COMPARISON AND CLUSTERING OF SEVERAL SPECTRAL DENSITIES

By

Alexios Savvides

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF CYPRUS
NICOSIA, CYPRUS
FEBRUARY 2008

To my family.

Table of Contents

List of Tables

List of Figures

Abstract

Consider the problem of estimating and comparing several spectral densities, say G, and assume that the first $G-1$ of them are related with the last one by an exponential model. Based on the asymptotic properties of periodogram ordinates, we develop parametric likelihood inference for this model. More specifically, we study in detail the asymptotic behavior of the maximum likelihood estimator under the semiparametric model. The results can be applied in a variety of situations including linear processes. Simulations and data analysis support further the theoretical findings.

The new methodology is applied to time series clustering. Methods for clustering are based on the calculation of suitable similarity measures which identifies the distance between two or more time series. New measures of distance are proposed and they are based on the so–called cepstral coefficients which carry information about the log spectrum of a stationary time series. These coefficients are estimated by means of the semiparametric model which was discussed earlier. After estimation, the estimated cepstral distance measures are given as an input to a clustering method to produce the disjoint groups of data. Simulated examples show that the method yields good results, even when the processes are not necessarily linear.

Περίληψη

Θεωρούμε το πρόβλημα εκτίμησης και σύγκρισης πολλών φασματικών συναρτήσεων πυκνοτήτων πιθανοτήτας, έστω G. Υποθέτουμε ότι οι $G-1$ από αυτές συνδέονται με την τελευταία μέσω ενός κατάλληλου ημιπαραμετρικού εκθετικού μοντέλου. Από τις ασυμπτωτικές ιδιότητες του περιοδογράμματος, αναπτύσσουμε παραμετρική συμπερασματολογία με βάση την σχετική πιθανοφάνεια για ένα ημιπαραμετρικό μοντέλο. Τα αποτελέσματα μπορούν να εφαρμοστούν σε μια πληθώρα καταστάσεων συμπεριλαμβανομένου των γραμμικών διαδικασιών. Προσομοιώσεις και ανάλυση δεδομένων υποστηρίζουν περαιτέρω τα θεωρητικά αποτελέσματα.

Η νέα μεθοδολογία εφαρμόζεται σε χρονοσειρές για ανάλυση κατά συστάδες. Γενικές μεθόδοι για την ανάλυση κατά συστάδες στηρίζονται στον υπολογισμό μιας απόστασης μεταξύ δύο ή περισσότερων χρονοσειρών. Νέες αποστάσεις προτείνονται και ο υπολογισμός τους βασίζεται στους συντελεστές Fourier της λογαριθμικής φασματικής συνάρτησης πυκνότητας πιθανότητας. Οι συντελεστές εκτιμώνται με βάση το προαναφερθέν ημιπαραμετρικό μοντέλο. Οι καινούργιες μετρικές προσδιορίζουν το διαχωρισμό των δεδομένων καλύτερα σε σχέση με τις ήδη υπάρχουσες.

Acknowledgements

I would like to thank Konstantinos Fokianos, my supervisor, for his many suggestions and constant support during this research and also for his guidance through the early years of chaos and confusion. He provided to me many useful references and friendly encouragement and never discourage me. In addition I would like to thank all the other members of the examining committee, Prof. M. Pourahmadi, Prof. H. Ombao, Prof. E. Paparoditis and Prof. T. Sapatinas for several useful comments and suggestions that improved the thesis.

Professor Vasilis Probonas expressed his interest in my work and supplied me with biological data which gave me a better and a new perspective on my own results. Professor Pashalides shared with me chemical data which enhance the theoretical part of my thesis. I also would like to thank Prof. S. Holan for providing his software for data generation from the exponential model and Prof. T. Christofides, A. Karagrigoriou and I. Vonta for support and encouragement.

Of course, I am grateful to my parents for their patience and *love*. Without them this work would never have come into existence (literally).

Nicosia, Cyprus — Alexios Savvides
February 29, 2008

Chapter 1

Introduction

A time series is a collection of random variables, say $\{Y_t\}$, $t = 1, \ldots, N$, ordered in time. The seminal texts by Priestley (1981), Brockwell and Davis (1991) and Shumway and Stoffer (2006) to name a few, provide an introduction to the subject while discussing statistical analysis and forecasting. The assumptions of stationarity, Gaussianity and linearity dominate the results in this area. In particular, stationarity implies the existence of the so called spectral density function under some assumptions–an important tool for time series data analysis.

The main goal of this thesis is to show how to compare several independent spectral density functions. This problem occurs frequently in diverse applications. Consider for instance Figure 1.1 which depicts the first 150 photometric measurements–out of 1000–that have been obtained under some certain chemical conditions, for determining the absorbance of a Cu(II) solution at three different wavelengths during a time period of 100 seconds. We denote by Y_{it} the three observed time series where $i = 1, 2, 3$ corresponds to the three distinct wavelengths, that is 465nm, 665nm and 865nm, respectively. The scientific question that is posed is whether or not there exists stability of the emission intensity of the source at different wavelengths. Equivalently, the problem can be posed as whether or not the dynamics of the three observed processes share certain characteristics. More details for these data are available in Section 3.7.2 but it is pointed out that similar questions arise naturally in many other different disciplines.

These data motivate this part of research, namely whether or not certain second order properties of several independent processes can be identified and shown to be similar. It is customary, in time series analysis, that second order properties of a stationary process to be studied by means of the spectral density function. However, to explore how second order

properties of several processes can be compared, we propose a model which connects the "spectral density ratio" of the last spectral density to the rest. To be more specific, consider the previous photometry data example and suppose that $\lambda_i(\omega)$, $\omega \in (-\pi, \pi)$, stands for the corresponding spectral density function of the process Y_{it}, $i = 1, 2, 3$. Setting as reference the spectral density $\lambda_3(\omega)$ we consider the ratios $\lambda_1(\omega)/\lambda_3(\omega)$ and $\lambda_2(\omega)/\lambda_3(\omega)$, for modeling, estimation and inference through a finite dimensional parameter. Notice that within the proposed framework, it is not important what denominator will be used to form these ratios; it is rather a matter of convenience since the inferential results are not altered. The approach is quite general and it will be shown that processes like the well-known ARMA processes fall within this framework. In addition, the assumption of normality does not affect the results. Based on this method, the data reveals processes with similar second order structure by using standard estimation and testing theory. Additionally, the problem of testing whether two or more spectral densities are equal reduces to a parametric problem whose solution is given by the asymptotic properties of periodogram ordinates and related theory, see also (Taniguchi and Kakizawa, 2000, Ch. 6).

The use of spectral ratio modeling has been initiated by Coates and Diggle (1986) and discussed further in the text by Diggle (1990), see also Brillinger (1981, Sec. 7.9) who has a good discussion on several series arising from designs of experiments. In addition, Diggle and Fisher (1991) proposed a graphical procedure for the comparison of two periodograms and they suggested a non parametric test for the equality of two spectra. Some related work can be found in Cameron and Turner (1987) and Beran (1993) who show the problem of spectral density estimation can be reformulated in terms of a generalized linear model. Likelihood ratio modeling of two or more probability density functions has been also considered by Fokianos (2004) in the context of biased sampling models. Some other related works to the comparison of spectral densities are given by Timmer et al. (1999) who concentrate on spectral peaks, by Maharaj (2002) who compares evolutionary spectra of non-stationary processes using randomization tests and by Dette and Paparoditis (2007) who propose a test based on an appropriate L_2-distance measure between the nonparametrically estimated individual spectral densities and an overall pooled spectral density estimator. Also Biau et al. (2005) develops a procedure for classifying curves based

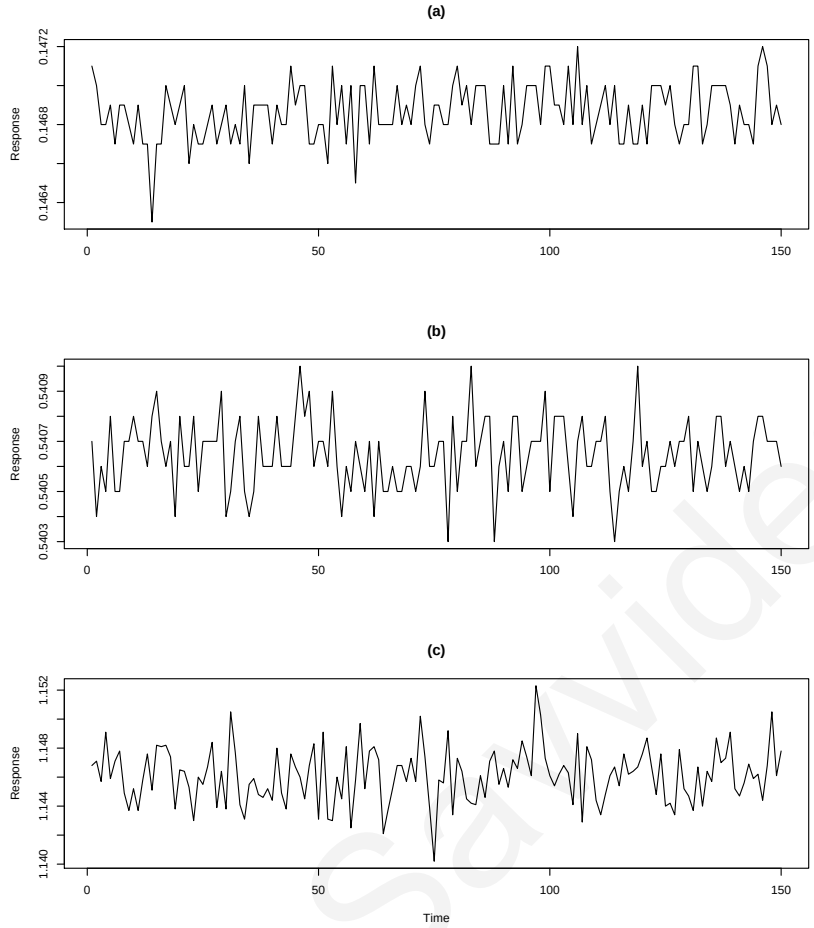

Figure 1.1: Time series plots of the first 150 photometric measurements of the absorbance of Cu (II) solution at three different wavelengths. (a) Y_{1t}, (b) Y_{2t}, (c) Y_{3t}.

on their Fourier transforms. They show that when used in conjunction with the k-nearest-neighbors approach their classification procedure is consistent. Finally Choi et al. (2008) uses the log ratio of periodograms (symmetrized) for testing changes in adjacent blocks of time series. This approach is based on a frequency-by-frequency comparison approach. In contrast, the proposed approach is based on the direct comparison of the spectral densities. The approach which is based on a frequency-by-frequency comparison leads to the identification of frequencies which contribute the most to any foreseen differences between the spectral densities. However, the approach which is based on comparing the spectral densities directly offers an overall dissimilarity measure between several time series.

This work extends the research of Coates and Diggle (1986) and Diggle (1990) to the case of multiple time series in several directions. The proposed method does not rely to any specific parametric forms and it only assumes that the log–likelihood ratio of the spectral densities is linear in a sense that will be described in the next sections. In addition thorough asymptotic theory is developed along the lines of (Taniguchi and Kakizawa, 2000,

Ch. 6).

Parametric testing and the spectral density ratio model allows for clustering of time series. Classification and clustering are data reduction methods for massive data sets that so frequently occur in applications. The purpose of clustering, which is the main interest of the second part of the thesis, is to obtain an assignment rule which divides the data set into homogeneous groups. Here, the notion of homogeneity means that objects within the group are similar while objects between groups are not that similar with respect to second order properties. In the context of independent data these concepts have been studied extensively, see Johnson and Wichern (1992) and Hastie et al. (2001), for instance. However, for massive time series data sets, which appear very often in applications, there does not seem to be such an extensive literature. The aim of this work is to contribute towards this research by introducing similarity (or dissimilarity) measures based on the so called cepstral coefficients. These coefficients are simply the Fourier coefficients of the log spectrum of a stationary process. Hence, the methodology is developed in the frequency domain which simplifies calculations because of the asymptotic independence of periodogram ordinates. In addition, focusing on spectral domain yields clustering of time series with similar second order structure–that is a feature based approach. This is one aspect of our work. The additional feature, which can be called as model based approach, is the employment of the semi parametric model that was discussed before. Using this model, we estimate a distance between two or more time series and then apply a specific clustering algorithm. Some different approaches have been suggested in the literature but a detailed review of the topic will be given later. We only mention here that the closest approach to ours is that taken by Kalpakis et al. (2001) who have studied the euclidian distance between cepstral coefficients of two or more time series in the context of ARIMA modeling. However there are fundamental differences between their approach and the method that is suggested in this work. More details and further references will be given later but the interested reader should see Liao (2005) and Shumway and Stoffer (2006) for a comprehensive review on the topic of clustering time series.

Motivation for studying this problem comes from the need of identifying similar physicochemical properties–such as hydrophobicity–of amino acid sequences, as proclaimed in the spectral domain. Identifying similar structures is useful for successful discrimination

among protein sequences which belong into distinct biologically relevant classes. Consider for instance, bacterial pathogenicity. It is well known that this physiological process, requires the export of certain proteins. These proteins are initially synthesized at the ribosomes which are located in the bacterial cytoplasm, and transported to the periplasmic space or the surrounding medium. This functionality is mediated by specialized transport molecules that help selected proteins to cross the otherwise impermeable to them lipid bilayer of the bacterial plasma membrane. Although several secretion pathways exist, it has been shown that–at least in the vast majority of known cases–all the information required for the initial export of bacterial proteins from the plasma membrane resides in short sequence segments. These segments are termed as signal peptides and are of variable length. They are located in the N-terminal region of the nascent polypeptide chain, see Emanuelsson et al. (2000). More specifically, these N-terminal signals are recognized by special protein molecules that assist their translocation through dedicated transmembrane protein channels located on the bacterial plasma membrane. The best characterized secretion pathway so far is the so–called Sec-dependent pathway. A general feature of signal peptides is a tripartite sequential structure consisting of a positively charged n-region, followed by a hydrophobic h-region and a polar uncharged c-region of variable respective lengths, see Emanuelsson et al. (2000). A novel pathway, namely the Tat (Twin arginine translocation) pathway, has been recently discovered in bacteria, Berks (1996); Berks et al. (2005). Although there seems to exist significant differences in the molecular mechanism of protein export via those two pathways, proteins entering the Tat pathway have signal peptides with a tripartite structure resembling the one of proteins exported by the Sec molecular machinery. Nevertheless, they often possess two (initially thought invariant) consecutive arginine amino acid residues at the border between the n- and h-regions (Bendtsen et al. (2005)). Aiming to reveal hidden protein sequence features, we have performed an analysis of fixed-length N-terminal amino acid sequences (50 residues long) from different bacterial proteins with experimental evidence for the presence of a secretory N-terminal signal peptide, either of the classical or of the Tat form. To carry out the identification of similar hidden features in amino acid sequences of well characterized proteins bearing either Tat- or Sec-machinery specific secretory signal peptides we consider standard spectral domain clustering techniques together with the cepstral distance

which was discussed earlier. It is not aimed to develop a prediction method neither to benchmark different clustering techniques, but rather to illustrate the power of novel cepstral coefficient–based spectral analysis tools in biological sequence analysis. As a general remark, spectral analysis tools have been extensively used in biological and biomedical research for more than two decades. In particular, there have been many successful applications in Computational Biology in diverge areas such as gene finding (Issac et al. (2002)), periodicity analysis (Pasquier et al. (2001)), proteomics (Yates (2004)), study of fibrous proteins (McLachlan and Stewart (1976)) and protein functional classification (Pasquier et al. (1998)). Very recently, cepstral–based measures have been used in an application of protein classification, in a proof–of–principle demonstration (Pham (2006)).

In summary, this thesis suggests a new model for the comparison of several independent spectral density functions. Estimation and testing are discussed in detail while the new method points out to new cepstral based clustering techniques.

Chapter 2

The Exponential Model for the Spectrum of a Stationary Process

2.1 Introduction

In this section we set up the basic notation and model assumptions. We define stationarity, the spectral density function of a stationary process and the periodogram, an essential tool for the sequel. The basic model we refer to is the so called exponential model for the spectrum of a stationary process which was introduced by Bloomfield (1973). Also we have some examples from the exponential model and we discuss a way to generate data from this class, of models following Pourahmadi (1983), Pourahmadi (1984) and Holan (2004). Finally note that the monograph Britton (1983) is also relevant to the exponential spectrum theory.

2.2 Notation and Model Assumptions

Most of following material is based on the texts of Priestley (1981) and Brockwell and Davis (1991).

Definition 2.2.1. (The Autocovariance Function) If $\{X_t, t \in \mathbb{Z}\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$, is a process such that $Var(X_t) < \infty$ for each $t \in \mathbb{Z}$, where $\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$, then the autocovariance function is defined by

$$\gamma_x(r, s) = Cov(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)] \quad , r, s \in \mathbb{Z}$$

The concept of stationarity is of central importance to the field of time series. Basically, the main notion is that the second order properties of the process do not depend on time but rather on the lag separation between two different time points.

Definition 2.2.2. (Stationarity) A time series $\{X_t, t \in \mathbb{Z}\}$ is said to be stationary if:

(i) $E|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$,

(ii) $EX_t = m$ for all $t \in \mathbb{Z}$,

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

In particular, property (iii) from above shows that $\gamma_x(r, s) = \gamma_x(r - s)$ which is a function of one variable.

Every stationary stochastic process with absolutely summable autocovariance function possess its spectral density function which is simply the inverse Fourier transform of the autocovariance function, see Brockwell and Davis (1991, Ch. 4.3).

Definition 2.2.3. (Spectral density function) Suppose that $\{X_t,\ t \in \mathbb{Z}\}$ is a stationary process with $E(X_t) = 0$ and autocovariance function $\gamma_x(.)$ satisfying

$$\sum_{h=-\infty}^{+\infty} |\gamma_x(h)| < \infty.$$

Then, the spectral density of $\{X_t\}$ is defined by

$$\lambda_x(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_x(h) \exp(-ih\omega), \qquad -\pi \leq \omega \leq \pi.$$

Non parametric estimation of the spectral density function, is based on the following quantity which is called the periodogram, see Brockwell and Davis (1991, Ch. 10.1).

Definition 2.2.4. (Periodogram) Consider $X := (X_1, \cdots, X_N)$, where $X_1, \cdots, X_N$ is an arbitrary set of observations made at times $1, \ldots, N$. The value $I(\omega_i)$ of the periodogram of X at frequency $\omega_i = 2\pi i/N$, $i = 1, \ldots, [(N-1)/2]$ is defined by:

$$I_X(\omega_i) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} X_t \exp(-it\omega_i) \right|^2,$$

where $[y]$ denotes the integer part of y.

Definition 2.2.5. (Extension of the Periodogram) For any $\omega \in [-\pi, \pi]$ the periodogram is defined as follows:

$$I_X(\omega) = \begin{cases} I_X(\omega_k), & \text{if } \omega_k - \frac{\pi}{N} < \omega \leq \omega_k + \frac{\pi}{N} \text{ and } 0 \leq \omega \leq \pi \\ I_X(-\omega), & \text{if } \omega \in [-\pi, 0] \end{cases}$$

Proposition 2.2.1. (Brockwell and Davis (1991, Ch.10.3)). If $\{X_t\}$ is stationary with mean μ and absolutely summable autocovariance function $\gamma_x(.)$, then the following hold:

i) $EI_X(0) - \frac{n}{2\pi}\mu^2 \to \lambda(0)$,

ii) $EI_X(\omega) \to \lambda(\omega) \;\; if \;\; \omega \neq 0$.

In particular, when $\mu = 0$, then $EI_X(\omega)$ converges uniformly to $\lambda(\omega)$ on $[-\pi, \pi]$.

It turns out that the periodogram ordinates are asymptotically independent. More precisely, the following is true:

Proposition 2.2.2. (Brockwell and Davis (1991, Ch.10.3)). Suppose that $\{X_t\} \sim IID(0, \sigma^2)$ and let $I_X(\omega)$, $-\pi \leq \omega \leq \pi$, denote the periodogram of $\{X_1, \cdots, X_N\}$ as given by Definition 2.2.5. Then the following hold:
i) If $0 < \omega_1 < \cdots < \omega_m < \pi$ then the random vector $(I_X(\omega_1), \cdots, I_X(\omega_m))'$ converges in distribution as $N \to \infty$ to a vector of independent and exponentially distributed random variable, each with mean σ^2.

ii) If $EX_1^4 = \eta\sigma^4 < \infty$ and $\omega_j = 2\pi j/n \in [0, \pi]$, then

$$Var(I_X(\omega_j)) = \begin{cases} \frac{1}{4\pi^2}(N^{-1}(\eta - 3)\sigma^4 + 2\sigma^4), & \text{if } \omega_j = 0 \text{ or } \pi \\ \frac{1}{4\pi^2}(N^{-1}(\eta - 3)\sigma^4 + \sigma^4), & \text{if } 0 < \omega_j < \pi \end{cases}$$

and

$$Cov(I_X(\omega_j), I_X(\omega_k)) = \frac{1}{4\pi^2}(N^{-1}(\eta - 3)\sigma^4) \;\; if \;\; \omega_j \neq \omega_k$$

In particular, if X_t is a Gaussian time series, then $\eta - 3 = 0$ so that $I_X(\omega_j)$ and $I_X(\omega_k)$ are uncorrelated for $j \neq k$.

Theorem 2.2.1. (Priestley (1981, Ch. 6.1.3)). If $\{X_t\}$ is a Gaussian random process with zero mean and variance σ^2 then the $I_X(\omega_k)$, $k = 1, \ldots, [(N-1)/2]$, are independently distributed, and for each k,

$$I_X(\omega_k) = \begin{cases} \frac{1}{4\pi}\sigma^2\chi_2^2, & k \neq 0, \ \frac{N}{2} \ \ (N \text{ even}) \\ \frac{1}{2\pi}\sigma^2\chi_1^2, & k = 0, \ \frac{N}{2} \end{cases}$$

In what follows, we will focus on processes which possess absolutely continuous spectrum. In particular, we will be concerned with the general linear process of the form

$$X_t = \sum_{u=-\infty}^{\infty} g_u \epsilon_{t-u} \tag{2.1}$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with $E[\epsilon_t] = 0$ and variance $Var[\epsilon_t] = \sigma_\epsilon^2 < \infty$. In addition, assume that

$$\sum_{u=-\infty}^{\infty} |g_u| < \infty$$

Then according to Priestley (1981)[Ch. 10.1.1], the autocovariance function $\gamma_x(.)$ of X_t satisfies

$$\sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty.$$

The last condition implies the existence of the spectral density function of X_t, see Definition 2.2.3. A useful fact that holds for linear process of the general form (2.1) and will be utilized in the sequel is that the periodogram ordinates $I_X(\omega_j)$ and $I_X(\omega_k)$ are asymptotically independent for all $j \neq k$ such that

$$I_X(\omega) \sim \frac{\lambda_x(\omega)\mathcal{X}_2^2}{2}, \quad 0 < \omega < \pi,$$

where $\mathcal{X}_d^2$ is the chi–square distribution with d degrees of freedom, see Brockwell and Davis (1991) and Proposition 2.2.2. Notice that for the above distributional results and when N is even, the value $\omega = \pi$ is excluded since in this case the distribution is proportional to the $\mathcal{X}_1^2$. The following are true.

Theorem 2.2.2. (Brockwell and Davis (1991, Ch. 10.3)). Let $\{X_t\}$ be a general linear process of the form 2.1 in which the $\{\epsilon_t\}$ are independent with $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma^2$, $E(\epsilon_t^4) < \infty$, and $\sum_{u=-\infty}^{\infty} |g_u||u|^{1/2} < \infty$. Then,

$$I_{N,X}(\omega_k) = 2\pi\lambda(\omega_k)\frac{1}{\sigma^2}I_{N,\epsilon}(\omega_k) + R_N(\omega_k),$$

where,

$$\max_{\omega_k \in [0,\pi]} E\left|R_N(\omega_k)\right|^2 \to O\left(\frac{1}{N}\right),$$

Theorem 2.2.2, is of special importance since it allows us to obtain asymptotic expression for the sampling properties of $I_{N,X}(\omega)$ directly from the known results on $I_{N,\epsilon}(\omega)$. Thus if $\{\epsilon_t\}$ are normal and the $\{g_u\}$ satisfy the required conditions we may deduce from Theorem 2.2.1 that, asymptotically, the set of random variables $\{I_{N,X}/f(\omega_k)\}$, $k = 0, 1, \cdots, [(N-1)/2]$ are independently distributed, and for $k \neq 0,\ N/2$ (N even),we have that

$$I_{N,X}(\omega_k) \sim \begin{cases} \frac{1}{2}\lambda(\omega_k)\chi_2^2, & k \neq 0,\ \frac{N}{2} \quad (N \text{ even}), \\ \lambda(\omega_k)\chi_1^2 & k = 0,\ \frac{N}{2} \end{cases}. \tag{2.2}$$

A well known example of a linear process (2.1) that has been studied extensively in the literature is the zero mean ARMA(p, q) process.

Definition 2.2.6. The process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t = \sum_{r=1}^{p} \phi_r X_{t-r} + \epsilon_t - \sum_{r=1}^{q} \psi_r \epsilon_{t-r} \tag{2.3}$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$. We say that $\{X_t\}$ is an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

2.3 The Exponential Model

The exponential model for the spectrum of a stationary process was introduced by Bloomfield (1973). It is based on the observation that the logarithm of an estimated spectral

density is a fairly well behaved function and therefore can be well approximated by a truncated Fourier series. More specifically, suppose that, the spectral density can be expressed as

$$\begin{aligned} \lambda(\omega) &= \frac{\tau^2}{2\pi}\exp\left(2\sum_{r=1}^{p}\theta_r\cos(r\omega)\right), \qquad 0<\omega<\pi, \\ &= \frac{\tau^2}{2\pi}h(\omega;\theta) \end{aligned} \tag{2.4}$$

where τ^2 and $\theta = (\theta_1, \cdots, \theta_p)$ are unknown parameters. Notice that since the exponent consists of cosine functions, we have that $\lambda(\omega) = \lambda(-\omega)$. Model (2.4) is defined as the exponential model of order p and it will be denoted as EXP(p).

An equivalent definition is given by Holan (2004) and is based on the following argument. Suppose that $\lambda(\omega)$ is the true spectral density function and assume further that is defined on the unit interval $0 \leq \omega < 1$. Also $\lambda(\omega) = \lambda(1-\omega)$. To approximate the spectral density $\lambda(\omega)$ consider the exponential spectral models of order p which can be defined by:

$$\log\lambda_{\theta,p}(\omega) = \theta_0 + 2\sum_{k=1}^{p}\theta_k\cos(2\pi k\omega) \tag{2.5}$$

where,

$$\theta_0 = \int_0^1 \log\lambda_{\theta,p}(\omega)d\omega.$$

When $\log\lambda(\omega)$ is absolutely integrable on $(0,1)$, the Fourier coefficients of $\log\lambda(\omega)$ are defined by:

$$\theta_k = \int_0^1 \log\lambda(\omega)\cos(2\pi k\omega)d\omega, \quad k = 0, 1, \ldots,$$

and are referred as cepstral correlation coefficients. Because the system $C = \{1, \cos(2\pi\omega), \cos(2\pi 2\omega), \cdots\}$ of cosine functions is complete for $C[0,1]$, the class of continuous functions on $[0,1]$, we have that $\log\lambda_{\theta,p}(\omega)$ converges to $\log\lambda(\omega)$ in mean square as $p \to \infty$, see Hart (1997). In addition, the system C forms an orthogonal basis for $C[0,1]$, which implies that any continuous function on the interval $[0,1]$, can be well approximated by a finite linear combination of the elements of C. Often in practice, when fitting the EXP model to short memory spectra the θ_k decay to 0 quickly implying that there exist a small value of p such that

$$\log\lambda_{\theta,p}(\omega) \approx \log\lambda(\omega), \quad \forall\ \omega \in [0,1].$$

An alternative modeling approach can be based on the use of Legendre polynomials and Hermite polynomials functions of the standard normal quantile function–see Parzen (1993) who suggests (2.4) for testing goodness of fit of a spectral density model. Walker (1964) discusses estimation of the general model

$$\lambda(\omega, \theta) = \frac{\tau^2}{2\pi} h(\omega; \theta).$$

One way to estimate θ is by using a Newton-Raphson minimization procedure, for more details see Bloomfield (1973). However, we will resort to the theory of generalized linear models.

2.4 GLM Inference for the Exponential Model

Estimation of the exponential model EXP(p) is based on the methodology of generalized linear models as it was described by McCullagh and Nelder (1989) and Cameron and Turner (1987).

In a generalized model we observe a random variable y with mean μ and distribution F, see McCullagh and Nelder (1989). The mean μ depends on explanatory variables $u_1, \cdots, u_k$ through a link function v such that

$$v(\mu) = \beta_0 + \beta_1 u_1 + \cdots + \beta_k u_k.$$

We assume that $I_y(\omega_j)$ can be approximated by $\lambda(\omega_j; \theta)\xi_j$, where ω_j have been defined in Definition 2.2.4 and from Thm. 2.2.2, ξ_j are independent exponential random variables with mean 1. Therefore, set $y_j = I_y(\omega_j)$ to obtain that $E(y_j) = \mu = \lambda(\omega_j, \theta)$. Suppose now that there exists a link function v such that $v(\mu)$ is linear in the parameters $\theta_1, \cdots, \theta_M$. Thus, assume that

$$v(\mu) = \theta_1 u_1(\omega) + \theta_2 u_2(\omega) + \cdots + \theta_M u_M(\omega)$$

for suitably chosen functions $u_1, u_2, \cdots, u_M$. We are then in the situation of generalized linear models with y equal to the periodogram, exponential distribution F, explanatory variables $u_1(\omega), u_2(\omega), \cdots, u_M(\omega)$, and link function v. A natural choice of the link function is given by

$$v(\mu) = \log(\mu)$$

Hence, inference regarding the exponential model of order p is carried out by numerous statistical languages that include GLM fitting. A more general definition of the EXP(p) has been given by Beran (1994).

Definition 2.4.1. (Fractional EXP process) Let $g : [-\pi, \pi] \to R_+$ be a positive function such that

$$\lim_{\omega \to 0} \frac{g(\omega)}{\omega} = 1$$

and $g(\omega) = g(-\omega)$. Define $\lambda_0 \equiv 1$, and let $\lambda_1, \lambda_2, \cdots, \lambda_p$ be functions that are smooth in the whole interval $[-\pi, \pi]$. Also, assume that $\lambda_k(\omega) = \lambda_k(-\omega)$ and for any n, the $n^* \times (p+1)$ matrix H with column vectors $\{f_k(2\pi/n), f_k(2\pi 2/n), f_k(2\pi 3/n), h_f(2\pi n^*/n)\}^T$, $(k = 0, 1, \cdots, p)$ is nonsingular. Furthermore, let $\theta = (\eta_0, H, \eta_1, \cdots, \eta p)$ be a real vector with $1/2 \leq H < 1$. We call $\{X_t\}$ a fractional EXP process (or an FEXP process) with short-memory components $\lambda_1, \cdots, \lambda_p$ and long-memory component g, if its spectral density is given by

$$\lambda(\omega; \theta) = g(\omega)^{1-2H} \exp\left(\sum_{j=0}^{p} \eta_j \lambda_j(\omega)\right)$$

Two classes of FEXP models are especially useful.

1) $g(\omega) = |1 - e^{i\omega}|$, $\lambda_k(\omega) = \cos k\omega$ $(k = 0, 1, \cdots, p)$ and $H = \frac{1}{2}$, we obtain the model class proposed by Bloomfield (1973). The spectral density is a product of factors of the form $\exp(b_j \cos k\omega)$. The logarithm of the spectral density is assumed to be decomposable into a finite number of cosines:

$$\log \lambda(\omega) = \eta_0 + \eta_1 \cos(\omega) + \eta_2 \cos(2\omega) + \cdots + \eta_p \cos p\omega.$$

Therefore, in its generality, Bloomfield's class is comparable to ARMA models: any smooth spectral density can be approximated with arbitrary accuracy, when p is large enough.

2) $g(\omega) = |1 - e^{i\omega}|, h_k(\omega) = \omega^k$ $(k = 0, 1, \cdots, p)$. The logarithm of the spectral density is the sum of the long-memory component $(1 - 2H) \log \omega$ and a polynomial in ω. A data example with $H = 1/2$, where a second-order polynomial $(p = 2)$ makes sense intuitively, is given in Diggle (1990). In addition , Diggle multiplied the spectrum by an AR(1) spectrum of the form $1/(1 - 2\alpha \cos \omega + \alpha^2)$.

2.5 Examples

Here we will show that any ARMA process can be approximated by an exponential model by choosing its order arbitrarily large. Recall Definition 2.3. For these processes, if $\Psi(z) = 1 - \sum_{i=1}^{q} \psi_i z^i$ and $\Phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ have no common zeroes and $\Phi(z)$ has no zeroes on the unit circle, then their spectral density function is given by, (Brockwell and Davis (1991, Ch.4.4))

$$\lambda_x(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{1 - \sum_{r=1}^{q} \psi_r e^{ir\omega}}{1 - \sum_{r=1}^{p} \phi_r e^{ir\omega}} \right|^2 . \tag{2.6}$$

Lemma 2.5.1. Suppose that $\{X_t\}$ is a stationary ARMA(p, q) time series and let $\lambda_x(\omega)$ denote its spectral density. Then

$$\log \lambda_x(\omega) = \log \sigma^2 - \log 2\pi + 2 \sum_{r=1}^{\infty} \left(\sum_{k=1}^{p} b_k^r - \sum_{l=1}^{q} c_l^r \right) \frac{\cos r\omega}{r},$$

where c_l (respectively b_k) are the reciprocals of the roots of the polynomial $\Psi(z)$ (respectively $\Phi(z)$).

Proof: To show the result consider

$$1 - \sum_{l=1}^{q} \psi_l e^{il\omega} = (1 - c_1 z) \cdots (1 - c_q z),$$

where $z = e^{i\omega}$.

Then,

$$\begin{aligned} \log(1 - \sum_{l=1}^{q} \psi_l e^{il\omega}) &= \sum_{l=1}^{q} \log(1 - c_l z) \\ &= -\sum_{l=1}^{q} \sum_{r=1}^{\infty} \frac{c_l^r z^r}{r} \\ &= -\sum_{l=1}^{q} \sum_{r=1}^{\infty} \frac{c_l^r e^{ir\omega}}{r}. \end{aligned}$$

Similarly,

$$\log(1 - \sum_{l=1}^{q} \psi_l e^{-il\omega}) = -\sum_{l=1}^{q} \sum_{r=1}^{\infty} \frac{c_l^r e^{-ir\omega}}{r}.$$

Working along the previous lines we obtain that

$$1 - \sum_{k=1}^{p} \phi_k e^{ik\omega} = (1 - b_1 z) \cdots (1 - b_p z).$$

Then,

$$\begin{aligned}\log(1-\sum_{k=1}^{p}\phi_k e^{ik\omega}) &= \sum_{k=1}^{p}\log(1-b_k z)\\ &= -\sum_{k=1}^{p}\sum_{r=1}^{\infty}\frac{b_k^r z^r}{r}\\ &= -\sum_{k=1}^{p}\sum_{r=1}^{\infty}\frac{b_k^r e^{ir\omega}}{r}\end{aligned}$$

and

$$\log(1-\sum_{k=1}^{p}\phi_k e^{-ik\omega}) = -\sum_{k=1}^{p}\sum_{r=1}^{\infty}\frac{b_k^r e^{-ir\omega}}{r}.$$

Thus,

$$\begin{aligned}\log\lambda_x(\omega) &= \log\sigma^2 - \log 2\pi\\ &+ \log(1-\sum_{l=1}^{q}\psi_l e^{il\omega}) + \log(1-\sum_{l=1}^{q}\psi_l e^{-il\omega})\\ &- \log(1-\sum_{k=1}^{p}\phi_k e^{ik\omega}) - \log(1-\sum_{k=1}^{p}\phi_k e^{-ik\omega})\\ &= \log\sigma^2 - \log 2\pi + 2\sum_{r=1}^{\infty}\left(\sum_{k=1}^{p}b_k^r - \sum_{l=1}^{q}c_l^r\right)\frac{\cos r\omega}{r}.\end{aligned}$$

Some examples illustrate the results of Lemma 2.5.1.

Example 2.5.1. Suppose that $\{X_t\}$ is an AR(1) process, that is

$$X_t = \phi X_{t-1} + \epsilon_t$$

with $|\phi| < 1$ and $\{\epsilon_t\}$ a white noise sequence with variance σ_ϵ^2. Lemma 2.5.1 shows that

$$\log\lambda_x(\omega) = \log\sigma_\epsilon^2 - \log 2\pi + 2\sum_{r=1}^{\infty}\frac{\phi^r}{r}\cos(r\omega)$$

since $b_k = \phi, c_l = 0$. Hence, according to (2.5), the spectral density of an AR(1) process can be approximated by an EXP(p) model with $\theta_r = \phi^r/r$. Table 2.1 below shows the values of the parameters for selected values of ϕ. If we set $v = \omega/2\pi$ in the above representation we will have only changes in scale. Thus

$$\log\lambda_x(2\pi v) = \log\sigma_\epsilon^2 + 2\sum_{r=1}^{\infty}\frac{\phi^r}{r}\cos(2\pi r v),$$

that is $\theta_0 = \log \sigma_\epsilon^2$. If the errors are standard Gaussian random variables, then $\theta_0 = 0$.

Table 2.1: Coefficients of the EXP(7) model (2.5) for AR(1) process for different parameter values

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
$\phi = -0.8$	0	-0.8	0.32	-0.17	0.102	-0.065	0.043	-0.029
$\phi = -0.5$	0	-0.5	0.125	-0.041	0.015	-0.006	0.002	-0.001
$\phi = -0.3$	0	-0.3	0.045	-0.009	0.002	-4E-04	1E-04	-3E-05
$\phi = -0.1$	0	-0.1	0.005	-3E-04	2E-05	-2E-06	1E-07	-1E-08
$\phi = +0.1$	0	0.1	0.005	3E-04	2E-05	2E-06	1E-07	1E-08
$\phi = +0.3$	0	0.3	0.045	0.009	0.002	4E-04	1E-04	3E-05
$\phi = +0.5$	0	0.5	0.125	0.041	0.015	0.006	0.002	0.001
$\phi = +0.8$	0	0.8	0.32	0.17	0.102	0.065	0.043	0.029

Example 2.5.2. Suppose that $\{X_t\}$ is an AR(2) process, that is

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

with $\phi_1 + \phi_2 < 1$, $|\phi_2| < 1$ and $\phi_1 - \phi_2 < 1$ and $\{\epsilon_t\}$ a white noise sequence with variance σ_ϵ^2. Lemma 2.5.1 shows that:

$$\log \lambda_x(\omega) = \log \sigma_\epsilon^2 - \log 2\pi + 2 \sum_{r=1}^{\infty} \frac{(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})^r + (\phi_1 - \sqrt{\phi_1^2 + 4\phi_2})^r}{r 2^r} \cos(r\omega)$$

since

$$\begin{aligned} c_l &= 0 \\ b_1 &= -\frac{2\phi_2}{\phi_1 + \sqrt{\phi_1^2 - 4\phi_2}}, \\ b_2 &= -\frac{2\phi_2}{\phi_1 - \sqrt{\phi_1^2 - 4\phi_2}}. \end{aligned}$$

Hence, according to (2.5), the spectral density of an AR(2) process can be approximated by an EXP(p) model with $\theta_r = \frac{(\phi_1+\sqrt{\phi_1^2+4\phi_2})^r+(\phi_1-\sqrt{\phi_1^2+4\phi_2})^r}{r2^r}$. Table 2.2 below shows the values of the parameters for selected values of ϕ_1 and ϕ_2. Setting $v = \omega/2\pi$, yields the same observation as in Example 2.5.1.

Example 2.5.3. Suppose that $\{X_t\}$ is an ARMA(1, 1) process, that is

$$X_t = \phi X_{t-1} - \psi \epsilon_{t-1} + \epsilon_t$$

Table 2.2: Coefficients of the EXP(7) model (2.5) for AR(2) process with different parameter values

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
$(\phi_1, \phi_2) = (-0.3, 0.5)$	0	-0.3	0.545	-0.159	0.172	-0.088	0.079	-0.052
$(\phi_1, \phi_2) = (-0.8, 0.3)$	0	-0.8	0.62	-0.41	0.339	-0.291	0.261	-0.242
$(\phi_1, \phi_2) = (-0.1, 0.8)$	0	-0.1	0.8	-0.08	0.328	-0.064	0.18	-0.05
$(\phi_1, \phi_2) = (0.3, 0.6)$	0	0.3	0.645	0.189	0.236	0.124	0.125	0.085
$(\phi_1, \phi_2) = (0.4, 0.4)$	0	0.4	0.48	0.181	0.15	0.091	0.07	0.05

Table 2.3: Coefficients of the EXP(7) model (2.5) for ARMA(1,1) process and different parameter values

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
$(\phi, \psi) = (-0.3, 0.5)$	0	-0.8	-0.08	-0.05	-0.013	-0.006	-0.002	-0.001
$(\phi, \psi) = (-0.8, 0.3)$	0	-1.1	0.275	-0.179	0.1	-0.066	0.043	-0.029
$(\phi, \psi) = (-0.1, 0.8)$	0	-0.9	-0.315	-0.171	-0.102	-0.065	-0.043	-0.029
$(\phi, \psi) = (0.3, 0.6)$	0	-0.3	-0.135	-0.063	-0.03	-0.015	-0.007	-0.003
$(\phi, \psi) = (0.4, 0.4)$	0	0	0	0	0	0	0	0

with $|\phi| < 1$ and $|\psi| < 1$ and $\{\epsilon_t\}$ a white noise sequence with variance σ_ϵ^2. Lemma 2.5.1 shows that:

$$\log \lambda_x(\omega) = \log \sigma_\epsilon^2 - \log 2\pi + 2 \sum_{r=1}^{\infty} \frac{(\phi^r - \psi^r)}{r} \cos(r\omega)$$

since $c_l = \psi, b_k = \phi$. Hence, according to (2.5), the spectral density of an ARMA$(1, 1)$ process can be approximated by an EXP(p) model with $\theta_r = (\phi^r - \psi^r)/r$. Table 2.3 below shows the values of the parameters for selected values of ϕ and ψ.

2.6 Generating Data from the Exponential Model

There are several methods to generate data from the exponential model. Davies and Harte (1987) proposed an exact frequency domain method, Percival (1992) describes an exact time domain method. Holan (2004) proposes to make use of Pourhamadi's formula, see Pourahmadi (1983). The main essence is to use the MA(∞) representation for the EXP(p) process. Let

$$\log \lambda_{\theta,p}(\omega) = \theta_0 + 2 \sum_{k=1}^{p} \theta_k \cos(k2\pi\omega)$$

be the logarithmic EXP(p) representation associated with the process $\{Y_t\}$ and let

$$Y_t = \psi_0 Y^v(t) + \psi_1 Y^v(t-1) + \psi_2 Y^v(t-2) + \cdots$$

be its associated MA(∞) representation ($\psi_0 \equiv 1$). Here $Y^v(t)$, for all t, is the innovation process which is Gaussian. Pourahmadi's formula enables direct computation of the coefficients ψ_h from θ_k by employing the following recursions:

$$\begin{aligned} \psi_0 &= 1 \\ \psi_h &= \frac{1}{h}\sum_{k=1}^{h} k\theta_k \psi_{h-k}, \quad h = 1, 2, \cdots \end{aligned}$$

The following algorithm details how to simulate samples $\{Y_t\}$ from the EXP(p) model:

1. Given the EXP(p) model calculate $\psi_0, \psi_1, \cdots, \psi_q$ for q sufficiently large, for instance $q \geq 1000$), using Pourahmadi's formula.

2. For $k = 0, 1, \cdots, q$ compute

$$SS_{sim}(k) = \frac{SS(k)}{SS(q)} = \frac{\sum_{h=0}^{k} \psi_h^2}{\sum_{h=0}^{q} \psi_h^2}.$$

3. Find the largest value of k such that $SS_sim(k) < 1 - \epsilon$ for example $SS_{sim}(k) = 0.9999999$).

4. Form the process

$$Y(t) = \psi_0 Y^v(t) + \psi_1 Y^v(t-1) + \cdots + \psi_k Y^v(t-k).$$

   Then

$$\begin{aligned} \log \lambda_{\theta,p}(\omega) &= \theta_0 + 2\sum_{k=1}^{p} \theta_k \cos(k2\pi\omega) \\ &\approx \log\left(\sigma^2 \left|\sum_{h=0}^{k} \psi_h \exp(2\pi i h\omega)\right|^2\right), \end{aligned}$$

   where $\sigma^2 = \exp(\theta_0)$ *and* $\psi_0 \equiv 1$.

5. Simulate the process $\{X_t\}$ from its k^{th} order truncated MA(∞) representation using $N(0, \sigma^2 = \exp(\theta_0))$ innovations.

Figure 2.1 shows a plot of two realizations from the EXP(2) model (2.4), using 150 observations. In particular for plot (a) we choose parameter $\theta = (-0.5, -0.90, 0.40)$ and for plot (b) $\theta = (0.5, 0.30, 0.15)$.

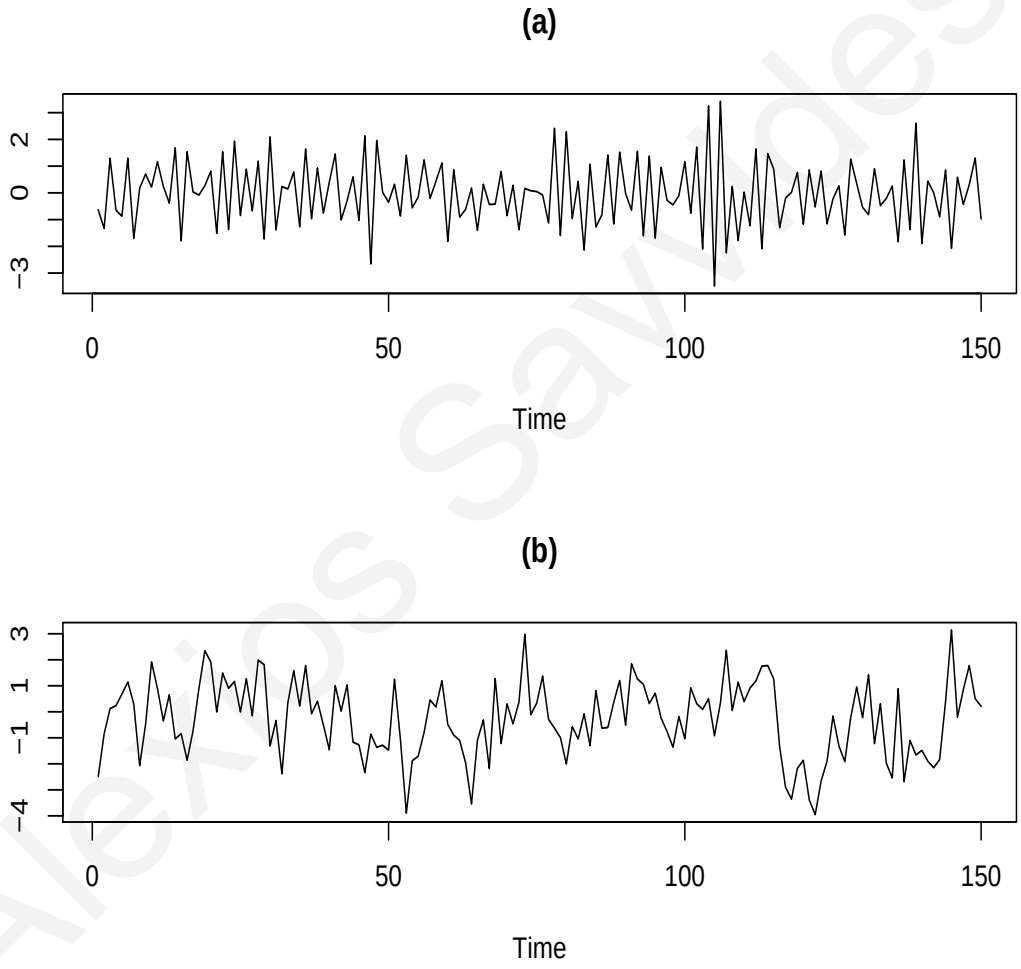

Figure 2.1: 150 observations from EXP(2) model (2.4) with (a) $\theta = (-0.5, -0.90, 0.40)$ and (b) $\theta = (0.5, 0.30, 0.15)$.

Chapter 3

Inference for two or more independent time series

3.1 Introduction

In this section we will develop methodology of comparing several independent time series. In addition we state some results that show how to proceed with likelihood inference and we develop large sample theory about testing the equality of spectral densities. The proposed method will be illustrated empirically by means of a simulated study which includes several examples. An analysis of photometric data that was presented in the Introduction integrates the presentation.

3.2 The case of two time series

The following lemma is of special importance in the sequel (see Johnson and Kotz (1970)).

Lemma 3.2.1. Suppose that $X \sim \theta_1 \mathcal{X}_2^2/2$, and $Y \sim \theta_2 \mathcal{X}_2^2/2$ and X, Y are independent. Then

$$W := \log \frac{X}{Y} \sim \text{Logistic}\left(\log \frac{\theta_1}{\theta_2}, 1\right),$$

where Logistic$(\alpha, 1)$ denotes the logistic random variable with distribution function

$$F(x) = \{1 + e^{-(x-\alpha)}\}^{-1}, \qquad -\infty < x < \infty.$$

Notice that for $X \sim$ Logistic$(\alpha, 1)$, we have that $E(X) = \alpha$ and $Var(X) = \pi^2/3$. Consider two independent stationary time series $\{X_t, t = 1, \ldots, n\}$ and $\{Y_t, t = 1, \ldots, n\}$

generated by a linear process (2.1) with spectral densities λ_x and λ_y respectively given by Definition 2.2.3. Now let $I_x(\omega)$ and $I_y(\omega)$ be the periodograms of $\{X_t\}$ and $\{Y_t\}$ respectively, given by Definition 2.2.4. Then, according to Theorem 2.2.2

$$I_x(\omega) \text{ converges assymtotically to } \lambda_x(\omega)\chi^2_2/2 \qquad (0 < \omega < \pi),$$
$$I_y(\omega) \text{ converges assymtotically to } \lambda_y(\omega)\chi^2_2/2 \qquad (0 < \omega < \pi).$$

Writing $J(\omega) = I_x(\omega)/I_y(\omega)$ and $\mu(\omega) = \lambda_x(\omega)/\lambda_y(\omega)$, it follows from Lemma 3.2.1 that

$$\log J(\omega) \text{ converges assymtotically to Logistic}(\log \mu(\omega), 1), \tag{3.1}$$

asymptotically. Hence, $\log J(\omega)$ is an asymptotically unbiased but inconsistent estimator for the log spectral ratio $\log \mu(\omega)$ and that its asymptotic variance is independent of ω.

On the basis of the above observation, Coates and Diggle (1986) and (Diggle, 1990, Ch.4.8) discuss tests of the following hypotheses

$$H_1 : \lambda_x(\omega) = \lambda_y(\omega), \qquad 0 < \omega < \pi,$$
$$H_2 : \lambda_x(\omega) = \kappa\lambda_y(\omega), \qquad 0 < \omega < \pi.$$

A generalized likelihood ratio test is constructed within the framework of the logistic distribution (3.1) by adapting the quadratic model

$$\log \mu(\omega) = \alpha + \beta\omega + \gamma\omega^2. \tag{3.2}$$

The quadratic model is a improvement over the linear model obtained by setting $\gamma = 0$. A further advantage of using the parametric approach is that H_1 and H_2 are nested hypotheses, corresponding to $\alpha = \beta = \gamma = 0$ and $\beta = \gamma = 0$ respectively. This means that standard likelihood inference can be carried out within this framework. Indeed, let $T_i := \log J(\omega_i), \quad i = 1, \cdots, m$. Then (3.1) and (3.2) together give the log-likelihood function

$$\text{log likelihood} = \sum_{i=1}^{m} \left(-t_i + \alpha + i\beta + i^2\gamma\right) - 2\sum_{i=1}^{m} \log\left\{1 + \exp(-t_i + \alpha + i\beta + i^2\gamma)\right\},$$

on which inference can be based. Statistical tests for H_1 and H_2 are computed by the standard asymptotic chi–square distribution of the likelihood ratio test.

3.3 The case of multiple time series

We extend the above results to the case of multiple time series. First, we generalize Lemma 3.2.1.

Lemma 3.3.1. Suppose that X_j, $j = 1, 2, \ldots, G$ are independent random variables which are distributed as $\lambda_j \mathcal{X}_2^2/2$, $\lambda_j > 0$. Then, the random variables

$$T_j = \log \frac{X_j}{X_G}, \quad j = 1, 2, \ldots, G-1$$

are distributed according to the following multivariate density function

$$f_{T_1,\ldots,T_{G-1}}(t_1, \ldots, t_{G-1}) = \frac{(G-1)! \exp\left(\sum_{i=1}^{G-1}(t_i - \log \mu_i)\right)}{\left(1 + \sum_{i=1}^{G-1} \exp(t_i - \log \mu_i)\right)^G}, \quad t_1, \ldots, t_{G-1} \in R,$$

where $\mu_i = \lambda_i/\lambda_G$, $i = 1, 2, \ldots, G-1$.

Proof: The joint distribution of $(X_1, \ldots, X_G)'$ is given by

$$f(x_1, x_2, \cdots, x_G) = \frac{1}{\lambda_1 \lambda_2 \cdots \lambda_G} \exp\left[-\left(\frac{x_1}{\lambda_1} + \frac{x_2}{\lambda_2} + \cdots + \frac{x_G}{\lambda_G}\right)\right], \; x_1, \ldots, x_G > 0.$$

Now let,

$$T_i = \log \frac{X_i}{X_G} \qquad \text{and} \qquad \mu_i = \frac{\lambda_i}{\lambda_G} \qquad \text{for} \;\; i = 1, \ldots, G-1.$$

Then, by introducing the variable $W = X_G$, the inverse transformation is given by

$$X_i = We^{T_i} \qquad \text{and} \qquad X_G = W \qquad \text{for} \;\; i = 1, \ldots, G-1.$$

The Jacobean of the transformation equals to

$$\begin{vmatrix} We^{T_1} & 0 & ... & 0 & e^{T_1} \\ 0 & we^{T_2} & ... & 0 & e^{T_2} \\ . & 0 & & 0 & . \\ . & . & ... & . & . \\ 0 & 0 & ... & We^{T_{G-1}} & e^{T_{G-1}} \\ 0 & 0 & ... & 0 & 1 \end{vmatrix} = W^{G-1} e^{T_1 + T_2 + \cdots + T_{G-1}}.$$

It follows that the joint distribution of $(T_1, T_2, \ldots, T_{G-1}, W)'$ is computed as

$$\begin{aligned} f(t_1, t_2, \cdots, t_{G-1}, w) &= \frac{1}{\lambda_1 \lambda_2 \cdots \lambda_G} e^{-\left(\frac{we^{t_1}}{\lambda_1} + \cdots + \frac{we^{t_{G-1}}}{\lambda_{G-1}} + \frac{w}{\lambda_G}\right)} \\ &\times \; w^{G-1} e^{t_1 + t_2 + \cdots + t_{G-1}}, \quad w > 0, t_1, \ldots, t_{G-1} \in R. \end{aligned}$$

Therefore, the marginal of $(T_1, T_2, \ldots, T_{G-1})'$ is computed by integrating out W :

$$\begin{aligned} f(t_1, t_2, \cdots, t_{G-1}) &= \int_0^\infty \frac{1}{\lambda_1\lambda_2\cdots\lambda_G} e^{-w\left(\frac{e^{t_1}}{\lambda_1}+\cdots+\frac{e^{t_{G-1}}}{\lambda_{G-1}}+\frac{1}{\lambda_G}\right)} w^{G-1} e^{t_1+t_2+\cdots+t_{G-1}} dw \\ &= \frac{e^{t_1+t_2+\cdots+t_{G-1}}}{\lambda_1\lambda_2\cdots\lambda_G} \int_0^\infty e^{-w\left(\frac{e^{t_1}}{\lambda_1}+\cdots+\frac{e^{t_{G-1}}}{\lambda_{G-1}}+\frac{1}{\lambda_G}\right)} w^{G-1} dw \\ &= \frac{\Gamma(G) e^{t_1+t_2+\cdots+t_{G-1}}}{\lambda_1\lambda_2\cdots\lambda_G \left(\frac{e^{t_1}}{\lambda_1}+\cdots+\frac{e^{t_{G-1}}}{\lambda_{G-1}}+\frac{1}{\lambda_G}\right)^G} \\ &= \frac{(G-1)! e^{t_1+t_2+\cdots+t_{G-1}}}{\lambda_1\lambda_2\cdots\lambda_{G-1}\frac{1}{\lambda_G^{G-1}} \left(\frac{e^{t_1}}{\frac{\lambda_1}{\lambda_G}}+\cdots+\frac{e^{t_{G-1}}}{\frac{\lambda_{G-1}}{\lambda_G}}+1\right)^G} \\ &= \frac{(G-1)! e^{[(t_1-\log\mu_1)+(t_2-\log\mu_2)+\cdots+(t_{G-1}-\log\mu_{G-1})]}}{\left(e^{t_1-\log\mu_1}+\cdots+e^{t_{G-1}-\log\mu_{G-1}}+1\right)^G}, \quad t_1\ldots, t_{G-1} \in R. \end{aligned}$$

The above Lemma in connection with the asymptotic distribution of periodogram shows that a reasonable way to model the spectral densities of several independent stationary stochastic processes can be based on their logarithmic ratio. Suppose that $\{Y_{jt}, t = 1, 2, \ldots, N\}$ are independent stationary time series and let $\lambda_j(\omega)$ be their corresponding spectral density function, for $j = 1, 2, \ldots, G$. Suppose that the following model holds,

$$\log\mu_j(\omega) \equiv \log\frac{\lambda_j(\omega)}{\lambda_G(\omega)} = \theta_j^T Z(\omega), -\pi < \omega < \pi, \tag{3.3}$$

for $j = 1, 2, \ldots, G-1$ where, $\theta_j = (a_{j0}, a_{j1}, \cdots, a_{jp})^T$ is an $(p+1)$–dimensional vector of unknown parameters to be estimated by the data and
$Z(\omega) = (1, 2\cos\omega, 2\cos 2\omega, \cdots, 2\cos p\omega)^T$. The form of the vector $Z(\omega)$ is motivated by the fact that we will be working with real valued spectral densities and by the representation of the exponential model. Furthermore, the order p is chosen in advance but in Section 3.7.2 we will see that real data can give us a guidance about its value by employing the so called AIC model selection criterion, see Akaike (1974). We argue that estimation and inference can be carried out within the framework of model (3.3).

It is worth considering some concrete examples. For instance, consider the hypotheses $\theta_j = 0$ for all $j = 1, 2, \ldots, G-1$ which imply that all the spectral densities are equal to the spectral density $\lambda_G(\omega)$, that is all the processes $\{Y_{jt}, t = 1, 2, \ldots, N\}$, $j = 1, 2 \ldots, G-1$ share the same second order structure to that of the process $\{Y_{Gt}, t = 1, 2, \ldots, N\}$. Another hypotheses of interest would be $a_{j1} = \ldots = a_{jp} = 0$ for $j = 1, 2, \ldots, G-1$ which

imply that the functions $\lambda_j(\omega)$ are proportional to $\lambda_G(\omega)$. Several other examples can be casted within this framework which is based on the system of cosine functions–a natural basis for time series analysis data. An alternative modeling approach can be based on the use of polynomials–see Coates and Diggle (1986)– or employment of Legendre and Hermite polynomial functions of the standard normal quantile function–see Parzen (1993). Using Lemma 3.3.1, we obtain the following result.

Lemma 3.3.2. Suppose that $\{Y_{jt}, t = 1, \cdots, N\}$, $j = 1, \cdots, G$, are independent stationary time series with absolutely continuous spectral densities, $\lambda_j(\omega)$, which are defined by Definition 2.2.3, where $j = 1, \cdots, G$ and G denotes the number of different time series. Suppose further that the condition of Theorem 2.2.2 are fulfilled. Let $X_{ji} \equiv I_j(\omega_i)$ be the value of the periodograms of each time series at the Fourier frequencies, $\omega_i = 2\pi i/N$, $i = 1, 2, \ldots, m$ which are defined by Definition 2.2.4. Now let,

$$\begin{aligned} T_{ji} &= \log \frac{X_{ji}}{X_{Gi}} \\ \log \mu_{ji} &\equiv \log \frac{\lambda_j(\omega_i)}{\lambda_G(\omega_i)} = \theta_j^T Z(\omega_i) \end{aligned} \tag{3.4}$$

for $j = 1, 2, \ldots, G-1$, $i = 1, 2, \ldots, m$ where the notation follows equation (3.3). For a fixed i, Lemma 3.3.1 shows that the joint distribution of T_{ji}, $j = 1, 2, \ldots, G-1$ is given by

$$f(t_{1i}, \ldots, t_{(G-1)i}) = \frac{(G-1)! \exp\left(\sum_{j=1}^{G-1}(t_{ji} - \log \mu_{ji})\right)}{\left(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \log \mu_{ji})\right)^G}$$

3.3.1 Estimation

The above result shows how to proceed with likelihood inference in the case of multiple time series. Specifically, recall the notation of Lemma 3.3.2 and define the following vectors. Notice that these are the observed data

$$\mathbf{T}_i = \left(T_{1i}, T_{2i}, \ldots, T_{(G-1)i}\right)^T,$$

for $i = 1, 2, \ldots, m$. Then, the likelihood function of $\mathbf{T}_1, \ldots, \mathbf{T}_m$ is

$$L(\theta) = \prod_{i=1}^{m} f(\mathbf{t}_i)$$

$$
\begin{aligned}
&= \prod_{i=1}^{m} f(t_{1i}, \ldots, t_{(G-1)i}) \\
&= \prod_{i=1}^{m} \frac{(G-1)! \exp\left(\sum_{j=1}^{G-1}(t_{ji} - \theta_j^T Z_i)\right)}{\left(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \theta_j^T Z_i)\right)^G},
\end{aligned}
$$

where $\theta = (\theta_1^T, \ldots, \theta_{G-1}^T)^T$ is a $(G-1) \times (p+1)$– dimensional vector and $Z_i = Z(\omega_i)$–see (3.3). Hence, the log–likelihood function is given up to a constant by

$$
l(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{G-1} (t_{ji} - \theta_j^T Z_i) - G \sum_{i=1}^{m} \log\left(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \theta_j^T Z_i)\right). \tag{3.5}
$$

The value of θ that maximizes $l(\theta)$, say $\hat{\theta}$ is the maximum likelihood estimator of θ.

With this notation, the score function is given by

$$
S(\theta) \equiv \frac{\partial l(\theta)}{\partial \theta} = \left(S_1^T(\theta), \ldots, S_{G-1}^T(\theta)\right)^T \tag{3.6}
$$

where

$$
S_j(\theta) = \frac{\partial l(\theta)}{\partial \theta_j} = \left(\frac{\partial l(\theta)}{\partial a_{j0}}, \ldots, \frac{\partial l(\theta)}{\partial a_{jp}}\right)^T
$$

for $j = 1, 2, \ldots, G-1$. Differentiation shows that the l'th element of the $(p+1)$–dimensional vector $S_j(\theta)$ is given by

$$
\begin{aligned}
\frac{\partial l}{\partial a_{jl}} &= -M_{(l)} \sum_{i=1}^{m} \cos l\omega_i + G M_{(l)} \sum_{i=1}^{m} \frac{e^{t_{ji} - \theta_j^T Z_i} \cos l\omega_i}{\left(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \theta_j^T Z_i)\right)} \\
&= -M_{(l)} \sum_{i=1}^{m} \cos l\omega_i + G M_{(l)} \sum_{i=1}^{m} \frac{e^{t_{ji} - \theta_j^T Z_i} \cos l\omega_i}{A_i}.
\end{aligned}
$$

where

$$
M_{(l)} = \begin{cases} 1, & l = 0 \\ 2, & \text{otherwise.} \end{cases}
$$

and

$$
A_i = \left(1 + \sum_{j=1}^{G-1} \exp(t_{ji} - \theta_j^T Z_i\right),
$$

for $j = 1, 2, \ldots, G-1$ and $l = 0, 1, 2, \ldots, p$. Furthermore,

$$
\begin{aligned}
\frac{\partial^2 l}{\partial a_{jl} \partial a_{jr}} &= -G M_{(l,r)} \sum_{i=1}^{m} \frac{A_i e^{t_{ji} - \theta_j^T Z_i} \cos l\omega_i \cos r\omega_i - \left(e^{t_{ji} - \theta_j^T Z_i}\right)^2 \cos l\omega_i \cos r\omega_i}{A_i^2} \\
&= -G M_{(l,r)} \sum_{i=1}^{m} \cos l\omega_i \cos r\omega_i \frac{A_i e^{t_{ji} - \theta_j^T Z_i} - \left(e^{t_{ji} - \theta_j^T Z_i}\right)^2}{A_i^2}
\end{aligned}
$$

and

$$\frac{\partial^2 l}{\partial a_{jl}\partial a_{kr}} = GM_{(l,r)}\sum_{i=1}^{m}\cos l\omega_i \cos r\omega_i \frac{\left(e^{t_{ji}-\theta_j^T Z_i}\right)\left(e^{t_{ki}-\theta_k^T Z_i}\right)}{A_i^2}$$

for $j,k = 1,2,\ldots,G-1$ with $j \neq k$ and $l,r = 0,1,\ldots,p$, where,

$$M_{(l,r)} = M_{(l)}M_{(r)} = \begin{cases} 1, & l = r = 0 \\ 2, & l = 0 \text{ and } r \neq 0, \text{or } r = 0 \text{ and } l \neq 0 \\ 4, & l \neq 0 \text{ and } r \neq 0 \end{cases}$$

In addition

$$\frac{\partial^3 l}{\partial a_{jl}\partial a_{jr}\partial a_{ju}} = GM_{(l,r,u)}\sum_{i=1}^{m} K_i \frac{A_i\left(e^{t_{ji}-\theta_j^T Z_i}\right)^2 + A_i^2\left(e^{t_{ji}-\theta_j^T Z_i}\right) - 4A_i\left(e^{t_{ji}-\theta_j^T Z_i}\right)^2 + 2\left(e^{t_{ji}-\theta_j^T Z_i}\right)^3}{A_i^3}$$

where,

$$M_{(l,r,u)} = M_{(l)}M_{(r)}M_{(u)} = \begin{cases} 1, & l = r = u = 0 \\ 2, & l = 0 \text{ and } r,u \neq 0, \text{or } r = 0 \text{ and } l,u \neq 0, \text{ or } u = 0 \text{ and } l,r \neq 0 \\ 4, & l = r = 0 \text{ and } u \neq 0 \text{ or } l = u = 0 \text{ and } r \neq 0 \text{ or } u = r = 0 \text{ and } l \neq 0 \\ 8, & l,r,u \neq 0 \end{cases}$$

and

$$K_i = \cos l\omega_i \cos r\omega_i \cos u\omega_i \qquad l,r,u = 0,1,\ldots,p.$$

Thus,

$$\frac{\partial^3 l}{\partial a_{jl}\partial a_{kr}\partial a_{ju}} = GM_{(l,r,u)}\sum_{i=1}^{m} K_i \frac{2\left(e^{t_{ji}-\theta_j^T Z_i}\right)^2\left(e^{t_{ki}-\theta_k^T Z_i}\right) - A_i\left(e^{t_{ji}-\theta_j^T Z_i}\right)\left(e^{t_{ki}-\theta_k^T Z_i}\right)}{A_i^3}, \quad (3.7)$$

for $j,k = 1,2,\ldots,G-1$ with $j \neq k$ and $l,r,u = 0,1,\ldots,p$. Note that these two expressions above, of third derivatives, are sufficient because of the symmetry. It will be shown that the above expressions are uniformly bounded, see assumption (K.5) of Section 3.4.

3.4 Large Sample Theory

The asymptotic properties of the maximum likelihood estimator $\hat{\theta}$ are studied by appealing to the asymptotic theory for the statistical analysis regarding functionals of spectra, see Taniguchi and Kakizawa (2000, Ch. 6.2). More specifically, suppose that $\mathcal{F}$ is the set of all $G \times G$ diagonal matrix valued functions $\mathbf{W}(\omega)$, $\omega \in [\pi, \pi]$ with $\mathbf{W}(.)$ hermitian and symmetric. Suppose that $\mathcal{L}$ denotes the set of all diagonal spectral density matrices, that is $\lambda(\omega) = \text{diag}(\lambda_1(\omega), \ldots, \lambda_G(\omega))$, a $G \times G$ diagonal matrix, whose elements satisfy the property that there exist constants $c_{i1}, c_{i2} > 0$ for $i = 1, 2, \ldots, G$ such that $c_{i1} \leq \lambda_i(\omega) \leq c_{i2}$ for $i = 1, 2 \ldots, G$.
In addition we define $\mathcal{N} = \left\{\lambda(\omega) \text{ such that (3.3) holds, for some } \theta \in \Theta \subset R^{(G-1)(p+1)}\right\}$.

Motivated by the preceding analysis and in particular by the log–likelihood equations (3.5), consider the following function

$$D(\theta, \lambda) = \frac{1}{4\pi} \int_{-\pi}^{\pi} K(\theta, \lambda(\omega), \omega) d\omega, \tag{3.8}$$

where

$$\begin{aligned} K(\theta, \mathbf{W}(\omega), \omega) &= -\sum_{j=1}^{G-1} \left(\log \frac{W_j(\omega)}{W_G(\omega)} - \theta_j^T Z(\omega) \right) \\ &+ G \log \left(1 + \sum_{j=1}^{G-1} \frac{W_j(\omega)}{W_G(\omega)} \exp(-\theta_j^T Z(\omega)) \right), \end{aligned} \tag{3.9}$$

with $Z(\omega) = (1, 2\cos(\omega), \ldots, 2\cos(p\omega))^T$. In addition, denote by

$$\mathbf{I}_N(\omega) = \text{diag}(I_1(\omega), \ldots, I_G(\omega)), \tag{3.10}$$

the $G \times G$ diagonal matrix whose entry at position j is the corresponding periodogram of the j'th series. Then, observe that the log–likelihood function (3.5) multiplied by $-1/m$, is the discrete approximation of (3.8) at the Fourier frequencies evaluated at $\mathbf{I}_N(\omega)$. However it is well known that inference based on periodogram might not lead to consistent estimates especially for non linear contrast functions such as (3.8). Therefore, consider $\hat{\lambda}$, a nonparametric kernel spectral density estimator of the form

$$\hat{\lambda}(\omega) = \int_{-\pi}^{\pi} W_n(\omega - \omega_1) \mathbf{I}_N(\omega_1) d\omega_1. \tag{3.11}$$

The asymptotic properties of the maximum likelihood estimator of θ will be studied by means of (3.8) evaluated at $\hat{\lambda}$. However simulations results show that the use of the raw periodogram estimator yield similar conclusions–see Section 3.7.

Define a functional T by the requirement that for every $\lambda \in \mathcal{L}$

$$D(T(\lambda), \lambda) = \min_{\theta \in \Theta} D(\theta, \lambda) \tag{3.12}$$

To study the asymptotic properties of the maximum likelihood estimator, consider the following set of assumptions:

Assumption A

A.1 The function $K(\theta, \mathbf{W}, \omega)$ is defined on $\Theta \times D \times [-\pi, \pi]$ where Θ is a compact subset of $R^{(G-1)(p+1)}$ and D is an open set of C^{G^2}–the set of all square matrices of dimension G with complex elements–which contains the whole range of $\mathcal{L}$.

A.2 There exists a positive constant r which does not depend upon θ and ω such that the ball
$C_\omega = \left\{ \mathbf{W} = (W_j)_{j=1}^G : |W_j - \lambda_j(\omega)| \leq r \right\}$ is contained in D, for all $\omega \in [-\pi, \pi]$.

A.3 The matrix D_λ defined by (3.13) below is nonsingular.

A.4 $\sum_h h^2 |\gamma_j(h)| < \infty$, where $\gamma_j(h)$ the autocovariance function of the process Y_{tj}, $j = 1, \ldots, G$.

A.5 The spectral window $W_n(\theta)$ is of the form

$$W_n(\theta) = M \sum_{\nu=-\infty}^{\infty} W(M(\theta + 2\pi\nu)),$$

where W and M satisfy the following:

1. $W(\theta)$ is real, bounded nonnegative even probability density function with finite second moment.
2. The function $w(x) = \int_{-\infty}^{\infty} W(\theta) \exp(i\theta x) d\theta$ is bounded by an even, integrable and monotonically decreasing function on $[0, \infty)$.
3. The bandwidth M depends on N in such a way that $M/N^{1/2} + N^{1/4}/M \to 0$, as $N \to \infty$.

Assumption B

B.1 If $\theta \neq \theta'$ then $\lambda(\omega) \neq \lambda'(\omega)$ where $\lambda(\omega)$ (respectively $\lambda'(\omega)$) refers to model (3.3) under θ (respectively under θ').

B.2 The parameter space Θ is a compact subset of $R^{(G-1)(p+1)}$.

B.3 The spectral density matrix $\lambda(\omega)$ belongs to $\mathcal{L} \cap \mathcal{N}$.

B.4 Every component of the spectral density matrix $\lambda(\omega)$ is three times continuously differentiable with respect to θ and these derivatives are continuous in $[-\pi, \pi]$.

B.5 The true spectral density matrix belong to $\mathcal{N}$ and θ belongs to the interior of Θ.

We first prove the existence of a minimum such that (3.12) holds.

Lemma 3.4.1. Suppose that (A.1) is true. For every $\lambda \in \mathcal{L}$, there exists a value $T(\lambda)$ such that (3.12) is true.

Proof: It is enough to verify Condition (K1) of Taniguchi and Kakizawa (2000, pp. 401) which is immediate consequence of (3.9) and assumption (A.1).

K.1 (i) The function $K(\theta, \mathbf{W}, \omega)$ is defined on $\Theta \times D \times [-\pi, \pi]$ where Θ is a compact subset of $R^{(G-1)(p+1)}$ and D is an open set of C^{G^2}–the set of all square matrices of dimension G with complex elements–which contains the whole range of $\mathcal{L}$.

(ii) for $\lambda \in \mathcal{L}$, $K(., \lambda(\omega), .)$ is real valued and satisfies $|K(\theta; \lambda(\omega), \omega)| \leq k(\lambda)$ where $\int_{-\pi}^{\pi} k(\lambda) < \infty$.

$$
\begin{aligned}
|K(\theta, \lambda(\omega), \omega)| &\leq \sum_{j=1}^{G-1} \left| \left(\log \frac{\lambda_j(\omega)}{\lambda_G(\omega)} - \theta_j^T Z(\omega) \right) \right| \\
&+ G \left| \log \left(1 + \sum_{j=1}^{G-1} \frac{\lambda_j(\omega)}{\lambda_G(\omega)} \exp(-\theta_j^T Z(\omega)) \right) \right| \\
&\leq \sum_{j=1}^{G-1} \left| \log \frac{\lambda_j(\omega)}{\lambda_G(\omega)} \right| + \sum_{j=1}^{G-1} \left| \theta_j^T Z(\omega) \right| \\
&+ G \left| \log \left(1 + \sum_{j=1}^{G-1} \frac{\lambda_j(\omega)}{\lambda_G(\omega)} \exp(-\theta_j^T Z(\omega)) \right) \right| \\
&\leq C
\end{aligned}
$$

where, C is a constant, because $\lambda_j(\omega)$ are bounded above and below by a constant and θ_j belongs to a compact set, from the definition of the space $\mathcal{L}$.

(iii) $K(\theta, ., .)$ is continuous with respect to θ. This assertion is verified by the form of (3.9).

Hence Lemma 3.4.1 is true.

Define the following $(G-1)(p+1)\times(G-1)(p+1)$ matrix

$$\begin{aligned} D_\lambda &= \int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta\partial\theta^T} K(\theta;\lambda(\omega);\omega)d\omega\Big|_{\theta=T(\lambda)}. \\ &= \int_{-\pi}^{\pi} C(\theta,\lambda,\omega)\otimes(Z(\omega)Z^T(\omega))_{\theta=T(\lambda)}d(\omega) \end{aligned} \tag{3.13}$$

with

$$\begin{aligned} C(\theta,\lambda,\omega)_{j,k} &= \frac{1}{(\lambda_G(\omega)+\sum_{l=1}^{G-1}\lambda_l(\omega)\exp(-\theta_l^T Z(\omega)))^2} \\ &\times \begin{cases} G\left(\lambda_j(\omega)\exp(-\theta_j^T Z(\omega))\left(\lambda_G(\omega)+\sum_{l=1,l\neq j}^{G-1}\lambda_l(\omega)\exp(-\theta_l^T Z(\omega))\right)\right), & j=k, \\ -G\lambda_j(\omega)\lambda_k(\omega)\exp(-(\theta_j+\theta_k)^T Z(\omega)) & j\neq k, \end{cases} \end{aligned}$$

where $\otimes$ denotes the Kronecker product.

Suppose that U_λ is a $(G-1)(p+1)\times(G-1)(p+1)$ where the (i,j) block, $i,j=1,2,\ldots,G-1$ has (s,l) element given by

$$[U_{ij}]_{sl} = 4\pi\int_{-\pi}^{\pi} tr\left\{K_{is}^{(1)}(\lambda(\omega);\omega)^T\lambda(\omega)K_{jl}^{(1)}(\lambda(\omega);\omega)^T\lambda(\omega)\right\}d\omega,\ \ s,l=0,\ldots,p \tag{3.14}$$

where

$$\begin{aligned} K_{is}^{(1)}(\mathbf{W};) &= \text{diag}\left(\frac{\partial K_{is}}{\partial W_1},\frac{\partial K_{is}}{\partial W_2},\ldots,\frac{\partial K_{is}}{\partial W_G}\right), \\ K_i(\theta;\mathbf{W};\omega) &= \frac{\partial K(\theta;\mathbf{W};\omega)}{\partial\theta_i}|_{\theta=T(\lambda)} \\ &= \left(Z(\omega)-G\frac{W_i(\omega)\exp(-\theta_i^T Z(\omega))Z(\omega)}{W_G(\omega)-\sum_{j=1}^{G-1}W_j\exp(-\theta_j^T Z(\omega))}\right)_{\theta=T(\lambda)} \end{aligned}$$

for $i=1,2,\ldots,G-1$ and K_{is} denotes the s'th element of the vector K_i. Then, the following theorem holds true.

Theorem 3.4.1. Suppose that Assumption A holds true. In addition, for the true spectral density matrix $\lambda\in\mathcal{L}$ suppose that $T(\lambda)$ exists uniquely and lies in the interior of the parameter space Θ, then

$$\sqrt{N}\left(T(\hat{\lambda})-T(\lambda)\right)\Rightarrow\mathcal{N}_{(G-1)(p+1)}\left(0,D_\lambda^{-1}U_\lambda D_\lambda^{-1}\right),$$

in distribution as $N\to\infty$.

To show the validity of the theorem it is enough to verify assumption (K1)–K(7) of Taniguchi and Kakizawa (2000, pp. 402–403). Assumption (K1) has been already verified by proving Lemma 3.4.1.

We state and prove the following.

K.2 (i) $K(.,\mathbf{W},.)$ is holomorphic in D.

It is enough to show that it is differentiable, but this holds true, because of the definition of K, see (3.9)

(ii) There exists a positive constant r (independent of θ and ω) such that for every $w \in [-\pi,\pi]$, the ball

$$C_\omega = \{\mathbf{W} = (W_j) : |W_j - \lambda_j(\omega)| \leq r\} \subseteq D$$

and

$$\sup_{\theta\in\Theta} \sup_{W\in\partial C_\omega} |K(\theta;\mathbf{W};\omega)| \leq k(\omega)$$

where,

$$\partial C_\omega = \{\mathbf{W} = (W_{ab}(\omega)) : W_{ab}(\omega) = \lambda_{ab}(\omega) + re^{i\theta_{ab}}, \quad -\pi \leq \theta_{ab} \leq \pi\}$$

and $\int_\pi^\pi k(\omega)d\omega < \infty$ and r has been defined by assumption A.2.

The result follows because we notice that

$$\begin{aligned} |K(\theta;\mathbf{W};\omega)| \quad &\leq \quad \sum_{j=1}^{G-1} \left|\log \frac{\lambda_j(\omega) + re^{i\theta_{ab}}}{\lambda_G(\omega) + re^{i\theta_{ab}}}\right| + \sum_{j=1}^{G-1} \left|\theta_j^T Z(\omega)\right| \\ &+ \quad G\left|\log\left(1 + \sum_{j=1}^{G-1} \frac{\lambda_j(\omega) + re^{i\theta_{ab}}}{\lambda_G(\omega) + re^{i\theta_{ab}}} e^{-\theta_j^T Z(\omega)}\right)\right|. \end{aligned}$$

Taking the sup for $W \in \partial C_\omega$ we obtain that,

$$\sup_{W\in\partial C_\omega} |K(\theta;\mathbf{W};\omega)| \leq (G-1)M + \sum_{j=1}^{G-1} |\theta_j^T Z(\omega)| + G\left|\log\left(1 + \sum_{j=1}^{G-1} Me^{-\theta_j^T Z(\omega)}\right)\right|,$$

where M is some constant.

Now taking the supremum over Θ, we have the result,since

$$|\theta_j(\omega)Z(\omega)| \leq \|\theta_j\|\|Z(\omega\| \leq M_1\|Z(\omega)\|$$

Therefore, define

$$k(\omega) = (G-1)M + (G-1)M_1\|Z(\omega)\| + G\log(1 + (G-1)M)$$

which is clearly integrable in $[-\pi,\pi]$.

K.3 $K(\theta,.,.)$ is three times continuously differentiable. This assertion is verified by the form of (3.7)

K.4 (i) The first and second derivatives

$$K_j(\mathbf{W};.;.) = \frac{\partial}{\partial\theta_j}K(\theta,\mathbf{W},)_{|\theta=T(f)}, \quad j=1,\ldots,G-1$$

and

$$K_{ij}(\mathbf{W}) = \frac{\partial^2}{\partial\theta_i\partial\theta_j^T}_{|\theta=T(f)}, \qquad i,j=1,\ldots,G-1.$$

are holomorphic in D.

(ii) There exists a positive constant r' (independent of ω such that for every $\omega\in[-\pi,\pi]$, the ball

$$C_\omega^{'} = \{\mathbf{W}=(W_{ab}) : |W_{ab}-\lambda_{ab}(\omega)| \le r'\} \subseteq D$$

and

$$\sup_{W\in\partial C_\omega^{'}} |K_j(\mathbf{W},\omega)| \le m_j(\omega) \text{ and } \sup_{w\in\partial C_W} |K_{ij}(\mathbf{W},\omega)| \le m_{ij}(\omega)$$

with

$$\partial C_\omega^{'} = \left\{\mathbf{W}=(W_{ab}) : W_{ab} = \lambda_{ab}(\omega) + r'e^{i\theta_{ab}}, \quad -\pi\le\theta_{ab}\le\pi\right\}$$

where $m_j(\omega)$ and $m_{ij}(\omega)$ are integrable with respect to $\omega\in[-\pi,\pi]$

The first derivative is given by

$$K_j(W;.;.) = \frac{\partial}{\partial\theta_j}K(\theta,W)_{|\theta=T(f)}, \quad j=1,\ldots,G-1$$

are $p+1$ dimensional vectors. These are given by

$$\begin{aligned} K_j(W;) &= Z(\omega) - G\frac{\frac{w_j(\omega)}{w_G(\omega)}e^{-\theta_j^T Z(\omega)}}{1+\sum_{j=1}^{G-1}\frac{w_j(\omega)}{w_G(\omega)}e^{-\theta_j Z(\omega)}}Z(\omega) \quad j=1,\ldots,G \\ &= \left(1-\frac{Ge^{-\theta_j^T Z(\omega)}w_j(\omega)}{w_G(\omega)+\sum_{j=1}^{G-1}w_j(\omega)e^{-\theta_j^T Z(\omega)}}\right)Z(\omega)_{|\theta=T(f)} \end{aligned}$$

We need to show that these are holomorphic in D, but this follows since, the derivative exists, because of the boundedness from below and above. Take r' as r, following the verification of (K.2). Then component-wise and M being a constant

$$\begin{aligned} \sup_{W\in\partial C_\omega}|K_j(\mathbf{W},\omega)| &= \sup_{W\in\partial C_\omega}\left|\left(1-G\frac{M(\lambda_j+r')}{(\lambda_j+r')+\sum_{j=1}^{G-1}(\lambda_j+r')\times(M)}\right)Z(\omega)\right| \\ &\le (1+G)Z(\omega), \end{aligned}$$

which is bounded and therefore integrable in $[-\pi, \pi]$. We consider now the second derivative.

$$K_{ij}(\mathbf{W}) = \frac{\partial^2}{\partial\theta_i \partial\theta_j^T}_{|\theta=T(f)}, \qquad i,j = 1,\ldots,G-1.$$

$$K_{ii}(\mathbf{W}) = \frac{\partial^2 K}{\partial\theta_i\partial\theta_i^T} = G\frac{\frac{W_i}{W_G}\exp[-\theta_i^T Z(\omega)]\left(1+\sum_{m=1,m\neq i}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)}{\left(1+\sum_{m=1}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)^2} Z(\omega)Z^T(\omega),$$

$$i = 1,\ldots,G-1.$$

And for $i \neq j$

$$K_{ij}(\mathbf{W}) = \frac{\partial^2 K}{\partial\theta_i\partial\theta_j^T} = -G\frac{\frac{W_i}{W_G}\exp[-\theta_i^T Z(\omega)]\frac{W_j}{W_G}\exp[-\theta_j^T Z(\omega)]}{\left(1+\sum_{i=1}^{G-1}\frac{W_i}{W_G}\exp[-\theta_i Z(\omega)]\right)^2} Z(\omega)Z^T(\omega),$$

$$i,j = 1,\ldots,G-1.$$

These are clearly holomorphic functions, again because of the assumptions. At the boundary of C'_ω

$$\sup|K_{ii}| \leq \frac{-G(\lambda_j+r)(\lambda_k+r)}{\left((\lambda_G+r)+\sum_{l=1}^{G-1}(\lambda_l+r)\right)^2} Z(\omega)Z^T(\omega)$$

$$\sup|K_{ij}| \leq \frac{-G(\lambda_j+r)(1+\sum_{l=1,l\neq j}^{G-1}(\lambda_l+r))}{\left((\lambda_G+r)+\sum_{l=1}^{G-1}(\lambda_l+r)\right)^2} Z(\omega)Z^T(\omega)$$

Choose $m_{jk}(\omega)$ as the jk element of $Z(\omega)Z^T(\omega)$ multiplied by G. Clearly this is integrable on $[-\pi,\pi]$.

K.5 for every $\lambda \in R$, there exists a function

$$K_{jkl}(\omega) \quad \text{with} \quad \left|\frac{\partial^3 K(\theta;\lambda;\omega)}{\partial\theta_j\theta_k\theta_l}\right| \leq k_{jkl}(\omega),$$

for θ in a neighborhood of $T(f)$ such that $\int_{-\pi}^{\pi} k_{jkl}(\omega)d\omega < \infty$. Recall (3.7) we observe that $|K_{jkl}(\omega)| \leq 3GmM(l,r,u) \leq 24Gm$

K.6 The first derivative of $K_i(\mathbf{W};)$,

$$K_i^{(1)}(\mathbf{W};), = \{K_{i,ab}^{(1)}(\mathbf{W};)\} = \left\{\frac{\partial}{\partial W_{ab}}K_i(\mathbf{W};)\right\}$$

satisfies

$$K_i^{(1)}(\lambda(\omega),\omega) = K_i^{(1)}(\lambda(\omega),\omega)^*$$

and

$$K_i^{(1)}(\lambda(-\omega),-\omega) = K_i^{(1)}(\lambda(\omega),\omega)'.$$

Furthermore, $K_{i,ab}^{(1)}(f(\omega);\omega)$ is a piecewise continuous function. Recall that, the first derivatives evaluated at $\theta = T(f)$ are given by

$$K_j(\mathbf{W}) = \left(1 - \frac{Ge^{-\theta_j^T Z(\omega)} w_j(\omega)}{w_G(\omega) + \sum_{j=1}^{G-1} w_j(\omega) e^{-\theta_j^T Z(\omega)}}\right) Z(\omega)_{|\theta=T(f)} \quad , j = 1,\ldots,G-1$$

In the sequel we will calculate $K_{is}^{(1)}(W;)$ for $i = 1,\ldots,G-1$ and $s = 1,\ldots,p+1$

$$\frac{\partial K_{is}}{\partial W_i} = -G\frac{\frac{1}{W_G}\exp[-\theta_i^T Z(\omega)]\left(1 + \sum_{m=1,m\neq i}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)}{\left(1 + \sum_{m=1}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1,\ldots,G-1, \quad s = 1,\ldots,p+1.$$

Now for $j \neq i, G$

$$\frac{\partial K_{is}}{\partial W_j} = G\frac{\frac{W_i}{W_G}\exp[-\theta_i^T Z(\omega)]\frac{1}{W_G}\exp[-\theta_j Z(\omega)]}{\left(1 + \sum_{m=1}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1,\ldots,G-1, \quad s = 1,\ldots,p+1.$$

and

$$\frac{\partial K_{is}}{\partial W_G} = G\frac{\frac{W_i}{(W_G)^2}\exp[-\theta_i^T Z(\omega)]}{\left(1 + \sum_{m=1}^{G-1}\frac{W_m}{W_G}\exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1,\ldots,G-1, \quad s = 1,\ldots,p+1.$$

Now concluding from the above we have that

$$K_{is}^{(1)}(W;) = \text{diag}\left(\frac{\partial K_{is}}{\partial W_1}, \frac{\partial K_{is}}{\partial W_2}, \ldots, \frac{\partial K_{is}}{\partial W_G}\right)$$

Thus, $K_{i,ab}^{(1)}(\lambda(\omega);\omega)$ is a piecewise continuous function. Also satisfies

$$K_i^{(1)}(\lambda(\omega),\omega) = K_i^{(1)}(\lambda(\omega),\omega)^*$$

and

$$K_i^{(1)}(\lambda(-\omega),-\omega) = K_i^{(1)}(\lambda(\omega),\omega)'.$$

K.7 The $(G-1)(p+1)\times(G-1)(p+1)$ matrix

$$D_\lambda = \int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta\partial\theta^T} K(\theta;\lambda(\omega);\omega)|_{\theta=T(\lambda)} d\omega.$$

defined by (3.13) is nonsingular, which holds by assumption. Notice that the matrix D_λ is hermitian and symmetric. Having verified the assumptions (K1)-(K7) the Theorem holds true.

Remark 3.4.1. Theorem 3.4.1 states the asymptotic distribution of the functional $T(\hat{\lambda})$ when model (3.3) does not necessarily hold but inference is based on the contrast function (3.8) in general. Therefore, the result is quite general and quantifies the effect of misspecification of model (3.3) to inference.

We will study the asymptotic properties of the maximum likelihood estimator $\hat{\theta}$ under the correct model (3.3). In addition, notice that when model (3.3) holds true, we obtain that

$$\begin{aligned} K(\theta,\mathbf{W}(\omega),\omega) &= -\sum_{j=1}^{G-1}\log\left(\frac{W_j(\omega)/\lambda_j(\omega)}{W_G(\omega)/\lambda_G(\omega)}\right) + G\log\left(1+\sum_{j=1}^{G-1}\frac{W_j(\omega)/\lambda_j(\omega)}{W_G(\omega)/\lambda_G(\omega)}\right) \\ &= H(\mathbf{W}\lambda^{-1}(\omega)) \end{aligned}$$

where $H(.)$ is a function that is defined on the set of all $G\times G$ diagonal matrices $\mathcal{F}$, by

$$H(\tilde{\mathbf{W}}) = -\sum_{j=1}^{G-1}\log\frac{\tilde{W}_j}{\tilde{W}_G} + G\log\left(1+\sum_{j=1}^{G-1}\frac{\tilde{W}_j}{\tilde{W}_G}\right) \tag{3.15}$$

Since we are considering positive spectral densities which are bounded above and below– recall the definition of $\mathcal{L}$–assumption (A.1) and (B.3), we obtain that the restriction of $H(.)$ to the diagonal matrices with real elements has a unique minimum at I_G–the identity matrix of dimension G.

Lemma 3.4.2. Recall (3.15). Then for all $G\times G$ diagonal matrices $\tilde{\mathbf{W}}$ with positive diagonal elements, $H(\tilde{\mathbf{W}}) \geq H(I_G)$.

Proof: We know that for any $x_1,\ldots,x_n \geq 0$

$$\frac{x_1+\ldots+x_n}{n} \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

In our case define $x_i = \frac{\psi_i}{\psi_G}$, for $i = 1, \ldots, G-1$ and $x_G = 1$.

Thus,

$$\frac{1 + \sum_{j=1}^{G-1} x_j}{G} \geq \left(\prod_{i=1}^{G-1} x_i \right)^{\frac{1}{G}}$$

$$\begin{aligned}
&\Rightarrow \quad \log\left(1 + \sum_{j=1}^{G-1} x_j\right) - \log G \geq \frac{1}{G} \sum_{j=1}^{G-1} \log x_j \\
&\Rightarrow \quad G \log\left(1 + \sum_{j=1}^{G-1} \frac{\psi_i}{\psi_G}\right) - \sum_{j=1}^{G-1} \log \frac{\psi_j}{\psi_G} \geq G \log G, \quad \forall \ \ \psi_1, \ldots, \psi_G \\
&\Rightarrow \quad H(\tilde{\mathbf{W}}) \geq H(I_G).
\end{aligned}$$

Therefore, we obtain the following theorem, regarding the asymptotic behavior of the maximum likelihood estimator under a correctly specified model, see Taniguchi and Kakizawa (2000, Cor. 6.2.5)

Theorem 3.4.2. Suppose that Assumption B holds true. Then

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \Rightarrow \mathcal{N}_{(G-1)(p+1)}\left(0, 4\pi D_0^{-1}\right)$$

in distribution as $N \to \infty$, where the matrix D_0 is defined by

$$D_0 = \int_{-\pi}^{\pi} \frac{1}{G}((G-1)I_{G-1} + (J_{G-1} - I_{G-1})) \otimes (Z(\omega)Z^T(\omega))d(\omega) \tag{3.16}$$

where I_{G-1} is the unit matrix of dimension $G-1$, J_{G-1} is the $(G-1) \times (G-1)$ matrix of ones. The above representation shows that the limiting variance matrix is not singular since both matrices forming the Kronecker product are not singular.

Proof: Define,

$$\begin{aligned}
K(\theta, \lambda_1(\omega), \ldots, \lambda_G(\omega), \omega)) = \ & - \sum_{j=1}^{G-1} \left(\log \frac{\lambda_j(\omega)}{\lambda_G(\omega)} - \theta_j^T(\omega) Z(\omega) \right) \\
& + G \log \left(1 + \sum_{j=1}^{G-1} \frac{\lambda_j(\omega)}{\lambda_G(\omega)} e^{-\theta_j^T Z(\omega)} \right)
\end{aligned}$$

Now consider all the matrices of the form $\mathbf{W} = \mathrm{diag}(W_1, W_2, \ldots, W_G)$

$$\begin{aligned}
K(\theta; \mathbf{W}; \omega) = \ & - \sum_{j=1}^{G-1} \left(\log \frac{W_j(\omega)}{W_G(\omega)} - \theta_j^T(\omega) Z(\omega) \right) \\
& + G \log \left(1 + \sum_{j=1}^{G-1} \frac{W_j(\omega)}{W_G(\omega)} e^{-\theta_j^T Z(\omega)} \right)
\end{aligned}$$

Recall (3.14) we calculate the following quantities:

$$K_{is}^{(1)}(\mathbf{W};) = \frac{\partial}{\partial \mathbf{W}} K_i(\theta; \mathbf{W}; \omega) \qquad i = 1, \ldots, G-1 \;\text{ and }\; s = 1, \ldots, p+1$$

and

$$K_i(\theta; \mathbf{W}; \omega) = \frac{\partial K(\theta; \mathbf{W}; \omega)}{\partial \theta_i}|_{\theta = T(\lambda)}, \quad T(\lambda) = \min_{\theta} D(\theta; \lambda) \qquad i = 1, \ldots, G-1$$

and

$$D_\lambda = \int_{-1/2}^{1/2} \frac{\partial^2}{\partial \theta \partial \theta^T} K(\theta; \lambda(\omega); \omega) d\omega \;|_{\theta = T(\lambda)}$$

Now we will calculate the D_λ matrix,

$$\frac{\partial K}{\partial \theta_i} = Z(\omega) - G \frac{\frac{W_i}{W_G} \exp[-\theta_i^T Z(\omega)] Z(\omega)}{1 + \sum_{m=1}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]}, \qquad i = 1, \ldots, G-1.$$

$$\begin{aligned} \frac{\partial^2 K}{\partial \theta_i \partial \theta_i^T} &= G \frac{\frac{W_i}{W_G} \exp[-\theta_i^T Z(\omega)] \left(1 + \sum_{m=1, m \neq i}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)^2} Z(\omega) Z^T(\omega), \\ &= C_{ii} Z(\omega) Z^T(\omega), \qquad i = 1, \ldots, G-1. \end{aligned}$$

And for $i \neq j$

$$\begin{aligned} \frac{\partial^2 K}{\partial \theta_i \partial \theta_j^T} &= -G \frac{\frac{W_i}{W_G} \exp[-\theta_i^T Z(\omega)] \frac{W_j}{W_G} \exp[-\theta_j^T Z(\omega)]}{\left(1 + \sum_{i=1}^{G-1} \frac{W_i}{W_G} \exp[-\theta_i Z(\omega)]\right)^2} Z(\omega) Z^T(\omega) \\ &= C_{ij} Z(\omega) Z^T(\omega), \qquad i, j = 1, \ldots, G-1. \end{aligned}$$

Now when the true model holds, then

$$C_{ii}|_{\theta = \theta_0} = \frac{G-1}{G}$$

and for $i \neq j$

$$C_{ij}|_{\theta = \theta_0} = -\frac{1}{G}.$$

Thus $D_\lambda = C \bigotimes Z(\omega) Z^T(\omega)$, where C is a matrix with elements $C_{ii} = \frac{G-1}{G}$ and $C_{ij} = -\frac{1}{G}$, for $j \neq i$. Now recall the vector $K_i = (K_{i1}, K_{i2}, \ldots, K_{i;p+1}), \quad i = 1, \ldots, G-1$

Then $K_{is} = \frac{\partial K}{\partial \theta_i} W_s(\omega)$. In the sequel we will calculate $K_{is}^{(1)}(\mathbf{W};)$ for $i = 1, \ldots, G-1$

and $s = 1, \ldots, p+1$

$$\frac{\partial K_{is}}{\partial W_i} = -G \frac{\frac{1}{W_G} \exp[-\theta_i^T Z(\omega)] \left(1 + \sum_{m=1, m\neq i}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

Now for $j \neq, i, G$

$$\frac{\partial K_{is}}{\partial W_j} = G \frac{\frac{W_i}{W_G} \exp[-\theta_i^T Z(\omega)] \frac{1}{W_G} \exp[-\theta_j Z(\omega)]}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

and

$$\frac{\partial K_{is}}{\partial W_G} = G \frac{\frac{W_i}{(W_G)^2} \exp[-\theta_i^T Z(\omega)]}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m}{W_G} \exp[-\theta_m Z(\omega)]\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

Now concluding from the above we have that

$$K_{is}^{(1)}(\mathbf{W};) = \text{diag}\left(\frac{\partial K_{is}}{\partial W_1}, \frac{\partial K_{is}}{\partial W_2}, \ldots, \frac{\partial K_{is}}{\partial W_G}\right)$$

Substituting the above into (3.14),

$$\begin{aligned} K_{is}^{(1)}(\lambda(\omega);\omega)^T \lambda(\omega) K_{jl}^{(1)}(\lambda(\omega);\omega)^T \lambda(\omega) &= \text{diag}\left(\frac{\partial K_{is}}{\partial W_1} W_1, \frac{\partial K_{is}}{\partial W_2} W_2, \ldots, \frac{\partial K_{is}}{\partial W_G} W_G\right) \\ &\times \text{diag}\left(\frac{\partial K_{jl}}{\partial W_1} W_1, \frac{\partial K_{jl}}{\partial W_2} W_2, \ldots, \frac{\partial K_{jl}}{\partial W_G} W_G\right) \end{aligned}$$

The result is a diagonal matrix :

$$\text{diag}\left(\frac{\partial K_{is}}{\partial W_1}\frac{\partial K_{jl}}{\partial W_1}(W_1)^2, \frac{\partial K_{is}}{\partial W_2}\frac{\partial K_{jl}}{\partial W_2}(W_2)^2, \ldots, \frac{\partial K_{is}}{\partial W_G}\frac{\partial K_{jl}}{\partial W_G}(W_G)^2\right),$$

and this implies that

$$\text{tr}\left[K_{is}^{(1)}(\lambda(\omega);\omega)^T \lambda(\omega) K_{jl}^{(1)}(\lambda(\omega);\omega)^T \lambda(\omega)\right] = \sum_{t=1}^{G} \frac{\partial K_{is}}{\partial W_t}\frac{\partial K_{jl}}{\partial W_t}(W_t)^2$$

Let $A_i = \exp[-\theta_i^T Z(\omega)], \quad i = 1, \ldots, G-1$

If we rewrite the following expressions we have,

$$\frac{\partial K_{is}}{\partial W_i} = -G \frac{\frac{A_i}{W_G}\left(1 + \sum_{m=1, m\neq i}^{G-1} \frac{W_m A_m}{W_G}\right)}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

for $j \neq, i, G$

$$\frac{\partial K_{is}}{\partial W_j} = G \frac{\frac{W_i A_i A_j}{W_G}}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

and

$$\frac{\partial K_{is}}{\partial W_G} = G \frac{\frac{W_i A_i}{(W_G)^2}}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} W_s(\omega),$$

$$i = 1, \ldots, G-1, \quad s = 1, \ldots, p+1.$$

Now,

$$\begin{aligned}
\sum_{t=1}^{G} \frac{\partial K_{is}}{\partial W_t} \frac{\partial K_{il}}{\partial W_t} (W_t)^2 &= \left[-G \frac{\frac{A_i}{W_G}\left(1 + \sum_{m=1, m\neq i}^{G-1} \frac{W_m A_m}{W_G}\right)}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} \right]^2 W_s(\omega) W_l(\omega) W_i^2 \\
&+ \sum_{t=1, t\neq i}^{G-1} \left\{ \left[G \frac{\frac{W_i A_i A_t}{W_G}}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} \right]^2 W_s(\omega) W_l(\omega) W_t^2 \right\} \\
&+ \left[G \frac{\frac{W_i A_i}{(W_G)^2}}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^2} \right]^2 W_s(\omega) W_l(\omega) W_G^2
\end{aligned}$$

Now under the true model

$$\begin{aligned}
\sum_{t=1}^{G} \frac{\partial K_{is}}{\partial W_t} \frac{\partial K_{il}}{\partial W_t} (W_t)^2 &= \frac{G-1}{G}, \\
\sum_{t=1}^{G} \frac{\partial K_{is}}{\partial W_t} \frac{\partial K_{jl}}{\partial W_t} (W_t)^2 &= -\frac{1}{G}
\end{aligned}$$

And for $i \neq j$

$$\sum_{t=1}^{G} \frac{\partial K_{is}}{\partial W_t} \frac{\partial K_{jl}}{\partial W_t} (W_t)^2 = -G^2 \frac{\frac{A_i}{W_G}\left(1 + \sum_{m=1, m\neq i}^{G-1} \frac{W_m A_m}{W_G}\right) \frac{W_j A_j A_t}{W_G}}{\left(1 + \sum_{m=1}^{G-1} \frac{W_m A_m}{W_G}\right)^4} W_s(\omega) W_l(\omega) W_i^2$$

$$
\begin{aligned}
&- \quad G^2 \frac{\frac{A_j}{W_G}\left(1+\sum_{m=1,m\neq i}^{G-1}\frac{W_mA_m}{W_G}\right)\frac{W_iA_iA_t}{W_G}}{\left(1+\sum_{m=1}^{G-1}\frac{W_mA_m}{W_G}\right)^4} W_s(\omega)W_l(\omega)W_j^2 \\
&+ \quad \sum_{t=1,t\neq i,j}^{G-1}\left\{ G^2 \frac{\frac{W_iA_iA_t}{W_G}\frac{W_jA_jA_t}{W_G}}{\left(1+\sum_{m=1}^{G-1}\frac{W_mA_m}{W_G}\right)^4} W_s(\omega)W_l(\omega)W_t^2 \right\} \\
&+ \quad G^2 \frac{\frac{W_iA_i}{(W_G)^2}\frac{W_jA_j}{(W_G)^2}}{\left(1+\sum_{m=1}^{G-1}\frac{W_mA_m}{W_G}\right)^4} W_s(\omega)W_l(\omega)W_G^2
\end{aligned}
$$

Now when the true model holds,

$$
\sum_{t=1}^{G}\frac{\partial K_{is}}{\partial W_t}\frac{\partial K_{il}}{\partial W_t}(W_t)^2 = \frac{G-1}{G} \quad \text{and} \quad \sum_{t=1}^{G}\frac{\partial K_{is}}{\partial W_t}\frac{\partial K_{jl}}{\partial W_t}(W_t)^2 = -\frac{1}{G}
$$

Thus $U = \Delta \otimes Z(\omega)Z^T(\omega)$, where Δ is a matrix with elements $\Delta_{ii} = \frac{G-1}{G}$ and $\Delta_{ij} = -\frac{1}{G}$, for $j \neq i$. Comparing the matrices U and D_λ, we can see that they are the same.

3.5 Testing

In this section we will be concerned with the composite hypothesis

$$
\mathrm{H}_0 : A\theta = 0 \quad \text{against} \quad \mathrm{H}_1 : A\theta \neq 0,
$$

for model (3.3), where A is a matrix of dimension $d \times (G-1)(p+1)$, $d \leq (G-1)(p+1)$, and assumed to be of full rank. To test the null hypothesis consider the likelihood ratio test

$$
LR = 2n\left(D(\tilde{\theta},\hat{\lambda}) - D(\hat{\theta},\hat{\lambda})\right) \tag{3.17}
$$

where $D(.,.)$ has been defined by (3.8) and $\tilde{\theta}$ denotes the maximum likelihood estimator of θ under the null hypothesis. Theorem 6.2.7 of Taniguchi and Kakizawa (2000) shows that the asymptotic distribution of the quantity LR is a chi–square distribution with degrees of freedom equal to d. In the same spirit, testing equality of several spectral density functions can be carried out by means of the score and Wald tests. To be more specific,

$$
W = n\left\{A'T(\hat{\lambda})\right\}'\left[A'D_0^{-1}A\right]\left\{A'T(\hat{\lambda})\right\} \tag{3.18}
$$

and

$$LM = n\left[\frac{\partial}{\partial\theta}D(\tilde{\theta},\hat{\lambda})\right]' D_0^{-1}\left[\frac{\partial}{\partial\theta}D(\tilde{\theta},\hat{\lambda})\right] \tag{3.19}$$

where $D(.,.)$ has been defined above, D_0 has been defined by (3.16),

$$T(\hat{\lambda}) = \arg\min_{\theta} D(\tilde{\theta},\hat{\lambda}) \qquad T_0(\hat{\lambda}) = \arg\min_{\theta:A'\theta=0} D(\tilde{\theta},\hat{\lambda}).$$

All these test statistics are the same, the asymptotic distribution of these quantities are a chi–square distribution with degrees of freedom equal to d. More specifically,

Lemma 3.5.1. (Taniguchi and Kakizawa, 2000, Thm.6.2.7). Under assumption B and under $A\theta = 0$, the tests LR, W and LM tend to a chi–square distribution with degrees of freedom equal to d.

Proof: The proof follows from Theorem 6.2.7 of Taniguchi and Kakizawa (2000).

A natural question that arises after rejecting the equality of all spectral densities is that of multiple comparisons among the time series under consideration for identification of similar groups. In this section of the thesis we take a simple approach by suggesting a Bonferroni correction to adjust the p–values of all the $\binom{G}{2}$ sets of hypotheses $\theta_i = \theta_j$, $i, j = 1, 2, \ldots, G$, $i \neq j$ and $\theta_G \equiv 0$. For the data examples that is considered next, see subsection 3.7.2, it is straightforward to carry out the adjustment of the p–values, since $G = 3$. However, when G is large other methods might be preferable, see Miller (1981), but also subsection 4.3.2.

3.5.1 A Note on Computation

To carry out the numerical calculations involved, recall the notation of (3.4). In addition, define

$$\tilde{T}_{ji} = \log\frac{\tilde{X}_{ji}}{\tilde{X}_{Gi}}, \tag{3.20}$$

where $\tilde{X}_{ji} = \hat{\lambda}_j(\omega_i)$, that is the smoothed periodogram ordinate. Then fit the spectral density ratio model (3.3) by maximizing (3.5) with either T_{ji} or $\tilde{T}_{ji}$. Although Theorem 3.4.2 states that the smoothed periodogram based estimator is asymptotically normal, several simulation experiments reported below are indicating that normality holds for the raw periodogram based estimator. Details about the smooth spectral density estimator are given in Section 3.7. Maximization of the log–likelihood function (3.5) is carried out by

using the statistical language R–see R Development Core Team (2004) and in particular it is implemented by using the function `optim()` after choosing the option "BFGS"' which corresponds to a quasi-Newton method. The score function that is employed with this option is given by

$$\frac{\partial l(\theta)}{\partial \theta} = \left(\frac{\partial l(\theta)}{\partial \theta_1}, \ldots, \frac{\partial l(\theta)}{\partial \theta_{G-1}} \right)^T$$

where

$$\frac{\partial l(\theta)}{\partial \theta_j} = -\sum_{i=1}^{m} Z_i + G \sum_{i=1}^{m} \frac{\exp(T_{ji} - \theta_j^T Z_i) Z_i}{\left(1 + \sum_{j=1}^{G-1} \exp(T_{ji} - \theta_j^T Z_i)\right)}, \quad j = 1, 2, \ldots, G-1$$

following the notation of (3.5). The above score equations are employed with either T_{ji} or $\tilde{T}_{ji}$.

Similar calculations show that for testing the following hypotheses

$$H_0 : \theta_1 = \ldots = \theta_{G-1} = 0$$

which points to the equality of all spectral distributions to the reference , the likelihood ratio test is equal to

$$LR = -2 \sum_{i=1}^{m} \sum_{j=1}^{G-1} \widehat{\theta}_j^T Z_i + 2G \sum_{i=1}^{m} \log \left(\frac{1 + \sum_{j=1}^{G-1} \exp(T_{ji})}{1 + \sum_{j=1}^{G-1} \exp(T_{ji} - \widehat{\theta}_j^T Z_i)}, \right) \tag{3.21}$$

and its distribution is a chi–square random variable with $(G-1)(p+1)$ degrees of freedom. Formula (3.21) is used with either T_{ji} or $\tilde{T}_{ji}$.

In summary we have studied the asymptotic inference for the regression parameters under model (3.3). Theorem 3.4.1 shows that the maximum likelihood estimator converges to a normal distribution with a covariance matrix that can be calculated and estimated consistently. The asymptotic normality holds true regardless the validity of model (3.3) when using the contrast function (3.8). However, when (3.3) holds true, the asymptotic covariance matrix has a much simpler form, see Theorem 3.4.2 and (3.16). Both results are true for independent linear processes, but not necessarily Gaussian. This implies that a vast collection of examples falls within this framework and therefore the proposed methodology is quite general. The asymptotic results are based on the data (3.20), that is the smoothed periodogram ordinates are employed for inference. But it is demonstrated next that the raw periodogram based estimator, see (3.4) has almost identical behavior to that of the smoothed periodogram.

3.6 Goodness of Fit test

For examining the adequacy of the model defined by (3.3) we use the idea of smooth goodness of fit tests. To be specific, we can embed this model into a larger parametric family, according to Rayner and Best (1989)

$$\log \mu_j(\omega) = \log \frac{\lambda_j(\omega)}{\lambda_G(\omega)} = \theta_j^T Z + \gamma_j^T Z^* \qquad \text{for} \quad j = 1, \cdots, G-1$$

where,

$$\begin{aligned} \theta_j &= (a_{j0}, a_{j1}, \ldots, a_{jp})^T, \\ Z &= (1, 2\cos\omega, 2\cos 2\omega, \ldots, 2\cos p\omega)^T, \\ \gamma_j &= (\gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jq})^T \end{aligned}$$

and

$$Z^* = (2\cos(p+1)\omega, 2\cos(p+2)\omega, \ldots, 2\cos(p+q)\omega)^T$$

For testing the adequacy of the model defined by (3.3) is equivalent to testing the hypothesis $H_0 : \gamma_j = 0$.

3.7 Examples and Data Analysis

3.7.1 Simulations

The proposed method is illustrated empirically by means of a simulated study which includes several examples. The first two examples deal with the case that all observed time series are generated with the same second order structure. The third example shows estimation results when the data are generated by an exponential model. The fourth example examines the power of the likelihood ratio test (3.21) when the data have different second order properties while the last example–which studies ARCH processes–is included to show the robustness of the methodology to non standard situations. The data fitting process is applied by rescaling the frequencies to the interval (0, 1/2). All simulations are based on 1000 runs. To calculate the values of $\tilde{T}_{ji}$, see (3.20), we use the Daniel window with a 5 point discrete spectral average estimator.

Example 3.7.1. Suppose that three time series data sequences are generated by the AR(1) model, see (2.3),

$$Y_{it} = 0.50Y_{i(t-1)} + \epsilon_{it}$$

for $i = 1, 2, 3$ and ϵ_{it} are independent and identically distributed standard normal variables. In other words, all three time series share the same second order structure. Model (3.3) is applied for various values of p. For estimation purposes, consider the spectral density of the third sequence as the reference function in (3.3) but notice that any other sequence would have yielded the same results since in this case representation (3.3) is identical to zero for all ω. Table 3.1 shows estimation results when $p = 2$ and for different length sequences. Notice that when $p = 2$ and $G = 3$, then the total number of parameters that needs to be estimated equals to six. More specifically, following the notation of (3.3), notice that $\theta_j = (a_{j0}, a_{j1}, a_{j2})'$ for $j = 1, 2$. The estimation method performs reasonably well for both the raw and smoothed periodogram based maximum likelihood estimators. The approximate standard errors reported in the last column of Table 3.1 are approximated by means of (3.16) evaluated at the Fourier frequencies. Evidentally the true standard errors are better approximated when the sample size tends to larger values. Further evidence of the asymptotic validity of the proposed methodology is illustrated by means of Table 3.2 which reports achieved significance levels of the likelihood ratio test statistic (3.21) for testing equality of the three corresponding spectral density functions. The table has been constructed for different sample sizes and by varying the value of p in (3.3) for either the raw periodogram based procedure or the smooth periodogram based method. The reference distribution is the chi–square with degrees of freedom equal to $2(p + 1)$. Notice that for the case of the raw periodogram and when N is large, the achieved significance levels are close to the nominal. However for the smoothed periodogram estimator the approximation is better when N is large and $p \leq 5$. This is a consequence of smoothing, that is a much smoother estimator will perform reasonably better.

N	Coefficients	True Param.	Estim. Param.[1]	Simul. Stand Err.[1]	Estim. Param.[2]	Simul. Stand. Err.[2]	Approx. Stand. Err.
50	a_{10}	0	-0.002	0.341	-0.010	0.309	0.283
	a_{11}	0	-0.020	0.241	-0.004	0.215	0.201
	a_{12}	0	0.003	0.245	0.002	0.200	0.201
	a_{20}	0	0.003	0.348	0.001	0.301	0.283
	a_{21}	0	-0.009	0.256	-0.001	0.210	0.201
	a_{22}	0	0.000	0.243	0.007	0.207	0.201
100	a_{10}	0	0.000	0.229	-0.011	0.215	0.200
	a_{11}	0	-0.002	0.170	0	0.154	0.141
	a_{12}	0	0.004	0.173	-0.006	0.150	0.141
	a_{20}	0	0.001	0.232	-0.005	0.193	0.200
	a_{21}	0	-0.002	0.165	0.001	0.154	0.141
	a_{22}	0	0.003	0.169	-0.002	0.152	0.141
500	a_{10}	0	-0.001	0.104	0.002	0.093	0.089
	a_{11}	0	-0.002	0.073	0	0.064	0.063
	a_{12}	0	0	0.074	-0.001	0.065	0.063
	a_{20}	0	0.003	0.105	-0.001	0.095	0.089
	a_{21}	0	0	0.074	0	0.068	0.063
	a_{22}	0	-0.001	0.074	0	0.063	0.063

Table 3.1: Three independent time series from the AR(1) model with different lengths and $\phi_1 = 0.5$. Model (3.3) holds true with $\theta_j = 0$ for $j = 1, 2$. Results are based for $p = 2$ and 1000 simulations. [1]: Estimation based on T_{ji}, see (3.4). [2]: Estimation based on $\tilde{T}_{ji}$, see (3.20).

Raw Periodogram

N	Significance Levels	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
50	0.01	0.008	0.007	0.013	0.012	0.018
	0.05	0.053	0.070	0.058	0.061	0.066
	0.10	0.106	0.140	0.127	0.121	0.129
100	0.01	0.010	0.008	0.008	0.007	0.010
	0.05	0.060	0.039	0.056	0.065	0.053
	0.10	0.118	0.082	0.099	0.111	0.106
500	0.01	0.014	0.009	0.013	0.013	0.012
	0.05	0.051	0.040	0.051	0.058	0.046
	0.10	0.097	0.095	0.106	0.095	0.096

Smoothed Periodogram

50	0.01	0.007	0.004	0.007	0.001	0.001
	0.05	0.042	0.029	0.030	0.011	0.004
	0.10	0.091	0.068	0.057	0.039	0.018
100	0.01	0.013	0.003	0.009	0.012	0.010
	0.05	0.058	0.040	0.045	0.046	0.045
	0.10	0.110	0.082	0.088	0.086	0.074
500	0.01	0.011	0.011	0.007	0.090	0.004
	0.05	0.043	0.050	0.053	0.040	0.030
	0.10	0.091	0.096	0.099	0.102	0.090

Table 3.2: Achieved significance levels of the likelihood ratio test statistic (3.21) for testing equality of three spectral density functions. The data are generated by the same three independent AR(1) process with $\phi_1 = 0.5$ for different N and model (3.3) is fitted for different p. Results are based on 1000 simulations.

Example 3.7.2. Suppose now that there are four sequences of observations from the ARMA(1,1) model

$$Y_{it} = 0.50Y_{i(t-1)} + \epsilon_{it} + 0.10\epsilon_{i(t-1)}$$

for $i = 1, 2, 3, 4$ and ϵ_{it} are independent and identically distributed standard normal variables. For this example, notice that $G = 4$ and therefore the total number of unknown parameters equals to $3(p + 1)$ which is determined by the choice of p. For instance, when $p = 2$, then the total number of parameters is nine. Figure 3.1 demonstrates this situation in the case of $N = 100$ by depicting qq-plots of the maximum likelihood estimators of θ_j, $j = 1, 2, 3$ for the raw periodogram based estimators while Figure 3.2 illustrates the same information for the smooth periodogram based estimator. In all cases, the asymptotic normality seems an adequate approximation to the asymptotic distribution of the maximum likelihood estimators. Furthermore, Figures 3.3 and 3.4 show qq–plots of the test statistic (3.21) for $N = 100$ for different values of p. So the left plot of Figure 3.3 illustrates qq-plots against the chi–square distribution with degrees of freedom equal to nine. Similarly for all the other graphs. We reconfirm the observation that was made before, that is the asymptotic approximation seems quite satisfactory in all cases considered.

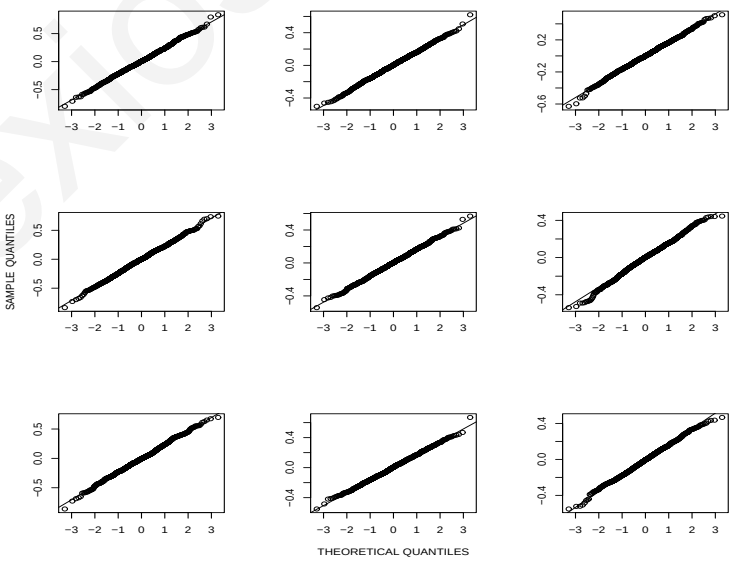

Figure 3.1: QQ–plots of $\hat{\theta}_1$ (upper level), $\hat{\theta}_2$ (middle level) and $\hat{\theta}_3$ (lower level) for four time series from the same ARMA(1,1) processes with $N = 100$. Model (3.3) in connection with (3.4) is fitted for $p = 2$ and results are based on 1000 simulations.

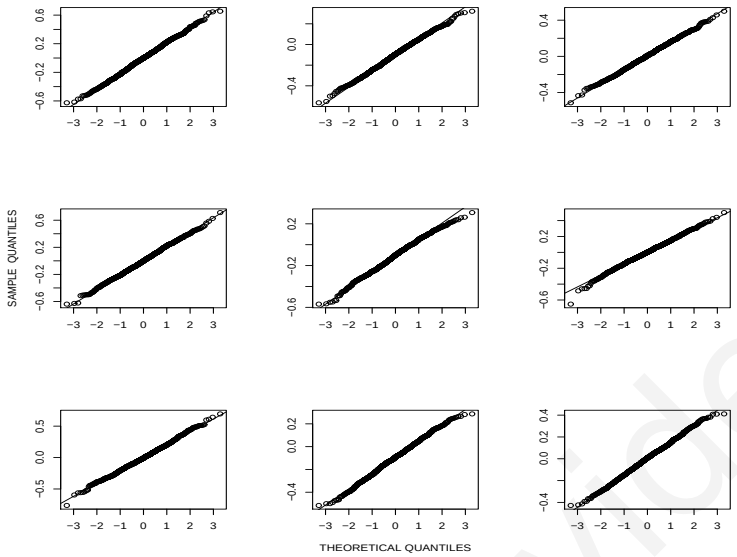

Figure 3.2: QQ–plots of $\hat{\theta}_1$ (upper level), $\hat{\theta}_2$ (middle level) and $\hat{\theta}_3$ (lower level) for four time series from the same ARMA(1,1) processes with $N = 100$. Model (3.3) in connection with (3.20) is fitted for $p = 2$ and results are based on 1000 simulations.

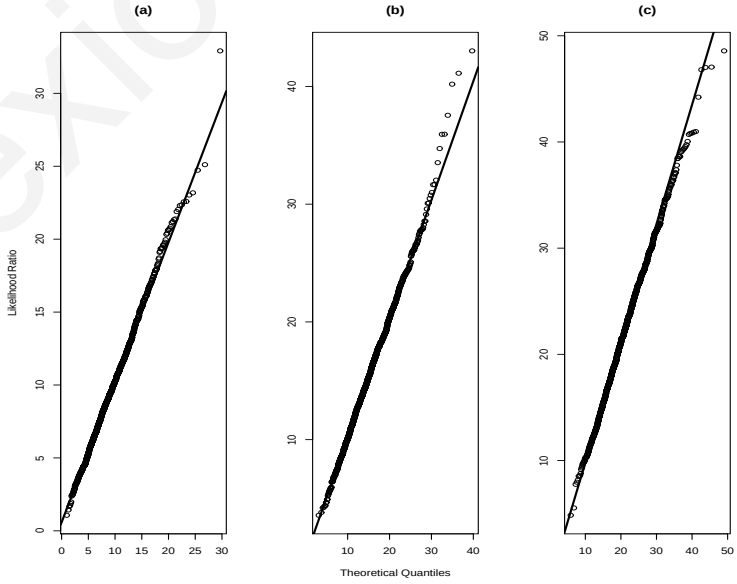

Figure 3.3: QQ–plots of the test statistic (3.21) for testing the equality of spectral density functions for four time series from the same ARMA(1,1) processes with $N = 100$. Model (3.3) in connection with (3.4) is fitted for different p and results are based on 1000 simulations. (a) $p = 2$, (b) $p = 4$, (c) $p = 6$.

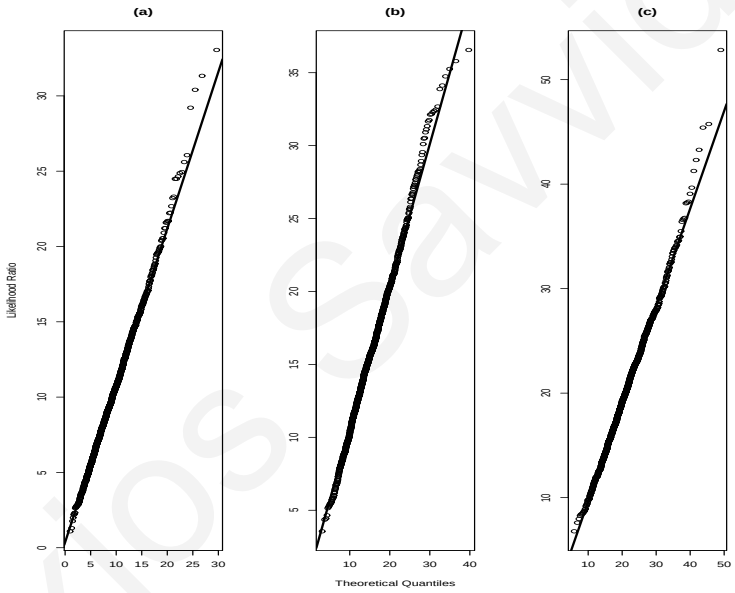

Figure 3.4: QQ–plots of the test statistic (3.21) for testing the equality of spectral density functions for four time series from the same ARMA(1,1) processes with $N = 100$. Model (3.3) in connection with (3.20) is fitted for different p and results are based on 1000 simulations. (a) $p = 2$, (b) $p = 4$, (c) $p = 6$.

Example 3.7.3. Two sequences of the $EXP(2)$ model with parameters $\theta = (-0.5, -0.9, 0.4)$ and one sequence of the $EXP(2)$ model with parameters $\theta = (0.5, 0.3, 0.15)$ were either drown. Evidentally model (3.3) holds true. Table 3.3 reports the estimation results for both smoothed and unsmoothed periodogram based estimators. As we can see we have better results as N increases.

N	Coefficients	True Param.	Estim. Param.[1]	Simul. Stand Err.[1]	Estim. Param.[2]	Simul. Stand. Err.[2]	Approx. Stand. Err.
50	a_{10}	-1	-0.931	0.367	-0.897	0.320	0.283
	a_{11}	-1.2	-1.157	0.249	-1.155	0.218	0.201
	a_{12}	0.25	0.225	0.246	0.207	0.205	0.201
	a_{20}	-1	-0.912	0.375	-0.881	0.314	0.283
	a_{21}	-1.2	-1.160	0.251	-1.148	0.219	0.201
	a_{22}	0.25	0.230	0.248	0.225	0.215	0.201
100	a_{10}	-1	-0.959	0.236	-0.959	0.220	0.200
	a_{11}	-1.2	-1.183	0.166	-1.180	0.144	0.142
	a_{12}	0.25	0.229	0.174	0.237	0.153	0.142
	a_{20}	-1	-0.965	0.247	-0.957	0.215	0.200
	a_{21}	-1.2	-1.187	0.170	-1.187	0.145	0.142
	a_{22}	0.25	0.235	0.173	0.236	0.154	0.142
500	a_{10}	-1	-0.992	0.102	-0.996	0.095	0.089
	a_{11}	-1.2	-1.197	0.073	-1.198	0.067	0.063
	a_{12}	0.25	0.248	0.075	0.247	0.067	0.063
	a_{20}	-1	-0.994	0.106	-0.991	0.095	0.089
	a_{21}	-1.2	-1.195	0.070	-1.200	0.063	0.063
	a_{22}	0.25	0.248	0.073	0.249	0.064	0.063

Table 3.3: True and estimated coefficients (together with simulated and true standard errors) for model (3.3) when $p = 2$ when the data are generated by three EXP(2) time series. Results are based on 1000 simulations.[1]: Estimation based on T_{ji}, see (3.4). [2]: Estimation based on $\tilde{T}_{ji}$, see (3.20).

Example 3.7.4. We now study the estimation problem, when the observed data does not share the same second order structure. In particular, for this example, consider four time series according to the following specification

$$\begin{aligned} Y_{it} &= 0.50Y_{i(t-1)} + \epsilon_{it}, \quad i = 1, 2 \\ Y_{3t} &= 0.20Y_{3(t-1)} + 0.60Y_{3(t-2)} + \epsilon_{3t}, \\ Y_{4t} &= 0.40Y_{4(t-1)} + \epsilon_{4t} - 0.10\epsilon_{4(t-1)} \end{aligned} \tag{3.22}$$

and ϵ_{it} are independent and identically distributed standard normal variables. Without repeating any estimation arguments, notice that Table 3.4 shows close agreement between the true parameter values and the estimated parameters. Recall that the approximate standard errors reported in the last column of Table 3.4 are approximated by means of (3.16) evaluated at the Fourier frequencies.

For this example, the spectral density of Y_{4t} is employed as reference for fitting model (3.3). That is we use the ratios $\lambda_i(\omega)/\lambda_4(\omega)$, $i = 1, 2, 3$, to fit model (3.3) using the representation (2.4). Therefore the true parameter values are calculated according to Lemma 2.5.1. Larger values of N indicate more adequate approximation, as it should be expected. In addition, Table 3.5 shows simulated power of the test statistic (3.21) for testing the hypotheses that all four spectral densities are equal. Note that both raw and smooth periodogram based tests share a large power as N increases but the likelihood ratio test based on (3.20) has larger power than that based on (3.4). The raw periodogram based test does not show any great variation among the simulated power values when p varies. In contrast the smooth periodogram test's power reduces as p assumes larger values–this is consequence of smoothing. For larger N, say $N = 500$, the power was identical to one for both tests and therefore it is not reported in the table.

N	Coefficients	True Param.	Estim. Param.[1]	Simul. Stand Err.[1]	Estim. Param.[2]	Simul. Stand. Err.[2]	Approx. Stand. Err.
50	a_{10}	0.00	0.004	0.326	0.025	0.299	0.283
	a_{11}	0.20	0.190	0.245	0.190	0.217	0.201
	a_{12}	0.05	0.046	0.243	0.043	0.202	0.201
	a_{20}	0.00	-0.004	0.330	0.014	0.300	0.283
	a_{21}	0.20	0.193	0.244	0.191	0.215	0.201
	a_{22}	0.05	0.060	0.239	0.056	0.209	0.201
	a_{30}	0.00	0.036	0.335	0.060	0.314	0.283
	a_{31}	-0.10	-0.099	0.251	-0.107	0.231	0.201
	a_{32}	0.545	0.501	0.246	0.476	0.218	0.201
100	a_{10}	0.00	0.007	0.238	0.002	0.208	0.200
	a_{11}	0.20	0.191	0.162	0.196	0.145	0.141
	a_{12}	0.05	0.045	0.166	0.053	0.149	0.141
	a_{20}	0.00	0.015	0.235	0.002	0.207	0.200
	a_{21}	0.20	0.188	0.167	0.194	0.141	0.141
	a_{22}	0.05	0.039	0.169	0.058	0.143	0.141
	a_{30}	0.00	0.042	0.234	0.031	0.215	0.200
	a_{31}	-0.10	-0.090	0.165	-0.090	0.149	0.141
	a_{32}	0.545	0.527	0.163	0.529	0.147	0.141
500	a_{10}	0.00	0.004	0.102	-0.002	0.091	0.089
	a_{11}	0.20	0.197	0.070	0.198	0.064	0.063
	a_{12}	0.05	0.047	0.074	0.054	0.062	0.063
	a_{20}	0.00	0.003	0.102	-0.004	0.091	0.089
	a_{21}	0.20	0.200	0.067	0.199	0.065	0.063
	a_{22}	0.05	0.049	0.074	0.051	0.065	0.063
	a_{20}	0.00	0.023	0.101	0.029	0.089	0.089
	a_{21}	-0.10	-0.086	0.072	-0.082	0.067	0.063
	a_{22}	0.545	0.546	0.075	0.554	0.063	0.063

Table 3.4: True and estimated coefficients (together with simulated and true standard errors) for model (3.3) when $p = 2$ when the data are generated by four time series according to (3.22). Results are based on 1000 simulations.[1]: Estimation based on T_{ji}, see (3.4). [2]: Estimation based on $\tilde{T}_{ji}$, see (3.20).

Raw Periodogram

N	α	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
50	0.01	0.296	0.227	0.230	0.230	0.205
	0.05	0.504	0.453	0.464	0.435	0.434
	0.10	0.636	0.585	0.615	0.595	0.571
100	0.01	0.734	0.683	0.727	0.663	0.618
	0.05	0.889	0.853	0.879	0.854	0.814
	0.10	0.949	0.913	0.934	0.920	0.881

Smoothed Periodogram

50	0.01	0.561	0.529	0.499	0.410	0.337
	0.05	0.811	0.767	0.739	0.658	0.575
	0.10	0.884	0.860	0.852	0.756	0.728
100	0.01	0.956	0.940	0.943	0.934	0.920
	0.05	0.988	0.988	0.988	0.983	0.987
	0.10	0.997	0.996	0.996	0.995	0.993

Table 3.5: Power of the likelihood ratio test statistic (3.21) when data are generated according to (3.22). Model (3.3) is fitted for different p and results are based on 1000 simulations.

Example 3.7.5. The last example refers to the estimation method when the data are generated by three independent autoregressive conditional heteroscedastic (ARCH) models

$$\begin{aligned} Y_{it} &= \sigma_{it}\epsilon_{it}, \\ \sigma_{it}^2 &= 0.20 + 0.50 Y_{i(t-1)}^2, \end{aligned}$$

for $i = 1, 2, 3$ and ϵ_{it} independent draws from the standard normal distribution. The ARCH models have been introduced by Engle (1982) in order to account for dependency of the second moment on past values of the processes. Further treatment of ARCH– and more general GARCH models–is given by Taniguchi and Kakizawa (2000), for instance. In particular periodogram based inference has been discussed in the recent contribution of Giraitis and Robinson (2001). Here, we apply the proposed method to examine the robustness of the proposed approach to estimation. To fit model (3.3), notice that $G = 3$ and set the spectral density of Y_{3t} as the reference. Table 3.6 (respectively, 3.7) shows the estimated coefficients obtained by maximization of the log likelihood function (3.5) for different N and different order p by using the data in the form (3.4) (respectively, (3.20)). In all cases the estimated parameters are close to zero and their simulated standard error decreases as N increases. The proposed method therefore discovers these time series with the same second order characteristics including the case of ARCH models.

N	p	$\hat{a}_{10}$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{15}$	$\hat{a}_{20}$	$\hat{a}_{21}$	$\hat{a}_{22}$	$\hat{a}_{23}$	$\hat{a}_{24}$	$\hat{a}_{25}$
	2	0.048	0.013	0.002				-0.008	0.010	-0.007			
		(0.607)	(0.322)	(0.261)				(0.617)	(0.306)	(0.266)			
50	3	0.046	0.006	0.002	0.001			0.027	0.001	0.003	0.002		
		(0.592)	(0.314)	(0.266)	(0.258)			(0.584)	(0.329)	(0.258)	(0.249)		
	4	-0.005	-0.002	-0.002	0.001	-0.002		-0.009	0.003	0.010	0.008	0.001	
		(0.575)	(0.317)	(0.275)	(0.262)	(0.239)		(0.576)	(0.317)	(0.273)	(0.259)	(0.242)	
	5	0.003	-0.003	0.001	0.001	0.004	-0.006	-0.004	-0.007	-0.004	0.004	0.007	-0.017
		(0.586)	(0.325)	(0.269)	(0.251)	(0.255)	(0.247)	(0.586)	(0.321)	(0.265)	(0.251)	(0.252)	(0.237)
	2	-0.005	-0.006	-0.003				-0.015	-0.007	-0.006			
		(0.478)	(0.228)	(0.196)				(0.450)	(0.242)	(0.185)			
100	3	-0.020	-0.001	-0.006	-0.005			-0.037	0.008	-0.003	-0.008		
		(0.460)	(0.233)	(0.196)	(0.173)			(0.470)	(0.228)	(0.185)	(0.174)		
	4	-0.007	0.001	-0.004	-0.002	-0.008		-0.009	-0.001	-0.004	0.003	-0.001	
		(0.472)	(0.236)	(0.199)	(0.169)	(0.174)		(0.450)	(0.230)	(0.190)	(0.172)	(0.173)	
	5	-0.006	-0.008	-0.003	0.009	-0.003	0.003	-0.015	-0.009	-0.005	0.003	0.005	0.008
		(0.485)	(0.230)	(0.191)	(0.180)	(0.164)	(0.169)	(0.485)	(0.234)	(0.188)	(0.176)	(0.169)	(0.169)
	2	0.006	-0.002	0.003				-0.006	-0.001	0.001			
		(0.240)	(0.114)	(0.095)				(0.229)	(0.117)	(0.094)			
500	3	-0.001	0.001	-0.001	-0.001			0.006	0.004	-0.001	-0.001		
		(0.239)	(0.116)	(0.097)	(0.081)			(0.238)	(0.114)	(0.092)	(0.082)		
	4	0.001	0.001	-0.001	-0.003	-0.001		0.002	-0.001	0.002	0.004	-0.002	
		(0.253)	(0.118)	(0.097)	(0.082)	(0.079)		(0.251)	(0.120)	(0.092)	(0.083)	(0.079)	
	5	0.009	-0.001	-0.005	-0.003	-0.003	-0.001	-0.001	-0.002	0.001	0.003	0.001	-0.001
		(0.243)	(0.116)	(0.091)	(0.083)	(0.081)	(0.074)	(0.233)	(0.117)	(0.094)	(0.085)	(0.078)	(0.075)

Table 3.6: Estimated coefficients (together with simulated standard errors) for model (3.3) in connection with (3.4) for different values of p when the data are generated by three ARCH(1) processes with the same parameters and for different sample sizes. Results are based on 1000 simulations.

N	p	$\hat{a}_{10}$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{15}$	$\hat{a}_{20}$	$\hat{a}_{21}$	$\hat{a}_{22}$	$\hat{a}_{23}$	$\hat{a}_{24}$	$\hat{a}_{25}$
	2	-0.020	-0.005	-0.004				-0.015	-0.006	-0.002			
		(0.564)	(0.294)	(0.214)				(0.571)	(0.277)	(0.214)			
50	3	-0.016	0.003	-0.012	-0.006			-0.023	0.006	-0.004	-0.010		
		(0.572)	(0.291)	(0.215)	(0.188)			(0.574)	(0.288)	(0.223)	(0.193)		
	4	-0.024	0.005	-0.006	0.010	0.004		-0.008	0.002	0.007	0.001	0.010	
		(0.576)	(0.289)	(0.211)	(0.191)	(0.163)		(0.556)	(0.288)	(0.225)	(0.191)	(0.176)	
	5	-0.012	0.002	-0.004	0.004	-0.003	-0.006	0.006	-0.008	-0.001	0.005	0.004	-0.005
		(0.564)	(0.288)	(0.220)	(0.194)	(0.175)	(0.148)	(0.550)	(0.286)	(0.225)	(0.192)	(0.170)	(0.146)
	2	-0.001	0.004	-0.002				0.000	0.007	-0.005			
		(0.440)	(0.212)	(0.163)				(0.432)	(0.214)	(0.164)			
100	3	0.012	-0.001	-0.006	0.002			0.026	-0.013	-0.005	0.004		
		(0.417)	(0.219)	(0.178)	(0.149)			(0.431)	(0.219)	(0.174)	(0.146)		
	4	0.005	-0.006	0.000	0.002	0.006		-0.009	-0.002	-0.005	-0.006	0.004	
		(0.455)	(0.220)	(0.164)	(0.150)	(0.142)		(0.432)	(0.221)	(0.166)	(0.150)	(0.147)	
	5	0.007	-0.004	-0.003	-0.004	-0.003	0.000	-0.001	-0.007	0.002	-0.001	0.006	0.001
		(0.422)	(0.215)	(0.169)	(0.157)	(0.145)	(0.135)	(0.444)	(0.214)	(0.175)	(0.152)	(0.138)	(0.136)
	2	0.003	0.001	0.005				0.000	0.002	0.002			
		(0.234)	(0.105)	(0.087)				(0.237)	(0.111)	(0.087)			
500	3	0.008	0.007	0	-0.002			0.009	0.001	0	-0.001		
		(0.228)	(0.111)	(0.086)	(0.075)			(0.245)	(0.111)	(0.086)	(0.075)		
	4	0.013	0.001	-0.002	0.003	-0.001		0.006	0.002	-0.002	0.000	0.000	
		(0.231)	(0.110)	(0.085)	(0.074)	(0.071)		(0.231)	(0.111)	(0.083)	(0.072)	(0.071)	
	5	0.013	0.005	0.001	0.001	-0.003	0.001	0.013	0.002	0.000	-0.002	-0.003	-0.001
		(0.236)	(0.113)	(0.084)	(0.073)	(0.068)	(0.066)	(0.239)	(0.111)	(0.085)	(0.074)	(0.069)	(0.065)

Table 3.7: Estimated coefficients (together with simulated standard errors) for model (3.3) in connection with (3.20) for different values of p when the data are generated by three ARCH(1) processes with the same parameters and for different sample sizes. Results are based on 1000 simulations.

3.7.2 Data Analysis

Photometry is an analytical technique, which is often used for determination of chemical species. The mathematical equation that describes the relationship between the instrumental signal (e.g. absorbance, A) and the amount of chemical species in the system under investigation (e.g. concentration C in mol/L) is given by Lambert-Beer Law: $A = \epsilon C\delta$, where ϵ is the molar extinction coefficient (in Lcm/mol) and δ the length of the cell (in cm). In order to control the stability and sensitivity of the photometer at different wavelengths and secure precision and repeatability of the data, statistical analysis of absorbance data obtained under certain conditions is required. The absorbance of a Cu(II) solution has been determined at three different wavelengths, namely, 465 nm (A=0.14), 665 nm (A=0.538) and 865 nm (A=1.14). The photometric measurements have been performed under standard conditions (spectroscopic parameters, temperature, etc.) as a function of time, after warming-up the spectrophotometer for an hour. The time span of each measurement was set at 100 seconds and the recording of the absorbance every 0.1 second, which corresponds to 1000 data points at each of the three different wavelengths. Observations were taken at different times and therefore can be considered as independent time series. Evaluation of the stability of the photometric system as a function of time at the three different wavelengths is tested by means of model (3.3).

Let Y_{1t}, Y_{2t}, Y_{3t} be the three time series of length 1000, as described above and recall that Figure 1.1 shows a plot of the first 150 observations from each time series. First we investigate whether or not a usual autoregressive model, see (2.3), can be fitted to the data. Table 3.8 shows the results of maximum likelihood estimation after fitting autoregressive modes for each of the three time series separately together with the associated standard errors.

Response	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
Y_{1t}	0.031	0.155	0.028	0.131	0.003	0.11	0.027	0.147	0.061	0.092	0.018	0.076
	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.031)	(0.031)
Y_{2t}	0.096	0.005	0.01	0.052	-0.026	0.094	-0.042	0.079	-0.02	0.139		
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)		
Y_{3t}	0											

Table 3.8: AR representations of photometric time series measurements of the absorbance of Cu (II) solution at three different wavelengths.

Table 3.8 shows that an AR process with different orders is sensible model for the first two time series. In fact Y_{1t} (respectively Y_{2t}) fits an autoregressive process of order 12 (10, respectively). The last process is a white noise sequence and therefore the coefficients are all 0. The results are obtained by employing the function `ar.mle` of `R` directly to each time series. The AR models seem to fit the data and further support is provided by the estimated autocorrelation function of the residuals–see plot 3.5. Hence, model (3.3) is applicable for testing whether the three processes posses the same second order structure. To fit the model, we use the spectral density of Y_{3t} as the reference. To apply our inferential procedure, we need to choose the order p. A reasonable way is to choose the parameter p by using the AIC criterion:

$$AIC(p) = -2l(\hat{\theta}) + 2(G-1)(p+1)$$

where, $l(.)$ denotes the log–likelihood function defined by (3.5). The results of the parameter selection are depicted in Figure 3.6 which shows the plot of $AIC(p)$ against p when all three time series are under consideration for different estimation methods. Based on Figure 3.6, the selected order is $p = 2$ and in this case the corresponding estimators are reported in the first six lines of Table 3.9. In addition, Table 3.9 reports p–values for testing the hypotheses that all spectral densities are equal by employing the likelihood ratio test (3.21). The p–values are computed by the chi-square approximation. The hypothesis that all spectral densities are equal is rejected at any given level of significance–p–value equals to 0.

In order to identify differences among the three groups, we perform multiple comparisons as shown in the last 3 entries of Table 3.9. The p–values are again computed by the chi–square approximation and it is seen that their size is relatively small. After a simple Bonferroni adjustment, we conclude that all three time series posses different spectral density functions. These differences are explained as follows; either there are some instabilities in source intensity across time or the sensitivity of the measurement detector is disturbed.

Estimation			
$\hat{\theta}$	Raw Periodogram	Five PDSAE	Seven PDSAE
$\hat{a}_{10}$	-5.161	-5.152	-5.151
$\hat{a}_{11}$	-0.024	-0.014	-0.011
$\hat{a}_{12}$	0.029	0.047	0.047
$\hat{a}_{20}$	-4.996	-4.982	-4.979
$\hat{a}_{21}$	0.100	0.085	0.090
$\hat{a}_{12}$	-0.142	-0.082	-0.077
Testing			
$\lambda_1(\omega) = \lambda_2(\omega) = \lambda_3(\omega)$	0	0	0
$\lambda_1(\omega) = \lambda_2(\omega)$	0.000673	0.000616	0.000446
$\lambda_1(\omega) = \lambda_3(\omega)$	0	0	0
$\lambda_2(\omega) = \lambda_3(\omega)$	0	0	0

Table 3.9: Results of model (3.3) applied to photometry data.

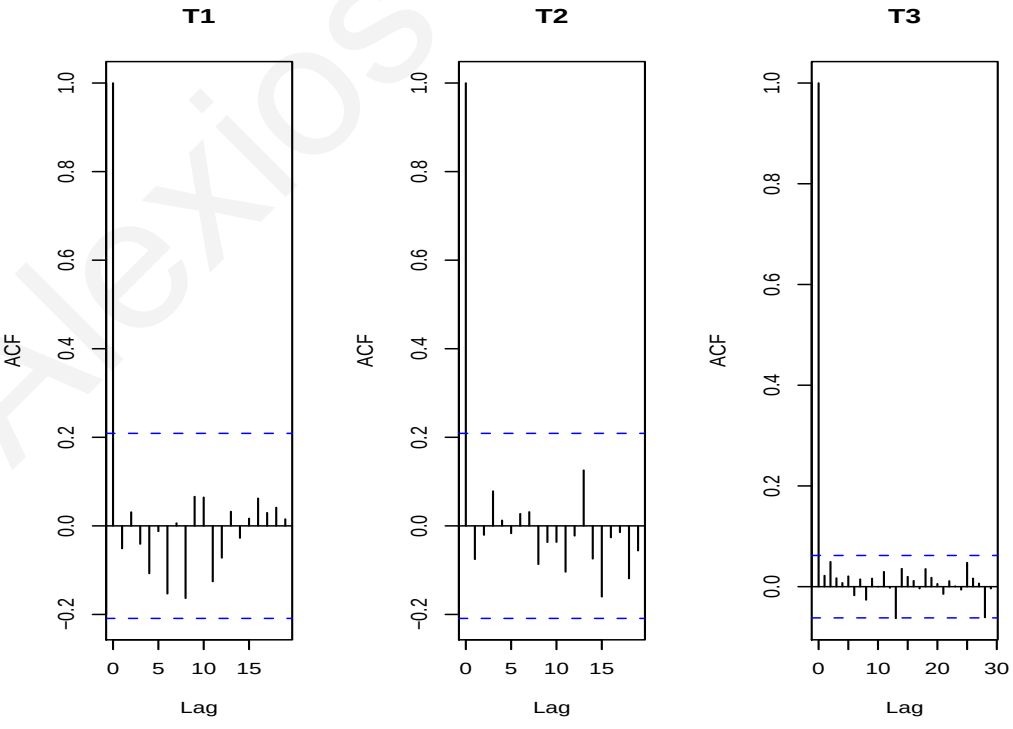

Figure 3.5: Sample autocorrelation function of the residuals after fitting the AR processes to the data according to Table 3.8.

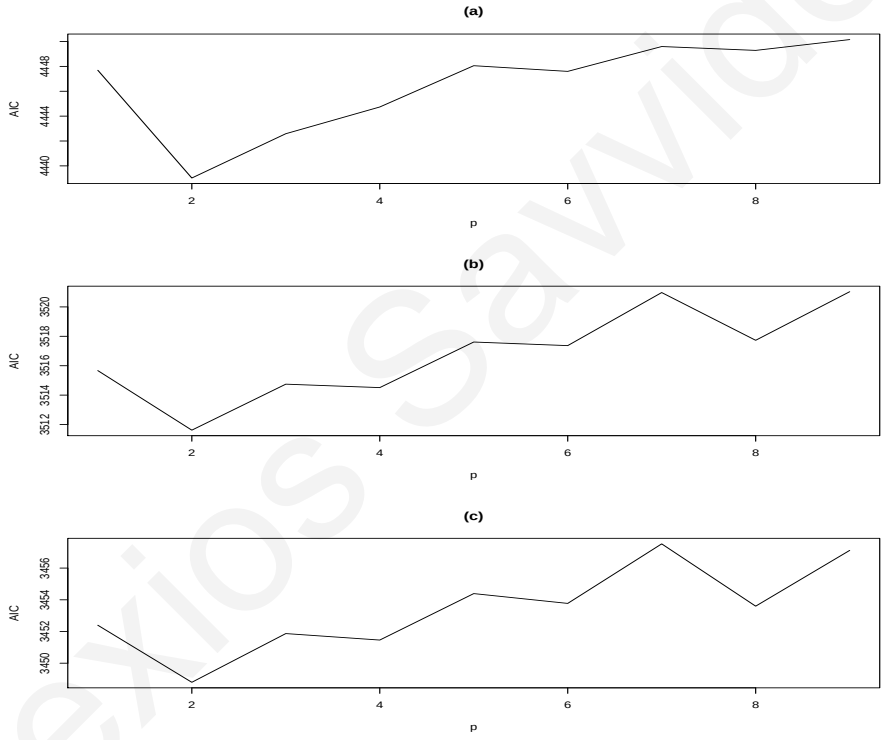

Figure 3.6: Plot of AIC(p) against p for all time series data. (a) Raw Periodogram, (b) Five point discrete spectral average estimator, (c) Seven point discrete spectral average estimator.

Chapter 4

Cepstral Based Clustering of Stationary Time Series

4.1 Introduction

In this section, we employ model (3.3) for clustering of several independent stationary time series. We propose three distinct methods for clustering independent stationary time series and examine their performance by simulations. The results are applied to clustering of biological time series.

4.2 Techniques for Time Series Clustering

In what follows, suppose that $\{Y_{jt}, t = 1, \cdots, N\}$ are G independent stationary time series with absolutely continuous spectral densities, $\lambda_j(\omega)$ for $j = 1, \cdots, G$. Time series clustering methodology is based on the identification of a distance (or similarity) metric between two time series which can be used as a raw input into the well known clustering procedures, see Johnson and Wichern (1992), for instance. In principle, the value of the distances implies whether two time processes share some common features. Then the idea is to cluster time series data according to these particular features. These methods can be divided into two classes, generally speaking. In the first class, there are the methods which are based on time domain characteristics while the second class of methods consists of spectral domain distances, for more see the comprehensive article by Liao (2005) and the text by Shumway and Stoffer (2006, Ch.7). In the following, we briefly review time and spectral domain distance measures.

4.2.1 Time Domain Distances

An initial attack to the problem of obtaining a distance measure between two stationary time series in the time domain among others was suggested by Piccolo (1990) who considers the case of invertible ARMA process. According, if Y_{jt} follows the ARMA process (2.3) and it is invertible, then it can be represented as an AR(∞) process. Thus, the coefficients of the expansion of $\pi_j(B) = \Psi_j^{-1}(B)\Phi_j(B)$, $j = 1, 2, \ldots, G$ can be used to define a distance. It is pointed out that this distance is based on modeling the data by means of (2.3), for more see Piccolo (1990). Without resorting to any parametric specification, Galeano and Peña (2000) suggest the use of the estimated autocorrelation function. More specifically if $\rho_j(h)$ denotes the autocorrelation function of the j'th series–that is $\rho_j(h) = \gamma_j(h)/\gamma_j(0)$, and $\hat{\rho}_j(h)$ stands for the respective estimated autocorrelation function, for $h = 0, 1, \ldots, m$ then the distance between the i and j sequence is given by

$$d_{ACF}(i,j) = \sqrt{(\hat{\rho}_i - \hat{\rho}_j)' W (\hat{\rho}_i - \hat{\rho}_j)}. \tag{4.1}$$

Here m is a truncation point, $\hat{\rho}_i = (\hat{\rho}(1), \ldots, \hat{\rho}(m))'$ and W is some weight matrix. In particular, choosing $W = I$, the identity matrix, yields the Euclidean distance.

4.2.2 Spectral Domain Distances

Spectral domain distances are based on the periodogram ordinates. Their statistical properties are studied by means of the asymptotic independence of the periodogram ordinates, see Section 2.2. Recall the Definition 2.2.4 of the periodogram and denote by $I_j^*(\omega_i) = I_j(\omega_i)/\hat{\gamma}_j(0)$ the normalized periodogram of the j'th series. Then a distance between two time series is defined by

$$d_{NP}(i,j) = \sqrt{\sum_{k=1}^{[N/2]} \left(I_i^*(\omega_k) - I_j^*(\omega_k)\right)^2}, \tag{4.2}$$

following Caiado et al. (2006). Notice that the use of such a distance is useful when interest is focused only on the correlation structure between the two processes. In addition, the fact that the variance of the periodogram ordinates is proportional to the spectrum value at the corresponding frequencies. yields the consideration, of the logarithm of the normalized periodogram, that is

$$d_{LNP}(i,j) = \sqrt{\sum_{k=1}^{[N/2]} \left[\log I_i^*(\omega_k) - \log I_j^*(\omega_k)\right]^2} \tag{4.3}$$

Evidentally, both (4.2) and (4.3) are distances. In fact, it can be shown that, see Caiado et al. (2006)

$$\begin{aligned} d_{NP}(i,j) &= (2\sqrt{N})\sqrt{\sum_{k=1}^{N-1}(\widehat{\rho}_i(k) - \widehat{\rho}_j(k))^2} \\ &= (2\sqrt{N}) d_{ACF}(i,j), \end{aligned}$$

by using expression (4.1) with W equal to the identity matrix. Hence measures (4.2) and (4.1) are equivalent with the choice $W = I$ but their application yields different inferential results when varying the truncation point m in (4.1). Furthermore, distance measures based on absolute values can be defined analogously to (4.2) and (4.3) by

$$d_{ABSNP}(i,j) = \sum_{k=1}^{[N/2]} \left| I_i^*(\omega_k) - I_j^*(\omega_k) \right|, \tag{4.4}$$

and

$$d_{ABSLNP}(i,j) = \sum_{k=1}^{[N/2]} \left| \log I_i^*(\omega_k) - \log I_j^*(\omega_k) \right|. \tag{4.5}$$

A comparison of all the above distances plus the measure introduced by (4.6) and discussed next, is carried out in Section 4.4. An important measure of similarity between two stationary time series is given by the Kullback-Leibler information distance, see Kakizawa et al. (1998) who prove that its time domain form is given by

$$d_{KLTD}(i,j) = \frac{1}{2N}\left(\mathrm{tr}\left(\Gamma_i \Gamma_j^{-1}\right) - \log\frac{|\Gamma_i|}{|\Gamma_j|} - 1\right).$$

Here the matrix Γ_i (respectively Γ_j) is the autocovariance matrix of the i'th series (respectively j'th series). In addition Kakizawa et al. (1998) suggest the use of the more general Chernoff information measure but this will not be the focus of the presentation since these measures depend upon an additional parameter. The previous equation has been derived in time domain and its application to real data results in several computational difficulties, including inversion of large dimensional matrices. It can be shown that the spectral domain analogue of the above formulae reduces to, (see Shumway and Stoffer (2006))

$$d_{KLFD}(i,j) = \frac{1}{N}\sum_{k=1}^{[N/2]}\left[\frac{I_i(\omega_k)}{I_j(\omega_k)} - \log\frac{I_i(\omega_k)}{I_j(\omega_k)} - 1\right].$$

The above quantity is greater or equal to zero, with equality if and only if $I_i(\omega_k) = I_j(\omega_k)$, for all k, almost everywhere. However it is not symmetric and therefore for the purpose of clustering we introduce the so called J–divergence which is defined as

$$\begin{aligned} J(i,j) &= d_{KLFD}(i,j) + d_{KLFD}(j,i) \\ &= \frac{1}{N}\sum_{k=1}^{[N/2]}\left[\frac{I_i(\omega_k)}{I_j(\omega_k)} + \frac{I_j(\omega_k)}{I_i(\omega_k)} - 2\right]. \end{aligned} \tag{4.6}$$

4.3 Clustering Methodology Based on Cepstral Coefficients

4.3.1 Distance Measures Based on Cepstral Coefficients

Recall (2.5) and let θ_{jk} denote the k'th cepstral coefficient of the j'th time series. It is natural to define a spectral distance measure between two or more time series by the following.

Definition 4.3.1. Let $\{Y_{jt}, t = 1, 2, \ldots, N\}$, $j = 1, 2, \ldots, G$ be G independent stationary times series of length N from the linear process (2.1) with $\sum_{u=-\infty}^{\infty} |g_u| < \infty$. In addition suppose that θ_{jr} are the cepstral coefficients of the j'th series for $r = 0, 1, 2, \ldots.$ defined by (2.5). Define the euclidian distance among the cepstral coefficients as

$$d_2(i,j) = \sqrt{\sum_{r=0}^{\infty}(\theta_{ir} - \theta_{jr})^2},$$

for $i, j = 1, 2, \ldots, G$.

The above definition introduces a distance between the log spectral densities of the (i, j) pair in the L^2 space, with the usual convention that we identify functions which differ only on a set of measure zero. It can be generalized further by introducing a suitable weight matrix. We focus however on the above definition to understand what are the properties of this particular distance. Consider the following two simple examples.

Example 4.3.1. Suppose that $G = 2$ and assume that $Y_{jt} = \phi_j Y_{j(t-1)} + \epsilon_{jt}$ are two AR(1) processes with $\mid \phi_j \mid < 1$ and ϵ_{jt} are independent iid sequences with zero mean and variance σ_j^2, $j = 1, 2$. Then

$$d_2^2(1,2) = (\log\sigma_1^2 - \log\sigma_2^2)^2 + \sum_{r=1}^{\infty}\frac{(\phi_1^r - \phi_2^r)^2}{r^2},$$

by means of Lemma 2.5.1. Hence, when ϕ_1 and σ_1^2 are close to ϕ_2 and σ_2^2, respectively, then the cepstral based distance of these process will be close to zero. Notice that the term $(\phi_1^r - \phi_2^r)^2/r^2$ tends to zero sufficiently fast and therefore Definition 4.3.1 is applicable to real data by truncating the infinite sum at some index value. Observations along these lines can be made for the case $G > 2$.

Example 4.3.2. Suppose that $G = 2$ again and assume that $Y_{jt} = \phi_j Y_{j(t-1)} + \epsilon_{jt} - \psi_j \epsilon_{j(t-1)}$ are two ARMA$(1,1)$ processes with $\mid \phi_j \mid < 1$, $\mid \psi_j \mid < 1$ and ϵ_{jt} are independent iid sequences with zero mean and variance σ_j^2, $j = 1, 2$. Then

$$d_2^2(1,2) = (\log \sigma_1^2 - \log \sigma_2^2)^2 + \sum_{r=1}^{\infty} \frac{((\phi_1^r - \phi_2^r) - (\psi_1^r - \psi_2^r))^2}{r^2},$$

using again (2.5.1). Similar remarks as in the case of the AR(1) example can be made.

Both of the above examples make the point that the quantity $d_2(i,j)$ can be approximated by

$$\tilde{d}_2(i,j) = \sqrt{\sum_{r=0}^{p} (\theta_{ir} - \theta_{jr})^2}, \tag{4.7}$$

where p is a fixed number chosen by the data analyst. When considering the log spectrum estimation a small value of p usually suffices since the corresponding cepstrum coefficients tend to zero exponentially fast. For fixed p, expression (4.7) introduces a distance. Instead of using a different analysis of each given time series we resort to the proposed model (3.3) which yields the following result:

$$\begin{aligned}
\tilde{d}_2^2(i,j) &= \sum_{r=0}^{p} (\theta_{ir} - \theta_{jr})^2 \\
&= \sum_{r=0}^{p} \left(\int_0^1 \log \lambda_i(\omega) \cos(2\pi r\omega) d\omega - \int_0^1 \log \lambda_j(\omega) \cos(2\pi r\omega) d\omega \right)^2 \\
&= \sum_{r=0}^{p} (a_{ir} - a_{jr})^2
\end{aligned}$$

for $i, j = 1, \cdots, G-1$, $i \neq j$, and

$$\begin{aligned}
\tilde{d}_2^2(i,G) &= \sum_{r=0}^{p} (\theta_{ir} - \theta_{Gr})^2 \\
&= \sum_{r=0}^{p} \left(\int_0^1 \log \lambda_i(\omega) \cos(2\pi r\omega) d\omega - \int_0^1 \log \lambda_G(\omega) \cos(2\pi r\omega) d\omega \right)^2 \\
&= \sum_{r=0}^{p} a_{ir}^2
\end{aligned}$$

when $j = G$. The above two expressions can be written compactly as

$$\tilde{d}_2^2(i,j) = \sum_{r=0}^{p}(a_{ir} - a_{jr})^2, \tag{4.8}$$

for $i, j = 1, 2, \ldots, G$, $i \neq j$ and $a_{Gr} = 0$ for all r. To carry out the above computation, it is not necessary to have the same value of p for all time series. It is implicitly assumed that for applying model (3.3) we choose the same value p for all time series. Furthermore, the definition of model (3.3) implies that the ratio of the i'th and j'th spectral densities remains the same, irrespective of the total number of spectral densities, that is

$$\log \frac{\lambda_i(\omega)}{\lambda_j(\omega)} = (a_i - a_j)^T Z(\omega).$$

This shows that distance defined by (4.7) is not affected by the reference density since it is a sum of the squared contrasts. In all the examples we use the spectral density of the last simulated time series as reference for fitting the proposed model.

In a related approach Kalpakis et al. (2001) have studied the euclidian distance between cepstral coefficients of two or more time series in the context of ARMA processes. However their approach is based on identifying the estimators of cepstral coefficients by means of Lemma 2.5.1 for each time series separately and then form an estimate for (4.7). In other words, they assume that each observed time series follows an ARMA model of the form (2.3) and then they estimate the unknown coefficients $\phi_i, i = 1, 2, \ldots, p$ and $\psi_j, j = 1, 2, \ldots, q$. The estimated model is employed to produce estimates of the cepstral coefficients which in turn are used for obtaining the euclidian distance between them. A natural question that arises is whether each time series follows an ARMA model. If not, then the inferential results obtained by Kalpakis et al. (2001) will not be in general reliable. Our approach is quite different since we do include more general process than ARMA models and when estimation is performed–a topic that is going to be discussed next–model (3.3) is employed which is semiparametric in the sense that the ratio of two or more spectral density functions is related by a parametric form. In addition, we introduce the following dissimilarity measure which is based on the absolute value of the difference between the cepstral coefficients. It is defined by the following:

Definition 4.3.2. Let $\{Y_{jt}, t = 1, 2, \ldots, N\}$, $j = 1, 2, \ldots, G$ be G independent stationary times series of length N and suppose that the assumptions of Definition 4.3.1 hold true. In

addition suppose that θ_{jr} are the cepstral coefficients of the j'th series for $r = 0, 1, 2, ...$ Define the absolute value distance among the cepstral coefficients as

$$d_1^2(i,j) = \sum_{r=0}^{\infty} \mid \theta_{ir} - \theta_{jr} \mid$$

for $i, j = 1, 2, \ldots, G$.

Under the spectral density ratio model (3.3) and given a truncation point p, the above quantity is estimated by

$$\tilde{d}_1^2(i,j) = \sum_{r=0}^{p} \mid a_{ir} - a_{jr} \mid, \tag{4.9}$$

Both (4.8) and (4.9) are employed after estimation in what follows and their performance is examined in both real and simulated data. More specifically, given the maximum likelihood estimator vector $\hat{\theta}$ for model (3.3) we estimate (4.8) by plugging in the corresponding estimators, that is

$$d_{EUCLCEP}^2(i,j) = \sum_{r=0}^{p} (\hat{a}_{ir} - \hat{a}_{jr})^2. \tag{4.10}$$

Similarly for (4.9) we use

$$d_{ABSCEP}^2(i,j) = \sum_{r=0}^{p} \mid \hat{a}_{ir} - \hat{a}_{jr} \mid . \tag{4.11}$$

These distances will be used in the sequel. In addition, we describe another distance measure which is based on the p-value of the test statistic for the hypothesis $H_o : \tilde{d}_2^2(i,j) = 0$, for $i, j = 1, 2, \ldots, G$, see Alonso and Maharaj (2006).

4.3.2 Distance measures based on p-values

Consider testing of the hypotheses $H_o : \tilde{d}_2^2(i,j) = 0$, for $i, j = 1, 2, \ldots, G-1$. Equivalently the hypotheses state that all process have identical cepstral coefficients up to order p. We approximate the asymptotic distribution of (4.10) to compute a p-value for testing the above hypotheses. Recall Theorem 3.4.2 and use the notation

$$F^{1/2} \left(\hat{\theta} - \theta \right) \Rightarrow \mathcal{N}_{(G-1)(p+1)} \left(0, I \right), \tag{4.12}$$

where $F = ND_0/4\pi$. As a first step we will find the distribution of $\theta_i - \theta_j$ for $i, j = 1, \ldots, G-1$. Let K be a diagonal matrix of dimensions $(p+1) \times (G-1)(p+1)$ of the form

$$K = [O_{p+1} \text{ everywhere}, I_{p+1} \text{ at the } i\text{-place}, -I_{p+1} \text{ at the } j\text{-place}]$$

where I_{p+1}, O_{p+1} are the identity and the zero matrix of dimension $(p+1)$ respectively. It is $K\theta = \theta_i - \theta_j$.

Proposition 4.3.1.

$$\left(KF^{-1}K^T\right)^{-1/2} K\left(\hat{\theta}-\theta\right) \overset{D}{\Rightarrow} \mathcal{N}_{(p+1)}(0, I) \tag{4.13}$$

Proof: Let $W = F^{1/2}\left(\hat{\theta}-\theta\right)$ and $X = \left(KF^{-1}K^T\right)^{-1/2} KF^{-1/2}$,. From (4.12) we have that $W \overset{D}{\Rightarrow} \mathcal{N}_{(G-1)(p+1)}(0, I)$ Therefore

$$\begin{aligned}\left(KF^{-1}K^T\right)^{-1/2} K\left(\hat{\theta}-\theta\right) &= \left(KF^{-1}K^T\right)^{-1/2} KF^{-1/2}F^{1/2}\left(\hat{\theta}-\theta\right)\\ &= \left(KF^{-1}K^T\right)^{-1/2} KF^{-1/2}W\\ &= XW,\end{aligned}$$

and

$$Var(XW) = XX^T = \left(KF^{-1}K^T\right)^{-1/2} KF^{-1/2}F^{-1/2}K^T\left(KF^{-1}K^T\right)^{-1/2} = I_{(p+1)}$$

Define now $\Sigma_0 = KF^{-1}K^T$ and $\delta = \Sigma_0^{-1/2}K\hat{\theta}$. Then

$$d^2_{EUCLCEP} = \delta^T\Sigma_0\delta.$$

From Proposition 4.3.1 and under the hypotheses H_0 since $K\theta = 0$, we have that $\delta \overset{D}{\Rightarrow} \mathcal{N}_{p+1}(0, I)$ As result, the asymptotic distribution of $d^2_{EUCLCEP}$ is derived applying known results on quadratic forms in standard Normal random variables, since

$$d^2_{EUCLCEP} = \delta^T\Sigma_0\delta = \sum_j \lambda_j\chi^2_{g_j},$$

where $\lambda_j, j \leq p+1$ are the non zero eigenvalues of Σ_0 and $\chi^2_{g_j}$ are independent chi -square random variables with g_j degrees of freedom given by the multiplicity of each eigenvalue (usually, $g_j \equiv 1,\ \forall j$). To facilitate the use of the distance an approximation of the distribution of a linear combination of chi-squared random variables is needed, see Corduas and Piccolo (2008)). In particular, we consider the approximation:

$$d^2_{EUCLCEP} \sim a\chi^2_v + b$$

where a, b, v are determined by the method of moments:

$$a = t_3/t_3, \qquad b = t_1 - (t_2^2/t_3), \qquad v = t_2^3/t_3^2,$$

with $t_k = \sum_j \lambda_j^k = tr(\Sigma_0^k), k = 1, 2, 3$. The distance measure that is introduced is defined by the following, see also Alonso and Maharaj (2006).

Definition 4.3.3. Let $\{Y_{jt}, t = 1, 2, \ldots, N\}$, $j = 1, 2, \ldots, G$ be G independent stationary times series of length N and suppose that representation (2.5) holds true, with θ_{jk} being the cepstral coefficients of the j'th series for $k = 0, 1, 2, \ldots$. Let $p(i, j)$ be the p-values for testing the hypotheses $H_o : \tilde{d}_2^2(i, j) = 0$, for $i, j = 1, 2, \ldots, G-1$ based on the asymptotic approximation of the quantity (4.10). Define the d_{PVAL} distance as

$$d_{PVAL}(i, j) = 1 - p(i, j), \qquad i, j = 1, 2, \ldots, G, \tag{4.14}$$

after suitable adjustment for taking into account multiple comparisons.

Remark 4.3.1. In the above definition the vector of the p-values is adjusted for taking into account multiple comparisons by using either a bonferroni correction method when the number of hypotheses to be tested is small. When the number of comparisons is large we can employ the method of Benjamini and Hochberg (1995).

4.3.3 Applying the cepstral distance measure for clustering

To apply the cepstral distance measure to real data an estimator of (4.7) is needed. The method which is going to be used is based on maximum likelihood estimation of the vector of coefficients θ_j for $j = 1, 2, \ldots, G-1$. We use (4.10), (4.11) and (4.14) as described above.

Based on the above distances, clustering of the time series is carried out by the method diana which yields a computing a divisive hierarchy, whereas most other procedures for hierarchical clustering is agglomerative. The complete algorithm consists of $G-2$ successive splits. In each step, we select the cluster C with the largest diameter, where

$$diam(C) := \max_{i,j \in C} d(i, j) \tag{4.15}$$

Assuming $diam(C) > 0$ we then split up C into two clusters A and B , according to a variant of the method Macnaughton Smith et al. (1965). Below we describe in pseudocode how such a split is performed. At first $A := C$ and $B := \emptyset$, and then

1) Move one object from A to B. For each object $i \in A$ we calculate $a(i)$, the average dissimilarity of i to all other objects of A.

$$a(i) := \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \tag{4.16}$$

The object m of A for which $a(m)$ is the largest, is moved to B:

$$A := A \setminus \{m\}, B := \{m\}. \tag{4.17}$$

2) Move other object from A to B, which is called the "splinter group". if $|A| = 1$, stop. Otherwise calculate $a(i)$ for all $i \in A$, and the average dissimilarity of i to all objects of B, denoted as $d(i, B)$

$$d(i, B) := \frac{1}{|B|} \sum_{j \in B} d(i, j) \tag{4.18}$$

Select the object $h \in A$ for which $a(h) - d(h, B) = \max_{i \in A}(a(i) - d(i, B))$. If $a(h) - d(h, B) > 0 \Rightarrow$ move h from A to B, and go to step 2.

If $a(h) - d(h, B) \leq 0 \Rightarrow$ the process stops. Keep A and B as they are now.

In brief the diana-algorithm constructs a hierarchy of clusterings, starting with one large cluster containing all $G - 1$ observations. Clusters are divided until each cluster contains only a single observation. At each stage, the cluster with the largest diameter is selected. (The diameter of a cluster is the largest dissimilarity between any two of its observations.) To divide the selected cluster, the algorithm first looks for its most disparate observation (i.e., which has the largest average dissimilarity to the other observations of the selected cluster). This observation initiates the "splinter group". In subsequent steps, the algorithm reassigns observations that are closer to the "splinter group" than to the "old party". The result is a division of the selected cluster into two new clusters.

The function diana also provides the divisive coefficient which measures the clustering structure of the data set. This coefficient is obtained as follows: for each object i, denote by $d(i)$ the diameter of the last cluster to which it belongs (before being split off as a single object), divided by the diameter of the whole data set. The divisive coefficient (DC) is then defined as the average of all $d(i)$.

In summary, we have the following algorithm for cepstral based clustering of time series:

1. Given G independent stationary time series, fit model (3.3) to obtain an estimator of the parameter θ based on the maximum likelihood method that was described above.

2. Based on the estimated coefficients $\hat{\theta}$ compute (4.10), (4.11) and (4.14) as described above.

3. Form a distance matrix with the (i, j) element be given by (4.10), (4.11) and (4.14) and use the diana clustering method to classify the data.

Other clustering methods can be used as well but our focus will be on the method described above, for more see Hastie et al. (2001).

4.4 Simulations

We apply the proposed method to simulated data. In particular, comparison of spectral domain clustering is implemented by means of all the metrics that were discussed in Sections 4.2 and 4.3. We consider examples that are non standard including simulated data that arises from non linear processes. All the calculations are carried out by the statistical language R, R Development Core Team (2004), and a program is given in the Appendix.

To compare the performance of the various clustering methods a cluster similarity measure is computed throughout the simulations. It is based on known ground truth, that is it can be defined when the total number of clusters is known. More specifically, following Liao (2005) suppose that $F = F_1, F_2, \ldots, F_k$ are the sets of k ground truth clusters and let $C = C_1, C_2, \ldots, C_k$ be those sets which are obtained by a clustering method under evaluation. The cluster similarity measure is defined as

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

where

$$\mathrm{Sim}(F_i, C_j) = \frac{2 \mid F_i \bigcap C_j \mid}{\mid F_i \mid + \mid C_j \mid},$$

where $\mid . \mid$ denotes the cardinality of a set. Obviously the quantity Sim takes values between 0 and 1, where 0 corresponds to the case that F and G are exhaustively disjoint and 1 corresponds to the case of perfect clustering. The simulations are carried out one thousands times. In each run, we evaluate the distances introduced by the J–divergence (4.6), the squared distances (4.2), (4.3) and (4.10), and the absolute distance (4.4), (4.5) and (4.11). In addition clustering based on the distance (4.14) is calculated.

For each distance a dissimilarity matrix is calculated and is given as input to diana. Then the similarity index is calculated and the results are averaged out for the whole process. We simulate one thousand time series replications. The sample sizes were taken equal to 50,100 and 500. For fitting model (3.3) the very last spectral density–that is the spectral density which corresponds to the last simulated time series–was kept as reference, see Section 3.3. We give two ways of estimating all distances, namely maximum likelihood estimation based at the raw periodogram and on smoothed periodogram, see Hitchcock et al. (2006). The smoothed periodogram is obtained by smoothing the result with a series of modified Daniell smoothers (moving averages giving half weight to the end values). Each time, a different order p is used to fit the model and consequently compute (4.10), (4.11) and (4.14). When changing the order all the other distances are recalculated again and therefore the resulting output is reported for all these runs. However we will see that there is no much variation of the results as it should be expected. When simulating from a time series model, the coefficients are chosen away from the boundary of stationarity/invertibility region of ARMA models and they are drawn from a uniform distribution with small variance instead of leaving the true parameters as fixed constants. This is a more realistic scenario in many applications and we use the notation $X \sim U(a, b)$ to denote that the random variable X follows the uniform distribution in the interval (a, b). Note that if we want to check the performance of the proposed clustering method using ARMA models 'randomly' or their coefficients near the boundary, we can use the method of Jones (1987).

Example 4.4.1. Consider three clusters of a total of thirteen time series according to the following specification

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \epsilon_{it}$$

where $\phi_{i1} \sim U(-0.61, -0.59)$ for $i = 1, 2, 3, 4$,

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \phi_{i2} Y_{i(t-2)} + \epsilon_{it},$$

where $\phi_{i1} \sim U(-0.61, -0.59)$, $\phi_{i2} \sim U(0.29, 0.31)$ for $i = 5, 6, 7, 8$, and

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \epsilon_{it} - \psi_{i1} \epsilon_{i(t-1)}$$

where $\phi_{i1} \sim U(-0.51, -0.49)$ and $\psi_{i1} \sim U(-0.18, -0.16)$ for $i = 9, 10, \ldots, 13$. The error sequences are independent and identically distributed normal random variables with standard deviation equal to 1 for $i = 1, 2, 3, 4$, 2 for $i = 5, 6, 7, 8$ and 0.5 for $i = 9, 10, 11, 12, 13$. In this case, we have four AR(1) processes, four AR(2) processes and 5 ARMA(1,1) process. Table 4.1 shows the cluster similarity index together with its standard deviation (in parentheses) for different metrics. The results clearly illustrate that clustering based on cepstral coefficients by means of the proposed model outperform all the other methods, even for relatively small sample sizes. In addition the order of the model fitted to the data does not influence the final results. Distances based on the logarithm of periodogram, that is (4.3) and (4.5), appear to give better results when compared to the other distances. The same results appear in Table 4.2 for the smoothed estimator. We notice that the proposed distance is not affected a lot by smoothing. However the other distances are improving.

Example 4.4.2. Consider two clusters of a total of thirteen time series according to the following specification

$$\begin{aligned} Y_{it} &= \cos(\omega_i t) + \sin(\omega_i t) + \epsilon_{it} \quad i = 1, 2, \ldots, 6 \\ Y_{it} &= \epsilon_{it} \quad i = 7, 8, \ldots, 13 \end{aligned}$$

where $\omega_i \sim U(0.97, 1.03)$. Here the error sequences are generated again by the normal distribution with standard deviation equal to 1.3 for $i = 1, 2, \ldots, 6$, and 0.7 for the rest. This is a case where the data are divided into two groups, namely one group is periodic and the other group has a flat spectrum. This simulation shows that model (3.3) is robust even when the spectral density is not continuous. Tables 4.3 and 4.4 report similar results to those discussed earlier.

Example 4.4.3. Consider twelve time series according to the following specification

$$\begin{aligned} Y_{it} &= \sigma_{it}\epsilon_{it} \\ \sigma_{it}^2 &= \omega_i + \phi_i Y_{i(t-1)}^2 \end{aligned}$$

where $\phi_i \sim U(0.69, 0.71)$ for $i = 1, 2, 3, 4$,

$$\begin{aligned} Y_{it} &= \sigma_{it}\epsilon_{it}, \\ \sigma_{it}^2 &= \omega_i + \phi_i Y_{i(t-1)}^2 + \theta_i \sigma_{i(t-1)}^2, \end{aligned}$$

where $\phi_i \sim U(0.39, 0.41)$, $\theta_i \sim U(0.49, 0.51)$ for $i = 5, 6, 7, 8$,

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \epsilon_{it},$$

where $\phi_{i1} \sim U(-0.61, -0.59)$ for $i = 9, 10, \ldots, 13$. The noise sequence ϵ_{it} is generated from the standard normal distribution and $\omega_i = 1e-6$. For this example, we have four ARCH(1) processes, four GARCH(1,1) processes and 5 AR(1) process. Tables 4.5 and 4.6 show that our method perform well even when the data contain non linear time series.

Example 4.4.4. The final example is about a more complicated situation where there are six clusters and the number of observed time series is twenty six. These data have been generated according to the following:

$$Y_{it} = \cos(\omega_i t) + \sin(\omega_i t) + \epsilon_{it} \quad i = 1, 2, 3, 4$$

where $\omega_i \sim U(0.19, 0.21)$, for $i = 1, 2, 3, 4$,

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \epsilon_{it} - \psi_{i1} \epsilon_{i(t-1)},$$

where $\phi_{i1} \sim U(0.07, 0.09)$, $\psi_{i1} \sim U(-0.01, 0.01)$ for $i = 5, 6, 7, 8$,

$$Y_{it} \sim \text{EXP}(2),$$

for $i = 9, 10, 11, 12$, where the above notation mean that the data follow model (2.5) with $\theta_{i0} \sim U(0.09, 0.11)$, $\theta_{i1} \sim U(-0.20, -0.18)$ and $\theta_{i2} \sim U(0.17, 0.19)$,

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \phi_{i2} Y_{i(t-2)} + \epsilon_{it} - \psi_{i1} \epsilon_{i(t-1)},$$

where $\phi_{i1} \sim U(-0.71, -0.69)$, $\phi_{i2} \sim U(0.19, 0.21)$, $\psi_{i1} \sim U(0.16, 0.18)$ for $i = 13, 14, 15, 16$,

$$Y_{it} = \epsilon_{it}$$

for $i = 17, 18, 19, 20$ and

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \phi_{i2} Y_{i(t-2)} + \epsilon_{it} - \psi_{i1} \epsilon_{i(t-1)},$$

with $\phi_{i1} \sim U(0.29, 0.31)$ $\phi_{i2} \sim U(0.19, 0.21)$, and $\psi_{i1} \sim U(-0.01, 0.01)$ for $i = 21, 22, \ldots, 26$. The noise sequence are independent and identically distributed normal variables with standard deviation equal to 2 for $i = 1, 2, 3, 4$, 1.5 for $i = 5, 6, 7, 8$, 0.7 for $i = 13, 14, 15, 16$, 1.1 for $i = 17, 18, 19, 20$ and 1.3 for $i = 21, 22, \ldots, 26$. Tables 4.7 and 4.8 show again that our method is more reliable even in this case where there are various types of processes under consideration.

Table 4.1: Simulation results for the cluster similarity measure of Example 4.4.1.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.616	0.620	0.642	0.976	0.645	0.664	0.981	0.994
		(0.097)	(0.076)	(0.096)	(0.054)	(0.085)	(0.097)	(0.05)	(0.034)
50	3	0.617	0.625	0.641	0.975	0.646	0.664	0.977	0.993
		(0.094)	(0.076)	(0.093)	(0.058)	(0.084)	(0.089)	(0.057)	(0.033)
	4	0.621	0.624	0.649	0.974	0.646	0.671	0.978	0.992
		(0.092)	(0.074)	(0.095)	(0.057)	(0.085)	(0.09)	(0.053)	(0.035)
	5	0.615	0.625	0.643	0.973	0.647	0.665	0.971	0.987
		(0.094)	(0.076)	(0.092)	(0.06)	(0.085)	(0.092)	(0.066)	(0.041)
	2	0.650	0.634	0.703	0.995	0.685	0.726	0.997	0.990
		(0.1)	(0.075)	(0.091)	(0.023)	(0.075)	(0.082)	(0.019)	(0.052)
100	3	0.648	0.637	0.702	0.996	0.686	0.723	0.997	0.989
		(0.093)	(0.075)	(0.088)	(0.023)	(0.075)	(0.084)	(0.022)	(0.054)
	4	0.648	0.637	0.703	0.995	0.686	0.726	0.997	0.991
		(0.101)	(0.075)	(0.091)	(0.024)	(0.074)	(0.089)	(0.018)	(0.048)
	5	0.647	0.634	0.700	0.995	0.684	0.723	0.996	0.993
		(0.095)	(0.075)	(0.091)	(0.023)	(0.081)	(0.088)	(0.02)	(0.044)
	2	0.706	0.616	0.807	1	0.744	0.840	1	0.815
		(0.098)	(0.077)	(0.094)	(0)	(0.083)	(0.11)	(0)	(0.110)
500	3	0.708	0.616	0.810	1	0.745	0.838	1	0.790
		(0.095)	(0.079)	(0.098)	(0)	(0.083)	(0.11)	(0)	(0.114)
	4	0.710	0.614	0.809	1	0.744	0.840	1	0.780
		(0.095)	(0.078)	(0.095)	(0)	(0.083)	(0.108)	(0)	(0.111)
	5	0.710	0.616	0.805	1	0.747	0.839	1	0.770
		(0.092)	(0.077)	(0.094)	(0)	(0.085)	(0.106)	(0)	(0.113)

Table 4.2: Simulation results for the cluster similarity measure of Example 4.4.1, based on five point discrete spectral estimator.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.724	0.717	0.72	0.983	0.733	0.721	0.988	0.992
		(0.099)	(0.09)	(0.098)	(0.047)	(0.097)	(0.097)	(0.042)	(0.045)
50	3	0.721	0.713	0.721	0.983	0.729	0.72	0.987	0.994
		(0.101)	(0.092)	(0.1)	(0.047)	(0.095)	(0.099)	(0.041)	(0.035)
	4	0.723	0.714	0.719	0.983	0.73	0.723	0.986	0.996
		(0.103)	(0.09)	(0.099)	(0.047)	(0.098)	(0.099)	(0.048)	(0.032)
	5	0.719	0.714	0.717	0.984	0.729	0.722	0.988	0.998
		(0.102)	(0.092)	(0.102)	(0.047)	(0.097)	(0.101)	(0.041)	(0.021)
	2	0.793	0.722	0.79	0.996	0.779	0.792	0.997	0.957
		(0.109)	(0.073)	(0.108)	(0.021)	(0.1)	(0.109)	(0.018)	(0.095)
100	3	0.789	0.722	0.787	0.996	0.779	0.786	0.998	0.96
		(0.107)	(0.08)	(0.106)	(0.021)	(0.098)	(0.108)	(0.017)	(0.091)
	4	0.789	0.724	0.786	0.996	0.782	0.788	0.998	0.967
		(0.111)	(0.078)	(0.11)	(0.021)	(0.102)	(0.109)	(0.014)	(0.084)
	5	0.793	0.725	0.79	0.996	0.782	0.79	0.997	0.964
		(0.113)	(0.078)	(0.113)	(0.018)	(0.101)	(0.112)	(0.018)	(0.086)
	2	0.953	0.703	0.952	1	0.861	0.948	1	0.759
		(0.101)	(0.04)	(0.103)	(0)	(0.132)	(0.107)	(0)	(0.103)
500	3	0.948	0.704	0.946	1	0.856	0.941	1	0.746
		(0.105)	(0.04)	(0.108)	(0)	(0.135)	(0.112)	(0)	(0.103)
	4	0.949	0.704	0.949	1	0.857	0.946	1	0.737
		(0.103)	(0.038)	(0.105)	(0)	(0.136)	(0.107)	(0)	(0.102)
	5	0.955	0.704	0.954	1	0.861	0.951	1	0.735
		(0.097)	(0.038)	(0.1)	(0)	(0.134)	(0.103)	(0)	(0.1)

Table 4.3: Simulation results for the cluster similarity measure of Example 4.4.2.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.653	0.875	0.670	0.999	0.878	0.695	0.996	0.999
		(0.066)	(0.106)	(0.083)	(0.011)	(0.114)	(0.105)	(0.031)	(0.010)
50	3	0.657	0.875	0.668	0.999	0.882	0.695	0.996	0.997
		(0.077)	(0.105)	(0.079)	(0.015)	(0.114)	(0.107)	(0.029)	(0.022)
	4	0.653	0.88	0.664	0.998	0.881	0.693	0.990	0.998
		(0.076)	(0.106)	(0.076)	(0.012)	(0.115)	(0.107)	(0.052)	(0.014)
	5	0.656	0.884	0.668	0.998	0.883	0.695	0.985	0.997
		(0.075)	(0.105)	(0.08)	(0.02)	(0.117)	(0.108)	(0.065)	(0.017)
	2	0.660	0.809	0.686	1	0.876	0.707	1	0.997
		(0.075)	(0.108)	(0.097)	(0)	(0.102)	(0.112)	(0)	(0.029)
100	3	0.656	0.804	0.674	1	0.87	0.704	1	0.993
		(0.074)	(0.112)	(0.085)	(0)	(0.108)	(0.109)	(0.011)	(0.047)
	4	0.656	0.805	0.679	1	0.874	0.705	0.999	0.996
		(0.073)	(0.109)	(0.091)	(0)	(0.104)	(0.112)	(0.012)	(0.034)
	5	0.658	0.815	0.679	1	0.875	0.706	1	0.997
		(0.076)	(0.108)	(0.092)	(0)	(0.108)	(0.112)	(0.003)	(0.028)
	2	0.669	0.667	0.765	1	0.795	0.823	1	0.849
		(0.088)	(0.062)	(0.147)	(0)	(0.135)	(0.148)	(0)	(0.138)
500	3	0.667	0.668	0.763	1	0.798	0.819	1	0.799
		(0.084)	(0.059)	(0.144)	(0)	(0.133)	(0.151)	(0)	(0.137)
	4	0.669	0.664	0.76	1	0.79	0.815	1	0.781
		(0.088)	(0.057)	(0.146)	(0)	(0.134)	(0.15)	(0)	(0.137)
	5	0.670	0.670	0.767	1	0.792	0.825	1	0.767
		(0.086)	(0.062)	(0.148)	(0)	(0.133)	(0.147)	(0)	(0.131)

Table 4.4: Simulation results for the cluster similarity measure of Example 4.4.2, based on five point discrete spectral estimator.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.803	0.872	0.805	1	0.87	0.821	0.999	0.997
		(0.126)	(0.106)	(0.128)	(0.005)	(0.11)	(0.126)	(0.017)	(0.03)
50	3	0.804	0.868	0.804	0.999	0.864	0.819	0.999	0.998
		(0.126)	(0.104)	(0.126)	(0.007)	(0.108)	(0.125)	(0.009)	(0.019)
	4	0.812	0.874	0.812	0.999	0.869	0.83	0.998	0.999
		(0.132)	(0.111)	(0.132)	(0.008)	(0.116)	(0.128)	(0.016)	(0.014)
	5	0.797	0.869	0.8	0.999	0.864	0.818	0.999	0.999
		(0.13)	(0.11)	(0.13)	(0.007)	(0.111)	(0.126)	(0.014)	(0.014)
	2	0.86	0.862	0.858	1	0.888	0.855	1	0.971
		(0.117)	(0.106)	(0.118)	(0.003)	(0.103)	(0.117)	(0.003)	(0.09)
100	3	0.854	0.859	0.851	1	0.883	0.85	1	0.965
		(0.118)	(0.106)	(0.12)	(0)	(0.105)	(0.118)	(0.011)	(0.097)
	4	0.859	0.862	0.856	1	0.885	0.852	1	0.964
		(0.117)	(0.107)	(0.118)	(0)	(0.102)	(0.118)	(0)	(0.101)
	5	0.865	0.865	0.862	1	0.891	0.864	1	0.972
		(0.116)	(0.106)	(0.118)	(0)	(0.101)	(0.115)	(0)	(0.09)
	2	0.976	0.889	0.976	1	0.968	0.97	1	0.774
		(0.057)	(0.134)	(0.052)	(0)	(0.069)	(0.055)	(0)	(0.129)
500	3	0.973	0.893	0.972	1	0.964	0.967	1	0.745
		(0.058)	(0.128)	(0.053)	(0)	(0.072)	(0.058)	(0)	(0.123)
	4	0.975	0.889	0.974	1	0.967	0.968	1	0.734
		(0.054)	(0.132)	(0.054)	(0)	(0.068)	(0.06)	(0)	(0.121)
	5	0.973	0.885	0.97	1	0.965	0.967	1	0.732
		(0.059)	(0.132)	(0.062)	(0)	(0.07)	(0.06)	(0)	(0.123)

Table 4.5: Simulation results for the cluster similarity measure of Example 4.4.3.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.587	0.624	0.616	0.846	0.664	0.651	0.837	0.844
		(0.092)	(0.08)	(0.097)	(0.088)	(0.078)	(0.091)	(0.085)	(0.088)
50	3	0.595	0.630	0.618	0.844	0.667	0.651	0.835	0.848
		(0.091)	(0.078)	(0.096)	(0.087)	(0.077)	(0.09)	(0.087)	(0.088)
	4	0.593	0.628	0.627	0.845	0.666	0.655	0.832	0.846
		(0.094)	(0.08)	(0.096)	(0.085)	(0.077)	(0.092)	(0.083)	(0.087)
	5	0.597	0.628	0.622	0.843	0.669	0.656	0.831	0.853
		(0.093)	(0.08)	(0.095)	(0.082)	(0.079)	(0.092)	(0.08)	(0.090)
	2	0.615	0.634	0.678	0.866	0.710	0.708	0.858	0.831
		(0.1)	(0.078)	(0.096)	(0.088)	(0.066)	(0.079)	(0.088)	(0.113)
100	3	0.611	0.635	0.674	0.869	0.710	0.709	0.858	0.834
		(0.102)	(0.078)	(0.097)	(0.091)	(0.062)	(0.08)	(0.093)	(0.111)
	4	0.615	0.639	0.677	0.873	0.709	0.707	0.861	0.837
		(0.095)	(0.077)	(0.095)	(0.09)	(0.062)	(0.075)	(0.091)	(0.107)
	5	0.61	0.636	0.677	0.868	0.706	0.704	0.856	0.846
		(0.094)	(0.08)	(0.094)	(0.089)	(0.066)	(0.078)	(0.089)	(0.108)
	2	0.655	0.634	0.747	0.934	0.735	0.746	0.924	0.702
		(0.097)	(0.081)	(0.038)	(0.088)	(0.052)	(0.044)	(0.093)	(0.099)
500	3	0.654	0.635	0.751	0.937	0.737	0.751	0.922	0.702
		(0.1)	(0.082)	(0.041)	(0.085)	(0.057)	(0.049)	(0.094)	(0.093)
	4	0.658	0.628	0.747	0.942	0.737	0.749	0.926	0.702
		(0.097)	(0.082)	(0.041)	(0.084)	(0.058)	(0.048)	(0.095)	(0.093)
	5	0.657	0.631	0.748	0.940	0.734	0.747	0.922	0.705
		(0.097)	(0.083)	(0.042)	(0.084)	(0.052)	(0.046)	(0.095)	(0.099)

Table 4.6: Simulation results for the cluster similarity measure of Example 4.4.3, based on five point discrete spectral estimator.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.715	0.695	0.713	0.85	0.71	0.713	0.843	0.851
		(0.07)	(0.069)	(0.071)	(0.087)	(0.07)	(0.072)	(0.086)	(0.092)
50	3	0.708	0.695	0.709	0.841	0.708	0.71	0.833	0.847
		(0.074)	(0.068)	(0.072)	(0.083)	(0.074)	(0.072)	(0.081)	(0.086)
	4	0.71	0.694	0.708	0.845	0.706	0.709	0.835	0.846
		(0.071)	(0.066)	(0.07)	(0.086)	(0.073)	(0.071)	(0.086)	(0.088)
	5	0.709	0.692	0.707	0.842	0.707	0.71	0.836	0.845
		(0.074)	(0.067)	(0.073)	(0.085)	(0.073)	(0.074)	(0.085)	(0.089)
	2	0.743	0.712	0.744	0.87	0.737	0.744	0.858	0.833
		(0.058)	(0.054)	(0.054)	(0.09)	(0.059)	(0.053)	(0.089)	(0.111)
100	3	0.743	0.714	0.743	0.873	0.736	0.744	0.863	0.845
		(0.054)	(0.053)	(0.053)	(0.091)	(0.059)	(0.056)	(0.092)	(0.108)
	4	0.74	0.71	0.742	0.871	0.734	0.743	0.858	0.841
		(0.058)	(0.054)	(0.055)	(0.088)	(0.06)	(0.057)	(0.089)	(0.108)
	5	0.742	0.711	0.743	0.871	0.736	0.743	0.858	0.847
		(0.058)	(0.055)	(0.056)	(0.09)	(0.061)	(0.055)	(0.09)	(0.108)
	2	0.758	0.717	0.757	0.937	0.753	0.757	0.928	0.698
		(0.041)	(0.033)	(0.039)	(0.084)	(0.043)	(0.04)	(0.09)	(0.094)
500	3	0.762	0.715	0.762	0.944	0.756	0.761	0.928	0.705
		(0.047)	(0.032)	(0.046)	(0.082)	(0.052)	(0.049)	(0.094)	(0.095)
	4	0.76	0.715	0.761	0.944	0.753	0.759	0.929	0.708
		(0.047)	(0.032)	(0.046)	(0.081)	(0.048)	(0.046)	(0.092)	(0.098)
	5	0.758	0.715	0.758	0.944	0.755	0.76	0.927	0.709
		(0.045)	(0.032)	(0.042)	(0.083)	(0.048)	(0.046)	(0.093)	(0.097)

Table 4.7: Simulation results for the cluster similarity measure of Example 4.4.4.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.501	0.493	0.516	0.726	0.545	0.526	0.726	0.590
		(0.052)	(0.064)	(0.053)	(0.068)	(0.062)	(0.053)	(0.07)	(0.055)
50	3	0.498	0.491	0.512	0.724	0.543	0.523	0.720	0.582
		(0.055)	(0.065)	(0.053)	(0.066)	(0.059)	(0.055)	(0.07)	(0.053)
	4	0.499	0.491	0.515	0.722	0.544	0.525	0.715	0.575
		(0.051)	(0.064)	(0.054)	(0.066)	(0.062)	(0.054)	(0.07)	(0.055)
	5	0.501	0.493	0.511	0.714	0.545	0.526	0.707	0.573
		(0.052)	(0.063)	(0.052)	(0.068)	(0.062)	(0.056)	(0.07)	(0.055)
	2	0.515	0.509	0.536	0.817	0.614	0.550	0.825	0.811
		(0.055)	(0.076)	(0.057)	(0.067)	(0.069)	(0.061)	(0.072)	(0.064)
100	3	0.515	0.514	0.536	0.817	0.62	0.554	0.822	0.807
		(0.055)	(0.073)	(0.059)	(0.068)	(0.072)	(0.061)	(0.069)	(0.062)
	4	0.516	0.514	0.536	0.814	0.616	0.555	0.816	0.803
		(0.056)	(0.077)	(0.06)	(0.065)	(0.071)	(0.06)	(0.068)	(0.061)
	5	0.513	0.512	0.534	0.810	0.611	0.550	0.809	0.796
		(0.056)	(0.076)	(0.056)	(0.066)	(0.071)	(0.062)	(0.067)	(0.056)
	2	0.529	0.435	0.627	0.978	0.745	0.66	0.993	0.846
		(0.057)	(0.049)	(0.069)	(0.049)	(0.077)	(0.067)	(0.027)	(0.089)
500	3	0.532	0.435	0.625	0.981	0.738	0.658	0.990	0.808
		(0.055)	(0.049)	(0.067)	(0.045)	(0.079)	(0.067)	(0.031)	(0.092)
	4	0.532	0.435	0.629	0.979	0.737	0.659	0.988	0.783
		(0.058)	(0.048)	(0.069)	(0.047)	(0.076)	(0.069)	(0.034)	(0.092)
	5	0.533	0.433	0.628	0.979	0.743	0.658	0.986	0.755
		(0.056)	(0.05)	(0.067)	(0.048)	(0.078)	(0.067)	(0.037)	(0.094)

Table 4.8: Simulation results for the cluster similarity measure of Example 4.4.4, based on five point discrete spectral estimator.

N	p	J–divergence	d_{NP}	d_{LNP}	$d_{EUCLCEP}$	d_{ABSNP}	d_{ABSLNP}	d_{ABSCEP}	d_{PVAL}
	2	0.573	0.588	0.573	0.731	0.611	0.576	0.729	0.591
		(0.065)	(0.063)	(0.064)	(0.069)	(0.062)	(0.063)	(0.071)	(0.056)
50	3	0.568	0.586	0.57	0.733	0.606	0.575	0.727	0.576
		(0.061)	(0.063)	(0.061)	(0.066)	(0.062)	(0.061)	(0.069)	(0.055)
	4	0.572	0.591	0.573	0.73	0.61	0.577	0.725	0.572
		(0.062)	(0.062)	(0.061)	(0.067)	(0.062)	(0.062)	(0.065)	(0.055)
	5	0.571	0.585	0.572	0.729	0.608	0.577	0.718	0.566
		(0.061)	(0.061)	(0.06)	(0.064)	(0.061)	(0.06)	(0.066)	(0.057)
	2	0.631	0.622	0.635	0.825	0.677	0.639	0.833	0.814
		(0.068)	(0.072)	(0.069)	(0.069)	(0.066)	(0.07)	(0.068)	(0.062)
100	3	0.636	0.621	0.639	0.825	0.675	0.643	0.83	0.814
		(0.071)	(0.073)	(0.071)	(0.068)	(0.068)	(0.071)	(0.069)	(0.061)
	4	0.632	0.622	0.634	0.826	0.678	0.635	0.83	0.807
		(0.068)	(0.07)	(0.068)	(0.069)	(0.071)	(0.067)	(0.068)	(0.056)
	5	0.629	0.622	0.633	0.817	0.676	0.638	0.818	0.797
		(0.069)	(0.073)	(0.071)	(0.067)	(0.072)	(0.068)	(0.069)	(0.059)
	2	0.817	0.611	0.819	0.982	0.834	0.808	0.991	0.841
		(0.059)	(0.09)	(0.058)	(0.045)	(0.047)	(0.062)	(0.032)	(0.09)
500	3	0.819	0.607	0.82	0.983	0.83	0.807	0.991	0.813
		(0.058)	(0.089)	(0.056)	(0.043)	(0.05)	(0.058)	(0.031)	(0.093)
	4	0.816	0.612	0.819	0.983	0.831	0.81	0.987	0.786
		(0.059)	(0.09)	(0.058)	(0.044)	(0.05)	(0.061)	(0.037)	(0.093)
	5	0.817	0.609	0.819	0.982	0.831	0.806	0.986	0.76
		(0.059)	(0.091)	(0.058)	(0.045)	(0.052)	(0.061)	(0.037)	(0.093)

4.5 Clustering of Biological Time Series

We turn now back to the biological example from the Introduction. As it was mentioned, the goal is to reveal hidden protein sequence features. Towards this direction, we perform an analysis of fixed-length N-terminal amino acid sequences–that is 50 residues long–from different bacterial proteins with experimental evidence for the presence of a secretory N-terminal signal peptide, either of the classical or of the Tat form. More specifically, we collect 454 such sequences (104 Tat and 350 Sec signal peptides after filtering for sequence length) based on previous works to develop relevant classifiers (see Nielsen and Krogh (1998); Bendtsen et al. (2005)). It is worth mentioning that those sequences have been selected in a way that no pair exhibits significant sequence similarity. We have choose 104 sequences from each set. In order to apply the proposed cepstral coefficient-based metrics, we transformed those N-terminal fragments to time series data by replacing each amino acid residue with its hydrophobicity value according to one of three different experimentally derived scales (octanol, Wimley et al. (1996); interface Wimley and White (1996); Kyte-Doolittle Kyte and Doolittle (1982)). In addition consider the average hydrophobicity scale, where the hydrophobicity value for each amino acid residue was computed as the mean of the respective values of the aforementioned experimentally derived scales. For each sequence encoding we followed the methodology described above, with the last time series (N-terminal protein sequence) serving as a reference. Even though several numerical scales exist for representing different physicochemical properties of amino acid residues (Wilkins et al. (1999)), the choice of one of the hydrophobicity scales is based on the tripartite structure of signal peptides. Nevertheless, we have performed the protein sequence to time series transform using other scales (such as bulkiness, Zimmerman et al. (1968)) and the results were clearly inferior to the hydrophobicity scales. The cluster similarity index is computed for both cepstral coefficient–based clustering and for the rest of the metrics presented in 4.2.2. The results are displayed in Tables 4.9 and 4.10. Model (3.3) is fitted to the data for values of p from 1 up to 9. These results show that in general there is no any uniform better criterion but the J–divergence seems robust for these data. The cepstral based distances perform equally well with the rest of the distances. For instance, for the Kyte–Doolittle scale and when $p = 5$, we obtain that the cluster similarity index is 0.650 (0.652, respectively) after applying the distances (4.10) ((4.11), respectively). These

number are in direct comparison with the corresponding entries of Table 4.10 and in some instances they are better. In addition, we define binary hydrophobicity scales by clipping the original data according to their sign. In other words original hydrophobicity values are transformed to 1 if they are positive and 0 otherwise. The analysis was rerun and the final results regarding the cluster similarity index are summarized in both Tables 4.11 and 4.12. Even though the information content of the 'protein time series' is reduced by this transformation, the results do illustrate that this representation seems to preserve some useful second order features, leading to better clustering compared to the ground truth clusters. For example when $p = 7$ and the scale is binary Octanol, we note that our method performs better than the rest. The results show that our method performs equally well with other spectral domain clustering methods. The computation can be carried out by standard statistical software and we plan to post a webpage with programs and results of the data analysis. The analysis was now rerun using a five point discrete spectral estimator and results are displayed as above in both Tables 4.13 and 4.14 and for the binary data they are summarized in both Tables 4.15 and 4.16. As a general remark, the method here does depend on the value of p. A practical guide to real applications for choosing its value is to use a model selection criterion, like AIC, BIC or some other variants.

Table 4.9: Clustering similarity index for signal peptide data, based on raw scales and cepstral–based distances.

scale or *metric*	p								
Octanol	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.550	0.584	0.591	0.590	0.595	0.585	0.573	0.574	0.578
d_{ABSCEP}	0.555	0.589	0.579	0.577	0.590	0.596	0.576	0.574	0.579
d_{PVAL}	0.557	0.583	0.581	0.594	0.605	0.582	0.575	0.566	0.598
Interface	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.525	0.518	0.540	0.558	0.569	0.559	0.579	0.568	0.566
d_{ABSCEP}	0.523	0.528	0.547	0.568	0.571	0.568	0.570	0.566	0.555
d_{PVAL}	0.521	0.521	0.545	0.555	0.565	0.567	0.595	0.596	0.598
Kyte-Doolittle	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.583	0.605	0.587	0.613	0.650	0.633	0.622	0.589	0.580
d_{ABSCEP}	0.584	0.615	0.602	0.613	0.652	0.629	0.622	0.622	0.622
d_{PVAL}	0.567	0.605	0.615	0.607	0.603	0.618	0.622	0.622	0.622
Bulkiness	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.478	0.478	0.507	0.537	0.532	0.547	0.556	0.588	0.606
d_{ABSCEP}	0.487	0.566	0.552	0.550	0.541	0.542	0.550	0.493	0.535
d_{PVAL}	0.586	0.582	0.596	0.608	0.599	0.613	0.566	0.593	0.611

Table 4.10: Clustering similarity index for signal peptide data (raw scales).

scale	J–divergence	d_{NP}	d_{LNP}	d_{ABSNP}	d_{ABSLNP}
Octanol	0.581	0.571	0.533	0.559	0.540
Interface	0.545	0.552	0.534	0.551	0.509
Kyte-Doolittle	0.622	0.516	0.622	0.535	0.622
Bulkinnes	0.621	0.510	0.586	0.543	0.582

Table 4.11: Clustering similarity index for signal peptide data, based on binary scales and cepstral–based distances.

scale or *metric*	p								
Octanol	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.519	0.608	0.497	0.581	0.645	0.581	0.560	0.642	0.643
d_{ABSCEP}	0.521	0.595	0.521	0.674	0.665	0.579	0.642	0.634	0.643
d_{PVAL}	0.578	0.599	0.594	0.608	0.616	0.614	0.617	0.611	0.619
Interface	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.596	0.564	0.592	0.640	0.560	0.593	0.601	0.602	0.597
d_{ABSCEP}	0.584	0.625	0.586	0.662	0.602	0.601	0.586	0.607	0.608
d_{PVAL}	0.618	0.617	0.611	0.603	0.602	0.612	0.615	0.604	0.610
Kyte-Doolittle	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.498	0.576	0.545	0.581	0.593	0.574	0.562	0.583	0.555
d_{ABSCEP}	0.499	0.584	0.581	0.593	0.583	0.564	0.543	0.582	0.561
d_{PVAL}	0.549	0.604	0.598	0.604	0.608	0.606	0.608	0.601	0.615

Table 4.12: Clustering similarity index for signal peptide data (binary scales).

scale	J–divergence	d_{NP}	d_{LNP}	d_{ABSNP}	d_{ABSLNP}
Octanol	0.594	0.596	0.594	0.565	0.594
Interface	0.622	0.578	0.622	0.554	0.622
Kyte-Doolittle	0.595	0.575	0.595	0.482	0.595

Table 4.13: Clustering similarity index for signal peptide data, based on raw scales and cepstral–based distances, using five point discrete spectral estimator

scale or *metric*	*p*								
Octanol	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.540	0.549	0.551	0.555	0.551	0.545	0.557	0.554	0.556
d_{ABSCEP}	0.536	0.545	0.545	0.549	0.549	0.549	0.555	0.559	0.559
d_{PVAL}	0.546	0.593	0.559	0.586	0.587	0.610	0.611	0.613	0.615
Interface	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.515	0.525	0.503	0.511	0.514	0.518	0.519	0.522	0.520
d_{ABSCEP}	0.510	0.525	0.522	0.506	0.545	0.520	0.519	0.539	0.538
d_{PVAL}	0.510	0.518	0.513	0.521	0.507	0.520	0.508	0.510	0.567
Kyte-Doolittle	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.557	0.557	0.556	0.609	0.529	0.596	0.528	0.535	0.610
d_{ABSCEP}	0.448	0.544	0.560	0.614	0.632	0.604	0.599	0.609	0.608
d_{PVAL}	0.571	0.587	0.606	0.618	0.618	0.616	0.618	0.616	0.614
Bulkiness	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.494	0.498	0.490	0.500	0.517	0.526	0.526	0.531	0.528
d_{ABSCEP}	0.506	0.475	0.506	0.495	0.505	0.525	0.527	0.524	0.524
d_{PVAL}	0.598	0.609	0.600	0.618	0.613	0.614	0.607	0.615	0.619

Table 4.14: Clustering similarity index for signal peptide data (raw scales).

scale	J–divergence	d_{NP}	d_{LNP}	d_{ABSNP}	d_{ABSLNP}
Octanol	0.502	0.553	0.510	0.564	0.509
Interface	0.543	0.537	0.529	0.507	0.555
Kyte-Doolittle	0.598	0.540	0.576	0.559	0.550
Bulkinnes	0.507	0.489	0.500	0.485	0.490

Table 4.15: Clustering similarity index for signal peptide data, based on binary scales and cepstral–based distances, using five point discrete spectral estimator

scale or *metric*	p								
Octanol	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.584	0.605	0.512	0.614	0.610	0.615	0.624	0.618	0.618
d_{ABSCEP}	0.586	0.544	0.538	0.617	0.606	0.628	0.627	0.600	0.600
d_{PVAL}	0.605	0.606	0.603	0.616	0.610	0.616	0.617	0.618	0.618
Interface	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.570	0.585	0.583	0.584	0.586	0.587	0.587	0.588	0.600
d_{ABSCEP}	0.563	0.568	0.572	0.584	0.586	0.598	0.591	0.586	0.592
d_{PVAL}	0.614	0.619	0.616	0.618	0.617	0.619	0.619	0.618	0.620
Kyte-Doolittle	1	2	3	4	5	6	7	8	9
$d_{EUCLCEP}$	0.486	0.526	0.515	0.521	0.493	0.500	0.525	0.525	0.504
d_{ABSCEP}	0.488	0.494	0.494	0.492	0.501	0.525	0.521	0.522	0.508
d_{PVAL}	0.518	0.591	0.605	0.609	0.613	0.616	0.616	0.616	0.619

Table 4.16: Clustering similarity index for signal peptide data (binary scales).

scale	J–divergence	d_{NP}	d_{LNP}	d_{ABSNP}	d_{ABSLNP}
Octanol	0.512	0.516	0.506	0.546	0.511
Interface	0.529	0.506	0.527	0.520	0.531
Kyte-Doolittle	0.520	0.513	0.524	0.522	0.525

Chapter 5

Conclusions and Further Research

The first part of the thesis suggests a method for testing the similarity of G spectral density functions from independent stationary processes. The main assumption is model (3.3) which connects the last spectral density to the remaining ones by a linear function. The methodology is based on the asymptotic independence of periodogram ordinates. The analysis was illustrated by means of real data. An advantage of the present approach is that it provides different tests of similarity of G spectral distributions by using different covariate configurations. As it was suggested earlier, an alternative modeling approach can be based on the use of Legendre polynomials and Hermite polynomials functions of the standard normal quantile function–see Parzen (1993). Furthermore, the choice of the covariate vectors Z_i can be approached in several ways. One possibility is to use non parametric methods, like spline regression–see Kooperberg et al. (1995), for instance. In principle, it does not matter which spectral density is taken as the reference density since the difference in the parameter values remains constant. The methodology is promising in the framework of non-linear stationary time series models, in particular the ARCH example shows that the estimated parameters of model (3.3) fluctuate around zero, or in other words pointing to the similarities among the processes. Some further comparisons with the approach suggested by Dette and Paparoditis (2007) will be also worthwhile considering to understand the advantages and disadvantages of each method. The exponential model-class described in Chapter 2 , can be viewed under two different perspectives, namely as a parametric model class for the spectral density leading to a representation of its logarithmic transformation as a finite linear combination of cosine functions and as an approximation of a class of continuous log-spectral densities satisfying certain smoothness

conditions by linear combinations of the cosine basis functions. One main theoretical question for possible future research is the following. Since the approach proposed in the thesis is based on the exponential model with a fixed and finite number of unknown parameters, the asymptotic results obtained are $\sqrt{N}$-consistent. However, the approach needs further study under the alternative perspective. In this case, asymptotic valid statistical inference will require for the order p of the exponential-type approximation of the log-spectral density to increase to infinity with the sample size.

The second part of this work put forward several distance measures based on cepstral coefficients for time series data clustering. The main instrument for carrying out this task is again the semiparametric model (3.3) which assumes that the log ratio of two spectral densities is linear in some parameters. Based on this specific model and the asymptotic independence of periodogram ordinates, we show that robust clustering procedures can be developed for time series data mining. The approach is quite robust even when the data do not necessarily follow a linear process.

Several extensions of this work can be considered. The methodology can be extended in the multivariate time series setting by appealing to the properties of the multivariate periodogram based estimator. Accordingly, the periodogram ordinates of a multivariate linear process have asymptotically the complex Wishart distribution, see Brockwell and Davis (1991, Prop. 11.7.4). Hence, in principle, the methodology can be extended towards this direction but the calculations will be more elaborate. Furthermore, the simulations give evidence of good behavior of the algorithm when the number of clusters is known. However in applications the number of groups are usually unknown and should be estimated. We anticipate that an E-M type algorithm should give some reasonable results provided that the proportion of data that belongs to each cluster can be well approximated. Although the method was applied to a set of long stationary time series with the same length some other extensions can be highlighted. For instance, to cope with series of unequal length an averaging scheme of blocks of periodogram ordinates has been suggested by Coates and Diggle (1986). The issue of non stationarity might be addressed by the recent progress which has been made in this area, see Shumway (2003) and Huang et al. (2004).

The exponential model-class described in Chapter 2 , can be viewed under two different perspectives: (a) As a parametric model class for the spectral density leading to a representation of its logarithmic transformation as a finite linear combination of cosine functions. (b) As an approximation of a class of continuous log-spectral densities satisfying certain smoothness conditions by linear combinations of the cosine basis functions. One main question that arises and which is worth discussing is the following. Since the approach proposed in the thesis is based on perspective (a) of the exponential model with a fixed and finite number of unknown parameters, the asymptotic results obtained are $\sqrt{N}$-consistent (N is the sample size), and, therefore, statistical inference and testing procedures based on these estimators seem to offer some advantages compared to alternative methods based on direct (non-parametric) frequency domain comparisons of spectral densities. However, the approach needs further study under perspective (b). In this case, asymptotic valid statistical inference will require for the order p of the exponential-type approximation of the log-spectral density to increase to infinity with the sample size (at some appropriate rate). This will lead to a loss of the $\sqrt{N}$-consistency of the asymptotic results obtained since p will act in this case as a smoothing-type parameter.

Chapter 6

Appendix

Below we include the program we have used to perform the simulations at example 4.4.1, using the statistical language R.

```
library(GeneCycle)
library(cluster)

#########################################################
correct4zeros<-function(mat)
{
    min2<-0
    smat<-sort(mat)
    if( max(mat) == 0)
    {
        print ("ERROR ... Maximum value equals to zero")
        return(mat)
    }
    for(i in 1:length(smat))
    {
        if( smat[i]>0 )
        {
            min2=smat[i]
            break
```

```
        }
    }
    for(i in 1:length(mat))
    {
        if(mat[i]==0) mat[i] = min2
    }
    return(mat)
}
############################################################
```

The Fisher Information Matrix.

```
############################################################
Fishinfmatr<-function(G,N,p)
{
A<-rep(NA,(p+1)*(p+1))
dim(A)<-c((p+1),(p+1))
B<-rep(NA,(p+1)*(p+1))
dim(B)<-c((p+1),(p+1))
d<-floor(N/2)
x<-rep(NA,(p+1)*d)
dim(x)<-c(p+1,d)
for(j in 0:p)
{
x[j+1,]<-cos(j*2*pi*(1:d)/N)
}
a<-(G-1)/(G)
b<--1/(G)
A[1,1]<-a*sum(x[1,]*x[1,])
B[1,1]<-b*sum(x[1,]*x[1,])
for(m in 2:(p+1))
```

```
{
A[1,m]<-2*a*sum(x[1,]*x[m,])
A[m,1]<-2*a*sum(x[m,]*x[1,])
B[1,m]<-2*b*sum(x[1,]*x[m,])
B[m,1]<-2*b*sum(x[m,]*x[1,])
}
for(l in 2:(p+1))
{
for(k in 2:(p+1))
{
A[l,k]<-4*a*sum(x[l,]*x[k,])
B[l,k]<-4*b*sum(x[l,]*x[k,])
}
}
return(list(x,A,B))
}
# G: the number of timeseries to be analysed
# N: the length of each timeseries
# p+1: the number of coefficients (a0, a1, ..., ap)
Fishinfmatr(G,N,p)->k;
k[[2]]->A;
k[[3]]->B;

J=matrix(1, nrow=G-1, ncol=G-1)
FISH<-kronecker(diag(1,G-1), A)+kronecker(J-diag(1,G-1), B)
############################################################



# This is the main simulation function
# G: the number of timeseries to be analysed
# N: the length of each timeseries
```

```
# p+1: the number of coefficients (a0, a1, ..., ap)
# ss: the number of simulations (repetitions)
############################################################
generalsim<-function(G,N,p,ss,K1,K2,K3)
{
############################################################
Dmatrix<-function(i,j)
{
x<-rep(0,G-1)
x[i]<-1
x[j]<--1
kronecker(t(x),diag(1,(p+1)))
}
####Dmatrix calculates the matrix Dij of the theory chapter5

############################################################
Fishinfmatr(G,N,p)->k;
k[[2]]->A;
k[[3]]->B;

J=matrix(1, nrow=G-1, ncol=G-1)
FISH<-kronecker(diag(1,G-1), A)+kronecker(J-diag(1,G-1), B)

K<-Dmatrix(1,2)%*%(solve(FISH))%*%t(Dmatrix(1,2))
EIGEN<-eigen(K)$values
tk<-rep(NA,3)
for(k in 1:3)
{
tk[k]<-sum(EIGEN^k)
}
degfreed<-((tk[2])^3)/((tk[3])^2)
```

```
acoef<-((tk[3]))/((tk[2]))
bcoef<-tk[1]-((tk[2])^2)/((tk[3]))

# lik3 functions calculate likelihood estimators
# der.lik3 calculate derivatives of the above likelihood
# samples VS samplesnocov: only a0 VS all coefficients

lik3.samplesnocov<-function(theta, data)
{
s<-rep(NA,G-1)
u<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
a<-rep(NA,floor(N/2)-1)
linear<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
for (i in 1:(G-1))
{
linear[,i] <- data[,i]
s[i]<--sum(linear[,i])
u[,i]<- exp(linear[,i])
}
for (j in 1:(floor(N/2)-1))
{
a[j]<-sum(u[j,])
}
ll     <-  sum(s)+G*sum(log(1+a))
}


der.lik3.samplesnocov <- function(theta, data)
{
s<-rep(NA,G-1)
u<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
```

```
a<-rep(NA,floor(N/2)-1)
linear<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
for (i in 1:(G-1))
{
linear[,i] <- data[,i]
s[i]<--sum(linear[,i])
u[,i]<- exp(linear[,i])
}
for (j in 1:(floor(N/2)-1))
{
a[j]<-sum(u[j,])
}
b<-rep(NA,G-1)
for (i in 1:(G-1))
b[i]<-dim(data)[1] -G*sum( exp(linear[,i])/(1+a))
{
c( b )
}
}



lik3.samples<-function(theta, data)
{
s<-rep(NA,G-1)
u<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
a<-rep(NA,floor(N/2)-1)
linear<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)

for (i in 1:(G-1))
{
```

```
d<-matrix(NA, nrow=floor(N/2)-1, ncol=p)
for (m in 1:p){

d[,m]<- theta[(i-1)*p+i+m]*data[,(G-1)+m]
}

l<-rep(NA,floor(N/2)-1)
for(k in 1:floor(N/2)-1)
{
l[k]<- sum(d[k,])
}
linear[,i]<- data[,i]-theta[(i-1)*p+i]-l



s[i]<--sum(linear[,i])
u[,i]<- exp(linear[,i])
}

for (j in 1:(floor(N/2)-1))
{
a[j]<-sum(u[j,])
}
ll    <-  sum(s)+G*sum(log(1+a))
}

der.lik3.samples <- function(theta, data)
{s<-rep(NA,G-1)
u<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
a<-rep(NA,floor(N/2)-1)
linear<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1)
```

```
for (i in 1:(G-1))
{

d<-matrix(NA, nrow=floor(N/2)-1, ncol=p)
for (m in 1:p){

d[,m]<- theta[(i-1)*p+i+m]*data[,(G-1)+m]
}

l<-rep(NA,floor(N/2)-1)
for(k in 1:floor(N/2)-1)
{
l[k]<- sum(d[k,])
}
linear[,i]<- data[,i]-theta[(i-1)*p+i]-l


s[i]<--sum(linear[,i])
u[,i]<- exp(linear[,i])
}

for (j in 1:(floor(N/2)-1))
{
a[j]<-sum(u[j,])
}

b<-rep(NA,G-1)
v<-matrix(NA, nrow=G-1, ncol=p)

for (i in 1:(G-1))
{
```

```
b[i]<-dim(data)[1] -G*sum( exp(linear[,i])/(1+a))
for(h in 1:p)
{
v[i,h]<-sum(data[,G-1+h])-G*sum(exp(linear[,i])*data[,G-1+h]/(1+a))
}
}
w<-matrix(NA, nrow=G-1, ncol=p+1)
for(m in 1:(G-1))
{
w[m,]<-c(b[m],v[m,])
}
A<-t(w)
{
c(as.vector(A))
}
}



coef <- matrix(NA, nrow=ss, ncol=(G-1)*(p+1))
coefnew <- matrix(NA, nrow=ss, ncol=(G-1)*(p+1))
dev  <- rep(NA,ss)
dev.new <- rep(NA,ss)
message <- rep(NA, ss)

SIM<-matrix(NA, nrow=ss, ncol=8)



#####################################################################
#This is the core of the simulation!
#####################################################################
for(m in 1:ss)
```

```
{
Y<-matrix(NA, nrow=G, ncol=N)

for (i in 1:4){
Y[i,]<-arima.sim(list(ar=runif(1,-0.61,-0.59),ma=0), N,sd=1)
}
for (i in 5:8){
Y[i,]<-arima.sim(list(ar=c(runif(1,-0.61,-0.59),runif(1,0.29,0.31)), ma=0),
N, sd=2)
}
for (i in 9:13){
Y[i,]<-arima.sim(list(ar=c(runif(1,-0.51,-0.49)),
ma=runif(1,0.16,0.18)), N, sd=0.5)
}



ffts<-matrix(NA, nrow=G, ncol=floor(N/2)-1)
nperiodogram<-matrix(NA, nrow=G, ncol=floor(N/2))
periodogramma<-matrix(NA, nrow=G, ncol=floor(N/2))
for(i in 1:G)
{
#periodogramma[i,]<-periodogram(Y[i,])$spec
#periodogramma[i,]<-spectrum(Y[i,],plot=F,taper=0)$spec
periodogramma[i,]<-spectrum(Y[i,],spans=5,taper=0,plot=F)$spec
periodogramma[i,]<-correct4zeros(periodogramma[i,])
nperiodogram[i,]<-( periodogramma[i,]  )/var(Y[i,])
ffts[i,]<- (periodogramma[i,])[-floor(N/2)]
}

############################################################
```

```
DNPSIM<-matrix(NA, nrow=G-1, ncol=G-1)
DABSNPSIM<-matrix(NA, nrow=G-1, ncol=G-1)

DLNPSIM<-matrix(NA, nrow=G-1, ncol=G-1)
DABSLNPSIM<-matrix(NA, nrow=G-1, ncol=G-1)

DKLFDSIM<-matrix(NA, nrow=G-1, ncol=G-1)



for(i in 1:(G-1))
{
for (j in 1:(G-1))
{
DNPSIM[i,j]<-sqrt(sum((nperiodogram[i,]-nperiodogram[j,])^2))
DABSNPSIM[i,j]<-sqrt(sum(abs(nperiodogram[i,]-nperiodogram[j,])))

DLNPSIM[i,j]<-sqrt(sum((log(nperiodogram[i,])-log(nperiodogram[j,]))^2))
DABSLNPSIM[i,j]<-sqrt(sum(abs(log(nperiodogram[i,])-log(nperiodogram[j,]))))

DKLFDSIM[i,j]<-(1/N)*sum((nperiodogram[i,]/nperiodogram[j,])+
(nperiodogram[j,]/nperiodogram[i,])-2)
}
}

dimnames(DNPSIM) <- list(c(1:(G-1)),c(1:(G-1)))
dimnames(DABSNPSIM) <- list(c(1:(G-1)),c(1:(G-1)))

dimnames(DLNPSIM) <- list(c(1:(G-1)),c(1:(G-1)))
dimnames(DABSLNPSIM) <- list(c(1:(G-1)),c(1:(G-1)))

dimnames(DKLFDSIM) <- list(c(1:(G-1)),c(1:(G-1)))
```

```
X<-matrix(NA, nrow=p, ncol=floor(N/2)-1)

for (i in 1:p){
X[i,]<-2*cos((2*i)*pi*spectrum(Y[1,], plot=F)$freq[-floor(N/2)])
}

stuff<-matrix(NA, nrow=floor(N/2)-1, ncol=G-1+p)

for (j in 1:G-1)
{
stuff[,j]<-log(ffts[j,]/ffts[G,])
}
for(k in G:(G-1+p))
{
stuff[,k]<-X[k-G+1,]
}


a<-rep(1,(G-1)*(p+1))
b<-rep(0,G-1)

outputs     <- optim(lik3.samples, p=c(a), data=stuff,
der.lik3.samples, method="BFGS" )
outputsnew  <- lik3.samples(rep(0, (G-1)*(p+1)), stuff)
coef[m,]    <- outputs$par
message[m] <- outputs$convergence
dev.new[m] <- 2*(outputsnew-outputs$value)
```

```
summarycoefsim<-matrix(NA, nrow=G-1, ncol=p+1)
for(i in 1:G-1)
{
summarycoefsim[i,]<-coef[m,((i-1)*p+i):(i*p+i)]
}

DABSCEPSIM<-matrix(NA, nrow=G-1, ncol=G-1)
for(i in 1:G-1)
{
for (j in 1:G-1)
{
DABSCEPSIM[i,j]<-sqrt(sum(abs(summarycoefsim[i,]-summarycoefsim[j,])))
}
}


dimnames(DABSCEPSIM) <- list(c(1:(G-1)),c(1:(G-1)))

DEUCLCEPSIM<-matrix(NA, nrow=G-1, ncol=G-1)
for(i in 1:G-1)
{
for (j in 1:G-1)
{
DEUCLCEPSIM[i,j]<-sqrt(sum((summarycoefsim[i,]-summarycoefsim[j,])^2))
}
}
dimnames(DEUCLCEPSIM) <- list(c(1:(G-1)),c(1:(G-1)))

############################################################

DEUCLPVALBCSIM<-matrix(NA, nrow=G-1, ncol=G-1)
```

```
for(i in 1:G-1)
{
for (j in 1:G-1)
{
DEUCLPVALBCSIM[i,j]<-1-pchisq((DEUCLCEPSIM[i,j]-bcoef)/acoef,df=degfreed)
}
}
mm<-DEUCLPVALBCSIM[1,2]
 for(i in 1:(G-2))
 {
 for(j in (i+1):(G-1))
 {mm<-c(mm,DEUCLPVALBCSIM[i,j])
 }
 }
mm<-mm[-1]

PADJUST<-p.adjust(mm,method ="bonferroni",n = length(mm))



DEUCLPVALSIM<-matrix(NA, nrow=G-1, ncol=G-1)
DEUCLPVALSIM[G-1,G-1]<-1
ui<-0
ki<-2
for(i in 1:(G-2))
{
ui<-ui-(i+1)
ki<-ki-i
DEUCLPVALSIM[i,(i+1):(G-1)]<-PADJUST[((i-1)*G+ki):(i*G+ui)]
for(j in 1:G-1)
{
DEUCLPVALSIM[j,i]<-DEUCLPVALSIM[i,j]
```

```
}
DEUCLPVALSIM[i,i]<-1
}
DEUCLPVALSIM<-1-DEUCLPVALSIM


dimnames(DEUCLPVALSIM) <- list(c(1:(G-1)),c(1:(G-1)))
#######DEUCLPVALSIM calculates the 1-pvalue distances of the deuclecep


dianasim1<-diana(DABSNPSIM,diss=T)

dianasim2<-diana(DNPSIM,diss=T)

dianasim3<-diana(DLNPSIM,diss=T)

dianasim7<-diana(DABSLNPSIM,diss=T)

dianasim8<-diana(DEUCLPVALSIM,diss=T)

dianasim4<-diana(DKLFDSIM,diss=T)

dianasim5<-diana(DEUCLCEPSIM,diss=T)

dianasim6<-diana(DABSCEPSIM,diss=T)

dv1 <- cutree(as.hclust(dianasim1), k = 3)
dv2 <- cutree(as.hclust(dianasim2), k = 3)
dv3 <- cutree(as.hclust(dianasim3), k = 3)
dv4 <- cutree(as.hclust(dianasim4), k = 3)
dv5 <- cutree(as.hclust(dianasim5), k = 3)
```

```
dv6 <- cutree(as.hclust(dianasim6), k = 3)
dv7 <- cutree(as.hclust(dianasim7), k = 3)
dv8 <- cutree(as.hclust(dianasim8), k = 3)

C1<-rownames(DABSNPSIM)[dv1 == 1]
D1<-rownames(DABSNPSIM)[dv1 == 2]
E1<-rownames(DABSNPSIM)[dv1 == 3]

C2<-rownames(DNPSIM)[dv2 == 1]
D2<-rownames(DNPSIM)[dv2 == 2]
E2<-rownames(DNPSIM)[dv2 == 3]

C3<-rownames(DLNPSIM)[dv3 == 1]
D3<-rownames(DLNPSIM)[dv3 == 2]
E3<-rownames(DLNPSIM)[dv3 == 3]

C7<-rownames(DABSLNPSIM)[dv7 == 1]
D7<-rownames(DABSLNPSIM)[dv7 == 2]
E7<-rownames(DABSLNPSIM)[dv7 == 3]

C8<-rownames(DEUCLPVALSIM)[dv8 == 1]
D8<-rownames(DEUCLPVALSIM)[dv8 == 2]
E8<-rownames(DEUCLPVALSIM)[dv8 == 3]

C4<-rownames(DKLFDSIM)[dv4 == 1]
D4<-rownames(DKLFDSIM)[dv4 == 2]
E4<-rownames(DKLFDSIM)[dv4 == 3]

C5<-rownames(DEUCLCEPSIM)[dv5 == 1]
D5<-rownames(DEUCLCEPSIM)[dv5 == 2]
E5<-rownames(DEUCLCEPSIM)[dv5 == 3]
```

```
C6<-rownames(DABSCEPSIM)[dv6 == 1]

D6<-rownames(DABSCEPSIM)[dv6 == 2]

E6<-rownames(DABSCEPSIM)[dv6 == 3]


##############################################

SIMI1<-rep(NA,3)

SIMILA1<-rep(NA,3)

SIMILB1<-rep(NA,3)

SIMILC1<-rep(NA,3)


SIMILA1[1]<-(2*(length(intersect(K1,C1))))/(length(K1)+length(C1))

SIMILA1[2]<-(2*(length(intersect(K1,D1))))/(length(K1)+length(D1))

SIMILA1[3]<-(2*(length(intersect(K1,E1))))/(length(K1)+length(E1))


SIMILB1[1]<-(2*(length(intersect(K2,C1))))/(length(K2)+length(C1))

SIMILB1[2]<-(2*(length(intersect(K2,D1))))/(length(K2)+length(D1))

SIMILB1[3]<-(2*(length(intersect(K2,E1))))/(length(K2)+length(E1))


SIMILC1[1]<-(2*(length(intersect(K3,C1))))/(length(K3)+length(C1))

SIMILC1[2]<-(2*(length(intersect(K3,D1))))/(length(K3)+length(D1))

SIMILC1[3]<-(2*(length(intersect(K3,E1))))/(length(K3)+length(E1))



SIMI1[1]<-max(SIMILA1)

SIMI1[2]<-max(SIMILB1)

SIMI1[3]<-max(SIMILC1)


SIM1<-mean(SIMI1)

#########################################################################

SIMI2<-rep(NA,3)
```

```
SIMILA2<-rep(NA,3)
SIMILB2<-rep(NA,3)
SIMILC2<-rep(NA,3)

SIMILA2[1]<-(2*(length(intersect(K1,C2))))/(length(K1)+length(C2))
SIMILA2[2]<-(2*(length(intersect(K1,D2))))/(length(K1)+length(D2))
SIMILA2[3]<-(2*(length(intersect(K1,E2))))/(length(K1)+length(E2))

SIMILB2[1]<-(2*(length(intersect(K2,C2))))/(length(K2)+length(C2))
SIMILB2[2]<-(2*(length(intersect(K2,D2))))/(length(K2)+length(D2))
SIMILB2[3]<-(2*(length(intersect(K2,E2))))/(length(K2)+length(E2))

SIMILC2[1]<-(2*(length(intersect(K3,C2))))/(length(K3)+length(C2))
SIMILC2[2]<-(2*(length(intersect(K3,D2))))/(length(K3)+length(D2))
SIMILC2[3]<-(2*(length(intersect(K3,E2))))/(length(K3)+length(E2))


SIMI2[1]<-max(SIMILA2)
SIMI2[2]<-max(SIMILB2)
SIMI2[3]<-max(SIMILC2)

SIM2<-mean(SIMI2)

############################################################
SIMI3<-rep(NA,3)
SIMILA3<-rep(NA,3)
SIMILB3<-rep(NA,3)
SIMILC3<-rep(NA,3)

SIMILA3[1]<-(2*(length(intersect(K1,C3))))/(length(K1)+length(C3))
SIMILA3[2]<-(2*(length(intersect(K1,D3))))/(length(K1)+length(D3))
```

```
SIMILA3[3]<-(2*(length(intersect(K1,E3))))/(length(K1)+length(E3))

SIMILB3[1]<-(2*(length(intersect(K2,C3))))/(length(K2)+length(C3))
SIMILB3[2]<-(2*(length(intersect(K2,D3))))/(length(K2)+length(D3))
SIMILB3[3]<-(2*(length(intersect(K2,E3))))/(length(K2)+length(E3))

SIMILC3[1]<-(2*(length(intersect(K3,C3))))/(length(K3)+length(C3))
SIMILC3[2]<-(2*(length(intersect(K3,D3))))/(length(K3)+length(D3))
SIMILC3[3]<-(2*(length(intersect(K3,E3))))/(length(K3)+length(E3))



SIMI3[1]<-max(SIMILA3)
SIMI3[2]<-max(SIMILB3)
SIMI3[3]<-max(SIMILC3)

SIM3<-mean(SIMI3)

##########################################################
SIMI4<-rep(NA,3)
SIMILA4<-rep(NA,3)
SIMILB4<-rep(NA,3)
SIMILC4<-rep(NA,3)

SIMILA4[1]<-(2*(length(intersect(K1,C4))))/(length(K1)+length(C4))
SIMILA4[2]<-(2*(length(intersect(K1,D4))))/(length(K1)+length(D4))
SIMILA4[3]<-(2*(length(intersect(K1,E4))))/(length(K1)+length(E4))

SIMILB4[1]<-(2*(length(intersect(K2,C4))))/(length(K2)+length(C4))
SIMILB4[2]<-(2*(length(intersect(K2,D4))))/(length(K2)+length(D4))
SIMILB4[3]<-(2*(length(intersect(K2,E4))))/(length(K2)+length(E4))
```

```
SIMILC4[1]<-(2*(length(intersect(K3,C4))))/(length(K3)+length(C4))
SIMILC4[2]<-(2*(length(intersect(K3,D4))))/(length(K3)+length(D4))
SIMILC4[3]<-(2*(length(intersect(K3,E4))))/(length(K3)+length(E4))


SIMI4[1]<-max(SIMILA4)
SIMI4[2]<-max(SIMILB4)
SIMI4[3]<-max(SIMILC4)

SIM4<-mean(SIMI4)

#################################################################
SIMI5<-rep(NA,3)
SIMILA5<-rep(NA,3)
SIMILB5<-rep(NA,3)
SIMILC5<-rep(NA,3)

SIMILA5[1]<-(2*(length(intersect(K1,C5))))/(length(K1)+length(C5))
SIMILA5[2]<-(2*(length(intersect(K1,D5))))/(length(K1)+length(D5))
SIMILA5[3]<-(2*(length(intersect(K1,E5))))/(length(K1)+length(E5))

SIMILB5[1]<-(2*(length(intersect(K2,C5))))/(length(K2)+length(C5))
SIMILB5[2]<-(2*(length(intersect(K2,D5))))/(length(K2)+length(D5))
SIMILB5[3]<-(2*(length(intersect(K2,E5))))/(length(K2)+length(E5))

SIMILC5[1]<-(2*(length(intersect(K3,C5))))/(length(K3)+length(C5))
SIMILC5[2]<-(2*(length(intersect(K3,D5))))/(length(K3)+length(D5))
SIMILC5[3]<-(2*(length(intersect(K3,E5))))/(length(K3)+length(E5))


SIMI5[1]<-max(SIMILA5)
```

```
SIMI5[2]<-max(SIMILB5)

SIMI5[3]<-max(SIMILC5)

SIM5<-mean(SIMI5)


#########################################################################

SIMI6<-rep(NA,3)

SIMILA6<-rep(NA,3)

SIMILB6<-rep(NA,3)

SIMILC6<-rep(NA,3)


SIMILA6[1]<-(2*(length(intersect(K1,C6))))/(length(K1)+length(C6))

SIMILA6[2]<-(2*(length(intersect(K1,D6))))/(length(K1)+length(D6))

SIMILA6[3]<-(2*(length(intersect(K1,E6))))/(length(K1)+length(E6))


SIMILB6[1]<-(2*(length(intersect(K2,C6))))/(length(K2)+length(C6))

SIMILB6[2]<-(2*(length(intersect(K2,D6))))/(length(K2)+length(D6))

SIMILB6[3]<-(2*(length(intersect(K2,E6))))/(length(K2)+length(E6))


SIMILC6[1]<-(2*(length(intersect(K3,C6))))/(length(K3)+length(C6))

SIMILC6[2]<-(2*(length(intersect(K3,D6))))/(length(K3)+length(D6))

SIMILC6[3]<-(2*(length(intersect(K3,E6))))/(length(K3)+length(E6))



SIMI6[1]<-max(SIMILA6)

SIMI6[2]<-max(SIMILB6)

SIMI6[3]<-max(SIMILC6)

SIM6<-mean(SIMI6)



#################################################################

SIMI7<-rep(NA,3)
```

```
SIMILA7<-rep(NA,3)
SIMILB7<-rep(NA,3)
SIMILC7<-rep(NA,3)

SIMILA7[1]<-(2*(length(intersect(K1,C7))))/(length(K1)+length(C7))
SIMILA7[2]<-(2*(length(intersect(K1,D7))))/(length(K1)+length(D7))
SIMILA7[3]<-(2*(length(intersect(K1,E7))))/(length(K1)+length(E7))

SIMILB7[1]<-(2*(length(intersect(K2,C7))))/(length(K2)+length(C7))
SIMILB7[2]<-(2*(length(intersect(K2,D7))))/(length(K2)+length(D7))
SIMILB7[3]<-(2*(length(intersect(K2,E7))))/(length(K2)+length(E7))

SIMILC7[1]<-(2*(length(intersect(K3,C7))))/(length(K3)+length(C7))
SIMILC7[2]<-(2*(length(intersect(K3,D7))))/(length(K3)+length(D7))
SIMILC7[3]<-(2*(length(intersect(K3,E7))))/(length(K3)+length(E7))


SIMI7[1]<-max(SIMILA7)
SIMI7[2]<-max(SIMILB7)
SIMI7[3]<-max(SIMILC7)
SIM7<-mean(SIMI7)


#####################################################################
SIMI8<-rep(NA,3)
SIMILA8<-rep(NA,3)
SIMILB8<-rep(NA,3)
SIMILC8<-rep(NA,3)

SIMILA8[1]<-(2*(length(intersect(K1,C8))))/(length(K1)+length(C8))
SIMILA8[2]<-(2*(length(intersect(K1,D8))))/(length(K1)+length(D8))
```

```
SIMILA8[3]<-(2*(length(intersect(K1,E8))))/(length(K1)+length(E8))

SIMILB8[1]<-(2*(length(intersect(K2,C8))))/(length(K2)+length(C8))
SIMILB8[2]<-(2*(length(intersect(K2,D8))))/(length(K2)+length(D8))
SIMILB8[3]<-(2*(length(intersect(K2,E8))))/(length(K2)+length(E8))

SIMILC8[1]<-(2*(length(intersect(K3,C8))))/(length(K3)+length(C8))
SIMILC8[2]<-(2*(length(intersect(K3,D8))))/(length(K3)+length(D8))
SIMILC8[3]<-(2*(length(intersect(K3,E8))))/(length(K3)+length(E8))


SIMI8[1]<-max(SIMILA8)
SIMI8[2]<-max(SIMILB8)
SIMI8[3]<-max(SIMILC8)
SIM8<-mean(SIMI8)


#########################################################################

SIM[m,]<-c(SIM4,SIM2,SIM3,SIM5,SIM1,SIM7,SIM6,SIM8)
dimnames(SIM) <- list(1:ss,c("DKLFD","DNP","DLNP","DEUCLECEP","DABSNP",
"DABSLNP","DABSCEP", "DEUCLPVAL") )
}
#########################################################################

SIMILAR<-matrix(NA, nrow=1, ncol=8)
SIMILAR[1,]<-apply(SIM,2,mean)
dimnames(SIMILAR) <- list("1",c("DKLFD","DNP","DLNP","DEUCLECEP","DABSNP",
"DABSLNP","DABSCEP","DEUCLPVAL"))

SDOFSIMILAR<-sqrt(diag(var(SIM)))
```

```
COEF<-apply(coef,2,mean)
COEFNEW<-apply(coefnew,2,mean)

summary(dev.new)


summarycoef<-matrix(NA, nrow=G-1, ncol=p+1)
for(i in 1:G-1)
{
summarycoef[i,]<-COEF[((i-1)*p+i):(i*p+i)]
}
summarycoef

sqrt(diag(var(coef)))

DABSCEP<-matrix(NA, nrow=G-1, ncol=G-1)
for(i in 1:G-1)
{
for (j in 1:G-1)
{
DABSCEP[i,j]<-sqrt(sum(abs(summarycoef[i,]-summarycoef[j,])))
}
}
dimnames(DABSCEP) <- list(c(1:(G-1)),c(1:(G-1)))


DEUCLCEP<-matrix(NA, nrow=G-1, ncol=G-1)
for(i in 1:G-1)
{
```

```
for (j in 1:G-1)
{
DEUCLCEP[i,j]<-sqrt(sum((summarycoef[i,]-summarycoef[j,])^2))
}
}
dimnames(DEUCLCEP) <- list(c(1:(G-1)),c(1:(G-1)))

return(list(SIM=SIM,SIMILAR=SIMILAR,SDOFSIMILAR=SDOFSIMILAR,
DABSCEP=DABSCEP, DEUCLCEP=DEUCLCEP ,G=G,N=N,p=p,ss=ss,
coef=coef,summarycoef=summarycoef,sdcoef=sqrt(diag(var(coef))),
dev=dev.new,message=message))
}
```

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC–19*, 716–723.

Alonso, A. and E. Maharaj (2006). Comparison of time series using subsampling. *Computational Statistics & Data Analysis 50*, 2589–2599.

Bendtsen, J. D., H. Nielsen, D. Widdick, T. Palmer, and S. Brunak (2005). Prediction of twin-arginine signal peptides. *BMC Bioinformatics 6*, 167.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate:a practical and powerfull approach to multiple testing. *Journal of the Royal Statist. Soc. B 57*(1), 289–300.

Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrica 80*, 817–822.

Beran, J. (1994). *Statistics for long-memory processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall.

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol 22*.

Berks, B. C., T. Palmer, and F. Sargent (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Curr Opin Microbiol 8*.

Biau, G., F. Bunea, and M. Wegkamp (2005). Consistent functional classification. *IEEE Trans Information Theory 51*, 2163–2172.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrica 60*, 217–226.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Springer. Expanded Edition.

Britton, W. (1983). *Conjugate Duality and the Exponential Fourier Spectrum*, Volume 18. Springer. Lecture Notes in Statistics.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer. Second Edition.

Caiado, J., N. Crato, and D. Peña (2006). A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis 50*, 2668–2684.

Cameron, M. and R. Turner (1987). Fitting models to spectra using regression packages. *Applied Statistics 36*, 47–57.

Choi, H., H. Ombao, and B. Ray (2008). Sequential change-point detection for non-stationary time series. *Technometrics 50*, 40–52.

Coates, D. S. and P. J. Diggle (1986). Tests for comparing two estimated spectral densities. *J. Time Ser. Anal. 7*, 7–20.

Corduas, M. and D. Piccolo (2008). Time series clustering and classification by the autoregressive metric. *Computational Statistics & Data Analysis*, 1860–1872.

Davies, R. and D. Harte (1987). Tests for hurst effect. *Biometrika 74*, 95–101.

Dette, H. and E. Paparoditis (2007). Testing equality of spectral densities.

Diggle, P. J. (1990). *Time series*. New York: Oxford University Press.

Diggle, P. J. and N. I. Fisher (1991). Nonparametric comparison of cumulative periodograms. *Appl. Statist. 40*, 423–434.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1008.

Fokianos, K. (2004). Merging information for semiparametric density estimation. *J. R. Stat. Soc. Ser. B 66*, 941–958.

Galeano, P. and D. Peña (2000). Multivariate analysis in vector time series. *Resenhas 4*, 383–403. Special number on time series analysis.

Giraitis, L. and P. M. Robinson (2001). Whittle estimation of ARCH models. *Econometric Theory 17*, 608–631.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer Series in Statistics. New York: Springer-Verlag.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. New York: Springer-Verlag.

Hitchcock, D., G. Casella, and J. Booth (2006). Improved estimation of dissimilarities by pre-smoothing functional data. *Journal of the American Statistical Association 101*, 211–222.

Holan, S. H. (2004). *Time Series Exponential Models:Theory and Methods*. Ph. D. thesis, Texas A & M University, Department of Statistics.

Huang, H.-Y., H. Ombao, and D. S. Stoffer (2004). Discrimination and classification of nonstationary time series using the SLEX model. *Journal of the American Statistical Association 99*, 763–774.

Issac, B., H. Singh, H. Kaur, and G. P. S. Raghava (2002). Locating probable genes using fourier transform approach. *Bioinformatics 18*, 196–197.

Johnson, N. and S. Kotz (1970). *Continuous Univariate Distribution*. 2 Wiley:New York.

Johnson, R. and D. W. Wichern (1992). *Applied Multivariate Statistical Analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Jones, M. (1987). Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models. *Applied Statistics 36*, 134–138.

Kakizawa, Y., R. H. Shumway, and M. Taniguchi (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association 93*, 328–340.

Kalpakis, K., D. Gada, and V. Puttagunta (2001, 29 November–2 December). Distance measures for the effective clustering of ARIMA time–series. In *Proceedings of the IEEE International Conference on Data Mining*, San Jose, CA, USA, pp. 273–280.

Kooperberg, C., C. J. Stone, and Y. K. Truong (1995). Logspline estimation of a possibly mixed spectral distribution. *J. Time Ser. Anal. 16*, 359–388.

Kyte, J. and R. F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*.

Liao, T. W. (2005). Clustering of time series data–a survey. *Pattern Recognition 38*, 1857–1874.

Macnaughton Smith, P., W. Williams, M. Dale, and L. Mockett (1965). Dissimilarity analysis: a new technique of hierarchical subdivision. *Nature 202*, 1034–1035.

Maharaj, E. (2002). Comparison of non-stationary time series in the frequency domain. *Computational Statistics & Data Analysis 40*, 131–141.

McCullagh, P. and J. A. Nelder (1989). *Generalized linear models*. London: Chapman & Hall.

McLachlan, A. D. and M. Stewart (1976). The 14-fold periodicity in alpha-tropomyosin and the interaction with actin. *J Mol Biol 103*.

Miller, R. G. (1981). *Simultaneous statistical inference* (Second ed.). New York: Springer-Verlag.

Nielsen, H. and A. Krogh (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol 6*, 122–130.

Parzen, E. (1993). Stationary time series analysis using information and spectral analysis. In T. S. Rao (Ed.), *Developments in Time Series Analysis. In Honour of M. B. Priestley*, pp. 139–148. Chapman & Hall.

Pasquier, C., V. J. Promponas, and S. J. Hamodrakas (2001). PRED-CLASS: cascading neural networks for generalized protein classification and genome-wide applications. *Proteins 44*.

Pasquier, C. M., P. V. J, V. N. J., and H. S. J. (1998). A web server to locate periodicities in a sequence. *Bioinformatics 14*.

Percival, D. (1992). Simulating gaussian random processes with specified spectra. *Computing Science and Statistics*, 534–538.

Pham, T. D. (2006). LPC cepstral distortion measure for protein sequence comparison. *IEEE Trans Nanobioscience 5*.

Piccolo, D. (1990). A distance measure for classifying arma models. *Journal of Time Series Analysis 11*, 152–164.

Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Comm. Statist. A—Theory Methods 12*, 2085–2094.

Pourahmadi, M. (1984). Taylor expansion of $exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. *Amer. Math. Monthly 91*, 303–307.

Priestley, M. B. (1981). *Spectral Analysis And Time Series*. Academic Press, London.

R Development Core Team (2004). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. 3-900051-07-0.

Rayner, J. and D. Best (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press.

Shumway, R. H. (2003). Time-frequency clustering and discriminant analysis. *63*, 307–314.

Shumway, R. H. and D. S. Stoffer (2006). *Time series analysis and its applications* (Second ed.). New York: Springer.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic theory of statistical inference for time series*. New York: Springer.

Timmer, J., M. Lauk, W. Vach, and L. C.H. (1999). A test for a difference between spectral peak frequencies. *Computational Statistics & Data Analysis 30*, 45–55.

Walker, A. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary non-deterministic time series. *Australian Mathematical Society 4*, 363–384.

Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol 112*, 531–552.

Wimley, W. C., T. P. Creamer, and S. H. White (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry 35*.

Wimley, W. C. and S. H. White (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol 3*.

Yates, J. R., r. (2004). Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct 33*, 297–316.

Zimmerman, J. M., N. Eliezer, and R. Simha (1968). The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol 21*.