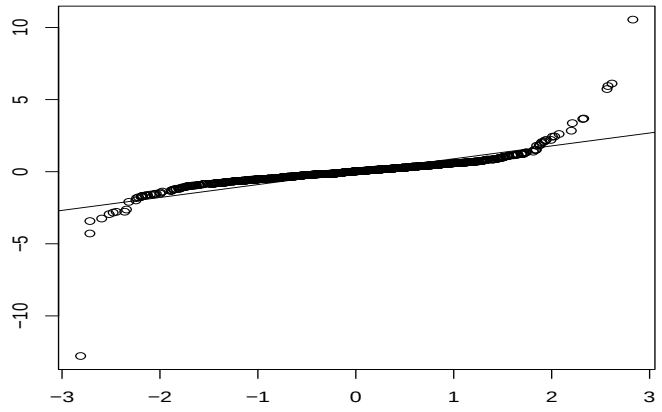

Figure 3.13: QQ- plot of Model 4 (optimal) residuals vs N(0,0.897)

Similar results were observed for all the other cases. As a second example, we compare the five models for the case of Subject 5, movement 9 under the "Watch" condition. Table (3.4) summarizes the results. Comparing the five models, Model 4 has the lowest AIC and lowest residual standard error. The R^2 value is highest

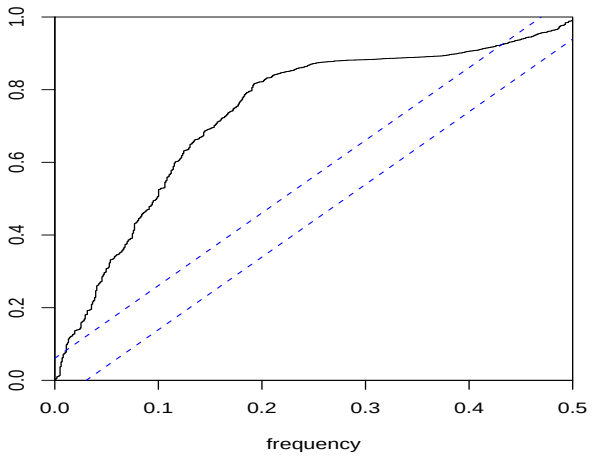

Figure 3.14: Cumulative periodogram of residuals from Eqn. (3.12) excluding (x_{t-1}, x_{t-2})

across the models but is comparable to Models 2, 4 and 5. Now comparing the values of the average absolute prediction error (AAPE), we see the Model 4 performs the best. In terms of the average squared prediction error (ASPE), Model 4 is not the lowest but comparable with corresponding values of Models 3 and 6. Hence, Model 4 was the optimal model and is given by,

$$\begin{aligned} x_t &= x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4} + x_{t-6} + x_{t-7} + x_{t-11} + w_{t-2} + w_{t-3} \quad (3.13) \\ &+ e_{t-2} + e_{t-3} + sh_t + sh_{t-1} + sh_{t-3} + \eta_t \end{aligned}$$

Again the residuals of Model 4 behave as a white noise process (Figure (3.15)). The histogram plot of residuals in Figure (3.16) and qq-plot in Figure (3.17) suggests non-normal distribution. This was observed for residuals across all fits.

In both Tables (3.5) and (3.6) compare Models 1 to 5 for two cases under "Imitate condition". Results in these tables are in line with the discussion above.

Table 3.4: Comparison between models for Subject 5 Movement 9 ("Watch" Condition)

Model	AIC	R^2	Resid SE	AAPE	ASPE
Model 1	6209.8360	0.4694	2.1500	1.2810	2.4659
Model 2	1685.9260	0.9781	0.4378	2.2426	7.3191
Model 3	1695.4720	0.9783	0.4350	0.8448	1.0475
Model 4	1674.6030	0.9784	0.4338	0.8350	1.1410
Model 5	1697.4860	0.9783	0.4350	0.8421	1.0680

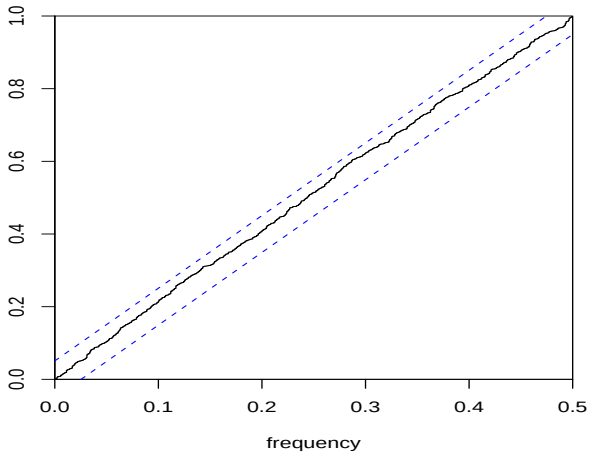

Figure 3.15: S5W9: Cumulative periodogram plot for Model 4 (optimal) residuals

In both the Tables, again Model 4 is the optimal model.

3.4 Conclusion

In conclusion, the analysis presented in this chapter points to the presence of a long memory component in eye gaze data. We have shown that for modeling and prediction purposes, both the arm component and the long autoregressive component should be taken into account. The exclusion of the arm component results in

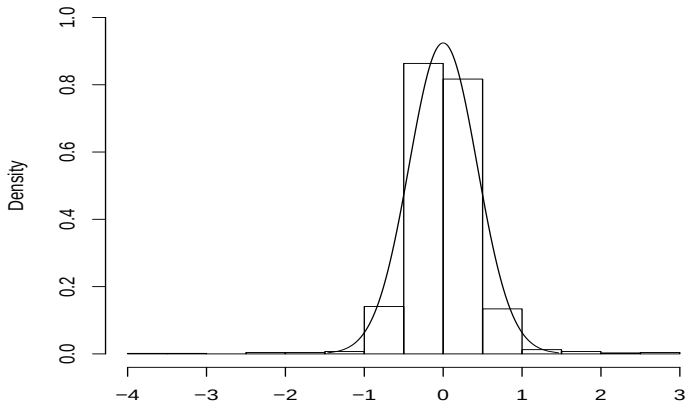

Figure 3.16: Histogram plot for Model 4 (optimal) residuals vs N(0,0.432)

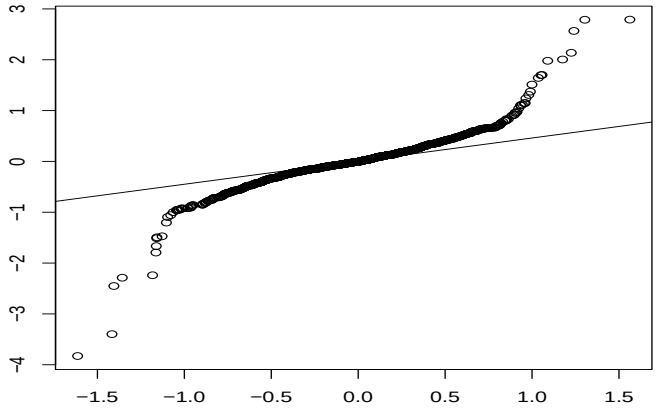

Figure 3.17: QQ- plot of Model 4 (optimal) residuals vs N(0,0.432)

bad prediction errors and excluding the long memory part resulted in poorer fits. *In particular, the first two lag terms in the long memory part play a significant role and is included in all the models. Hence previous two time points i.e. information in the last 8 to 16 milliseconds is crucial for predicting future eye gaze positions.*

Table 3.5: Comparison between models for Subject 5 Movement 8 ("Imitate" Condition).

Model	AIC	R^2	Resid SE	AAPE	ASPE
Model 1	3178.6810	0.6361	2.4440	33.844	22.229
Model 2	784.7760	0.9888	0.4282	47.627	45.172
Model 3	795.8265	0.9891	0.4230	33.647	24.973
Model 4	771.4968	0.9892	0.4203	32.964	24.083
Model 5	797.2301	0.9891	0.4232	34.206	25.478

Table 3.6: Comparison between models for Subject 6 Movement 8 ("Imitate" Condition).

Model	AIC	R^2	Resid SE	AAPE	ASPE
Model 1	2980.999	0.6153	2.1160	2.032	4.365
Model 2	873.4108	0.9822	0.4570	2.509	11.497
Model 3	895.922	0.9829	0.4550	2.459	10.464
Model 4	868.979	0.9825	0.4516	2.439	10.178
Model 5	894.2741	0.9823	0.4542	2.417	9.710

Chapter 4

HOC for long memory processes

Higher order crossings (HOC) for the case of a discrete time signal are defined as the zero-crossing counts of filtered versions of time series. A detailed study of HOC and their statistical properties can be found in [19]. In this section, we are particularly interested in HOC sequences corresponding to the difference filter referred to as the *simple* HOC. We investigate properties of HOC when a long memory ARFIMA process is considered. We introduce a method to estimate the long memory parameter d using zero-crossing count. Further, we suggest an algorithm to estimate paramters in an ARFIMA(p, d, q) process. Finally, we discuss the application of HOC in discrimination of the eye gaze data.

4.0.1 Simple HOC

Let us consider $Z_1, Z_2, ..., Z_N$ to be a real-valued zero mean stationary time series of length N. Since $Z_t, t = 1, 2, ..., N$ is a discrete time series oscillating about level zero, we can count the number of sign changes to which we refer as "zero-crossings". This count can be considered as a measure of oscillation exhibited by a time series. A simple way to obtain the zero-crossing count is from a clipped binary

process X_t given by the nonlinear transformation

$$X_t = \begin{cases} 1 & \text{if } Z_t \geq 0 \\ \\ 0 & \text{if } Z_t < 0 \end{cases} \tag{4.1}$$

The zero-crossing count denoted by D, is defined in terms of X_t as

$$D = \sum_{t=1}^{N} [X_t - X_{t-1}]^2 \tag{4.2}$$

This process of finding the zero-crossings can be repeated subsequently on filtered versions of the original time series Z_t. A detailed study of the zero-crossings counts obtained from the application of various filters can be found in [19], [23]. In this section, we will consider taking repeated differences (of a time series) as a family of filters.

Let ∇ be the difference operator defined by Eqn. (4.4).

$$\nabla Z_t \equiv Z_t - Z_{t-1} \tag{4.3}$$

The second difference is then

$$\nabla(\nabla Z_t) \equiv \nabla^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \tag{4.4}$$

In general, for $k = 0,\ 1,\ 2,...,\ \nabla^k Z_t$ is given by

$$\nabla^k Z_t = \sum_{j=0}^{k} \binom{k}{j} (-1)^j Z_{t-j} \tag{4.5}$$

where $\nabla^0 Z_t \equiv Z_t$. For each k, we further obtain the binary clipped process $X_t(k)$ and the corresponding HOC counts D_k using Eqs. (4.6) and (4.7), respectively,

$$X_t(k) = \begin{cases} 1, & \text{if } \nabla^{k-1} Z_t \geq 0 \\ \\ 0, & \text{if } \nabla^{k-1} Z_t < 0 \end{cases} \tag{4.6}$$

$$D_k = \sum_{t=1}^{N} [X_t(k) - X_{t-1}(k)]^2 \tag{4.7}$$

It is important to note here that the HOC sequence D_1, D_2,... is obtained through very simple calculations and also are interpretable in terms of the visual features of the time series $\{Z_t\}$. D_1 by definition is the zero-crossing count, D_2 essentially is the number of peaks and troughs, D_3 is the number of inflection points in the original time series $\{Z_t\}$ and so on. In addition to this property, there are other useful results some of which are stated below.

1. For any k, we have $0 \leq D_k \leq N-1$.

2. Let Z_t, $t = 1, 2, 3, ..., N$, be a zero-mean stationary process. Then regardless of spectrum type, $0 \leq E[D_1] \leq E[D_2] \leq ..... \leq N-1$.

3. (HOC Theorem) Let Z_t, $t = 1, 2, 3, ..., N$, be a zero-mean stationary process, and assume that π is included in the spectral support. Then,
   (a) $X_t(k) \Rightarrow \begin{cases} ...01010101...., & \text{with probability } \frac{1}{2} \\ ...10101010...., & \text{with probability } \frac{1}{2} \end{cases}$
   (b) $\lim_{k\to\infty} \lim_{N\to\infty} (D_k/N-1) = 1$ with probability 1.

4. For a zero-mean Gaussian process Z_t, $t = 1, 2, 3,...., N$, the sequence $E[D_k]$ determines the spectrum upto a constant.

5. For a stationary Gaussian process or a stationary random sinusoid process $\{Z_t\}$, the relationship between first order autocorrelation ρ_1 and $E(D)$ is given as the cosine formula,

$$\rho_1 = \cos(\frac{\pi E(D)}{N-1}) \tag{4.8}$$

From the above results it is evident that the HOC sequence D_k is monotone and bounded, and as k increases $E(D_k)$ eventually attains an upper bound. In addition, the rate of increase of the D_k is fast for small values of k, however, as k increases the rate of increase becomes negligible. This fact is supported by a fundamental theoretical result, the higher order crossing theorem (HOC Theorem). The HOC theorem mainly tells us that under a mild assumption the binary clipped series $X_t(k)$ obtained from subsequent differenced series $\nabla^k Z_t$ converges to a steady state of 01010101... or 1010101010... as k increases. For most processes the steady state is approached quite rapidly. Hence for moderate and large values of k, the oscillatory features D_k of any two processes will not be very different. On the other hand, the discrimination can be carried out by calculating very few early D_k. In practice, $D_1, ..., D_8$ have been found useful for discrimination purposes.

4.1 HOC of long memory processes

In this section, we look at the application of simple HOC on long memory processes and study. Recall from Chapter 1, the spectral density of a stationary ARFIMA (p, d, q) process $\{Y_t\}$ is given by,

$$f_{ARFIMA}(\omega) = \left|1 - e^{(i\omega)}\right|^{-2d} f_{ARMA}(\omega) \tag{4.9}$$

where $f_{ARMA}(\omega)$ is obtained from Eqn. (2.29). Specifically for long memory processes, consider $d \in (0, 0.5)$. As mentioned in the previous section, application of the difference operator (Eqn. (4.4)) on a time series $\{Y_t\}$ gives the output series

as $\nabla Y_t = Y_t - Y_{t-1}$. Since difference operator from Eqn. (4.4) is a linear filter, its transfer function $h(\omega)$ and the squared gain $|h(\omega)|^2$ is given by,

$$h(\omega) = 1 - \exp(-i\omega) \tag{4.10}$$

$$|h(\omega)|^2 = |1 - \exp(-i\omega)|^2 \tag{4.11}$$

Hence the spectral density $f_{\nabla Y_t}$ of the output process ∇Y_t is,

$$\begin{aligned} f_{\nabla Y_t} &= |h(\omega)|^2 f_{ARFIMA}(\omega) \\ &= |1 - \exp(-i\omega)|^2 f_{ARFIMA}(\omega) \\ &= |1 - \exp(-i\omega)|^2 \left|1 - \exp^{(-i\omega)}\right|^{-2d} f_{ARMA}(\omega) \\ &= |1 - \exp(-i\omega)|^{2-2d} f_{ARMA}(\omega) \end{aligned} \tag{4.12}$$

Since for a stationary long memory processes $\{Y_t\}$, the fractional parameter $d \in (0, 0.5)$. This implies $2 - 2d > 0$. Hence the spectral density of the process $\{\nabla Y_t\}$ does not have a pole at zero frequency. In fact, at zero frequency the spectral density function of $\{\nabla Y_t\}$ has zero mass. Further if the highest positive frequency in the spectral support is ω^*, then the application of higher order difference operator ∇^j such that $j \to \infty$ will push the power towards that frequency ω^*. For a Gaussian long memory process, the following result holds.

Theorem 4: Suppose $\{Y_t\}$ is a zero-mean stationary long memory Gaussian ARFIMA(p, d, q) process and let ω^* be the highest frequency in the spectral support. Then, the sequence of simple expected normalised HOC $\{\pi E(D_j)/(N-1)\}$ converges to the highest frequency ω^* as $j \to \infty$.

Proof: Define $\nu(.)$, $j = 0, 1, 2, ...,$ by

$$\nu_j(d\omega) = \frac{\sin^{2j}(\omega/2)dF(\omega)}{\int_{\pi}^{-\pi} \sin^{2j}(\lambda/2)dF(\lambda)} \tag{4.13}$$

where F is the spectral distribution of $\{Y_t\}$ and $\omega \in (-\pi, \pi)$. $\nu_j(.)$ is the probability measure on $[-\omega^*, \omega^*]$ such that

$$\nu_j[-\omega^*, \omega^*] = 1 \tag{4.14}$$

Given that the spectral density exists for the ARFIMA process $\{Y_t\}$, it has the following form,

$$\begin{aligned} f_{ARFIMA}(\omega) &= \left|1 - e^{i\omega}\right|^{-2d} f_{ARMA}(\omega) \\ &= |2\sin(\omega/2)|^{-2d} f_{ARMA}(\omega) \end{aligned}$$

and,

$$dF(\omega) = f_{ARFIMA}(\omega)d\omega \tag{4.15}$$

where $f_{ARMA}(\omega)$ is as described before the spectral density of the ARMA component and it has no pole at zero frequency.

Consider,

$$\nu_j[0, \omega^* - \epsilon] = \frac{\int_{\omega^*-\epsilon}^{0} \sin^{2j}(\omega/2)dF(\omega)}{\int_{\pi}^{-\pi} \sin^{2j}(\lambda/2)dF(\lambda)} \tag{4.16}$$

or,

$$\begin{aligned} \nu_j[0, \omega^* - \epsilon] &= \frac{\int_{\omega^*-\epsilon}^{0} \sin^{2j}(\omega/2) f_{ARFIMA}(\omega)d\omega}{\int_{\pi}^{-\pi} \sin^{2j}(\lambda/2) f_{ARFIMA}(\lambda)d\lambda} \\ &= \frac{\int_{\omega^*-\epsilon}^{0} \sin^{2j-2d}(\omega/2) f_{ARMA}(\omega)d\omega}{\int_{\pi}^{-\pi} \sin^{2j-2d}(\lambda/2) f_{ARMA}(\lambda)d\lambda} \end{aligned} \tag{4.17}$$

Since $\sin(\omega/2)$ is monotonically increasing on $[0,\pi]$ and $2j-2d>0$ for all $j=1,2,...,$ we have for $j\to 0$,

$$\begin{aligned}\nu_j[0,\omega^*-\epsilon] &\leq \frac{\int_{\omega^*-\epsilon}^{0}\sin^{2j-2d}(\omega/2)f_{ARMA}(\omega)d\omega}{\int_{\omega^*-\epsilon/2}^{\omega^*}\sin^{2j-2d}(\lambda/2)f_{ARMA}(\lambda)d\lambda}\\ &\leq \left|\frac{\sin(\frac{\omega^*-\epsilon}{2})}{\sin(\frac{\omega^*-\epsilon/2}{2})}\right|^{2j-2d}\frac{\int_{\omega^*-\epsilon}^{0}f_{ARMA}(\omega)d\omega}{\int_{\omega^*-\epsilon/2}^{\omega^*}f_{ARMA}(\lambda)d\lambda}\\ &\to 0\end{aligned}$$

By symmetry of $\nu_j(.)$, as $j\to 0$,

$$\nu_j[-\omega^*+\epsilon,\omega^*-\epsilon]\to 0$$

From (4.14) and (4.18), as $j\to 0$,

$$\nu_j[-\omega^*,-\omega^*+\epsilon)\cup(\omega^*-\epsilon,\omega^*]\to 1$$

As $j\to 0$,

$$\nu_j \Rightarrow \frac{1}{2}\delta_{-\omega^*}+\frac{1}{2}\delta_{\omega^*}$$

where δ_u is point mass at u.

Using the definition of ν_j, a generalization of the cosine formula in Eqn. (4.8) is given by,

$$\rho_1(j+1)=\int_{\pi}^{-\pi}\cos(\omega)\nu_j(d\omega)=\cos(\frac{\pi E[D_{j+1}]}{N-1}) \tag{4.18}$$

Using (4.18), as $j\to 0$,

$$\cos(\frac{\pi E[D_{j+1}]}{N-1})\to\cos(\omega^*) \tag{4.19}$$

This implies that sequence of simple expected normalised HOC $\{\pi E(D_j)/(N-1)\}$ converges to the highest frequency ω^* as $j\to\infty$.

Now, given that π is in the spectral support, the first autocorrelation of the process $\nabla^j Y_t$ converges to -1 ($\cos(\pi)$) for large j. This perfect negative autocorrelation implies that the process tends to change sign at every time point. This further implies that the clipped binary sequence given in (4.7) tends have either of the forms ....0101010101..... or ....1010101010..... each with probability 0.5. Rigorous proof of HOC Theorem is given in [20].

4.2 Estimation of ARFIMA parameters

In this section, we estimate the parameters of the ARFIMA(p, d, q) process using the zero-crossing count (Eqn. (4.2)). An estimator for the fractional parameter is developed by taking into account the relationship between zero-crossing count and the first order autocorrelation as given in Eqn. (4.8).

Particularly for an ARFIMA$(0, d, 0)$ process the autocovariance function is given by,

$$\gamma(h) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(1+h-d)} \tag{4.20}$$

Using Eqn. (4.20), the autocorrelation function at lag h is

$$\rho(h) = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(1+h-d)} \tag{4.21}$$

For lag $h = 1$,

$$\begin{aligned}\rho(1) &= \frac{\Gamma(1-d)}{\Gamma(d)}\frac{\Gamma(1+d)}{\Gamma(2-d)} \\ &= \frac{\Gamma(1-d)\Gamma(1+d)}{\Gamma(d)(1-d)\Gamma(1-d)} \\ &= \frac{\Gamma(1-d)d\Gamma(d)}{\Gamma(d)(1-d)\Gamma(1-d)} \\ &= \frac{d}{1-d}\end{aligned} \tag{4.22}$$

Hence by equating the first order autocorrelation (4.8), (4.22) and solving for the long memory parameter d, we have

$$\begin{aligned}\frac{d}{1-d} &= \cos(\frac{\pi E(D)}{N-1}) \\ d &= \frac{\cos(\frac{\pi E(D)}{N-1})}{1+\cos(\frac{\pi E(D)}{N-1})}\end{aligned} \tag{4.23}$$

Eqn. (4.23) suggests a possible estimator for long memory parameter d in ARFIMA$(0, d, 0)$ processes as,

$$\hat{d}_D = \frac{\cos(\frac{\pi D}{N-1})}{1+\cos(\frac{\pi D}{N-1})} \tag{4.24}$$

We next investigate the performance of the estimator $\hat{d}_D$. We also compare it with estimates from the maximum likelihood method (MLE) and Whittle's method. The details of these two methods have been discussed in Chapter 1. In the first set of simulations, estimator $\hat{d}_D$ (4.24) is evaluated by considering simulated series from an ARFIMA$(0, d, 0)$ process. The values of d considered are $0.2, 0.32, 0.45$. Different lengths of the simulated series considered are $N = 50, 100, 500$. This simulation study is carried out 1000 times. Table (4.1) summarizes the results of the simulation

Table 4.1: Comparison of parameter estimates of d in an ARFIMA$(0, d, 0)$ process

N	d	$\hat{d}_D$	$\hat{d}_{MLE}$	$\hat{d}_{Whittle}$
50	0.20	0.1927	0.1271	0.1862
		(0.1671)	(0.1034)	(0.1515)
100		0.1957	0.1572	0.1964
		(0.1163)	(0.0829)	(0.0936)
500		0.1990	0.1911	0.2002
		(0.0508)	(0.0366)	(0.0373)
50	0.32	0.3023	0.2185	0.3082
		(0.1225)	(0.1112)	(0.1528)
100		0.3045	0.2617	0.3126
		(0.0926)	(0.0837)	(0.0954)
500		0.3153	0.3064	0.3181
		(0.0466)	(0.0353)	(0.0368)
50	0.45	0.4233	0.3192	0.4447
		(0.0920)	(0.1071)	(0.1538)
100		0.4303	0.3763	0.4538
		(0.0706)	(0.0693)	(0.0953)
500		0.4377	0.4310	0.4536
		(0.0458)	(0.0300)	(0.0373)

in terms of the mean and standard deviations of the estimators. In general for all the estimators, as N increases the mean estimated values get closer to the true value and there is a decrease in their standard deviations. The MLE estimator $\hat{d}_{MLE}$ performs poorly in comparison to the other estimators. The Whittle estimator $\hat{d}_{Whittle}$ shows the best results. The results of the estimator $\hat{d}_D$ are comparable to that of $\hat{d}_{Whittle}$.

Next, we look at the general ARFIMA(p, d, q) process. In the MLE method, all the parameter estimates can be obtained as described in Chapter 1. For the Whittle method, an estimate of d is obtained by assuming that the process is an ARFIMA$(0, d, 0)$ process. The estimator $\hat{d}_D$ is also based on autocorrelation from an ARFIMA$(0, d, 0)$ process. The estimates thus obtained can be put in the following

algorithm to get the estimates of all the parameters in an ARFIMA(p, d, q) model. The algorithm has the following steps [28]. Here we specifically show the algorithm with the estimator $\hat{d}_D$. Let $Y_t \sim ARFIMA(p, d, q)$ as defined in (2.5).

1. Estimate $\hat{d}_D$ using Eqn. (4.24).
2. Calculate $X_t = (1 - B)^{\hat{d}_D} Y_t$.
3. Find the estimates of the parameters in polynomials $\phi(.)$ and $\psi(.)$.
4. Calculate
$$U_t = \frac{\hat{\phi}(B)}{\hat{\psi}(B)} Y_t \tag{4.25}$$
5. Obtain estimate of $\hat{d}_D$ from U_t.
6. Repeat steps 2-5, till the estimates converge.

We apply the algorithm to test and compare the performances of the estimators $\hat{d}_D$ and $\hat{d}_{Whittle}$. The MLE estimator $\hat{d}_{MLE}$ is not evaluated using the algorithm. Again, 1000 simulations of an ARFIMA(p, d, q) process are considered. For simplicity of calculations in Step 4 in the algorithm, ARFIMA$(1, d, 0)$ processes are considered. The maximum number of iterations considered in the algorithm is 10. The mean and standard deviations across 1000 simulations are reported in Table 4.2. Table 4.2 summarizes the results for three ARFIMA$(1, d, 0)$ processes. In all the cases, both estimators $\hat{d}_D$ and $\hat{d}_{Whittle}$ outperform the MLE estimator. As observed from the simulations, both the estimators $\hat{d}_D$ and $\hat{d}_{Whittle}$ are comparable in terms of their mean and standard deviations.

Table 4.2: Comparison of parameter estimates in an ARFIMA$(1, d, 0)$ process

N	d	ϕ	$\hat{d}_D$	$\hat{\phi}_D$	$\hat{d}_{MLE}$	$\hat{\phi}_{MLE}$	$\hat{d}_{Whittle}$	$\hat{\phi}_{Whittle}$
50	0.2	0.4	0.3092	0.2609	0.0284	0.4993	0.3362	0.2307
			(0.1052)	(0.1253)	(0.0751)	(0.1513)	(0.1492)	(0.0706)
100	0.2	0.4	0.2667	0.3222	0.0546	0.5074	0.2963	0.2765
			(0.0872)	(0.0848)	(0.0979)	(0.1423)	(0.0968)	(0.0398)
500	0.2	0.4	0.2267	0.3671	0.1280	0.4643	0.2563	0.3275
			(0.0455)	(0.0383)	(0.0990)	(0.1127)	(0.0394)	(0.0172)
50	-0.1	0.25	0.2968	-0.1589	0.0793	0.0320	0.2587	-0.1267
			(0.1134)	(0.1293)	(0.1096)	(0.1833)	(0.1575)	(0.0917)
100	-0.1	0.25	0.2739	-0.1263	0.1318	0.0009	0.2308	-0.0936
			(0.0915)	(0.0903)	(0.1146)	(0.1551)	(0.0967)	(0.0640)
500	-0.1	0.25	0.2509	-0.1010	0.2230	-0.0766	0.2489	-0.1001
			(0.0464)	(0.0415)	(0.0587)	(0.0711)	(0.0358)	(0.0290)
50	-0.3	0.4	0.3840	-0.2835	0.1574	-0.1031	0.3612	-0.2826
			(0.0935)	(0.1281)	(0.1380)	(0.1976)	(0.1568)	(0.1045)
100	-0.3	0.4	0.3885	-0.2911	0.2621	-0.1915	0.4005	-0.2980
			(0.0755)	(0.0985)	(0.1141)	(0.1444)	(0.0971)	(0.0699)
500	-0.3	0.4	0.3934	-0.2926	0.3684	-0.2777	0.4018	-0.3026
			(0.0476)	(0.0505)	(0.0459)	(0.0597)	(0.0377)	(0.0311)

4.3 Discrimination using HOC

As remarked in the section (4.0.1), the initial rate of increase in D_k can be used as a discriminator among processes. Since simple HOC sequences are in general monotone, the rate of increase in D_k can be quantified using the increments Δ_k as defined as

$$\Delta_k \equiv \begin{cases} D_1 & \text{if } k = 1 \\ D_k - D_{k-1} & \text{if } k = 2, 3, ..., K-1 \\ (N-1) - D_{K-1} & \text{if } k = K \end{cases} \tag{4.26}$$

To compute distances between two processes, a statistic based on the increments Δ_k from the first process and second processes can be computed using Eq. (4.27). Other variant forms of measures based on HOC have also be discussed [12], [1], [19]. Here, we are particularly interested to compute distances from a reference

series using the ψ^2 as given in Eq. (4.28) below. The ψ^2 statistic is based on the increments Δ_k from a process and the expected values of the increments $E[\Delta_k](\equiv m_k)$ from another (a reference) process. As a convenient reference, we can use the Gaussian white noise process for which $E[D_k]$ is given in Eq. (4.29). The $E[\Delta_k]$ values for Gaussian white noise process can be obtained from Eqs. (4.26) and (4.29).

$$\Delta(1,2) = \sqrt{\sum_{k=1}^{K}(\Delta_{1k}-\Delta_{2k})^2} \tag{4.27}$$

$$\psi^2 = \sum_{k=1}^{K}\frac{(\Delta_k - m_k)^2}{m_k} \tag{4.28}$$

$$E[D_k] = (N-1)[\frac{1}{2}+\frac{1}{\pi}\sin^{-1}(\frac{k-1}{k})] \tag{4.29}$$

4.3.1 Application of ψ^2 in discrimination

To illustrate the use of ψ^2 in discrimination between processes, we performed a simulation study using data from the first order autoregressive (AR) process,

$$z_t = \phi_1 z_{t-1} + \epsilon_t, \quad t = 1, ..., N \tag{4.30}$$

We generated two sets each containing 10 time series from (4.30) of length $N(= 100, 500, 700, 1000)$ as follows. In case one, set 1 contained 10 realizations from (4.30) with $\phi_1 = 0.3$, and set 2 contained 10 realizations from (4.30) with $\phi_1 = 0.4$. ψ^2 values were computed from each time series in the two sets, giving two ψ^2 samples of size 10 each corresponding to $\phi_1 = 0.3$ and $\phi_1 = 0.4$, respectively. On these two ψ^2 samples, the Wilcoxon rank-sum test was performed to test the hypothesis of

equidistribution versus the alternative that the distributions are different ($H_1 : 1 \neq 2$), and a p-value was computed. This was repeated 1000 times from which an average p-value ($\bar{p}$) was computed. From Table (4.3), we see that the discrimination between the two sets is evident for $N > 500$ and significance level 0.05.

The same was repeated in case two with the same $\phi_1 = 0.3$ in the two sets. Indeed, from Table 4.3 the averages $\bar{p}$ are relatively large as they should.

In case three, set 1 contained 8 realizations with $\phi_1 = 0.3$ and 2 from $\phi_1 = 0.4$, whereas set 2 contained 2 realizations with $\phi_1 = 0.3$ and 8 from $\phi_1 = 0.4$. In this mixture case, discrimination failed. But when the difference between the ϕ_1 values increases, the dissimilarity between the two sets becomes evident as expressed by small $\bar{p}$ even for $N = 100$. See Table 4.3 cases three and four.

Similar simulation experiments were run with long memory processes. Again, the two sets of processes from either ARFIMA$(0, d, 0)$, ARFIMA$(1, d, 0)$ or ARFIMA$(1, d, 1)$ were considered. Discrimination between the two sets of processes were even more challenging for these processes. For large differences in the value of the fractional parameter d in the ARFIMA$(0, d, 0)$, the discrimination was evident for lengths $N = 100, 500$. When the differeces in d were relatively small ($d = 0.1, d = 0.18$), the discrimination is evident only for large N. The same is observed for ARFIMA$(1, d, 1)$ processes, where the small differences in d result in p-values that suggest no differences in the two sets.

To summarize, ψ^2 as distance from white noise is capable of detecting relatively small differences for sufficiently long times series in case of short memory processes. For the long memory process, small differences are not detectable in most cases. In

Table 4.3: Results of Wilcoxon rank-sum test applied to ψ^2 samples of AR processes. Mean(ψ_1^2)& Mean(ψ_2^2) are averages of ψ^2 values for Set(1) and Set(2) respectively. The last column reports the average two-sided p-value for $H_0 : 1 = 2$.

Set(1)	Set(2)	N	Mean(ψ_1^2)	Mean(ψ_2^2)	$H_0 : 1 = 2$
AR(0.3) (10)	AR(0.4) (10)	100	24.58	28.85	0.394
		500	38.60	57.58	0.065
		700	47.44	73.51	0.032
		1000	60.73	97.99	0.008
AR(0.3) (10)	AR(0.3) (10)	100	24.528	24.537	0.530
		500	38.758	38.804	0.5104
		700	46.952	47.29	0.531
		1000	60.107	60.408	0.509
AR(0.3) (8)	AR(0.3) (2)	100	25.49	27.987	0.482
AR(0.4) (2)	AR(0.4) (8)	500	42.75	53.87	0.283
		700	52.54	68.505	0.1953
		1000	67.83	90.33	0.129
AR(0.3) (8)	AR(0.7) (2)	100	25.450	74.977	0.0001
AR(0.4) (2)	AR(0.8) (8)	500	42.85	303.93	< 0.0001
		700	52.39	420.06	<0.0001

general, mixture (or contaminated) cases are challenging for both short and long memory processes, unless the differences increase.

4.3.2 Application of ψ^2 statistic on eye gaze data

We now use univariate time series x_t, y_t r_t, θ_t and try to look for a possible difference in the features of these time series under the "Watch" and "Imitate" conditions using higher order crossings.

The HOC theorem can be illustrated using the data. Consider the x-coordinate of Subject 1, movement 5 and Subject 4, movement 6 under the "'Watch"' condition. The binary clipped series $X_t(k)$ is evaluated using Eqn. (4.7). Below is a section

Table 4.4: Results of Wilcoxon rank-sum test applied to ψ^2 samples of ARFIMA processes. Mean(ψ_1^2)& Mean(ψ_2^2) are averages of ψ^2 values for Set(1) and Set(2) respectively. The last column reports the average two-sided p-value for $H_0 : 1 = 2$.

Set(1)	Set(2)	N	Mean(ψ_1^2)	Mean(ψ_2^2)	$H_0 : 1 = 2$
$d = 0.2$	$d = 0.45$	100	21.7480	76.0713	0.0029
		500	31.1548	307.8371	1.1712E-05
$d = 0.1$	$d = 0.18$	100	17.9025	17.0154	0.4768
		500	20.7656	26.7550	0.1305
		700	19.2409	34.2657	0.0630
		1000	19.1558	45.7939	0.0225
$d = 0.2$	$d = 0.25$	100	63.0473	20.9857	0.0001
$\phi = 0.6$	$\phi = -0.2$	500	216.6297	23.4295	1.0923e-05
$d = 0.2$	$d = 0.25$	100	45.3910	49.6619	0.5186
$\phi = 0.6$	$\phi = 0.4$	500	172.1257	166.0504	0.5067
$\psi = 0.2$	$\psi = 0$	700	238.7899	214.7899	0.4932
		1000	324.8325	283.8343	0.4787

of the series $X_t(k)$ for $k = 1, 2, ..., 8$ and $t = 501, ..., 520$. The convergence to the limiting sequence is attained for $k = 4$ for the first case and $k = 7$ for the second case.

$k = 1$ - 11111111111111111111

$k = 2$ - 10101010101010101011

$k = 3$ - 10101010101010101011

$k = 4$ - 10101010101010101010

$k = 5$ - 10101010101010101010

$k = 6$ - 10101010101010101010

$k = 7$ - 10101010101010101010

$k = 8$ - 10101010101010101010

$k = 1$ - 11111111111111111111

$k = 2$ - 10101010101110100010

$k = 3$ - 10101010101100101010

$k = 4$ - 10101010101000101010

$k = 5$ - 10101010101011101010

$k = 6$ - 10101010101011101010

$k = 7$ - 10101010101010101010

$k = 8$ - 10101010101010101010

For purpose of discriminating between the two conditions "Watch" and "Imitate", we consider univariate time signals r_t and θ_t. We choose the said coordinates, as the results are more clearer with the radial and angular transformations. For each eye signal, the calculations of the D_k's are based on a 751-point segments, corresponding to $t = 50, 51, ..., 800$. The lower truncation reduces the transient effects at the beginning of each signal and the upper truncation makes all the signals of equal length corresponding to that of the shortest signal. Figure 4.3.2 shows the plot of the HOC sequence of the differenced series ∇r_t for a particular case. The plot clearly shows the monotone property of D_k's but as k increases the monotonicity tapers off gradually. We also observe that for this case the D_k's for the "Imitate" condition are further away from the $\mathrm{E}(D_k)$ of white noise as compared to the D_k's from the watch condition. But in general, the D_k's for the watch and imitate conditions are not clearly separated as shown for the case of Subject 2 Movement 5 in Figure (4.3.2). Since the discrimination of the two conditions is not evident from the

D_k's alone, we investigate their distance from the white noise using the ψ^2 metric in Eq. (4.26).

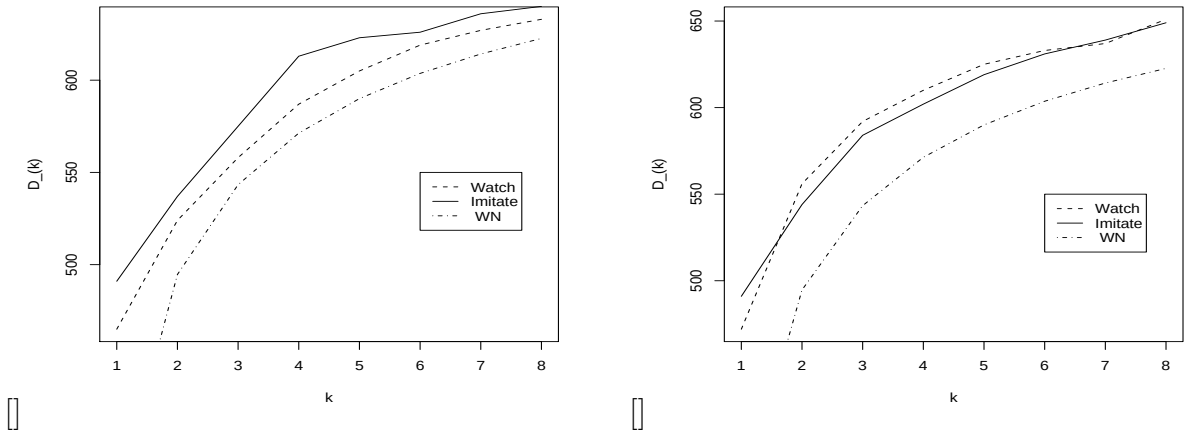

Figure 4.1: (a) D_k for watch and imitate conditions for Subject 1, Movement 9 and the E(D_k) from white noise. (b) D_k for "Watch" and "Imitate" conditions for Subject 2, Movement 5 and the E(D_k) from white noise.

First we perform data analysis on the radial component of the eye gaze data. For each time series, the respective ψ^2 distances from white noise were computed using D_k for $k = 1, 2, ..., 8$. Table (4.5) summarizes the results for all the subjects. The table lists the average ψ^2 values calculated for all the 7 subjects under the "Watch" and "Imitate" conditions. We see that 4 out of the 7 subjects have higher ψ^2 imitate averages under the 10 movements. To test the hypothesis that ψ^2 values differ significantly under the Watch and Imitate conditions for each subject, Wilcoxon rank-sum test was performed. Only subjects 1, 3, 6 and 7 show any significant difference (p-value < 0.05)in the 10 movements. For the series θ_t, only

subject 7 showed a significant difference. The same analysis was again performed on the differenced series ∇r_t. This differenced series does not take into account the long memory property. The results for this case are summarized in Table (4.6). Again, 4 out of the 7 subjects have higher ψ^2 imitate averages under the 10 movements. Subjects 1, 6 and 7 show significant difference (p-value < 0.05) between the two conditions. In addition, the mean across all subjects of the averaged ψ^2 values computed for the Watch and Imitate condition provide an indication of difference in their oscillatory behavior. Figure 4.2 shows the ψ^2 values for both conditions for Subjects 6 and 7. It is evident from Figure (4.2) that for Subject 7, the ψ^2 distances for the imitate condition are greater than those for watch condition whereas Subject 6 shows the opposite trend.

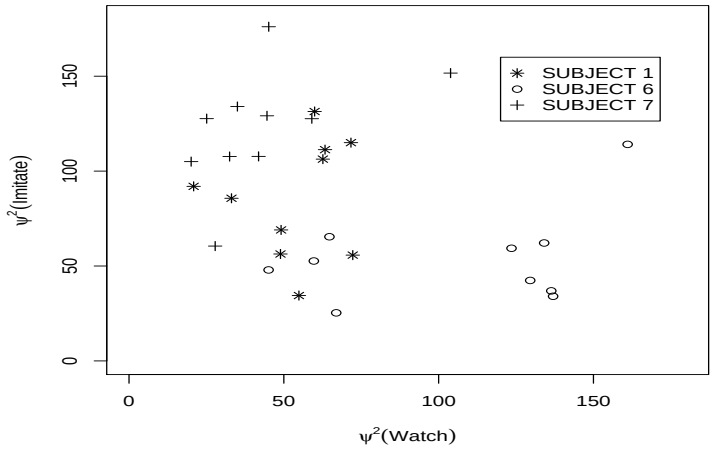

Figure 4.2: ψ^2 values for "Watch" and "Imitate" conditions for Subjects 1, 6 and 7 for the 10 movements.

To get further insight from the 70 cases for each of the two conditions, each

Table 4.5: r_t -Average ψ^2 distance from white noise across 10 movements. The last column reports the average two-sided p-value for $H_0 : W1 = I1$.

Subject	Watch1	Imitate1	$H_0 : W1 = I1$
1	1264.668	1396.186	0.0355
2	1189.076	1213.079	0.6842
3	1466.036	1277.540	0.0288
4	1379.241	1417.908	0.3149
5	1242.425	1259.798	0.7959
6	1514.970	1256.533	0.0038
7	1249.393	1053.805	0.0001

Table 4.6: ∇r_t -Average ψ^2 distance from white noise across 10 movements. The last column reports the average two-sided p-value for $H_0 : W1 = I1$.

Subject	Watch1	Imitate1	$H_0 : W1 = I1$
1	53.317	85.527	0.0232
2	37.064	48.627	0.2176
3	92.091	66.617	0.1051
4	75.271	79.421	0.5787
5	51.149	50.656	0.9118
6	105.241	53.506	0.0029
7	43.894	122.453	2.16E-05
Average	65.43	72.40	
Median	53.2	66.62	
S^2	666.6118	692.5498	

of the 70 time series was divided into two equal parts. The ψ^2 values were then computed for 140 time series under each condition. Essentially, this implies as if the subjects were viewing 20 movements instead of 10. The average ψ^2 values across these 20 movements for the two conditions were calculated and are listed in Table 4.7 under columns Watch2 and Imitate2. Clearly, any signifinant difference with the Wilcoxon test is observed for Subjects 1, 6 and 7 only. This supports the analysis from the previous table. *It is also interesting to observe that for both the 10 and 20 movement data, the average and median ψ^2 across all subjects are consistently*

higher for the "Imitate" condition.

A similar analysis was performed on the angular component time series $\nabla\theta_t$. Tables

Table 4.7: ∇r_t -Average ψ^2 distance from white noise across 20 movements. The last column reports the average two-sided p-value for $H_0 : W2 = I2$.

Subject	Watch2	Imitate2	$H_0 : W2 = I2$
1	33.237	50.629	0.0074
2	27.884	32.373	0.3834
3	52.357	44.101	0.1274
4	40.421	46.265	0.3834
5	32.863	33.162	0.6980
6	60.175	34.186	0.0052
7	26.643	74.597	< 0.005
Average	39.08	45.04	
Median	33.24	44.10	
S^2	162.6498	221.0974	

(4.8), (4.9) show the averaged ψ^2 values for 10 and 20 movement case. From both cases we see that four subjects (3, 4, 6, 7) show higher averaged ψ^2 values for the "Imitate" condition. Except for Subject 7 in the 20 movement case, the wilcoxon test does not show any significant difference. Except for the overall averaged ψ^2 being higher for the "Watch" condition in Table (4.8), again the overall average and the medians are higher for the "Imitate" condition.

Table 4.8: $\nabla\theta_t$ -Average ψ^2 distance from white noise across 10 movements. The last column reports the average two-sided p-value for $H_0 : W2 = I2$.

Subject	Watch1	Imitate1	$H_0 : W1 = I1$
1	84.105	73.023	0.2475
2	50.350	44.339	0.9118
3	61.165	73.764	0.3527
4	42.943	53.169	0.2475
5	59.336	48.105	0.1903
6	81.165	91.601	0.5288
7	67.876	92.803	0.2176
Average	84.105	68.115	
Median	61.165	73.023	
S^2	228.870	401.028	

Table 4.9: $\nabla\theta_t$ -Average ψ^2 distance from white noise across 20 movements

Subject	Watch2	Imitate2	$H_0 : W2 = I2$
1	49.166	43.356	0.6980
2	33.420	30.636	0.7584
3	37.815	47.539	0.1493
4	29.005	32.503	0.2110
5	36.282	29.156	0.1207
6	47.757	52.202	0.4945
7	38.296	79.894	0.0051
Average	38.820	45.041	
Median	37.815	43.356	
S^2	53.355	315.355	

Chapter 5

Classification of long memory processes

In this chapter, we perform classification of long memory processes using clustering techniques. We consider distance measures that are based on distances from a reference time series. In the spectral domain, we look at the second order properties of a time series via their spectral density functions. We compare the spectral densities of several processes with respect to a reference spectral density. Two model forms of the spectral density, EXP and FEXP as mentioned in Chapter 1 are considered. The EXP model has been extensively studied in clustering [31], [32]. We investigate the performance of the EXP model in classification of long memory processes and further extend the method to FEXP model. Finally, we apply the methods to classify the eye gaze data.

5.1 Likelihood Function

In this section, we construct a likelihood function for the ratios of periodogram ordinates of long memory processes. We use the distribution of periodogram ordinates as mentioned in Theorem 2 for the frequencies near zero. Recall that the distribution of the last (not close to zero) periodogram ordinates are considered to be $(1/2)\chi^2_2$ distributed. We assume independence of the periodogram ordinates across all the Fourier frequencies. Here consider a set of G stationary long memory

processes $Y_{jt}, t = 1, 2, ..., N$ and assume that each of the G time series possesses a spectral density function $f_j(\omega)$. Let the Gth time series Y_{Gt} be the reference time series. Let $\{d_j, j = 1, 2, ..., G\}$ be the memory parameters of the G time series processes respectively. Let $M = [\sqrt{N}]$ be the number of Fourier frequencies that are considered close to zero. Consider k independent chi-squared random variables $X_1, X_2, ..., X_k$ each with one degree of freedom. The linear combination of the k random variables is given by

$$X = c_1X_1 + c_2X_2 + ... + c_kX_k \tag{5.1}$$

We consider the following approximation of the distribution of X [19].

$$X \approx h\chi^2_{d^*} \tag{5.2}$$

where $h = \frac{Var(X)}{2E(X)}$ and $d^* = \frac{2(EX)^2}{Var(X)}$.

Using the above moment approximations for chi-square distribution, we consider the first M normalized periodogram ordinates to have the following approximate distributions. Using Δ_j defined in 2.20

$$I_j(\omega_i) \approx f_j(\omega_i)h_{ij}\chi^2_{d_{ij}} \tag{5.3}$$

where,

$$d_{ji} = \frac{2(\Delta_j(d_j, -1))^2}{(\Delta_j(d_j, -1))^2 + 4(\Delta_j(d_j, 1))^2}$$

And,

$$h_{ji} = \frac{(\Delta_j(d_j, -1))^2 + 4(\Delta_j(d_j, 1))^2}{2\Delta_j(d_j, -1)}$$

Particularly for the case where $d_j = 0, \forall j$, that is the short memory process, we have

$$\begin{aligned} \Delta_j(0,-1) &= 1 \\ \Delta_j(0,1) &= 0 \\ d_{ji} &= 2 \\ h_{ji} &= \frac{1}{2} \end{aligned} \tag{5.4}$$

Hence for this case $d_j = 0$, the distribution of the normalized periodogram reduces $(1/2)\chi^2_2$ distribution.

For different values of d_j, Figure (5.1) shows the values of approximate degrees of freedom d_{ji}. Here $N = 500$ and $M = 22$. As evident from the plot, the approximated values of d_{ji} for the chi-square distribution is close to 2 for values of the long memory parameter d_j that are close to zero. Also as i becomes large, d_{ji} becomes close to 2 for all j. Using the approximate distribution of periodogram ordinate in (5.3) and independence across the G time series, for each $i \in \{1, 2.., M\}$ we have the joint density of $\{I_j(\omega_i), j = 1, 2, ...G\}$ at Fourier frequency ω_i as

$$f_i(I_1, I_2.., I_G) = \prod_{j=1}^{G} [\frac{h_{ji}}{f_j(\omega_i)} \frac{1}{2^{d_{ji}/2}} \frac{1}{\Gamma(d_{ji}/2)} (\frac{I_j(\omega_i)}{f_j(\omega_i)})^{d_{ji}/2-1} \exp(\frac{-I_j(\omega_i)}{2f_j(\omega_i)})] \tag{5.5}$$

Let us consider the transformation $T_{ji} = \frac{I_j(\omega_i)}{I_G(\omega_i)}$, for $j \in \{1, 2...., (G-1)\}$ and $T_{Gi} = I_G(\omega_i)$. Then joint density of $\{T_{1i}, T_{2i}, ..., T_{Gi}\}$ is

$$\begin{aligned} f_i(t_{1i}, t_{2i}, ..., t_{Gi}) &= \prod_{j=1}^{G} \frac{h_{ji}}{2^{d_{ji}/2}\Gamma(d_{ji}/2) f_j(\omega_i)^{d_{ji}/2}} \exp(\sum_{j=1}^{G-1} d_{ji}/2 t_{ji}) t_{Gi}^{\sum_{j=1}^{G} d_{ji}/2-1} \\ & \exp(\frac{-t_{Gi}}{2}(\sum_{j=1}^{G-1} \frac{\exp(t_{ji})}{f_j(\omega_i)} + \frac{1}{f_G(\omega_i)}) \end{aligned} \tag{5.6}$$

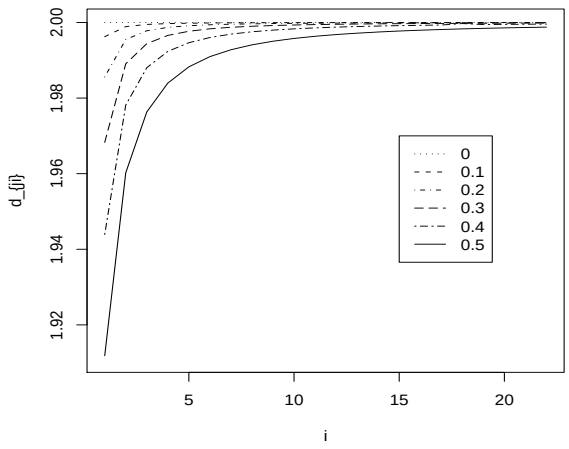

Figure 5.1: d_{ji} for different values of d_j in legend

The joint density of $\{T_{1i}, T_{2i}, ..., T_{(G-1)i}\}$ can be obtained by integrating w.r.t T_{Gi} and we have

$$f_i(t_{1i}, t_{2i}, ..., t_{(G-1)i}) = \Gamma(\sum_{j=1}^{G} d_{ji}/2) \prod_{j=1}^{G} [\frac{h_{ji}}{\Gamma(d_{ji}/2) f_j(\omega_i)^{d_{ji}/2}} \frac{\exp(\sum_{j=1}^{G-1} d_{ji}/2t_{ji})}{(\sum_{j=1}^{G-1} \frac{\exp(t_{ji})}{f_j(\omega_i)} + \frac{1}{f_G(\omega_i)})^{\sum_{j=1}^{G} d_{ji}/2}}] \tag{5.7}$$

Assuming the Exponential model in (2.47), the above function can be rewritten as

$$f_i(t_{1i}, t_{2i}, ..., t_{(G-1)i}) = \Gamma(\sum_{j=1}^{G} d_{ji}/2) \prod_{j=1}^{G} [\frac{h_{ji}}{\Gamma(d_{ji}/2)} \frac{\exp(\sum_{j=1}^{G-1} d_{ji}/2(t_{ji} - a_j^T Z_i))}{(\sum_{j=1}^{G-1} \exp(t_{ji} - a_j^T Z_i) + 1)^{\sum_{j=1}^{G} d_{ji}/2}}] \tag{5.8}$$

Assuming independence here, the approximate log likelihood for M frequencies

is

$$\begin{aligned}
\log(L_1) &= \log(\prod_{i=1}^{M} f_i(t_{1i}, t_{2i}, ..., t_{(G-1)i})) \qquad (5.9)\\
&= \sum_{i=1}^{M}[\log(\Gamma(\sum_{j=1}^{G} d_{ji}/2)) - \sum_{j=1}^{G} \log\Gamma(d_{ji}/2)\\
&+ \sum_{j=1}^{G} \log h_{ji} + \sum_{j=1}^{G} \frac{d_{ji}}{2}(t_{ji} - a_j^T Z_i)\\
&- (\sum_{j=1}^{G} \frac{d_{ji}}{2}) \log(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - a_j^T Z_i))]
\end{aligned}$$

Using all the Fourier frequencies and assuming that the periodogram ordinates for the jth time series are exponentially distributed with mean $f_j(\omega_i)$ for $i > M$, we have the following approximate log likelihood function as

$$\begin{aligned}
\log(L_2) &= \log(\prod_{i=1}^{[\frac{N-1}{2}]} f_i(t_{1i}, t_{2i}, ..., t_{(G-1)i})) \qquad (5.10)\\
&= const + \sum_{i=1}^{M}[\log(\Gamma(\sum_{j=1}^{G} d_{ji}/2)) - \sum_{j=1}^{G} \log\Gamma(d_{ji}/2)\\
&+ \sum_{j=1}^{G} \log h_{ji} + \sum_{j=1}^{G} d_{ji}/2(t_{ji} - a_j^T Z_i)\\
&- (\sum_{j=1}^{G} d_{ji}/2) \log(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - a_j^T Z_i))]\\
&+ \sum_{i=(M+1)}^{[(N-1)/2]} \sum_{j=1}^{(G-1)} (t_{ji} - a_j^T Z_i)\\
&- G \sum_{i=(M+1)}^{[(N-1)/2]} \log(1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - a_j^T Z_i))
\end{aligned}$$

The likelihood function given by Eqn. (5.10) takes into account the Exponential spectral density model in Eqn. (2.47). We note here through a simple example that for a long memory process, the likelihood function (5.10) is a monotone func-

tion of the long memory parameter d. Consider two independent time series (i.e. $G = 2$).

$Y_1 \approx ARFIMA(0, d = 0.3, 0)$

$Y_2 \approx N(0, 1)$

Given that p and $\mathbf{a}$ is chosen and fixed, Eqn. 5.10 can be evaluated as a function of d. Particularly for $p = 1$ and $\mathbf{a} = (1, 1)'$, the values of the likelihood function (5.10) is evaluated for different values of d. Figure (5.2) shows the increasing trend as d increases. By replacing the EXP(p) model by FEXP(p) model instead in the likelihood, we make a similar observation. Figure (5.3) shows the increasing trend as d increases for the FEXP(p) model case. It can be shown that assumptions on the distributions of the periodogram ordinates near zero frequencies can be applied to get alleviate this and is discussed in the next section.

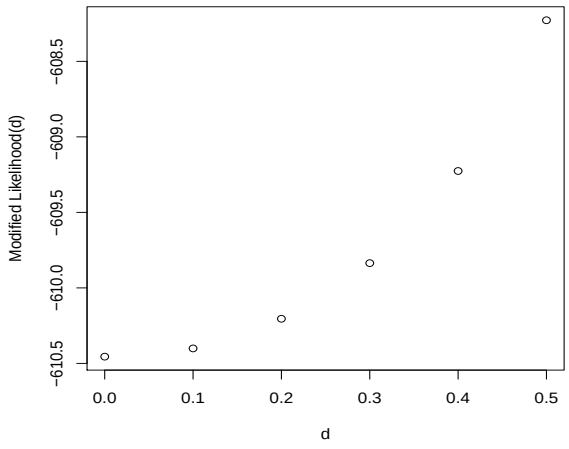

Figure 5.2: Example 1: Plot of d vs likelihood (EXP)

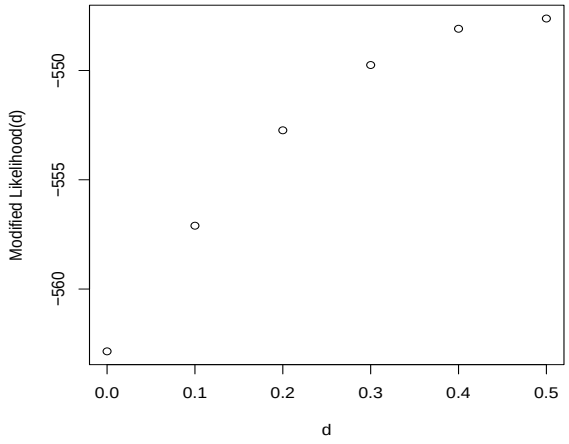

Figure 5.3: Example 1: Plot of d vs likelihood (FEXP)

5.1.1 Modified likelihood function

We can apply the same methodology as above and use further approximations on the distributions of periodogram ordinates to form the modified likelihood function. These approximations in fact provide better results. Considering the periodogram ordinates to behave independently and their normalized version $I^*(\omega_j)$ in (2.18) have the same $(1/2)\chi^2_2$ distribution for all Fourier frequncies, the likelihood function can be modified to have the following form,

$$\begin{aligned} l_{FEXP}(\theta) &= \sum_{i=1}^{[(N-1)/2]} \sum_{j=1}^{(G-1)} (t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i) \\ &- G \sum_{i=1}^{[(N-1)/2]} \log(1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i)) \end{aligned} \tag{5.11}$$

Where $\theta = \{d_j, \mathbf{a}_j, j = 1, 2...(G-1)\}$. We observe that the new likelihood involves (G-1)(p+2) parameters. The inclusion of parameter d takes into account the long

memory behaviour of the process. This likelihood $l_{FEXP}(\theta)$ is also based on the assumption that the ratio of the spectral densities is the same as in FEXP model. It is important to note here that the basis formed by the functions $\{\log|g(\omega)|, \cos(\omega), \cos(2\omega), ..., \cos(p\omega)\}$ are no longer orthogonal on $(0, \pi)$.

Specifically, for an ARFIMA process we have,

$$\begin{aligned} 2\log|g(\omega)| &= \log\left|1 - e^{i\omega}\right|^2 \\ &= \log(4\sin^2(\frac{\omega}{2})) \end{aligned} \tag{5.12}$$

In this case, the expansion set in the FEXP(p) model given by $\{\log(4\sin^2(\frac{\omega}{2})), \cos(\omega), \cos(2\omega), ..., \cos(p\omega)\}$ is not an orthogonal basis on $(0, \pi)$.

The estimates of the parameters, the score function and the likelihood test statistic for testing equality of spectral densities can be obtained as explained as above in the EXP(p) model case.

The maximum likelihood estimates of the coefficient vector θ can then be obtained by setting the system of equations $S(\theta) = 0$, where $S(\theta)$ is the score function corresponding to the likelihood function $l_{FEXP}(\theta)$.

$$S(\theta) = \frac{\partial l_{FEXP}(\theta)}{\partial \theta} \tag{5.13}$$

where for $j = 1, 2, .., (G-1)$ and $k = 1, 2, ..., p$,

$$\begin{aligned}
\frac{\partial l_{FEXP}(\theta)}{\partial d_j} &= -\sum_{i=1}^{[(N-1)/2]} \log|g(\omega_i)|^2 + G \sum_{i=1}^{[(N-1)/2]} \frac{\exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i) \log|g(\omega_i)|^2}{1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i)} \\
\frac{\partial l_{FEXP}(\theta)}{\partial a_{j0}} &= -\sum_{i=1}^{[(N-1)/2]} 1 + G \sum_{i=1}^{[(N-1)/2]} \frac{\exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i)}{1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i)} \\
\frac{\partial l_{FEXP}(\theta)}{\partial a_{jk}} &= -\sum_{i=1}^{[(N-1)/2]} 2\cos(k\omega_i) + G \sum_{i=1}^{[(N-1)/2]} \frac{\exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i) 2\cos(k\omega_i)}{1 + \sum_{j=1}^{(G-1)} \exp(t_{ji} - 2d_j \log|g(\omega_i)| - \mathbf{a}_j^T Z_i)}
\end{aligned}$$

An example is discussed below.

Example 1. Let us consider $G = 3$ time series, each of length N.

$Y_{jN} \approx$ ARFIMA$(0, d = 0.3, 0)$, for $j = 1, 2, 3$

Let the last time series be considered as reference. We perform a simulation study to obtain the estimates of the parameters $\{\theta_{01}, d_1, \theta_{02}, d_2\}$ in the likelihood function (5.11) and report their corresponding estimated standard deviations. This simulation study is carried out 1000 times. Table (5.1.1) shows the results for different values of N. In this example, the FEXP(p) model considered does not include the short memory component, that is the value of p is set as 0. The number of parameters to be estimated is 4 corresponding to the two spectral density ratios. Both raw and smoothed periodogram (width of modified Daniell window = 5) are used for estimation purposes. The table reports the average value of the estimates across 1000 simulations. Columns Estimate$_1$, Estimate$_2$ present the estimates of paramters calculated from raw peridogram and smoothed periodogram respectively. Notice the AIC statistic is minimised as it should when no short memory component is considered (i.e. when $p = 0$) as shown in Figure (5.4). For $N = 500$, Figure (5.5)

shows the qq-plot of the LRT statistic with chi-square distribution with 4 degrees of freedom.

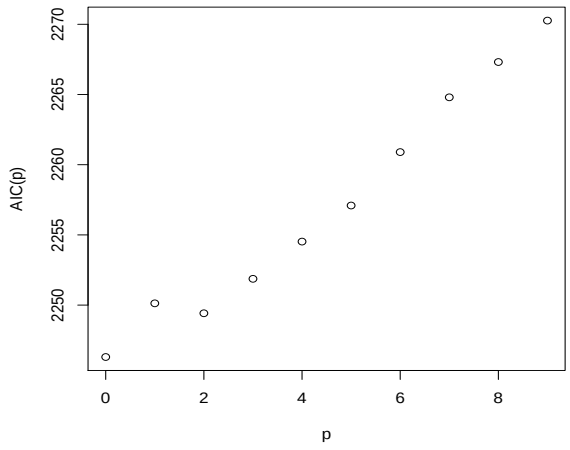

Figure 5.4: AIC for different values of p

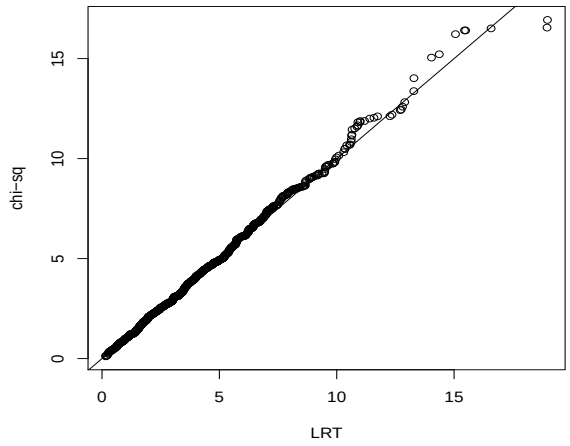

Figure 5.5: QQ-plot of the test statistic LRT

In the next section, we look at the problem of classification of long memory time series. This is done using clustering techniques. The parameters involved in

Table 5.1: Parameter estimates using raw_1 and smoothed_2 periodogram in the FEXP(p) cosine model

N	Parameter	True	Estimate_1	Std Error_1	Estimate_2	Std Error_2
50	θ_{01}	0	0.0122	0.2832	0.0044	0.2833
	d_1	0	0.0038	0.2003	0.0041	0.2003
	θ_{02}	0	0.0170	0.2832	-0.0085	0.2832
	d_2	0	0.0138	0.2003	0.0075	0.2003
100	θ_{01}	0	0.0003	0.2000	-0.0066	0.2000
	d_1	0	-0.0008	0.1414	0.0039	0.1414
	θ_{02}	0	0.0011	0.2000	0.0044	0.2000
	d_2	0	0.0019	0.1414	0.0034	0.1415
500	θ_{01}	0	-0.0019	0.0894	-0.0056	0.0894
	d_1	0	-0.0003	0.0632	-0.0027	0.0632
	θ_{02}	0	-0.0023	0.0894	-0.0019	0.0894
	d_2	0	-0.0011	0.0632	-0.0007	0.0632

the spectral density ratio models are considered for the construction of the distance metrices. Both EXP and FEXP models are included in the simulation studies as discussed in the next section.

5.2 Clustering

Given a dataset of G time series $\{\mathbf{Y}_1, \mathbf{Y}_2, ..., Y_G\}$ each of length N, the clustering problem is to identify classes or groups of time series such that group members are closer to each other in certain properties. For the purpose of clustering, there is a need for an algorithm that assigns a time series to a particular group based on some clustering criteria. This assignment is based on distances between each pair of time series. There are various algorithms available, and in this study we use single linkage, complete linkage algorithms (hierarchical, agglomerative), WARD, PAM and DIANA algorithm (hierarchical, divisive).

This section deals with the classification of long memory processes. The distance between any two time series can be computed using the parameters in the EXP or FEXP model. In particular, the spectral density ratio models for both cases are considered. By looking at these spectral ratios, the parameters computed are with respect to a reference process. The reference can be chosen as any one of the time series in the dataset or a convenient choice is a Gaussian white noise process. The distances mentioned below are however not affected by the choice of the reference process.

Using the parameter vector for EXP(p) in Eqn. (2.47), the absolute and Euclidean distances are as follows,

$$d^1_{EUCL}(j,k) = \sqrt{\sum_{r=0}^{p} (\hat{a}_{jr} - \hat{a}_{kr})^2} \tag{5.14}$$

$$d^1_{ABS}(j,k) = \sqrt{\sum_{r=0}^{p} |\hat{a}_{jr} - \hat{a}_{kr}|} \tag{5.15}$$

where $j,k \in 1,2...,(G-1)$. And using the spectral density ratio with FEXP (p) model, the distances are

$$d^2_{EUCL}(j,k) = \sqrt{(\hat{a}_{j0} - \hat{a}_{k0})^2 + (\hat{d}_j - \hat{d}_k)^2 + \sum_{r=1}^{p} (\hat{a}_{jr} - \hat{a}_{kr})^2} \tag{5.16}$$

$$d^2_{ABS}(j,k) = \sqrt{|\hat{a}_{j0} - \hat{a}_{k0}| + \left|\hat{d}_j - \hat{d}_k\right| + \sum_{r=1}^{p} |\hat{a}_{jr} - \hat{a}_{kr}|} \tag{5.17}$$

We compare the performance of the above measures in simulation study for long memory process using the similarity index described below. If true k clusters are $F = (F_1, F_2, ..., F_k)$, given that k is known and the clusters as obtained from the simulation are $C = (C_1, C_2, ..., C_k)$, then the similarity index is defined as

$$Sim(F, C) = \frac{1}{k} \sum_{i=1}^{k} max_{1 \leq j \leq k} Sim(F_i, C_j) \tag{5.18}$$

where,

$$Sim(F_i, C_j) = \frac{2\,|F_i \cap C_j|}{|F_i| + |C_j|}$$

where $|.|$ denotes the cardinality of a set. The similarity index $Sim(F, C)$ is bounded between 0 and 1, where $Sim(F, C) = 0$ implies $F_i \cap C_j = \Phi, i, j = 1, 2, ..., k$ (disjoint sets) and $Sim(F, C) = 1$ means perfect clustering. That is perfect identification of k true clusters $F_1, ..., F_k$.

The simulation study is based on a known number of clusters. In each of the examples below, the number of clusters is 2. The reference time series is chosen to be a Gaussian process with zero mean, standard deviation 1 and of the same length as the other time series.

Example 1. Consider two clusters of total 20 time series each of length N.

$Y_{jN} \approx$ ARFIMA$(0, d = 0.2, 0)$, for $j = 1, 2, ..., 10$

$Y_{jN} \approx$ ARFIMA$(0, d = 0.4, 0)$, for $j = 11, 2, ..., 20$

Here the length of each time series was selected as $N = 50, 100$ and 500. Tables 5.2, 5.3 show the computed similarity index values averaged over 1000 simulations and its corresponding standard deviations when using the EXP(p) model. Tables 5.4, 5.5 show the computed similarity index values for the FEXP(p) model. In addition, we perform simulations to investigate the performance of $d^2_{EUCL}(.)$ and

$d^2_{ABS}(.)$ when only the first M $(M < (\frac{N-1}{2}))$ Fourier frequencies are considered. As observed from the Tables, the absolute value and Euclidean value distances perform in a similar way. Especially in the case of $d^1_{ABS}(.)$ and $d^1_{EUCL}(.)$, we find an increase in similarity indices as p increases. For distances $d^2_{ABS}(.)$ and $d^2_{EUCL}(.)$ that uses the FEXP model, the similarity index is highest for $p = 0$. That is, when only fractional parameter d is taken into account. Further investigation was performed when only first $M(= \sqrt{N})$ frequencies were taken into account. Tables 5.6, 5.7 are similarity indicies corresponding to the first $M(= \sqrt{N})$ frequencies using the FEXP(p) model. As seen from the tables, the values of similarity indicies have greatly decreased.

Table 5.2: Average similarity indices using distance $d^1_{ABS}(.)$ in Eqn. (5.15)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.6430	0.6510	0.6390	0.0.6456	0.6421	0.6454
		(0.0437)	(0.0153)	(0.0709)	(0.0881)	(0.0731)	(0.0919)
	2	0.6495	0.6528	0.6447	0.6487	0.6486	0.6474
		(0.0446)	(0.0113)	(0.0714)	(0.0866)	(0.0697)	(0.0848)
	3	0.6490	0.6533	0.6386	0.6404	0.6469	0.6450
		(0.0346)	(0.0065)	(0.0647)	(0.0808)	(0.0674)	(0.0841)
100	1	0.6694	0.6538	0.6897	0.7091	0.6971	0.7140
		(0.0723)	(0.0272)	(0.1032)	(0.1132)	(0.1037)	(0.1152)
	2	0.6728	0.6541	0.6950	0.7236	0.7049	0.7263
		(0.0739)	(0.0180)	(0.1006)	(0.1113)	(0.1088)	(0.1130)
	3	0.6704	0.6539	0.6974	0.7239	0.7028	0.7114
		(0.0654)	(0.0145)	(0.0985)	(0.1119)	(0.1033)	(0.1080)
500	1	0.9397	0.7719	0.9357	0.9586	0.9527	0.9678
		(0.1010)	(0.1564)	(0.0974)	(0.0543)	(0.0900)	(0.0419)
	2	0.9634	0.7842	0.9603	0.9718	0.9748	0.9766
		(0.0751)	(0.1617)	(0.0699)	(0.0437)	(0.0595)	(0.0393)
	3	0.9753	0.7849	0.9668	0.9779	0.9818	0.9759
		(0.0571)	(0.1640)	(0.0669)	(0.0377)	(0.0499)	(0.0398)

Example 2. Consider two clusters of total 20 time series.

$Y_{jN} \approx$ ARFIMA$(0, d = 0.2, 0)$, for $j = 1, 2, ..., 10$

Table 5.3: Average similarity indices using distance $d^1_{EUCL}(.)$ in Eqn. (5.14)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.6435	0.6509	0.6390	0.6434	0.6427	0.6448
		(0.0435)	(0.0165)	(0.0705)	(0.0838)	(0.0736)	(0.0870)
	2	0.6467	0.6528	0.6406	0.6456	0.6449	0.6461
		(0.0429)	(0.0098)	(0.0697)	(0.0837)	(0.0708)	(0.0828)
	3	0.6482	0.6531	0.6394	0.6403	0.6442	0.6454
		(0.0372)	(0.0074)	(0.0655)	(0.0783)	(0.0684)	(0.0812)
100	1	0.6682	0.6551	0.6825	0.7078	0.6951	0.7099
		(0.0714)	(0.0328)	(0.0973)	(0.1101)	(0.1015)	(0.1074)
	2	0.6679	0.6532	0.6887	0.7172	0.7043	0.7229
		(0.0697)	(0.0118)	(0.0970)	(0.1106)	(0.1087)	(0.1144)
	3	0.6681	0.6533	0.6919	0.7198	0.7026	0.7120
		(0.0653)	(0.0111)	(0.0934)	(0.1101)	(0.1039)	(0.1078)
500	1	0.9471	0.7868	0.9395	0.9625	0.9527	0.9710
		(0.0916)	(0.1609)	(0.0871)	(0.0492)	(0.0638)	(0.0403)
	2	0.9665	0.7981	0.9612	0.9741	0.9808	0.9809
		(0.0720)	(0.1653)	(0.0688)	(0.0409)	(0.0475)	(0.0358)
	3	0.9759	0.8018	0.9668	0.9799	0.9866	0.9823
		(0.0582)	(0.1677)	(0.0660)	(0.0352)	(0.0379)	(0.0343)

$Y_{jN} \approx \text{ARFIMA}(0, d = 0.24, 0)$, for $j = 11, 2, ..., 20$

Tables 5.8, 5.9, 5.10, 5.11 show the computed similarity index values averaged over 1000 simulations and its corresponding standard deviations for EXP(p) and FEXP(p) models respectively. As observed from the Tables, none of the distances give us good classification.

Example 3. Consider two clusters of total 20 time series.

$Y_{jN} \approx \text{ARFIMA}(\phi = 0.2, d = 0.2, \psi = 0.4)$, for $j = 1, 2, ..., 10$

$Y_{jN} \approx \text{ARFIMA}(\phi = 0.4, d = 0.4, \psi = 0.2)$, for $j = 11, 2, ..., 20$

Tables 5.12, 5.13 show the computed similarity index values for absolute dis-

Table 5.4: Average similarity indices using distance $d^2_{ABS}(.)$ in Eqn. (5.17)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6727	0.6572	0.6893	0.6999	0.6939	0.0759
		(0.0660)	(0.0391)	(0.0798)	(0.0869)	(0.0805)	(0.0881)
	1	0.6375	0.6459	0.6292	0.6262	0.6280	0.6245
		(0.0453)	(0.0242)	(0.0595)	(0.0686)	(0.0568)	(0.0690)
	2	0.6305	0.6438	0.6147	0.6084	0.6162	0.6084
		(0.0430)	(0.0268)	(0.0543)	(0.0626)	(0.0546)	(0.0633)
100	0	0.7366	0.6845	0.7696	0.7900	0.7740	0.7986
		(0.0995)	(0.0771)	(0.0976)	(0.0905)	(0.1005)	(0.0903)
	1	0.6573	0.6552	0.6627	0.6721	0.6625	0.6693
		(0.0587)	(0.0302)	(0.0734)	(0.0828)	(0.0732)	(0.0846)
	2	0.6389	0.6474	0.6303	0.6308	0.6319	0.6271
		(0.0472)	(0.0249)	(0.0595)	(0.0713)	(0.0616)	(0.0713)
500	0	0.9794	0.9377	0.9726	0.9773	0.9843	0.9848
		(0.0424)	(0.1162)	(0.0517)	(0.0388)	(0.0336)	(0.0302)
	1	0.8440	0.7595	0.8505	0.8968	0.8821	0.8992
		(0.1280)	(0.1375)	(0.1035)	(0.0812)	(0.0981)	(0.0753)
	2	0.7530	0.7038	0.7655	0.8251	0.7991	0.8257
		(0.1110)	(0.0994)	(0.0975)	(0.0975)	(0.1060)	(0.0891)

tances $d^1_{ABS}(.)$ and $d^2_{ABS}(.)$ averaged over 1000 simulations and its corresponding standard deviations for EXP(p) and FEXP(p) models respectively. As observed, for a chosen $p = 2$, the similarity index attains the highest value for distance $d^1_{ABS}(.)$. In case of $d^2_{ABS}(.)$, again the highest similarity index is attained for $p = 0$.

5.3 Classification of eye gaze data

In this section, we apply the clustering methods to classify eye gaze data. As in Chapter 3, we look for differences in two conditions "Watch" and "Imitate" under which the eye gaze data was recorded. Again for each of the seven subjects and each condition, 10 eye gaze time series corresponding to 10 movements were considered. The reference time series is Gaussian white noise process with zero mean

Table 5.5: Average similarity indices using distance $d^2_{EUCL}(.)$ in Eqn. (5.16)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6728	0.6560	0.6895	0.6988	0.6950	0.6989
		(0.0658)	(0.0366)	(0.0792)	(0.0864)	(0.0813)	(0.0848)
	1	0.6383	0.6459	0.6307	0.6287	0.6310	0.6287
		(0.0473)	(0.0239)	(0.0607)	(0.0681)	(0.0602)	(0.0711)
	2	0.6317	0.6434	0.6164	0.6115	0.6160	0.6107
		(0.0425)	(0.0259)	(0.0557)	(0.0630)	(0.0558)	(0.0627)
100	0	0.7313	0.6819	0.7693	0.7837	0.7739	0.7919
		(0.0982)	(0.0742)	(0.0976)	(0.0915)	(0.0996)	(0.0921)
	1	0.6592	0.6519	0.6660	0.6730	0.6702	0.6743
		(0.0595)	(0.0311)	(0.0734)	(0.0811)	(0.0794)	(0.0858)
	2	0.6406	0.6480	0.6358	0.6338	0.6365	0.6357
		(0.0503)	(0.0249)	(0.0625)	(0.0717)	(0.0668)	(0.0739)
500	0	0.9746	0.9257	0.9666	0.9751	0.9811	0.9818
		(0.0474)	(0.1250)	(0.0569)	(0.0406)	(0.0355)	(0.0317)
	1	0.8573	0.7614	0.8695	0.8973	0.8991	0.9069
		(0.1184)	(0.1386)	(0.1057)	(0.0806)	(0.0844)	(0.0705)
	2	0.7725	0.7034	0.7941	0.8255	0.8233	0.8373
		(0.1149)	(0.0977)	(0.0990)	(0.0990)	(0.1002)	(0.0876)

and standard deviation 1. The number of clusters is two. The goal to to look for two clusters of 10 time series each, as in the above examples. A truncated segment of length 751 time points of eye gaze data was considered for each movement. Details can be obtained from Section (4.3.2) in Chapter 3. In this analysis, we specifically look at the radial r_t coordinate. In the first method for classification, we employ the EXP models of order p. The distance used is obtained from Eqn. (5.15). Table 5.14 reports the similarity incides for seven subjects for various clustering algorithms. The similarity is computed for all values of p between 1 and 10. The table reports similarity index correponding to the value of p that minimises the AIC criterion. A similarity index equal to 1 is perfect clustering that gives the true clusters. Clearly, from the table the maximum value is 0.7494 for Subject 3 using DIANA algorithm.

Table 5.6: Average similarity indices using distance $d^2_{ABS}(.)$ in Eqn. (5.17) with first M frequencies

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6288	0.6416	0.6152	0.6075	0.6187	0.6077
		0.0403	0.0263	0.0535	.0609	0.0533	0.0636
	1	0.6278	0.6439	0.6105	0.6021	0.6124	0.6025
		0.0368	0.0218	0.0490	0.0537	0.0479	0.0559
	2	0.6285	0.6424	0.6088	0.5984	0.6142	0.5984
		0.0361	0.0240	0.0477	0.0532	0.0454	0.0546
100	0	0.6334	0.6430	0.6207	0.6157	0.6234	0.6147
		0.0414	0.0285	0.0555	0.0651	0.0536	0.0648
	1	0.6268	0.6423	0.6081	0.5997	0.6111	0.5995
		0.0389	0.0246	0.0468	0.0514	0.0467	0.0525
	2	0.6273	0.6439	0.6115	0.5985	0.6135	0.5967
		0.0386	0.0238	0.0481	0.0521	0.0460	0.0545
500	0	0.6379	0.6454	0.62995	0.6256	0.6336	0.6272
		0.0460	0.0275	0.0590	0.0709	0.0597	0.0742
	1	0.6260	0.6419	0.6087	0.5990	0.6129	0.5970
		0.0404	0.0236	0.0499	0.0554	0.0497	0.0568
	2	0.6264	0.6427	0.6104	0.6039	0.6142	0.6012
		0.0414	0.0270	0.0510	0.0573	0.0495	0.0582

The division of the 20 times series per subject into two clusters of 10 each did not result in the original true clusters under two conditions.

The same analysis was repeated using the FEXP model of the spectral densities. The distance $d^2_{ABS}(.)$ was used. The value of p was set equal to zero for all cases. This implies that only the constant and fractional parameter d was used in the distance $d^2_{ABS}(.)$. The reference was Gaussian white noise with mean zero and standard deviation 1. Again, as evident from Table 5.15 perfect true clusters are not attained. The maximum value for the similarity index is 0.7 for Subject 2. From the results of the two methods discussed, there is no clear seperation of two clusters of "Watch" and "Imitate" for any of the seven subjects.

Table 5.7: Average similarity indices using distance $d^2_{EUCL}(.)$ in Eqn. (5.16) with first M frequencies

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6283	0.6416	0.6144	0.6078	0.6168	0.6081
		0.0417	0.0261	0.0538	0.0596	0.0549	0.0624
	1	0.6269	0.6438	0.6114	0.6039	2 0.6100	0.6032
		0.0372	0.0217	0.0460	0.0526	0.0486	0.0527
	2	0.6271	0.6423	0.6088	0.6017	0.6104	0.5988
		0.0372	0.0240	0.0475	0.0523	0.0476	0.0512
100	0	0.6323	0.6425	0.6198	0.6139	0.6196	0.6148
		0.0417	0.0269	0.0537	0.0611	0.0557	0.0635
	1	0.6257	0.6426	0.6093	0.6013	0.6089	0.6000
		0.0399	0.0243	0.0469	0.0499	0.0483	0.0522
	2	0.6268	0.6437	0.6121	0.6015	0.6097	0.6008
		0.0393	0.0238	0.0490	0.0516	0.0490	0.0512
500	0	0.6370	0.6443	0.6285	0.6245	0.6306	0.6251
		0.0465	0.0263	0.0599	0.0677	0.0628	0.0686
	1	0.6251	0.6420	0.6081	0.6017	0.6085	0.5992
		0.0404	0.0235	0.0494	0.0543	0.0529	0.0555
	2	0.6251	0.6430	0.6098	0.6046	0.6121	0.6013
		0.0419	0.0265	0.0513	0.0566	0.0530	0.0576

Table 5.8: Average similarity indices using distance $d^1_{ABS}(.)$ in Eqn. (5.15)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.6304	0.6493	0.6117	0.6036	0.6151	0.6007
		(0.0350)	(0.0150)	(0.0477)	(0.0556)	(0.0476)	(0.0557)
	2	0.6364	0.6519	0.6133	0.6005	0.6166	0.6013
		(0.0307)	(0.0103)	(0.0415)	(0.0560)	(0.0479)	(0.0577)
	3	0.6401	0.6531	0.6132	0.6011	0.6182	0.6007
		(0.0295)	(0.0080)	(0.0490)	(0.0529)	(0.0483)	(0.0584)
100	1	0.6319	0.6499	0.6115	0.6031	0.6139	0.6011
		(0.0369)	(0.0158)	(0.0516)	(0.0583)	(0.0498)	(0.0579)
	2	0.6355	0.6516	0.6119	0.6033	0.6144	0.5994
		(0.0313)	(0.0110)	(0.0494)	(0.0560)	(0.0503)	(0.0559)
	3	0.6389	0.6527	0.6147	0.6045	0.6138	0.6047
		(0.0298)	(0.0091)	(0.0482)	(0.0559)	(0.0485)	(0.0576)
500	1	0.6396	0.6504	0.6286	0.6266	0.6318	0.6268
		(0.0427)	(0.0151)	(0.0631)	(0.0749)	(0.0639)	(0.0771)
	2	0.6428	0.6526	0.6326	0.6309	0.6331	0.6318
		(0.0401)	(0.0143)	(0.6779)	(0.0786)	(0.0681)	(0.0797)
	3	0.6442	0.6527	0.6300	0.6306	0.6323	0.6261
		(0.0342)	(0.0091)	(0.0600)	(0.0715)	(0.0635)	(0.0747)

Table 5.9: Average similarity indices using distance $d^1_{EUCL}(.)$ in Eqn. (5.14)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.6303	0.6492	0.6129	0.6045	0.6112	0.6028
		(0.0360)	(0.0154)	(0.0466)	(0.0548)	(0.0493)	(0.0563)
	2	0.6358	0.6519	0.6152	0.6013	0.6144	0.6030
		(0.0319)	(0.0118)	(0.0469)	(0.0541)	(0.0491)	(0.0546)
	3	0.6390	0.6525	0.6045	0.6052	0.6169	0.6015
		(0.0292)	(0.0087)	(0.0548)	(0.0530)	(0.0471)	(0.0533)
100	1	0.6317	0.6491	0.6142	0.6046	0.6126	0.6019
		(0.0378)	(0.0136)	(0.0495)	(0.0544)	(0.0516)	(0.0567)
	2	0.6350	0.6509	0.6136	0.6051	0.6120	0.6014
		(0.0302)	(0.0126)	(0.0485)	(0.0565)	(0.0495)	(0.0548)
	3	0.6376	0.6524	0.6132	0.6044	0.6139	0.6026
		(0.0305)	(0.0092)	(0.0470)	(0.0526)	(0.0478)	(0.0560)
500	1	0.6363	0.6498	0.6259	0.6267	0.6298	0.6274
		(0.0415)	(0.0164)	(0.0614)	(0.0720)	(0.0644)	(0.0735)
	2	0.6415	0.6520	0.6298	0.6291	0.6320	0.6291
		(0.0406)	(0.0142)	(0.0614)	(0.0743)	(0.0688)	(0.0763)
	3	0.6423	0.6528	0.6274	0.6266	0.6277	0.6246
		(0.0331)	(0.0102)	(0.0571)	(0.0679)	(0.0618)	(0.0735)

Table 5.10: Average similarity indices using distance $d^2_{ABS}(.)$ in Eqn. (5.17)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6264	0.6426	0.6117	0.6021	0.6161	0.6032
		(0.0406)	(0.0249)	(0.0507)	(0.0558)	(0.0529)	(0.0571)
	1	0.6260	0.6422	0.6121	0.6025	0.6143	0.6026
		(0.0417)	(0.0252)	(0.0488)	(0.0545)	(0.0508)	(0.0563)
	2	0.6260	0.6428	0.6087	0.6002	0.6129	0.5994
		(0.0402)	(0.0259)	(0.0488)	(0.0546)	(0.0504)	(0.0574)
100	0	0.6299	0.6438	0.6191	0.6148	0.6208	0.6159
		(0.0457)	(0.0263)	(0.0548)	(0.0627)	(0.0590)	(0.0662)
	1	0.6253	0.6414	0.6093	0.6041	0.6124	0.6038
		(0.0412)	(0.0252)	(0.0486)	(0.0554)	(0.0510)	(0.0558)
	2	0.6251	0.6427	0.6101	0.5994	0.6097	0.5980
		(0.0398)	(0.0247)	(0.0472)	(0.0500)	(0.0482)	(0.0521)
500	0	0.6588	0.6512	0.6676	0.6728	0.6670	0.6767
		(0.0580)	(0.0298)	(0.0760)	(0.0844)	(0.0759)	(0.0851)
	1	0.6361	0.6465	0.6276	0.6258	0.6303	0.6250
		(0.0504)	(0.0259)	(0.0626)	(0.0725)	(0.0646)	(0.0703)
	2	0.6313	0.6452	0.6203	0.6143	0.6207	0.6110
		(0.0479)	(0.0259)	(0.0577)	(0.0670)	(0.0607)	(0.0660)

Table 5.11: Average similarity indices using distance $d^2_{EUCL}(.)$ in Eqn. (5.16)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	0	0.6253	0.6427	0.6113	0.6029	0.6131	0.6033
		(0.0414)	(0.0253)	(0.0514)	(0.0541)	(0.0533)	(0.0568)
	1	0.6256	0.6427	0.6106	0.6040	0.6102	0.6033
		(0.0423)	(0.0268)	(0.0492)	(0.0523)	(0.0523)	(0.0552)
	2	0.6250	0.6423	0.6094	0.6029	0.6105	0.6003
		(0.0407)	(0.0262)	(0.0487)	(0.0531)	(0.0524)	(0.0554)
100	0	0.6299	0.6442	0.6186	0.6148	0.6194	0.6140
		(0.0460)	(0.0267)	(0.0568)	(0.0608)	(0.0607)	(0.0645)
	1	0.6248	0.6412	0.6119	0.6056	0.6097	0.6059
		(0.0426)	(0.0259)	(0.0508)	(0.0548)	(0.0513)	(0.0568)
	2	0.6258	0.6431	0.6094	0.6006	0.6085	0.5985
		(0.0396)	(0.0237)	(0.0479)	(0.0502)	(0.0503)	(0.0519)
500	0	0.6579	0.6512	0.6635	0.6711	0.6666	0.6704
		(0.0559)	(0.0300)	(0.0759)	(0.0830)	(0.0774)	(0.0829)
	1	0.6367	0.6464	0.6310	0.6276	0.6302	0.6273
		(0.0503)	(0.0252)	(0.0635)	(0.0712)	(0.0627)	(0.0734)
	2	0.6323	0.6446	0.6208	0.6163	0.6220	0.6137
		(0.0487)	(0.0253)	(0.0586)	(0.0663)	(0.0639)	(0.0674)

Table 5.12: Average similarity indices using distance $d^1_{ABS}(.)$ in Eqn. (5.15)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.8549	0.7037	0.8712	0.9170	0.8936	0.9304
		(0.1365)	(0.1126)	(0.1252)	(0.0708)	(0.1205)	(0.0718)
	2	0.8616	0.6990	0.8858	0.9274	0.9078	0.9296
		(0.1407)	(0.1095)	(0.1179)	(0.0713)	(0.1160)	(0.0736)
	3	0.8367	0.6777	0.8720	0.9156	0.8915	0.9090
		(0.1448)	(0.0806)	(0.1232)	(0.0815)	(0.1214)	(0.0811)
100	1	0.9835	0.8912	0.9803	0.9863	0.9866	0.9894
		(0.0515)	(0.1555)	(0.0532)	(0.0301)	(0.0506)	(0.0242)
	2	0.9902	0.8935	0.9859	0.9902	0.9910	0.9907
		(0.0349)	(0.1557)	(0.0393)	(0.0249)	(0.0391)	(0.0223)
	3	0.9868	0.8640	0.9862	0.9895	0.9987	0.9880
		(0.0485)	(0.1654)	(0.0369)	(0.0261)	(0.0449)	(0.0248)
500	1	1	1	1	1	1	1
		(0)	(0)	(0)	(0)	(0)	(0)
	2	1	1	1	1	1	1
		(0)	(0)	(0)	(0)	(0)	(0)
	3	1	1	1	1	1	1
		(0)	(0)	(0)	(0)	(0)	(0)

Table 5.13: Average similarity indices using distance $d^2_{ABS}(.)$ in Eqn. (5.17)

N	p	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
50	1	0.9079	0.8044	0.9154	0.9279	0.9318	0.9392
		(0.1008)	(0.1508)	(0.0841)	(0.0676)	(0.0790)	(0.0575)
	2	0.6629	0.6575	0.6644	0.6861	0.6696	0.6847
		(0.0643)	(0.0416)	(0.0789)	(0.0930)	(0.0817)	(0.0939)
	3	0.6334	0.6475	0.6202	0.6189	0.6220	0.6154
		(0.0427)	(0.0214)	(0.0546)	(0.0704)	(0.0570)	(0.0676)
100	1	0.9884	0.9640	0.9853	0.9863	0.9911	0.9922
		(0.0278)	(0.0925)	(0.0394)	(0.0288)	(0.0307)	(0.0200)
	2	0.7344	0.6910	0.7418	0.8073	0.7717	0.8149
		(0.1079)	(0.0902)	(0.1038)	(0.1041)	(0.1087)	(0.0951)
	3	0.6503	0.6544	0.6438	0.6669	0.6529	0.6789
		(0.0579)	(0.0364)	(0.0727)	(0.0980)	(0.0781)	(0.0982)
500	1	1	1	1	1	1	1
		(0)	(0)	(0)	(0)	(0)	(0)
	2	0.9974	0.9964	0.9855	0.9965	0.9976	0.9962
		(0.0144)	(0.0300)	(0.0497)	(0.0156)	(0.0134)	(0.0150)
	3	0.9628	0.9699	0.8617	0.9868	0.9638	0.9751
		(0.0930)	(0.0946)	(0.1422)	(0.0360)	(0.0869)	(0.0396)

Table 5.14: Similarity indices using distance $d^1_{ABS}(.)$ applied on eye gaze data

Subject	p	Average	Single	Complete	Ward	Diana	Pam
1	4	0.5833	0.6154	0.5833	0.5833	0.5833	0.6000
2	6	0.5960	0.6552	0.5960	0.5960	0.5960	0.5960
3	8	0.6970	0.6552	0.6970	0.7494	0.7494	0.6154
4	2	0.6553	0.6552	0.6154	0.6000	0.6154	0.6154
5	2	0.6296	0.6296	0.6296	0.5833	0.5833	0.5833
6	6	0.6419	0.6552	0.6419	0.6419	0.5833	0.6419
7	10	0.6429	0.6429	0.5489	0.5489	0.6429	0.5960

Table 5.15: Similarity indices using distance $d^2_{ABS}(.)$ applied on eye gaze data

Subject	AVERAGE	SINGLE	COMPLETE	WARD	DIANA	PAM
1	0.5833	0.5833	0.5833	0.5833	0.5833	0.5833
2	0.5960	0.6552	0.7000	0.7000	0.6491	0.6491
3	0.6296	0.5833	0.5833	0.5833	0.5652	0.6000
4	0.6552	0.6552	0.6552	0.6000	0.6552	0.6154
5	0.6154	0.6154	0.5833	0.5833	0.5833	0.6000
6	0.6429	0.6552	0.6419	0.6875	0.6429	0.6970
7	0.6429	0.6429	0.6429	0.5489	0.6429	0.6419

Chapter 6

Conclusions

The main goal of this dissertation is to look for possible differences in the eye gaze under "Watch" and "Imitate" conditions. We first investigate statistical properties in eye gaze time series in response to simple one arm movements. The analysis suggests the eye gaze data to be a long memory process. Our modeling efforts indicate that this long memory property have to be taken into account. In particular, the first two lag terms in the long memory part play a significant role and is included in all the models. Hence the previous two time points i.e. information in the last 8 to 16 milliseconds is crucial for predicting future eye gaze positions. Also since the eye gaze is guided by the movements, the arm coordinates play an important role. The arm components are a part of the best model and are useful for predicting future eye gaze positions.

We approach the problem of discriminating between two conditions from two different directions. In the time domain, we look at higher order crossing (HOC) sequences of each of time series. The HOC sequence quantifies the oscillatory behaviour of a given time series which can be used as a discriminator between two time series. We study few properties of HOC for stationary long memory processes. We introduced an estimator for the fractional paramter d using zero-order crossings for ARFIMA processes. We compare this estimator with maximum likelihood and

Whittle's estimator. The performance is found to be comparable to Whittle's estimator. For purpose of discriminating, we use ψ^2 measure which is based on distances from white noise process. A simulation experiment suggested that this measure can be used to detect differences if the processes themselves are largely seperated. In fact, the application of ψ^2 measure on radial coordinate pointed out differences for some of the subjects.

In the spectral domain, we use exponential (EXP) and fractional exponential (FEXP) model forms of the spectral densities. The method used is based on comparing several spectral densities to a reference spectral density. In our case, we use the Gaussian white noise process as reference. The parameters involved in the logarithm of ratio of spectral densities are further used in distance measures for classification purposes. The EXP model has been extensively used in literature for classification of various linear and non-linear processes. We investigate the performance of EXP parameter based distance measures for ARFIMA models through simulation studies. It is found to be a good measure for processes that are well seperated in terms of their parameter values. This measure is dependent on the order p. The value of p can be determined as the one that minimizes the AIC. We extend this application by using the FEXP model form. To estimate the parameters in FEXP(p) we developed a novel likelihood method. This likelihood behaves well and is based on the assumption that for all Fourier frequencies the periodogram ordinates are independent and their distributions are the same. For classification in simulated studies, distance measure based on the d only was found to be performing well. Hence long memory parameter d is an important classifier and it is estimated

from all the frequencies.

The spectral domain based classification method were applied to eye gaze date. The application did not result in clear seperation of the two conditions. For all the seven subjects, perfect clustering into two true clusters or groups was not attained.

As an extension of this analysis, further investigation can be carried out using bivariate time series $\{x_t, y_t\}$ or $\{r_t, \theta_t\}$. In the bivariate setting, behaviour of eye gaze data, its modeling and classification can be studied. In fact, classification of long memory processes can be investigated in the bivariate/multivariate setting. Also, the use of FEXP model parameters in classification can be replaced by parameters in the orthogonal FEXP model and their asymptotic properties as well as performance can be studied in clustering.

Bibliography

[1] Dickstein P. A. (2009). Identification of changes in the entropy of seismic signals preceding an event through higher order zero-crossing analysis. Seismological Research Letters, 80(3):473-478.

[2] Barkoulas J., Baum C. and Oguz G. (1997). Stochastic long memory in traded goods prices. Boston college working papers in economics.

[3] Beran J. (1993). Fitting long-memory models by generalized linear regression. Biometrika, 80(4).

[4] Beran J. (1994). *Statistics for long-memory processes.* Chapman & Hall, New York.

[5] Bloomfield P. (1973). An exponential model for the spectrum of a scalar time series. Biometrika, 60(2):217-226.

[6] Box G.E.P. and Cox D.R. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

[7] Box G.E.P., Jenkins G.M., and Reinsel G. (1994). *Time Series Analysis, Forecasting, ad Control.* Prentice Hall, Upper Saddle River, N.J., 3rd edition.

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

[9] Caiado J., Crato N. and Pena D. (2006). A periodogram-based metric for time series classification. Computational Statistics & Data Analysis, 50(10):2668-2684.

[10] Coates D. S. and Diggle P. J. (1986). Tests for comparing two estimated spectral densities. Journal of Time Series Analysis, 7.

[11] Dahlhaus R. (1989). Efficient parameter estimation for self-similar processes. The Annals of Statistics, 17(4):1749-1766.

[12] Demirel Omer F., Willemain Thomas R. (2003). Nonparametric methods of process discrimination and model validation using zero crossings. Communications in statistics. Simulation and computation, 32(2):517-539.

[13] Diggle P. (1990). *Time series: a biostatistical introduction.* Oxford University Press, illustrated, reprint edition.

[14] Fokianos K. and Savvides A. (2008). On comparing several spectral densities. Tech- nometrics, 50(3):317-331.

[15] Giraitis L. and Surgailis D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittles estimate. Probability Theory Related Fields, 86:87-104.

[16] Hadjileontiadis L.J.(2003). Discrimination analysis of discontinuous breath sounds using higher-order crossings. Medical and Biological Engineering and Computing, 41(4):445-455.

[17] Han. S. (2008). Comparing spectral densities in replicated time series by smoothing spline anova. PhD thesis, University of Pittsburg, USA.

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

[19] Kedem B. (1994). *Time Series Analysis by Higher Order Crossings.* IEEE Press, Piscataway, NJ.

[20] Kedem, B. and Slud, E. (1982). Time series discrimination by higher order crossings. Annals of statistics, 10(3):786-794.

[21] Kong X., Qiu T. (2000). Injury detection for central nervous system via EEG with higher order crossing-based methods. Methods of Information in Medicine, 39(2).

[22] Lee D., Schmidt P. (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. Journal of Econometrics, 73:285-302.

[23] Li T. (1996). Discrimination of time series by parametric filtering. Journal of the American Statistical Association, 91(433):284-293.

[24] Mataric, M. J. and Pomplun, M. (1998). Fixation Behavior in observation and imitation of human movement. Cognitive Brain Research, 7(2).

[25] Dickstein P. A., Spelt J. K., Sinclair A. N. (1991). Application of higher order crossing feature to non-destructive evaluation: A sample demonstration of sensitivity to the condition of adhesive joints. Ultrasonics, 29(5):355-365.

[26] Noy L. (2009). From Action to Perception and Back Again: Studying the features of movement imitation. PhD thesis, Weizmann Institute of Science, Rehovot, Israel.

[27] Palma W. (2007). *Long-memory time series, Theory and methods.* Wiley-Interscience.

[28] Reisen V. A., Abraham B., Toscano E. M. (2000). Parametric and semiparametric estimators of stationary univariate ARFIMA models. Brazilian Journal of Probability and Statistics, 14:185-206.

[29] Reisen V.A. (1994). Estimation of the fractional difference parameter in the ARFIMA(p, d, q) model using the smoothed periodogram. Journal of Time Series Analysis, 15(3):335-350.

[30] Robinson P.M. (1995). Log-Periodogram regression of time series with long range dependence. The Annals of Statistics, 23(3):1048-1072.

[31] Savvides A.(2008). Comparison and clustering of several spectral densities. PhD thesis, University of Cyprus.

[32] Savvides A. and Promponas V. J. and Fokianos K. (2008). Clustering of biological time series by cepstral coefficients based distances. Pattern Recognition, 41:2398-2412.

[33] Shumway R.H. and Stoffer D.S. (2000). *Time Series Analysis and Its Applications.* Springer-Verlag, New York.

[34] Taqqu M. S. and Teverovsky V. (1996). Robustness of Whittle-type estimators for time series with long-range dependence. Technical Report, Boston University.

[35] Warrenliao T. (2005). Clustering of time series data survey. Pattern Recognition, 38(11):1857-1874.

[36] Whittle P. (1953). Estimation and information in stationary time series. Arkiv for Matematik, 2:423-434.

[37] Yarbus A.L. (1967). *Eye Movements and Vision.* Plenum, New York.