COMPARING SPECTRAL DENSITIES IN REPLICATED TIME SERIES BY SMOOTHING SPLINE ANOVA

by

Sangdae Han

BS, Hallym University, 1996

MS, Michigan State University, 1998

MA, University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

the Department of Statistics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

THE DEPARTMENT OF STATISTICS

This dissertation was presented

by

Sangdae Han

It was defended on

April 30, 2008

and approved by

David S. Stoffer, Professor

Stewart J. Anderson, Professor

Leon J. Gleser, Professor

Wesley K. Thompson, Professor

Dissertation Director: David S. Stoffer, Professor

COMPARING SPECTRAL DENSITIES IN REPLICATED TIME SERIES BY SMOOTHING SPLINE ANOVA

Sangdae Han, PhD

University of Pittsburgh, 2008

Comparing several groups of populations based on replicated data is one of the main concerns in statistical analysis. A specific type of data, time series data, such as waves of earthquakes present difficulties because of the correlations amongst the data. Spectral analysis solves this problem somewhat because the discrete Fourier transform transforms the data to near independence under general conditions.

The goal of our research is to develop general, user friendly, statistical methods to compare group spectral density functions. To accomplish this, we consider two main problems: How can we construct an estimation function from replicated time series for each group and what method can be used to compare the estimated functions? For the first part, we present smooth estimates of spectral densities from time series data obtained from replication across subjects (units) (Wahba 1990; Guo et al. 2003). We assume that each spectral density, $f(\omega)$, is in some reproducing kernel Hilbert space and apply penalized least squares methods to estimate $\hat{f}(\omega)$ in smoothing spline ANOVA. For the second part, we consider confidence intervals to determine the frequencies where the spectrum of one spectral density may differ from another. These confidence intervals are the independent simultaneous confidence interval and the bootstrapping confidence interval (Babu et al. 1983; Olshen et al. 1989). Finally, as an application, we consider the replicated time series data that consist of shear (S) waves of 8 earthquakes and 8 explosions (Shumway & Stoffer 2006).

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

I recognize my thesis is possible from a number of people who helped me while I was in Pittsburgh as a graduate student. First, I would like to thank Professor David S. Stoffer, my adviser, for his excellent advice and support during this research. With his help, I have learned valuable independent research skills. Without his help, I could not complete my research. I am very proud to be his student and respect him very much.

I also would like to thank my three committee members, Professor Leon J. Gleser, Professor Wesley K. Thompson, and Professor Stewart J. Anderson for their helpful suggestions and encouragement. Lastly, I wish to express my appreciation to the other faculty, staff and friends of the statistics department in the University of Pittsburgh for friendship and encouragement throughout my graduate years.

Finally, I would like to thank my wife, Ilhwa, and 11-month old son, Joonhee, for their unconditional and constant love and happiness. When I met difficulties and fell into slumps in my research, I was able to continue and finish my research thanks to their love and belief in me. I would like to dedicate my dissertation to my family.

1.0 INTRODUCTION

1.1 PROBLEM STATEMENT

The goal of this thesis is to develop a general statistical methodology for comparing spectral density functions across groups. To accomplish this, we will construct smooth and continuous estimates of the spectral density from time series data obtained from replication across subjects (units).

The basic idea of our thesis is as follows; suppose a time series, $y_1, y_2, \ldots, y_n$, is given; let $I(\omega_k)$ and $f(\omega)$ denote the periodogram and the spectrum of the time series at Fourier (fundamental) frequencies $\omega_k = k/n$ where $k = 0, 1, 2, \ldots, n-1$ and frequency ω, respectively. For a specific frequency ω, say ω_0, we assume $\omega_{k'}$ is the closest Fourier frequency to ω_0 and $\omega_{k'} \to \omega_0$ as $n \to \infty$. For $\omega_{k'} \in (0, 1/2)$, $I(\omega_{k'})$ is distributed asymptotically as $f(\omega_0)$ times a random variable with a Gamma distribution whose shape and scale parameters are 1 and 1. Moreover, at frequencies $\omega_{k_1} \neq \omega_{k_2}$, $I(\omega_{k_1})$ and $I(\omega_{k_2})$ are asymptotically independent under general conditions.

Suppose we have G groups and each group consists of replicated time series data from $f_g(\omega|\boldsymbol{x_i})$ where $g = 1, 2, \ldots, G$ and $\boldsymbol{x_i} = [x_{i1}, x_{i2}, \ldots, x_{ip}]$ is a vector of p group specific covariates for subject i. We assume that $\boldsymbol{x_i}$ consists of only group identifiers. If $\boldsymbol{x_i}$ contains other covariates such as age or temperature, then these covariates (if they differ between groups) would be confounded with differences between group spectral densities.

Now we assume each such spectral density is given by

$$f_g(\omega|\boldsymbol{x_i}) = \exp[\beta_{g0}(\omega) + \beta_{g1}(\omega)x_{i1} + \beta_{g2}(\omega)x_{i2} + \cdots + \beta_{gp}(\omega)x_{ip}], \tag{1.1}$$

where $[\beta_{g0}(\omega),\ldots,\ \beta_{gp}(\omega)]$ is a vector of regression coefficients specific to group g and frequency ω. The goal is to discover if there are any differences in the group spectra and to find the frequencies, ω, where the spectrum of one group differs from that of another. For this purpose, we can consider the confidence bands for pairwise log-differences for all groups, $\log f_{g_1}(\omega|\boldsymbol{x_i}) - \log f_{g_2}(\omega|\boldsymbol{x_i})$, where $g_1, g_2 = 1, 2, \ldots, G$ $(g_1 \neq g_2)$. Then, we can simply determine for each band whether or not zero is contained in the band at frequencies ω. For example, if zero is contained in the confidence band for the log-difference between group g_1 and g_2 at ω, conclusion $H_0 : f_{g_1}(\omega|\boldsymbol{x_i}) = f_{g_2}(\omega|\boldsymbol{x_i})$ is reached at ω. Note that these confidence bands should be simultaneous confidence bands for all $\binom{G}{2}$ pairs. In the simplest case in (1.2) and (1.3) below, $\log f(\omega|x_i = 1) - \log f(\omega|x_i = 0)$ can be expressed as $\beta_1(\omega)$. We will use the two different confidence intervals (bands) for $\beta_1(\omega)$ to determine the frequencies, ω, where the spectra of the two groups are different; these confidence intervals are the independent simultaneous confidence interval and the bootstrapping confidence interval for functions [6, 21].

Suppose we wish to compare the spectra of two groups where we observe l_1 subjects from group 1 and l_2 subjects from group 2. Then, we model the two group spectra as in (1.2),

$$f(\omega|x_i) = \exp[\beta_0(\omega) + \beta_1(\omega)x_i] \tag{1.2}$$

where $x_i = 1$ for $i = 1, 2, \ldots, l_1$ and $x_i = 0$ for $i = l_1 + 1, l_1 + 2, \ldots, l = l_1 + l_2$. Note that equation (1.2) is a decomposition of the two spectra, and that $\beta_1(\omega)$ is

$$\beta_1(\omega) = \log \frac{f(\omega|x_i = 1)}{f(\omega|x_i = 0)}. \tag{1.3}$$

$\beta_1(\omega)$ has an important interpretation: if $\beta_1(\omega) > 0$, then the spectral power of group 1 at ω is greater than that of group 2, and if $\beta_1(\omega) < 0$, then the spectral power of group 2 is greater than that of group 1 at ω. Therefore, our analysis will be focused on $\beta_1(\omega)$ for each ω.

The motivation of our research is from Wahba [29]. That is, suppose $\{y_t\}_{t=1}^n$ is a stationary series. Instead of indexing Fourier frequencies by $\omega_k = \frac{k}{n}$, we henceforth simply

index them by k only. Thus, $I(k) \equiv I(\omega_k)$. According to standard asymptotic theory (*e.g.* [15, 29]), and using (1.2),

$$I(k) = f(k)U_k \tag{1.4}$$

$$\log I(k) = \log f(k) + \epsilon_k \tag{1.5}$$

$$\log I(k) = \beta_0(k) + \beta_1(k)x_i + \epsilon_k \tag{1.6}$$

where $k = 1, 2, \ldots, [\frac{n}{2}]$ if n is odd, and $k = 1, 2, \ldots, [\frac{n-1}{2}]$ if n is even, $U_k \overset{ind}{\sim} \gamma(1,1)$ where γ denotes the Gamma distribution, and the ϵ_k are independent random variables having the distribution of $\log \gamma(1,1)$ so that $E(\epsilon_k) = -0.57721$ and $\mathrm{Var}(\epsilon_k) = \pi^2/6$.

We will use $\log I(k) = \log f(k) + \epsilon_k$ in (1.5) as the basis of our model. Because $\log I(k_1)$ and $\log I(k_2)$ are asymptotically independent for $k_1 \neq k_2$ as $n \to \infty$, we regard (1.6) as a simple regression for estimation of $\beta_0(k)$ and $\beta_1(k)$. The differences from a simple regression are that the estimates are smooth and $\epsilon_k \overset{ind}{\sim} \log \gamma(1,1)$.

We will finally analyze the replicated time series data that consist of shear (S) waves of 8 earthquakes (group 1) and 8 explosions (group 2). Each sequence of data consists of 1024 points (40 points per second) that are recorded in seismic recording stations in Scandinavia [28]. Figure 1.1 shows one series of earthquakes and explosions series each and the **R** code is shown in Appendix D.1.

1.2 THESIS OUTLINE

This thesis is organized in the following fashion: In Chapter 2, brief reviews of some related definitions for our research are given. In addition, we review the papers Wahba [29] and Diggle and Al Wasel [9] that provided motivation for our research in Section 2.2. Chapter 3 is devoted to our estimation method. This chapter contains the notions of smoothing spline estimation and the penalty functional in reproducing kernel Hilbert spaces. The model and the modified (centered) model for estimating spectral densities are also presented. In Chapter 4, we present simulation results and an application to the real time series data that

validate our proposed estimation method described in Chapter 3. In particular, we focus on two confidence intervals used to validate our estimation method. Finally, Chapter 5 gives a summary of the results of our thesis and issues for our future studies.

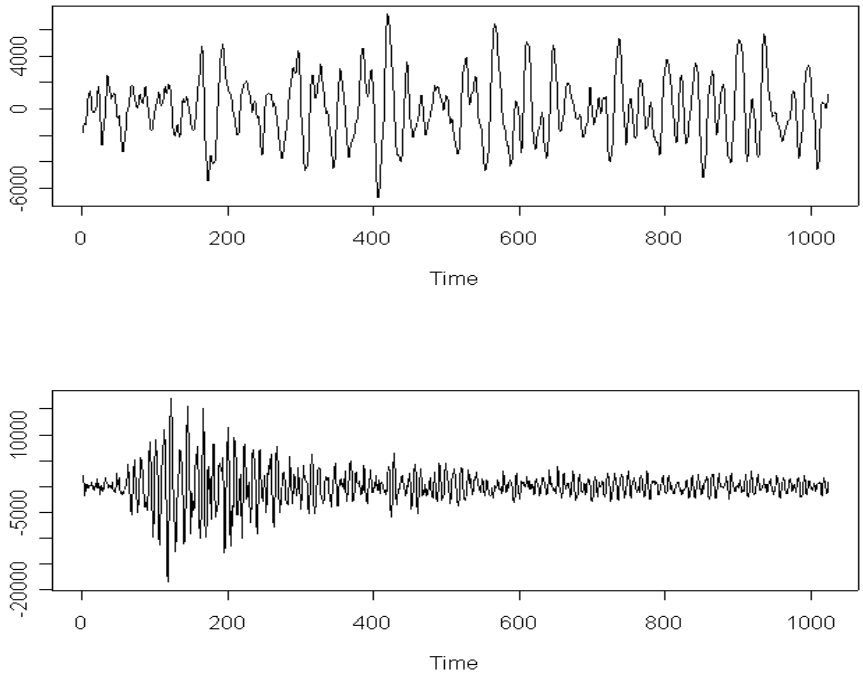


Figure 1.1: Arrival phases from S waves from an earthquake and explosion.

2.0 LITERATURE REVIEW

Many of the basic ideas which are the foundation of this thesis are from spectral analysis (*e.g.* [28]), Wahba's estimation method [29, 30], and spectral analysis of replicated time series [9]. These ideas are reviewed in this chapter.

2.1 BACKGROUND OF THE PROBLEM

Our problem in this thesis deals with spectral estimation which is based on periodograms (log-periodograms). Therefore, I briefly give an overview of the definitions and properties of the spectral density, the periodogram, the approximate confidence interval for spectral density at w, and smoothing spline analysis of variance in reproducing kernel Hilbert spaces.

2.1.1 Spectral Density

Given a stationary process $\{y_t\}$, $t = 0, \pm 1, \ldots$ with autocovariance function, $\gamma(h)$, which is absolutely summable[1], $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the spectral density, $f(\omega)$, of the process is defined as[2]

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h} \qquad -1/2 \leq \omega \leq 1/2, \tag{2.1}$$

[1] If this condition is not satisfied $f(\omega)$ is not continuous.

[2] A frequency is measured in cycles per time point. i.e., $\omega = \frac{1}{T}$ where the period, T, is measured the number of points in a cycle. Meanwhile, some books (*e.g.* [4]) use the frequency as $v = 2\pi\omega$ instead of ω. In this case, v is radians per unit time, and $-\pi \leq v \leq \pi$.

and

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} f(\omega) d\omega \qquad h = 0, \pm 1, \pm 2, \ldots$$

The spectral density is similar to a probability density in that $f(\omega) \geq 0$ for all ω, which holds because $\gamma(h)$ is a non-negative definite operator. The important property is that the $\text{Var}(y_t)$ can be expressed by $f(\omega)$ and $\gamma(h)$[3] in (2.2). It means that the variance of $\{y_t\}$ is the integrated spectral density over all of the frequencies.

$$\text{Var}[x_t] \equiv \gamma(0) = \int_{-1/2}^{1/2} f(\omega) d\omega. \tag{2.2}$$

The properties of $f(\omega)$ for all $-1/2 \leq \omega \leq 1/2$ are as follows.

1. **real-valued function**: $f(\omega)$ is a real value.
2. **nonnegative**: $f(\omega) \geq 0$.
3. **even function**: $f(\omega) = f(-\omega)$.
4. **period = 1**: $f(\omega + 1) = f(\omega)$.

2.1.2 Discrete Fourier Transform (DFT) & Periodogram.

If time series data $y_1, y_2, \ldots, y_n$ are given, the discrete Fourier transform of this data is defined as

$$d(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t e^{-2\pi i\omega_k t} \tag{2.3}$$

for $k = 0, 1, 2, \ldots, n-1$, where the frequencies $\omega_k = \frac{k}{n}$ are called *Fourier frequencies*.
The periodogram is then defined as

$$I(\omega_k) = |d(\omega_k)|^2 = \frac{1}{n} \left| \sum_{t=1}^{n} y_t e^{-2\pi i\omega_k t} \right|^2 \tag{2.4}$$

for $k = 0, 1, 2, \ldots, n-1$, with $I(0) = n\bar{x}^2$.

[3] $\gamma(h)$ is also the characteristic function of the $f(\omega)$. In other words, $\gamma(h)$ and $f(\omega)$ have the same information about the process. The difference is that $\gamma(h)$ is defined in the time domain and $f(\omega)$ is defined in the frequency domain.

$I(\omega_k)$ is also called the sample spectral density and it is the sample variance of $y_1, y_2, \ldots, y_n$ at frequency ω_k. The important property of $I(\omega_k)$ is that $I(\omega_{k_1})$ and $I(\omega_{k_2})$ are asymptotically independent for each k_1 and k_2 ($k_1 \neq k_2$).

$I(\omega_k)$ is distributed asymptotically as $f(\omega_k)$ times a random variable with a gamma or chi-squared distribution at each ω_k, and this fact is used to find an approximate confidence interval for $f(\omega_k)$ [28].

If $y_1, y_2, \ldots, y_n$ is a stationary Gaussian time series with $E(y_t) = 0$ and satisfies the condition, $\sum_{h=-\infty}^{\infty} |h||\gamma(h)| < \infty$, then as $n \to \infty$,

$$\frac{2I(\omega_k)}{f(\omega_k)} \sim \chi_2^2. \tag{2.5}$$

Because $E[\frac{2I(\omega_k)}{f(\omega_k)}] = \frac{2E[I(\omega_k)]}{f(\omega_k)} = 2 = E(\chi_2^2)$, $E[I(\omega_k)] = f(\omega_k)$ and because $\text{Var}[\frac{2I(\omega_k)}{f(\omega_k)}] = \frac{4\text{Var}[I(\omega_k)]}{f^2(\omega_k)} = 4 = \text{Var}(\chi_2^2)$, $\text{Var}[I(\omega_k)] = f^2(\omega_k)$. Thus, $I(\omega_k)$ is an asymptotically unbiased estimator of $f(\omega_k)$ but is not consistent since $f^2(\omega_k)$ does not converge to 0 as $n \to \infty$.

2.1.3 Approximate Confidence Interval for the Spectral Density

From equation (2.5), we can find an approximate $100(1-\alpha)\%$ confidence interval for the spectral density $f(\omega_k)$ at ω_k, that is,

$$\frac{2I(\omega_k)}{\chi_2^2(1-\alpha/2)} \leq f(\omega_k) \leq \frac{2I(\omega_k)}{\chi_2^2(\alpha/2)}. \tag{2.6}$$

where $\chi_v^2(\alpha)$ denote the lower α probability tail for the chi-squared distribution with v degrees of freedom. Moreover, in the case of the replicated time series, suppose the data $\{y_{it} : i = 1, 2, \ldots, N; t = 1, 2, \ldots, n\}$ with each series having the density, $f(\omega_k)$, constitute N mutually independent partial realizations of $\{Y_t\}$, and thus the observed $I_1(\omega_k)$, $I_2(\omega_k)$, $\ldots$, $I_N(\omega_k)$ are mutually independent periodograms at w_k. Because the average of N independent

$\gamma(1,1)$ random variables is $\gamma(N, \frac{1}{N})$, $\bar{I}_i(\omega_k) \sim f(\omega_k) \times \gamma(N, \frac{1}{N})$. Thus, an approximate $100(1-\alpha)\%$ confidence interval for the spectral density at w_k is

$$\frac{2N\bar{I}(\omega_k)}{\chi^2_{2N}(1-\alpha/2)} \leq f(\omega_k) \leq \frac{2N\bar{I}(\omega_k)}{\chi^2_{2N}(\alpha/2)} \tag{2.7}$$

In the case of an approximate confidence interval for $f_1(\omega_k)/f_2(\omega_k)$, suppose the data are $\{z_{it} : i = 1, 2, \ldots, N_1; t = 1, 2, \ldots, n\}$ with each series having the density, $f_1(\omega_k)$, and $\{y_{it} : i = 1, 2, \ldots, N_2; t = 1, 2, \ldots, n\}$ with each series having the density, $f_2(\omega_k)$. This data constitutes N_1 and N_2 mutually independent partial realizations of $\{Z_t\}$ and $\{Y_t\}$ respectively. The observed replicated periodograms, $I_{1,1}(\omega_k), I_{1,2}(\omega_k), \ldots, I_{1,N_1}(\omega_k)$ and $I_{2,1}(\omega_k), I_{2,2}(\omega_k), \ldots, I_{2,N_2}(\omega_k)$ are mutually independent at ω_k. Similar to (2.7),

$$\frac{2N_i\bar{I}_i(\omega_k)}{f_i(\omega_k)} \sim \chi^2_{2N_i} \quad i = 1, 2. \tag{2.8}$$

Thus,

$$\frac{\bar{I}_i(\omega_k)}{f_i(\omega_k)} \sim \frac{\chi^2_{2N_i}}{2N_i} \quad i = 1, 2. \tag{2.9}$$

Because $\bar{I}_i(\omega_k)/f_i(\omega_k)$ are independent for $i = 1, 2$, it follows from (2.9) that their ratio follows an F distribution with degrees of freedom $2N_1$ and $2N_2$:

$$\frac{\bar{I}_1(\omega_k)/f_1(\omega_k)}{\bar{I}_2(\omega_k)/f_2(\omega_k)} \sim F_{2N_1,2N_2}. \tag{2.10}$$

From (2.10), an approximate $100(1-\alpha)\%$ confidence interval for the ratio, $\frac{f_1(\omega_k)}{f_2(\omega_k)}$, of the spectral densities at ω_k is,

$$\frac{\bar{I}_1(\omega_k)}{\bar{I}_2(\omega_k)F_{2N_1,2N_2}(1-\alpha/2)} \leq \frac{f_1(\omega_k)}{f_2(\omega_k)} \leq \frac{\bar{I}_1(\omega_k)}{\bar{I}_2(\omega_k)F_{2N_1,2N_2}(\alpha/2)} \tag{2.11}$$

When $N_1 = N_2 = N$, we can use the fact that $\frac{1}{F_{N,N}(\alpha)} = F_{N,N}(1-\alpha)$, to show that (2.11) can be expressed as,

$$\frac{\bar{I}_1(\omega_k)}{\bar{I}_2(\omega_k)}F_{2N,2N}(\alpha/2) \leq \frac{f_1(\omega_k)}{f_2(\omega_k)} \leq \frac{\bar{I}_1(\omega_k)}{\bar{I}_2(\omega_k)}F_{2N,2N}(1-\alpha/2) \tag{2.12}$$

$F_{2N,2N}(\alpha/2)$ increases and $F_{2N,2N}(1-\alpha/2)$ decreases as $N \to \infty$. Hence, if enough replicated time series are available, we can get a useful approximate confidence interval from equation (2.12). Note that $\lim_{N\to\infty} \frac{\bar{I}_1(\omega_k)}{\bar{I}_2(\omega_k)}F_{2N,2N}(\alpha) = \frac{f_1(\omega_k)}{f_2(\omega_k)}$ because $F_{2N,2N}(\alpha) \xrightarrow{p} 1$ as $N \to \infty$.

2.1.4 Analysis of Variance in Reproducing Kernel Hilbert Spaces (RKHS)

We begin with the assumption f is in some RKHS, that is, a Hilbert space, $\mathcal{H}$, of functions in which all the point evaluations are bounded [2, 14, 32]. Let $\mathcal{H}$ be some RKHS of real-valued of functions of $\boldsymbol{t} = (t_1, t_2, \ldots, t_d)$ and $\boldsymbol{t} \in \mathcal{T} = \mathcal{T}^{(1)} \otimes \mathcal{T}^{(2)} \cdots, \otimes \mathcal{T}^{(d)}$, where t_a is the ath variable in $\mathcal{T}^{(a)}$ and $\mathcal{T}^{(a)}$ is some measurable space. Now we construct a probability measure $d\mu_a$ on $\mathcal{T}^{(a)}$ for each $a = 1, 2, \ldots, d$ with the symbol $(A_a f)(\boldsymbol{t})$, defined by

$$(A_a f)(\boldsymbol{t}) = \int_{\mathcal{T}^{(a)}} f(t_1, \ldots, t_d) d\mu_a(t_a) \tag{2.13}$$

is well defined and finite and for every $f \in \mathcal{H}$ and $\boldsymbol{t} \in \mathcal{T}$.

We can consider A_a as an operator from $\mathcal{H}$ to $\mathcal{H}$ and the decomposition of the identity, I, as

$$\begin{aligned} I &= \prod_a [A_a + (I - A_a)] \\ &= \prod_a A_a + \sum_a (I - A_a) \prod_{b \neq a} A_b \\ &+ \sum_{a<b} (I - A_a)(I - A_b) \prod_{c \neq a,b} A_c + \cdots + \prod_a (I - A_a). \end{aligned} \tag{2.14}$$

In general, f has an unique representation of the form in SS-ANOVA [14]

$$f(\boldsymbol{t}) = C + \sum_a f_a(t_a) + \sum_{a<b} f_{ab}(t_a, t_b) + \sum_{a<b<c} f_{abc}(t_a, t_b, t_c) + \cdots \tag{2.15}$$

where $C = (\prod_a A_a) f$ is the mean, $f_a = [(I - A_a) \prod_{b \neq a} A_b] f$ are the main effects, and $f_{ab} = [(I - A_a)(I - A_b) \prod_{c \neq a,b} A_c] f$ are the two-factor interactions in (2.14), etc.

We will use this idea with $d = 1$, that is, $a = 1$, and $f_0 = (A_a) f \in \mathcal{H}_0$ and $f_1 = (I - A_a) f \in \mathcal{H}_1$ in Section 3.1. The important thing is that the penalty functional that we wish to use, is $||(I - A_a) f||^2_{\mathcal{H}}$ where $(I - A_a)$ is the orthogonal projection of f onto $\mathcal{H}_1$ [the notation P_1 is used for $(I - A_a)$ and $\mathcal{H}_\alpha$ is used for $\mathcal{H}$ in Section 3.1].

Lastly, note that some papers describe the case where $d = 2$ (*e.g.* the time-frequency functional model). Thus, $\mathcal{H}$ can be expressed by tensor (direct) sum ($\oplus$) and tensor product ($\otimes$) of the corresponding subspaces. See [15, 24] for details.

2.2 LITERATURE REVIEW

We review Wahba [29] and Diggle and Al Wasel [9] in this section. These two papers provided motivation for our research.

2.2.1 Estimation Method of Wahba

Cogburn and Davis [7] suggested the log periodogram as an estimator for the log spectral density. This idea was extended by Wahba [29], which she refers to as an optimally smoothed spline (OSS). Her important idea was that the periodograms (log-periodograms) of a stationary time series are asymptotically independent at different Fourier frequencies. Therefore, we can apply classical statistical methods to time series data. The asymptotic distribution of a periodogram is shown in equation (2.5).

In her paper [29], the goal was to estimate $\log f(\omega)$ from the data $y_1, y_2, \ldots, y_{2n}$ (that is, the total number of data is even). Wahba noted that the plot of $I(\omega)$ is uselessly "wiggly", even though $f(\omega)$ is smooth. Therefore, she proposed a smoothing estimation method for $\log f(\omega)$ based on $\log I(\omega)$ obtained by fitting a smoothing spline through minimizing the expected mean square errors of the data.

Let $Y_k = \log I_k + C_k$, where $C_k = 0.57721$ for $k = \pm 1, 2, \ldots, n-1$ and $C_0 = C_n = (\log 2 + 0.57721)/\pi = 0.40437$. Then, $Y_k = \log f(k/2n) + \epsilon_k$, where $E(\epsilon_k) = 0, E(\epsilon_k^2) = \mathrm{Var}(\epsilon_k) = \pi^2/6$ for $k = -(n-1), \ldots, (n-1)$; for details, see Bateman [3].

Let $g(\omega) = \log f(\omega)$, and define the integrated mean square error $R_n(\lambda, m)$ where λ controls the width of smoothing splines and m controls the steepness, as

$$R_n(\lambda, m) = \int_{-1/2}^{1/2} [g_{n,m,\lambda}(\omega) - g(\omega)]^2 d\omega \tag{2.16}$$

The estimate $\hat{g}_{n,m,\lambda}(\omega)$ for $g(\omega)$ is obtained through minimizing $R_n(\lambda, m)$:

$$\hat{g}_{n,m,\lambda}(\omega) = \sum_{v=-(n-1)}^{n} \frac{\tilde{g}_v}{(1 + \lambda(2\pi v)^{2m})} \exp(2\pi i v \omega), \tag{2.17}$$

where

$$\tilde{g}_v = \frac{1}{2n} \sum_{k=-(n-1)}^{n} Y_k \exp(-2\pi i v k/2n). \tag{2.18}$$

The proof that (2.17) minimizes (2.16) is given in Wahba [29].

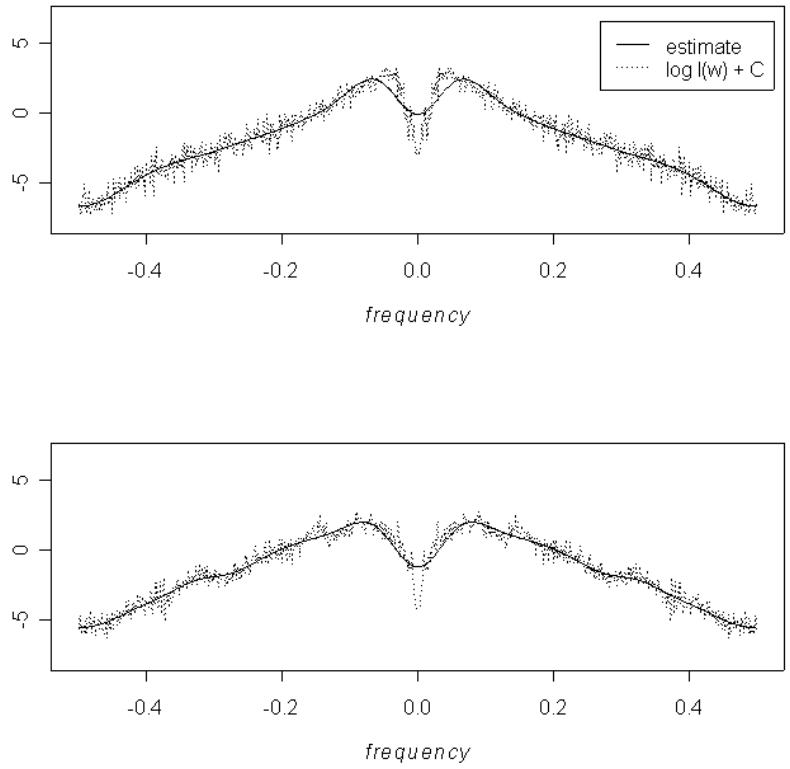


Figure 2.1: S components of earthquake and explosion.

We plotted S components of earthquakes and explosions [28] by (2.17) with **R** statistical package where $\hat{g}_{n,m,\lambda}(\omega)$ with $m = 4$ and $\lambda = 10^{-14}$ on the log-periodograms in Figure 2.1 and the **R** code is shown in Appendix D.2. One thing we have to concern is that we have to calculate the $C_k' = -E(\log Y)$ where $Y \sim \gamma(8, 1/8)$ because Figure 2.1 is based on the mean periodograms of 8 earthquakes and 8 explosions respectively. That is, $C_k' = 0.0637944$ by Bateman [3] instead of $C_k = 0.57721$, the Euler Mascheroni constant, is from $-E(\log X)$

where $X \sim \gamma(1,1)$.

The method proposed by Wahba seems to fit the data fairly well. However, her method used the simple mean periodogram and did not account for subject effects. Hence, that approach is suitable for single time series data but not for replicated data when the goal of analysis includes the subject effects.

2.2.2 Diggle and Al Wasel [DWA]'s Model

Most of the methodological development has been devoted toward a single time series as described in Subsection 2.2.1. The paper by [DWA] [9] provides a general framework for spectral analysis of replicated time series. If replicated time series are available, our concern is naturally about the estimation of population characteristics (parameters), rather than on the behavior of the individual time series.

Diggle and Al Wasel (1997) proposed the method of spectral analysis to interpret biomedical time series data involving observations of time series from random samples of subjects. Their data consist of luteinizing hormone concentrations in blood samples taken from each subject at 1 min intervals during 1 hour. The original data were filtered by a weighted seven-points moving average with weight of 1, 3, 6, 7, 6, 3 and 1. After filtering, $54 = 60 - (3 * 2)$ observations are left for each subject.

In their paper, they proposed the model (2.19) to estimate $f(\omega)$, using a log-linear transformation. There is no assumption of a common population spectrum for all subjects.
Let $f_i(\omega_k)$ denote the population spectrum corresponding to the ith subject and kth frequency.

$$\log f_i(\omega_k) = \sum_{j=1}^{p} d_{ikj}\beta_j \qquad i = 1, 2, \ldots, 8, \quad k = 1, 2, \ldots, 26 = [(54-1)/2], \tag{2.19}$$

where the d_{ikj} are known explanatory variables and β_j are parameters to be estimated.

The model (2.19) looks similar to our proposed model (3.8) to be presented Chapter 3, with respect to replicated spectral analysis. However, β_j does not depend on i or k. This means the coefficients do not vary with frequencies in (2.19). This paper gave us the motivation to propose a model which has coefficients which change with different frequencies. The approach and model of DWA do not give us enough information when our goal is to determine the frequencies where the spectrum of one group may differ from that of another.

This paper [9] also indicated that variability between subjects in periodograms at given frequencies is larger than the variability predicted by asymptotic distribution theory for replicated time series.

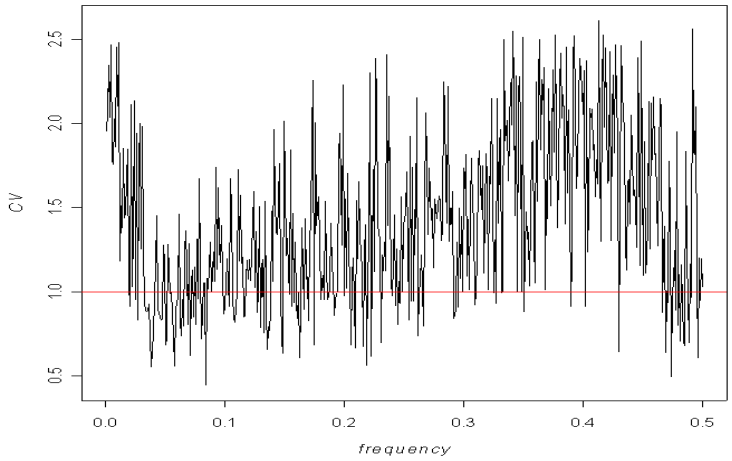


Figure 2.2: Sample coefficient of variation.

For example, we plotted the sample coefficients of variations (CVs) against frequencies. [The **R** code is shown in Appendix D.3.] The data consist of eight S components of earthquakes. Let I_{ik} denote the periodogram for the ith time series at the kth Fourier frequency, for $i = 1, 2, \ldots, 8$ and $k = 1, 2, \ldots, 512$. According to standard asymptotic theory, for each value of k the eight values $I_{1k}, I_{2k}, \ldots, I_{8k}$ are an independent random sample from the $\gamma(1, f_k)$ distribution. Therefore, we can expect for each sample coefficient of variation,

$CV_k = s_k/\bar{I}_k \approx 1$. In Figure 2.2, the averages of the sample CVs are substantially greater than 1 and this supports the argument of DAW that the standard theory does not hold in the homogeneous cases. Therefore, when we analyze replicated time series, instead of using mean periodogram, we will consider the variations between replicated time series at each frequency.

3.0 ESTIMATION METHOD

In this chapter, we present the model (modified model) and a new estimation method. We briefly describe the idea of smoothing spline ANOVA and the penalty functional in Section 3.1. We present our model and the modified model that we use for our estimation in Section 3.2 and the method to find estimation function in Section 3.3.

3.1 SMOOTHING SPLINE ESTIMATION AND THE PENALTY FUNCTIONAL

We briefly describe about smoothing spline ANOVA (SS-ANOVA) and the orthogonal projections onto the reproducing kernel Hilbert spaces. These ideas are used in smoothness for functions on the frequency domain in Section 3.3.

We begin with Taylor's theorem with remainder: If f is a real valued function on [0, 1] with $f', f'', \ldots, f^{(\alpha-1)}$ continuous derivatives and $f^{(\alpha)} \in \mathcal{L}_2[0,1]$, then f can be expressed as,

$$f(x) = \sum_{v=0}^{\alpha-1} \frac{x^v}{v!} f^{(v)}(0) + \int_0^1 \frac{(x-u)_+^{\alpha-1}}{(\alpha-1)!} f^{(\alpha)}(u) du, \tag{3.1}$$

where $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ if $x < 0$.

Now, we define Hilbert space $\mathcal{H}_\alpha$[1]: $\mathcal{H}_\alpha[0,1] = \{f : f, f', f'' \ldots, f^{(\alpha-1)}$ absolutely continuous and $f^{(\alpha)} \in \mathcal{L}_2\}$. Then, each function f in $\mathcal{H}_\alpha$ has a Taylor series expansion (3.1) to

[1]It is the so-called Sobolev Hilbert space. See [1] for details.

order α and has a decomposition as $f = f_0 + f_1$ with $f_0 \in \mathcal{H}_0$ and $f_1 \in \mathcal{H}_1$ given by the first and the second terms in (3.1). Moreover, $\int_0^1 [(D^\alpha f_0)(u)]^2 du = 0$ and $\sum_{v=0}^{\alpha-1} [(D^v f_1)(0)]^2 = 0$ where D^α denotes the αth derivative of function [32]. Therefore, we can represent $\mathcal{H}_\alpha$ using direct sum notation,

$$\mathcal{H}_\alpha = \mathcal{H}_0 \oplus \mathcal{H}_1. \tag{3.2}$$

If we consider the square norm of $f \in \mathcal{H}_\alpha$,

$$||f||^2 = \sum_{v=0}^{\alpha-1} [(D^v f)(0)]^2 + \int_0^1 [(D^\alpha f)(u)]^2 du, \tag{3.3}$$

This means that subspaces $\mathcal{H}_0$ and $\mathcal{H}_1$ of $\mathcal{H}_\alpha$ are perpendicular i.e., $\mathcal{H}_0 \perp \mathcal{H}_1$. Note that we use the penalty functional $\int_0^1 [f^\alpha(u)]^2 = ||P_1 f||^2_{\mathcal{H}_\alpha}$ where P_1 is the orthogonal projection of f onto $\mathcal{H}_1$ in $\mathcal{H}_\alpha$ in our research.

In general, given the series $y_1, y_2, \ldots y_n$, the estimation method of f ($f \in \mathcal{H}_\alpha$) by the penalized least squares is to minimize

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - D_i f)^2 + \lambda ||P_1 f||^2_{\mathcal{H}_\alpha}, \tag{3.4}$$

where P_1 is as before, and D_i is a design matrix of fixed values, λ is a smoothing parameter i.e., $\lambda \geq 0$ and as λ is increased, the estimated function is smoother.
We just extend the penalty functional to multi-covariate case in Section 3.3. Thus, we also need subsidiary smoothing parameters θ depending on covariates.

3.2 THE MODEL

In this section, $i = 1, 2, \ldots, l$ indexes the replicated time series, $k = 1, 2, \ldots, s$ indexes the indices for the Fourier frequencies, ω_k, and $p = 1, 2, \ldots, P$ indexes the covariates.

Let $g_i(\omega) = \log[f_i(\omega)]$. Our proposed functional model for $g_i(\omega)$ is

$$g_i(\omega_k) = D_i\boldsymbol{\beta}(\omega_k), \tag{3.5}$$

where $D_i = \{D_i[1], \ldots, D_i[P]\}$ is a design matrix that consists of measured covariates and dummy variables, and $\boldsymbol{\beta}(\omega_k)=\{\beta_1(\omega_k), \ldots, \beta_P(\omega_k)\}^T$ is a vector of frequency coefficients at ω_k. Note that the $\boldsymbol{\beta}(\omega_k)$ are periodic functions with period 1, i.e., $\boldsymbol{\beta}(\omega_k) = \boldsymbol{\beta}(\omega_k + 1)$, and are symmetric around $\omega_k = 0.5$ i.e., $\boldsymbol{\beta}(\omega_k) = \boldsymbol{\beta}(1 - \omega_k)$ in the frequency domain. For convenience, we only use half of the frequency domain in (3.5) to be free of the periodic constraint.

Let $y_{ik} = \log(I_i(\omega_k)) + C_{\omega_k}$, where $C_{\omega_k} = 0.57721$ for $\omega_k \neq 0, 1/2$ and $C_{\omega_k} = (\log 2 + 0.57721)/\pi = 0.4043$ for $\omega_k = 0, 1/2$. We have the following frequency functional model for the log-periodograms,

$$\log I_i(\omega_k) \approx g_i(\omega_k) + \delta_{ik}, \tag{3.6}$$

where $E(\delta_{ik}) = -C_{\omega_k}$, $\text{Var}(\delta_{ik}) = \pi^2/6$ and δ_{ik} are asymptotically independent for all i and k [15, 24]. Using equation (3.5), model (3.6) can be modified as follows:

$$y_{ik} \approx D_i\boldsymbol{\beta}(\omega_k) + \epsilon_{ik}, \tag{3.7}$$

where ϵ_{ik} are asymptotically independent with mean 0 and variance $\pi^2/6$ for all i and k.

3.3 ESTIMATION METHOD (PENALIZED LEAST SQUARES ESTIMATE)

We apply the following penalized least squares (PLS) method to estimate $\boldsymbol{\beta}(\boldsymbol{\Omega})$ on the reproducing kernel Hilbert spaces (RKHS):

Let us denote $\boldsymbol{Y} = \{\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_l^T\}^T$ where $\boldsymbol{y}_i = \{y_{ik}, k = 1, 2, \ldots, s\}$, $\boldsymbol{\beta}(\boldsymbol{\Omega}) = \{\beta_1^T(\boldsymbol{\Omega}), \ldots, \beta_P^T(\boldsymbol{\Omega})\}^T$, and $\boldsymbol{D}[p] = \text{diag}\{\boldsymbol{D}_1[p], \ldots, \boldsymbol{D}_l[p]\}$ where $\boldsymbol{D}_i[p] = \text{diag}\{D_{i1}[p], \ldots, D_{is}[p]\}$. The PLS criterion we will use is:

$$\frac{1}{ls}\left|\boldsymbol{Y} - \sum_{p=1}^{P} \boldsymbol{D}[p]\beta_p(\boldsymbol{\Omega})\right|^2 + \sum_{p=1}^{P} \frac{\lambda}{\theta_p}|P_p\beta_p|^2, \tag{3.8}$$

where λ is the main smoothing parameter, θ_p's are subsidiary smoothing parameters, and P_p represents the orthogonal projections onto the RKHS in the smoothing spline ANOVA (SS-ANOVA) model.

Conditional on the smoothing parameters (λ and θ_p), Gu and Wahba [13] give the estimates that minimize (3.8) as follows:

$$\hat{\beta}_p(\boldsymbol{\Omega}) = U_p h_p + Q_p c_p \tag{3.9}$$

where $U_p = (1, \omega_j)_{j=1}^{ls}$, $Q_p = \theta_p \boldsymbol{R}_p(\boldsymbol{\Omega}, \boldsymbol{\Omega})$, and $\boldsymbol{R}_p(\boldsymbol{\Omega}, \boldsymbol{\Omega}) = \{R_p(\omega_j, \omega_{j'})\}_{j=1,j'=1}^{ls,ls}$, where $R_p(\omega_j, \omega_{j'}) = -k_4(|\omega_j - \omega_{j'}|)$, where $k_4(x) = B_4(x)/4!$ and $B_4(x) = x^4 - 2x^3 + x^2 - 1/30$ is the fourth-order Bernoulli polynomial [15, 32].

Let us denote $D = \{\boldsymbol{D}[1], \ldots, \boldsymbol{D}[P]\}$, $T = \text{diag}\{U_1, \ldots, U_P\}$, $H = \{h_1^T, \ldots, h_P^T\}^T$, $Q = \text{diag}\{Q_1, \ldots, Q_P\}$ and $C = \{c_1^T, , \ldots, c_P^T\}^T$. Then, we can find the PLS estimate of $\boldsymbol{\beta}(\boldsymbol{\Omega})$ that can be estimated from the PLS of (3.10) and that is $\hat{\boldsymbol{\beta}}(\boldsymbol{\Omega}) = TH + QC$.

$$\frac{1}{ls}|\boldsymbol{Y} - DTH - DQC|^2 + \lambda C^T Q C. \tag{3.10}$$

where

$$H = \{T^T D^T M^{-1} D T\}^{-1} T^T D^T M^{-1} \boldsymbol{Y}, \tag{3.11}$$

and

$$C = DM^{-1}\{I - DT(T^T D^T M^{-1} DT)^{-1} T^T D^T M^{-1}\}\boldsymbol{Y}, \tag{3.12}$$

where $M = DQD^T + ls\lambda I$.

Moreover, when we only need an estimate for a specific ω_0, we can also use the PLSE at any given frequency point ω_0 from equation (3.9):

$$\hat{\beta}_p(\omega_0) = U_{p0} h_p + Q_{p0} c_p, \tag{3.13}$$

where $U_{p0} = (1, \omega_0), Q_{p0} = \theta_p \boldsymbol{R}_p(\omega_0, \boldsymbol{\Omega})$ and $\boldsymbol{R}_p(\omega_0, \boldsymbol{\Omega}) = \{R_p(\omega_0, \omega_j)\}_{j=1}^{ls}$, for $p = 1, \ldots, P$.

We can consider methods to estimate parameters (λ and θ_p). There are a few ways [*e.g.* generalized cross validation (GCV) [26]] beside generalized maximum likelihood (GLM) [15, 24, 31] to find parameters. The GML criterion is,

$$V(\lambda, \theta_p) = \frac{\boldsymbol{Y}^T F_2 (F_2^T DQD^T F_2 + ls\lambda I)^{-1} F_2^T \boldsymbol{Y}}{[det(F_2^T DQD^T F_2 + ls\lambda I)]^{1/(ls-4)}} \tag{3.14}$$

where F_2 is given by the QR decomposition of DT: $DT = (F_1, F_2)\binom{G}{0}$ where (F_1, F_2) is orthogonal and G is upper triangular. Thus, the dimensions are $F_1 = ls \times (2P - 1)$ and $F_2 = ls \times 1$.

Lastly, we consider the estimation of variances at given frequencies. Numerous authors [13, 15, 24, 30] discuss variances of $\hat{\beta}(\omega_0)$ conditioning on the estimates of the λ and θ_p and Bayesian confidence intervals for $\hat{\beta}(\omega_0)$. The Bayesian confidence intervals have pointwise interpretation. Nychka [18] pointed out that they are curvewise confidence intervals and averaging confidence intervals across the all frequencies. That is, $100(1-\alpha)\%$ confidence interval is,

$$\frac{1}{ls}\sum_{k=1}^{ls} \Pr\left[\beta(w_k) \in \hat{\beta}(w_k) \pm z_{\alpha/2}\sqrt{\text{var}(\hat{\beta}(w_k)|\boldsymbol{Y})}\right] \approx 1 - \alpha. \tag{3.15}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard Normal distribution. (3.15) shows informative regions depending on the design covariates in the frequency domain.

Here, in equations (3.10) - (3.12), we find $\hat{\boldsymbol{\beta}}(\boldsymbol{\Omega})$ and estimated the associated standard errors via a bootstrapping method [17]. Then, we use two confidence intervals for $\boldsymbol{\beta}(\boldsymbol{\Omega})$ to compare the group spectra for all frequencies. These are the independent simultaneous confidence interval and the bootstrapping confidence interval [21]. See details in Chapter 4.

4.0 SIMULATION STUDIES

We apply our proposed method to the simulated replicated data series [$P = 1$ in (3.8)] that consist of two groups. Thus, we assigned 1 for group 1 and 0 for group 2 as their group-covariate to estimate $\beta_1(\omega)$ in (1.2). In Section 4.1, we introduce the simulation procedure and the simulated data. In Section 4.2, we construct two confidence intervals. The bootstrapping confidence interval for functions, one of two confidence intervals, is our original work motivated by Olshen [21]. Lastly, we evaluate performance of our proposed method by determining whether the true $\beta_1(\omega) = \log[f_1(\omega)/f_2(\omega)]$ lies within the approximate 95% confidence interval for $\beta_1(\omega)$ that is estimated by using $\hat{\beta}_1(\omega) = \log[\hat{f}_1(\omega)/\hat{f}_2(\omega)]$ for all ω. in Section 4.3,

4.1 SIMULATION

We conduct the simulation and this procedure consists of four steps:

Step 1. Simulation of replicated time series from given $AR(2)$ model.

Step 2. Calculation of log-periodograms from simulated replicated data.

Step 3. Estimation of the true log-spectra through our proposed method.

Step 4. Validation of the estimated log-spectra function.

For step 1, we simulated data whose size is 500, four series are from $x_t = 0.9x_{t-1} + 0.05x_{t-2} + \epsilon_t$, four series are from $x_t = x_{t-1} - 0.9x_{t-2} + \epsilon_t$, and four series are from $x_t = -0.9x_{t-1} + 0.05x_{t-2} + \epsilon_t$ where $\epsilon_t \sim N(0, 1)$. For step 2, we used (2.4). For step 3, all procedures are described in Section 3.3. For step 4, our proposed estimation method

is evaluated by checking whether the true log-spectra function lies within 95% confidence interval for the true log-spectra. In particular, we need to find the true log-spectra for Step 4. These are, $f_1(\omega) = \frac{1}{1.8125-1.71\cos(2\pi\omega)-0.1\cos(4\pi\omega)}$, $f_2(\omega) = \frac{1}{2.81-3.8\cos(2\pi\omega)+1.8\cos(4\pi\omega)}$, and $f_3(\omega) = \frac{1}{1.8125+1.71\cos(2\pi\omega)-0.1\cos(4\pi\omega)}$, respectively. The mathematical procedures to find the spectra from time series [$AR(2)$ model] are shown in Appendix B. Note that we will construct confidence intervals based on the estimated log-spectra function in Step 3.

Three log-spectral densities are given in Figure 4.1 and the centered log-periodograms, $\log I(\omega_k) + 0.57721$, from $f_1(\omega)$, $f_2(\omega)$ and $f_3(\omega)$ and the corresponding smoothing spline estimates are shown in Figure 4.2. We used (3.8) with smoothing parameters $\lambda = 0.001$ and $\theta_1 = 1$ to provide the smoothed estimates. The **R** code is shown in Appendix D.4 and Appendix D.5 for Figure 4.1 and Figure 4.2 respectively.

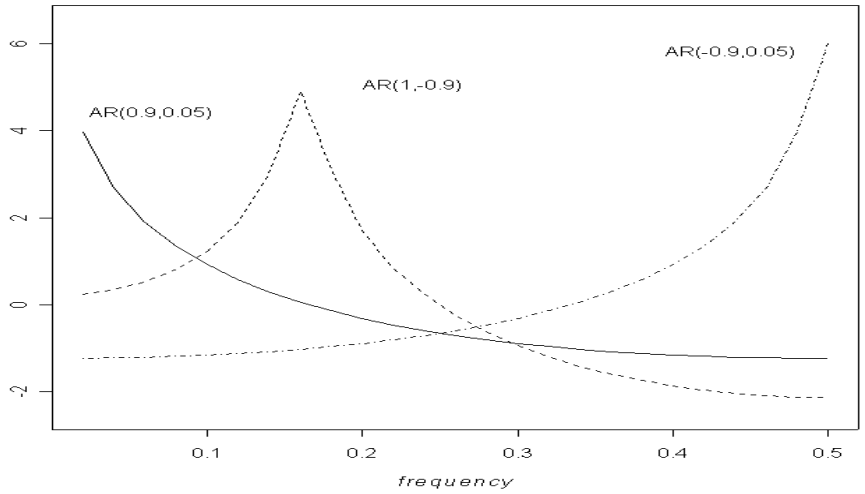


Figure 4.1: True simulated densities.

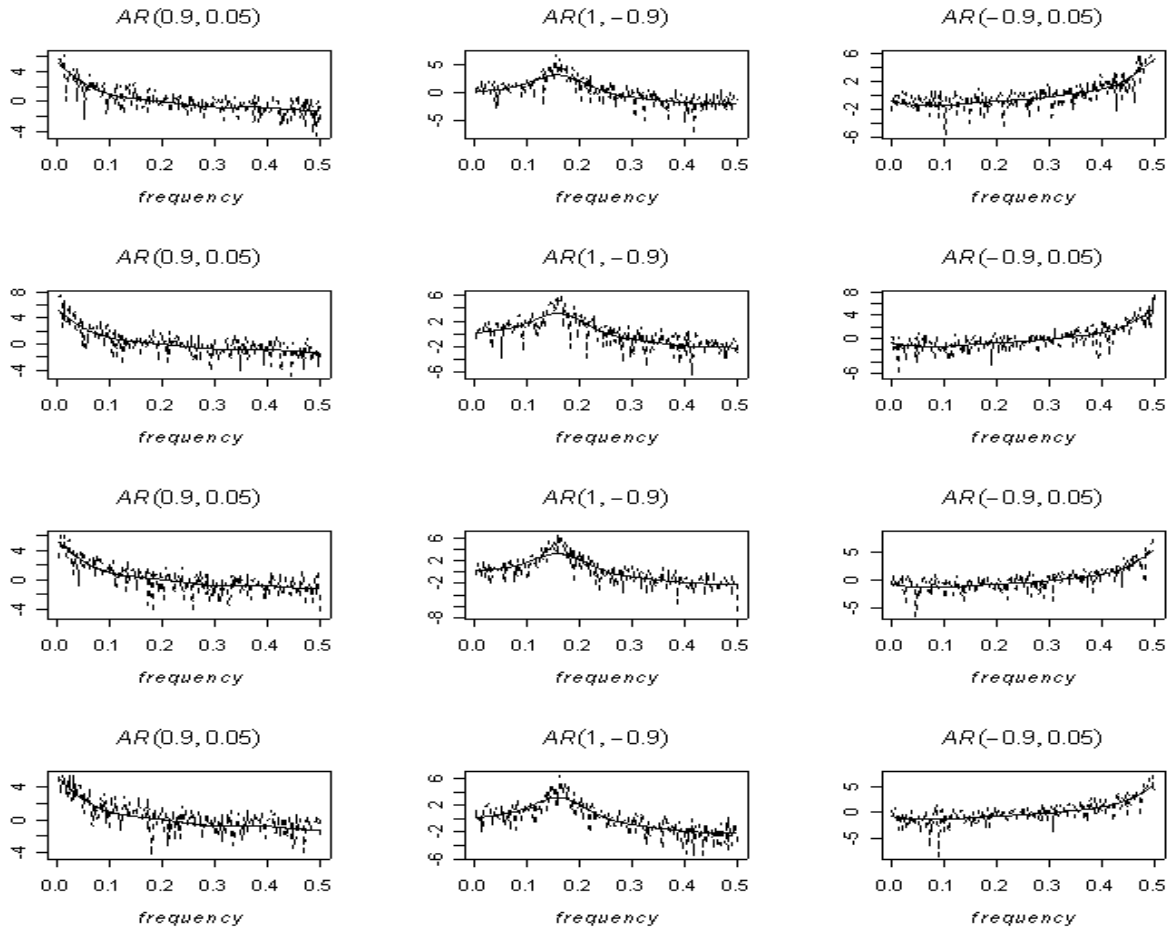


Figure 4.2: log-periodograms + 0.57721 from three simulated series and their fitted curves.

Our simulation study plans are as follows:

Simulation Study 1. We will compare $\log f_1(\omega)$ with $\log f_2(\omega)$ based on the estimated $\log f_1(\omega)$ and $\log f_2(\omega)$ for each ω. In particular, we will notice the estimates around $\omega = 0.17$. because there are sharp peak spectra around $\omega = 0.17$ of $\log f_2(\omega)[AR(1, -0.9)]$ in Figure 4.1 and Figure 4.2 and our estimation function is smooth.

Simulation Study 2. We will compare $\log f_1(\omega)$ with $\log f_3(\omega)$ based on the estimated $\log f_1(\omega)$ and $\log f_3(\omega)$ for each ω.

Before we continue our main analysis, we calculated the residuals in (3.7), their means and variances. Their means are very close to 0 and their variances are about 1.65 for all the simulation runs, which is close to the asymptotic variance $\sigma^2 = \pi^2/6$. Figure (4.3) shows the box plots of the residuals e_{ik} where $i = 1, 2, 3, 4$ and $k = 1, 2, \ldots, 250$ for three spectra settings.

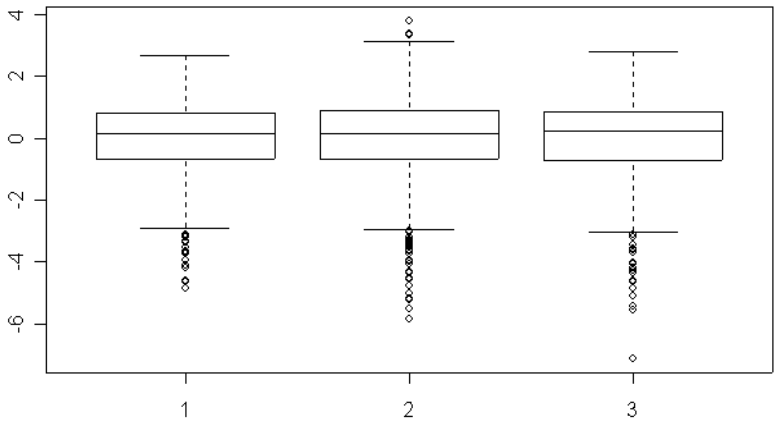


Figure 4.3: Panels from left to right: AR(0.9, 0.05), AR(1, -0.9) and AR(-0.9, 0.05).

4.2 APPROXIMATE CONFIDENCE INTERVAL

We develop two approximate confidence intervals and these confidence intervals will be used to determine the frequencies where the spectrum of one group may differ from that of another in Section 4.3. Now suppose we construct an approximate $100(1-\alpha)\%$ confidence interval for $\beta_1(\omega_k)$ [Note that we notate ω_k instead of ω in this section because our confidence intervals are constructed on all Fourier frequencies] in (1.3) using $\hat{\beta}_1(\omega_k) = \log[\hat{f}_1(\omega_k)] - \log[\hat{f}_2(\omega_k)] = \log[\hat{f}_1(\omega_k)/\hat{f}_2(\omega_k)]$ for each ω_k. Note that we choose the bootstrapping method for points [17] to find the standard errors of $\hat{\beta}_1(\omega_k)$ because we do not have any information about the

variations of $\log[\hat{f}_1(\omega_k)]$ and $\log[\hat{f}_2(\omega_k)]$.

4.2.1 Independent Simultaneous Confidence Interval

We can construct the confidence interval for $\beta_1(\omega_k)$ at each ω_k by

$$\hat{\beta}_1(\omega_k) \pm Z^* \sqrt{\text{Var}(\log \hat{f}_1(\omega_k)) + \text{Var}(\log \hat{f}_2(\omega_k))} \tag{4.1}$$

for $k = 1, 2, \ldots, s$, and Z^* can be predicted using the following procedure. We can find this confidence interval with the assumption that each spectrum (log-spectrum) is independent.

Let $A_k = \{\text{interval for } k\text{th frequency covers true value}\}$ and $P(A_k) = 1 - \alpha_0$ where α_0 is a marginal confidence interval for frequency ω_k. Since A_k are independent for $k = 1, 2, \ldots, s$, the Z^* for the $100(1-\alpha)\%$ confidence interval is,

$$1 - \alpha = P(\bigcap_{k=1}^{s} A_k) = \prod_{k=1}^{s} P(A_k) = (1 - \alpha_0)^s. \tag{4.2}$$

Thus, $\alpha_0 = 1-(1-\alpha)^{1/s}$ and $Z^* = Z_{1-\frac{\alpha_0}{2}} = Z_{(\frac{1}{2}+\frac{1}{2}(1-\alpha)^{1/s})}$ where Z_p is the $100(p)$ percentile of the standard normal distribution. Note that $\text{Var}(\log \hat{f}_i(\omega_k))$ for $i = 1, 2$, and $k = 1, 2, \ldots, s$, can be estimated by the bootstrapping method for points [17].

4.2.2 Bootstrapping Confidence Interval for Functions

This subsection is our original work in the sense that we use the bootstrapping confidence interval (for functions) for the differences between two groups $[\beta_1(\omega_k) = \log[f_1(\omega_k)] - \log[f_2(\omega_k)]]$ for all ω_k in the frequency domain. Previous study is by Olshen et al. [21] used this confidence interval for a group [gait analysis of children with the cycle by percent when they walk].

The simultaneous prediction regions are the form of

$$\bigcap_{\omega_k} \{\bar{\hat{\beta}}_1(\omega_k) - m\hat{\sigma}(\omega_k) \leq \tilde{\beta}_1(\omega_k) \leq \bar{\hat{\beta}}_1(\omega_k) + m\hat{\sigma}(\omega_k)\} \tag{4.3}$$

$$= \left\{ \tilde{\beta}_1(\omega_k) : \max_{\omega_k} \left| \frac{\tilde{\beta}_1(\omega_k) - \bar{\tilde{\beta}}_1(\omega_k)}{\hat{\sigma}(\omega_k)} \right| \leq m \right\}, \tag{4.4}$$

where $\beta_1(\omega_k)$ for $k = 1, 2, \ldots, s$ is estimated by $\bar{\hat{\beta}}_1(\omega_k)$, in which $\hat{\beta}_1(\omega_k)$ are unbiased estimates of $\beta_1(\omega_k)$ for replicated time series of the training samples, and $\bar{\hat{\beta}}_1(\omega_k)$ is the mean of $\hat{\beta}_1(\omega_k)$ for each ω_k [suppose we repeat the bootstrap sampling R times]. $\tilde{\beta}_1(\omega_k)$ are another unbiased estimates for $\beta_1(\omega_k)$, and m is a positive number. $\hat{\sigma}(\omega_k)$ is a standard error at ω_k and found estimated by the bootstrapping method for points [17]. Therefore, $\hat{\sigma}(\omega_k) = \sqrt{\text{Var}(\log \hat{f}_1(\omega_k)) + \text{Var}(\log \hat{f}_2(\omega_k))}$ for all ω_k in (4.1).

The important part to construct the bootstrapping confidence interval for functions is to estimate the probability of the event (4.3) depending on a fixed m by the bootstrapping method for functions [21].

Define $\hat{F}_B(m)$ with positive m and R repetition bootstrap sampling in (4.5). Suppose we have N random sample time series with replacement from our training samples,

$$\hat{F}_B(m) = \frac{1}{N} \# \left\{ \hat{\beta}_1(\omega_k) : \max_{\omega_k} \left| \frac{\hat{\beta}_1(\omega_k) - \bar{\hat{\beta}}_{1B}(\omega_k)}{\hat{\sigma}_B(\omega_k)} \right| \leq m \right\} \tag{4.5}$$

where $B = 1, 2, \ldots, R$. We repeat the bootstrap sampling R times, then estimate $\bar{\hat{F}}_B(m)$ by $(\hat{F}_1(m) + \hat{F}_2(m) + \cdots + \hat{F}_R(m))/R$ for a fixed m. $\bar{\hat{\beta}}_{1B}(\omega_k)$ is , as defined before, the average of the $R\hat{\beta}_1(\omega_k)$.

Again we define m_p by

$$m_p = \min\{m : \bar{\hat{F}}_B(m) \geq p\} \qquad \text{where} \quad 0 \leq p \leq 1. \tag{4.6}$$

p is our estimate of the probability of (4.3) if $m = m_p$. That is, $\tilde{\beta}_1(\omega_k)$ will lie between $\bar{\hat{\beta}}_1(\omega_k) - m_p\hat{\sigma}_B(\omega_k)$ and $\bar{\hat{\beta}}_1(\omega_k) + m_p\hat{\sigma}_B(\omega_k)$ for all ω_k.

Figure 4.4 shows the example of $\bar{\hat{F}}_B(m)$ as a function of m_p when group 1 is the simulated series from $AR(0.9, 0.05)$ and group 2 is the simulated series from $AR(1, -0.9)$ with $N = 20$

and $R = 10$ in (4.5). We can realize that the magnitude of the jump points of $\bar{\hat{F}}(m)$ is $1/NR$. Thus, we can get smoother curve as N and R are increased. We can notice $\bar{\hat{F}}(m) = 0.95$ when $m_{0.95} = 2.85$, and $m_{0.95} = 2.85$ will be used to construct 95% bootstrapping confidence interval for $\beta_1(\omega)$ when we compare $f_1(\omega)$ with $f_2(\omega)$ in Section 4.3. The **R** code is shown in Appendix D.7.

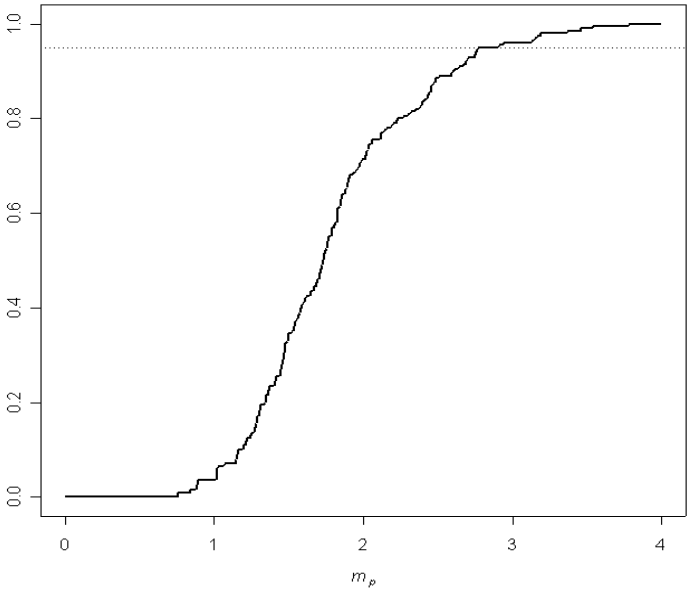


Figure 4.4: Equation 4.5 when group 1 = AR(0.9, 0.05) and group 2 = AR(1, -0.9).

$\bar{\hat{F}}_B(m)$ approximates uniformly in m the corresponding theoretical bootstrap probability with high probability as $R \to \infty$ [Babu and Singh [6]]. Therefore, we can predict the approximate $100(p)\%$ confidence interval for $\beta_1(\omega_k)$ between $\bar{\hat{\beta}}_1(\omega_k) - m_p\hat{\sigma}_B(\omega_k)$ and $\bar{\hat{\beta}}_1(\omega_k) + m_p\hat{\sigma}_B(\omega_k)$ as $R \to \infty$ for all ω_k.

The proof is based on the first theorem (see [6]). In order for the theorem to apply, we need to show that conditions $A1$, $A2$ and $A3$ below are satisfied. Let $F_1(\omega)$ and $F_2(\omega)$ be the distribution functions of the log spectral densities i.e., $\log f_1(\omega)$ and $\log f_2(\omega)$, where $f_1(\omega)$ = group 1 and $f_2(\omega)$ = group 2 respectively. We will take n_1 independent random samples series of $\log f_1(\omega)$ and n_2 from $\log f_2(\omega)$ with each series having size s. Let $R = n_1 + n_2$.

Define $\hat{F}_1(\omega_k)$ and $\hat{F}_2(\omega_k)$ be the empirical distribution functions of the random variables X_{1k} and X_{2k} where $X_{1k} = \log \hat{f}_1(\omega_k)$ and $X_{2k} = \log \hat{f}_2(\omega_k)$ for a specific ω_k. See [pp 999, [6]] for details.

A1: The n_i tends to infinity at the same rate. In other words, the $R/n_i \leq \lambda < \infty$ for $i = 1, 2$.

A2: For at least one i, $F_i(\omega)$ is continuous.

A3: $F_i(\omega)$ has finite 6th moment for $i = 1, 2$.

For $A1$, since $n_1 = n_2$ in our case, $R/n_i = 2n_i/n_i = 2 < \infty$ for $i = 1, 2$.

For $A2$, it is sufficient to show $f(\omega)$ is continuous. Given a stationary process $\{y_t\}_{t=1}^n$, $t = 0, \pm 1, \ldots$ with $\text{var}(y_t) = \sigma^2 < \infty$, suppose the autocovariance function, $\gamma(h)$, of a process satisfies the absolute summable condition, $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then by the inverse transform of the spectral density, $f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h} = \sum_{h=-\infty}^{\infty} \gamma(h)[\cos(2\pi\omega h) - i \sin(2\pi\omega h)]$ $= \sigma^2 + 2\sum_{h=1}^{\infty} \gamma(h) \cos(2\pi\omega h)$.

Given any $\epsilon > 0$, let a $\delta(\epsilon) = \epsilon$. Then for any frequencies, ω_0, in (-1/2, 1/2) if $0 < |2\sum_{h=1}^{\infty} \gamma(h)[\cos(2\pi\omega h) - \cos(2\pi\omega_0 h)]| < \delta(\epsilon)$, we trivially have

$$\begin{aligned} |f(\omega) - f(\omega_0)| &= \left| \sigma^2 + 2\sum_{h=1}^{\infty} \gamma(h) \cos(2\pi\omega h) - [\sigma^2 + 2\sum_{h=1}^{\infty} \gamma(h) \cos(2\pi\omega h)] \right| \\ &= \left| 2\sum_{h=1}^{\infty} \gamma(h)[\cos(2\pi\omega h) - \cos(2\pi\omega_0 h)] \right| \\ &< \delta(\epsilon) = \epsilon. \end{aligned} \tag{4.7}$$

For $A3$, assume $i = 1$ and let a random variable W be from $F_1(\omega)$, the distribution function of $\log f_1(\omega)$ (= group 1), where $W \in (-1/2, 1/2)$. By the condition $A2$, we can express the

6th moment of $F_1(\omega)$ as,

$$\begin{aligned} E[(W)^6] &= \int_{-1/2}^{1/2} W^6 \log f_1(\omega) d\omega < \int_{-1/2}^{1/2} W^6 f_1(\omega) d\omega \\ &\leq \int_{-1/2}^{1/2} (0.5)^6 f_1(\omega) d\omega = (0.5)^6 \int_{-1/2}^{1/2} f_1(\omega) d\omega \\ &= (0.5)^6 \sigma_1^2 < \infty. \end{aligned} \tag{4.8}$$

Since $f_1(\omega) \neq 0$ for all ω [if $f_1(\omega) = 0$ for all ω, then $\{y_{1t}\}_{t=1}^n$ have the same values because their variances, $\text{var}(y_{1t}) = \sigma_1^2 = 0$] and by the assumption of $\sigma_1^2 < \infty$. The same proof is for $i = 2$. Note that the spectral distribution function is bounded [28]. Thus, all moments of the spectral distribution function must exist and these moments are finite. This fact can be extended to the spectral distribution function of log density.

4.3 PERFORMANCE OUR ESTIMATION METHOD

To evaluate the performance of our proposed method in Chapter 3, we use the simulated data in Section 4.2 to construct the confidence intervals in Section 4.3.

Suppose we compare $f_1(\omega)$ with $f_2(\omega)$ based on replicated series from two groups. the procedure consists of 4 steps:

Step 1. We will calculate $\hat{\beta}_1(\omega)$ from $\log \hat{f}_1(\omega)$ and $\log \hat{f}_2(\omega)$ for all ω.

Step 2. We will construct the approximate 95% confidence intervals for $\beta_1(\omega)$ using the $\hat{\beta}_1(\omega)$ in Step 1.

Step 3. We will calculate the true $\beta_1(\omega)$ from $\log f_1(\omega)$ and $\log f_2(\omega)$.

Step 4. We will determine whether the true $\beta_1(\omega)$'s lie within the confidence intervals that are derived from Step 2 for all ω.

For Step 1, we calculate $\hat{\beta}_1(\omega_k) = \log[\hat{f}_1(\omega_k)/\hat{f}_2(\omega_k)]$ because $\omega_k \to \omega$ as the number of sample size, n, is increased. For Step 2, the independent simultaneous confidence interval (Subsection 4.2.1) and the bootstrapping confidence interval for functions (Subsection 4.2.2) are used. As previously mentioned in Subsection 4.2.2, we use $N = 20$ and $R = 10$ to construct the bootstrapping confidence interval for functions. For Step 3, $\beta_1(\omega) = \log[f_1(\omega)/f_2(\omega)]$ where $f_1(\omega)$ and $f_2(\omega)$ are given in page 23. For Step 4, if true $\beta_1(\omega)$'s lie within the confidence intervals means that our proposed estimation method will be useful to analyze real time series.

4.3.1 Evaluation of Performance 1:

We wish to compare $f_1(\omega)$ with $f_2(\omega)$. Thus, we calculate $\hat{\beta}_1(\omega_k)$ from 4 replicated time series data that are obtained from $AR(0.09, 0.05)$ = group 1, and from 4 replicated time series data that are from $AR(1, -0.9)$ = group 2 for all Fourier frequencies.

Figure 4.5 shows for all ω the true $\beta_1(\omega)$'s [solid line in the middle (red line)], $\hat{\beta}_1(\omega_k)$'s [dotted line in the middle], and the approximate 95% upper (lower) limit for $\beta_1(\omega_k)$'s for the

two confidence intervals described in Subsection 4.2.1 and Subsection 4.2.2 respectively. For Figure 4.5, we used $Z^* = Z_{(\frac{1}{2}+\frac{1}{2}(0.95)^{1/s})} = 3.712$ because $s = 250$ in (4.2) and $m_{0.95} = 2.85$ in (4.6). Thus, we can notice that the independent simultaneous confidence interval is wider than the bootstrapping confidence interval for functions. The **R** code is shown in Appendix D.6.

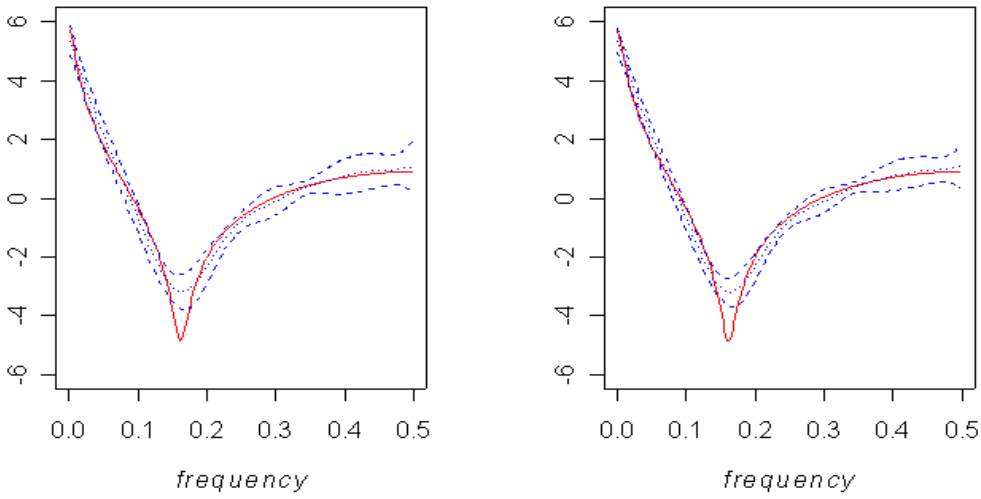


Figure 4.5: Confidence intervals when group 1 [2] = AR(0.9, 0.05) [AR(1, -0.9)].

As we can see, we observed similar confidence intervals using the independent simultaneous confidence interval and the bootstrapping confidence interval for functions. We also observed satisfactory results for most frequencies because $\beta_1(\omega)$'s lie between the approximate 95% confidence intervals for $\beta_1(\omega)$. However, this is not the case around $\omega = 0.17$. We have an issue to mention about both confidence intervals at this point. If we see the middle column in Figure 4.2, these are log-periodograms of simulated series from $AR(1, -0.9)$. We pay attention that four estimated functions, $\hat{f}_2(\omega)$, are underestimated around $\omega = 0.17$. This is because smooth estimated functions underestimate around sharp peak points, and overestimate around steep drop points. This reason results in higher $\hat{\beta}_1(\omega)$'s than real $\beta_1(\omega)$ around $\omega = 0.17$. Recall that $\hat{\beta}_1(\omega)$'s are obtained from $\log[\hat{f}_1(\omega)/\hat{f}_2(\omega)]$ for all ω.

4.3.2 Evaluation of Performance 2:

We wish to compare $f_1(\omega)$ with $f_3(\omega)$ and the procedure of this Subsection 4.3.2 is the same as Subsection 4.3.1 except the process in group 2 is $AR(-0.9, 0.05)$. In other words, we calculate $\hat{\beta}_1(\omega_k)$ from 4 replicated time series data that are obtained from $AR(0.09, 0.05)$ = group 1, and from 4 replicated time series data that are from $AR(-0.9, 0.05)$ = group 2 for all Fourier frequencies.

Figure 4.6 shows for all ω the true $\beta_1(\omega)$'s [solid line in the middle (red line)], $\hat{\beta}_1(\omega_k)$'s [dotted line in the middle], and the approximate 95% upper (lower) limit for $\beta_1(\omega_k)$'s for the two confidence intervals described in Subsection 4.2.1 and Subsection 4.2.2 respectively.

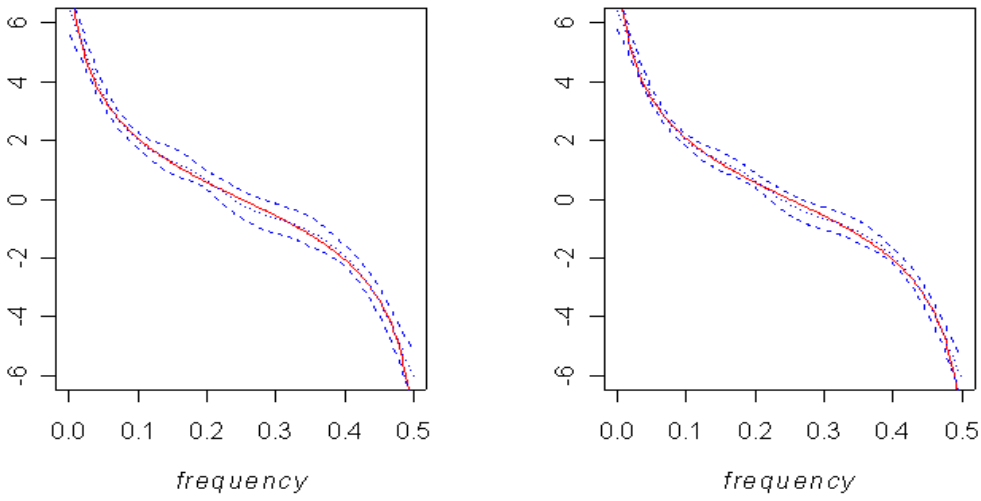


Figure 4.6: Confidence intervals when group 1 [2] = AR(0.9, 0.05) [AR(-0.9, 0.05)].

We observed satisfactory results using the two confidence intervals in the sense that the $\beta_1(\omega)$'s lie steadily between the approximate 95% confidence intervals for $\beta_1(\omega)$ for all ω. If we see Figure 4.2, centered log-periodograms from simulated series from $AR(0.09, 0.05)$ and $AR(-0.9, 0.05)$ are pretty smooth and predict the result in Figure 4.6. We used $Z^* = Z_{(\frac{1}{2}+\frac{1}{2}(0.95)^{1/s})} = 3.712$ because $s = 250$ in (4.1) and $m_{0.95} = 2.82$ [see Appendix C] in (4.6). The **R** code is shown in Appendix D.8.

4.4 ANALYSIS FOR REPLICATED EARTHQUAKES & EXPLOSIONS TIME SERIES DATA

We applied our proposed method as described in Chapter 3 to earthquakes (group 1), and explosions (group 2) time series data [Because of computation problems, we selected subsets of 4 series out of 8 series]. Each sequence of data consists of 512 points i.e., $s = 256$. Figure 4.7 shows the centered log-periodograms, $\log I(\omega) + 0.57721$, from earthquakes and explosions time series data and the corresponding smoothing spline estimates with smoothing parameters $\lambda = 0.0001$ and $\theta_1 = 1$ in (3.9).

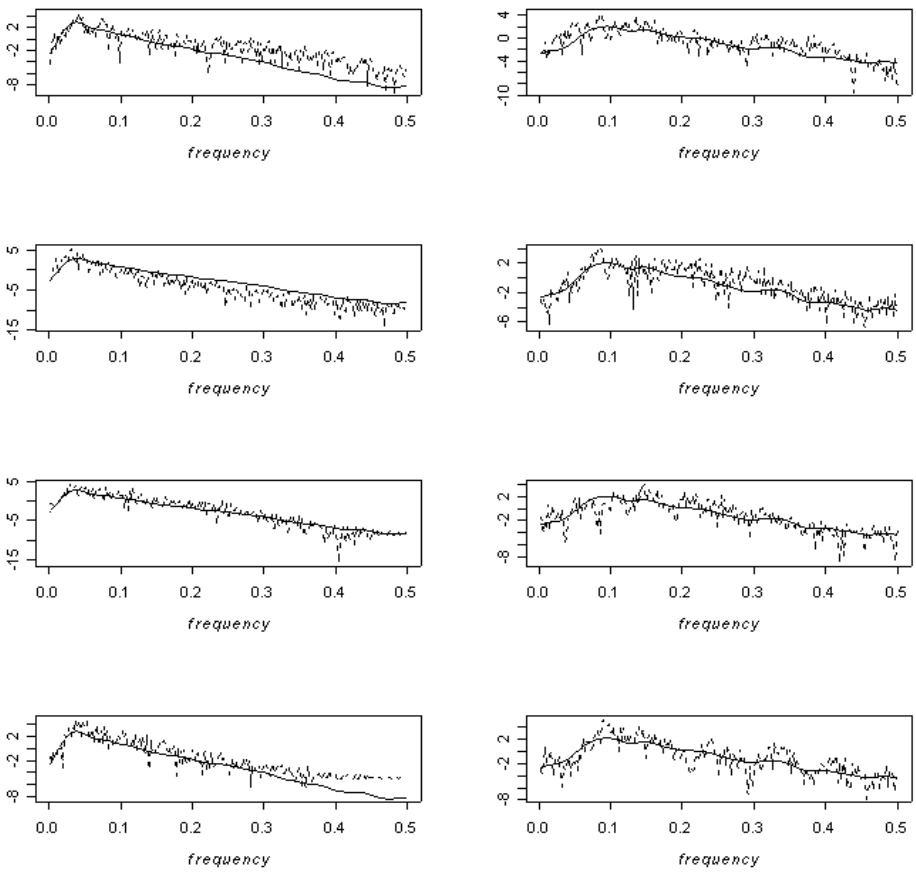


Figure 4.7: log-periodograms + 0.57721 and their fitted curves.

Note that because the series are not simulated series, at a glance, the fit is not as good as Figure 4.2. We used the smoothing parameter $\lambda = 0.0001$[1] instead of $\lambda = 0.001$ as in 4.2, because we wish to reduce underestimation (overestimation) so we can overcome the problem in Subsection 4.3.1 to some extent.

For our confidence intervals, we use $Z^* = Z_{(\frac{1}{2}+\frac{1}{2}(0.95)^{1/s})} = 3.718$ in (4.1) because $s = 256$ for the independent simultaneous confidence interval, and $m_{0.95} = 2.45$ in (4.6) for the bootstrapping confidence interval for functions. Figure 4.8 shows that $\bar{\hat{F}}_B(m)$ as a function of m_p when group 1 is earthquakes series and group 2 is explosions series with $N = 20$ and $R = 10$ in (4.5).

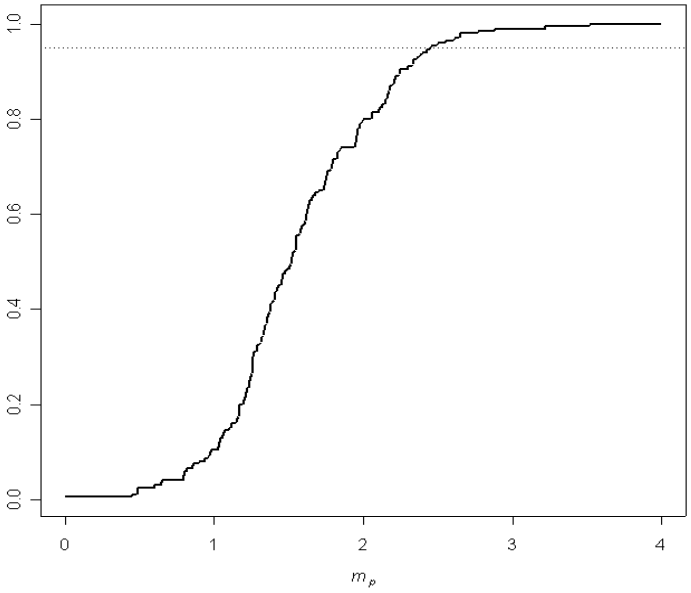


Figure 4.8: Equation 4.5 when group 1 = earthquakes and group 2 = explosions series.

Figure 4.9 shows the $\hat{\beta}_1(\omega_k)$'s [solid line in the middle (red line)] and the two approximate 95% confidence intervals with reference line, spectrum = 0, to analyze the two groups. The reason we are interseted in the reference line is because $\beta_1(\omega) = \log[f_1(\omega)/f_2(\omega)]$ where

[1] As $\lambda \to \infty$, the fit is getting smoother

$f_1(\omega)$ is a spectral density of earthquake and $f_2(\omega)$ is a spectral density of explosion. Thus, $\beta_1(\omega) > 0$ means the spectral power of earthquake is greater than that of explosion at ω and vice versa.

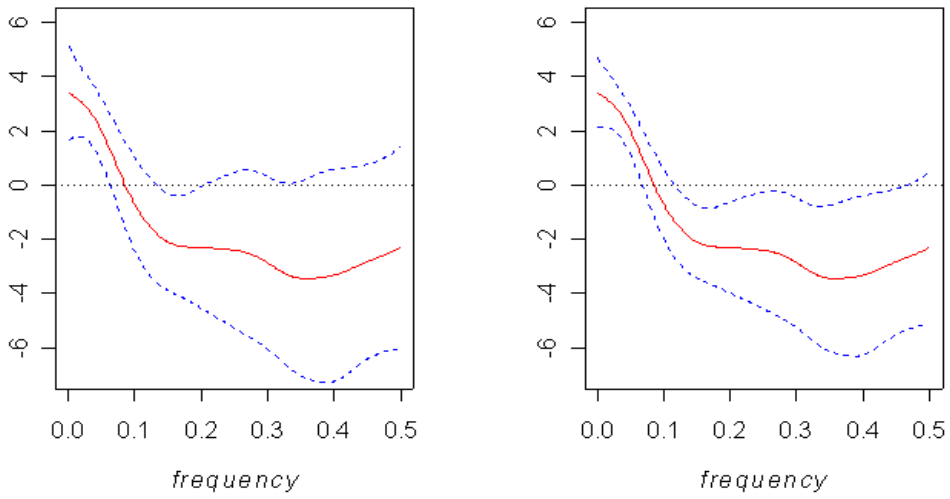


Figure 4.9: Confidence intervals when group 1 [2] = earthquakes [explosions] series.

We have wider confidence intervals than Figure 4.5 and Figure 4.6. This is because $\hat{\sigma}(\omega_k) = \sqrt{\text{Var}(\log \hat{f}_1(\omega_k)) + \text{Var}(\log \hat{f}_2(\omega_k))}$ is much larger, although $m_{0.95} = 2.45$ is somewhat smaller than 2.85 and 2.82 in the bootstrapping confidence interval for functions.

Our analysis may be somewhat different depending on the independent simultaneous confidence interval or the bootstrapping confidence interval for functions. These are, the spectral power of earthquakes is greater at low frequencies than that of explosions. In mid frequencies, the spectral power of explosions is greater than that of earthquakes: around $\omega = 0.15$ in the independent simultaneous confidence interval, and from $\omega = 0.12$ to $\omega = 0.4$ in the bootstrapping confidence interval. There is no difference between the two groups at higher frequencies under either method.

Remark

- We conducted a standardized transformation to earthquakes and explosions series before using them.
- Because **R** cannot handle the inversion of matrices of dimensions roughly higher than 1000*1000 in our design matrices, we chose 4 data series which have 500 points each in the simulation and 4 data series which have 512 points each in real data analysis.

5.0 CONCLUSION AND FUTURE WORK

5.1 SUMMARY AND CONTRIBUTIONS

This study is motivated from a simple but very important fact: $I(\omega) \sim f(\omega) \times$ Gamma(scale $= 1$, shape $= 1$), where $f(\omega)$ is a spectral density and $I(\omega)$ is the corresponding periodogram at frequency ω. With this motivation, we develop the method to compare several spectral densities in the spectral domain. This method consists of the estimation of spectral density in Chapter 3, and comparing the densities by the confidence intervals in Chapter 4.

In our research, we make the following contributions:

- We applied classical statistical methods in the frequency domain. For example, after constructing the periodograms from the time series data, we can use them to construct the confidence intervals for spectral density mentioned in Chapter 3.

- The bootstrapping confidence interval (for functions) for the difference of two groups in Subsection 4.2.2. This confidence interval can be useful for clinical research. For example, the analysis of an epileptic intracranial electroencephalogram (IEEG) data set [24] can help predict seizure with this confidence interval, and protect the patients.

- Our research is basically different from that of Wahba [29] in that it uses replicated series data instead of a single series, and differs from that of Diggle and Al Wasel [DAW] [9] in allowing the coefficients to vary with the frequencies.

5.2 FUTURE WORK

For our future research, we make the following future work:

- One drawback to our proposed estimation method is that it does not work well if the data series is not smooth. In other words, based on the two confidence intervals in Figure 4.5, our estimated function is overestimated around steep drop points. Naturally, our concern is about how to deal with data series that have steep drop (sharp peak) points.
- We will also consider other methods to compare spectral densities. For example, the Penalized Whittle Likelihood estimate (PWLE) [23] can be considered for the estimation method and the Bayesian confidence interval [18] can be considered in addition to the confidence intervals in Chapter 4.
- We have considered only group identifiers as covariates in (1.1). If other covariates differ between groups, the spectral density depends on these covariates as well as frequencies. We wish to compare spectral densities in the case when the spectral density also depends on these other covariates.
- When we have G groups to compare, we have to compute $\binom{G}{2}$ pairwise comparisons. We can consider the method to compare all groups at one time. Thus, we consider a generalized random effects model; $\log f_{gl}(\omega|\boldsymbol{x_i}) = \mu_g(\omega) + \epsilon_{gl}(\omega)$, where the $\mu_g(\omega)$ are independent $N(\mu(\omega), \sigma^2_\mu(\omega))$, the $\epsilon_{gl}(\omega)$ are independent $N(0, \sigma^2_\epsilon(\omega))$, $\mu_g(\omega)$ and $\epsilon_{gl}(\omega)$ are independent random variables for each ω for $g = 1, 2, \ldots, G$; $l = 1, 2, \ldots, L$, where L is the number of subjects. We need to test $H_o : \sigma^2_\mu(\omega) = 0$ for each ω.

APPENDIX A

NOTATION

I list the notations that often used in my thesis and their meanings.

- $y_1, y_2, \ldots, y_n$: time series.
- x_i : group indicator.
- n : number of data.
- ω : frequency.
- $\omega_k = k/n$: Fourier (fundamental) frequency.
- $I(\omega_k)$: periodogram at ω_k.
- $I_i(\omega_k)$: ith periodogram at ω_k in the replicated series.
- $i = 1, 2, \ldots, l$: indices for the replicated series [indices for the groups if x_i].
- $p = 1, 2, \ldots, P$: indices for covariates.
- $k = 1, 2, \ldots, s$: indices for the Fourier frequencies.
- $g = 1, 2, \ldots, G$: indices for groups.
- $f(\omega)$: spectral density.
- $f_1(\omega)$: the spectral densities of $AR(0.9, 0.05)$ in the simulation.
- $f_2(\omega)$: the spectral densities of $AR(1, -0.9)$ in the simulation.
- $f_3(\omega)$: the spectral densities of $AR(-0.9, 0.05)$ in the simulation.
- $\log \gamma(\alpha, \beta)$: log-gamma random variable with shape parameter α and scale parameter β.
- $\gamma(h)$: autovariance function with lag h.
- $\hat{\sigma}(\omega_k)$: standard error at Fourier frequency ω_k.

- λ : main smoothing parameter.
- θ : subsidiary smoothing parameter.
- Z_α : $100(\alpha)$ percentile of the standard normal distribution.
- $D_\alpha f$: αth derivative of function f.

APPENDIX B

SPECTRAL DENSITY OF AR(2) MODEL

$x_t = 0.9x_{t-1} + 0.05x_{t-2} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$. We start with the fact that the spectrum of ϵ_t, $f_\epsilon(\omega) = 1$ [28] and we use the uniqueness of the Fourier transformation.

$$\begin{aligned}
\gamma_\epsilon(h) &= E(\epsilon_{t+h}\epsilon_t) - E(\epsilon_{t+h})E(\epsilon_t) \qquad \text{(B.1)}\\
&= E[(x_{t+h} - 0.9x_{t+h-1} - 0.05x_{t+h-2})(x_t - 0.9x_{t-1} - 0.05x_{t-2})]\\
&= [1 + (0.9)^2 + (0.05)^2]\gamma_x(h) + [(0.9 \times 0.05) - 0.9][\gamma_x(h+1) + \gamma_x(h-1)]\\
&\quad -0.05[\gamma_x(h+2) + \gamma_x(h-2)]\\
&= 1.8125\gamma_x(h) - 0.855[\gamma_x(h+1) + \gamma_x(h-1)] - 0.05[\gamma_x(h+2) + \gamma_x(h-2)]\\
&= \int_{-1/2}^{1/2} [1.8125 - 0.855(e^{2\pi i\omega} + e^{-2\pi i\omega}) - 0.05(e^{4\pi i\omega} + e^{-4\pi i\omega})]e^{2\pi i\omega h} f_x(\omega)d\omega\\
&= \int_{-1/2}^{1/2} [1.8125 - 1.71\cos(2\pi\omega) - 0.1\cos(4\pi\omega)]e^{2\pi i\omega h} f_x(\omega)d\omega.
\end{aligned}$$

By the uniqueness, $f_\epsilon(\omega) = [1.8125 - 1.71\cos(2\pi\omega) - 0.1\cos(4\pi\omega)]f_x(\omega)$ Thus, $f_x(\omega) = \frac{1}{1.8125 - 1.71\cos(2\pi\omega) - 0.1\cos(4\pi\omega)}$. Similarly, we can find the spectral densities in cases of $x_t = x_{t-1} - 0.09x_{t-2} + \epsilon_t$ and $x_t = -0.9x_{t-1} + 0.05x_{t-2} + \epsilon_t$ where $\epsilon_t \sim N(0,1)$.

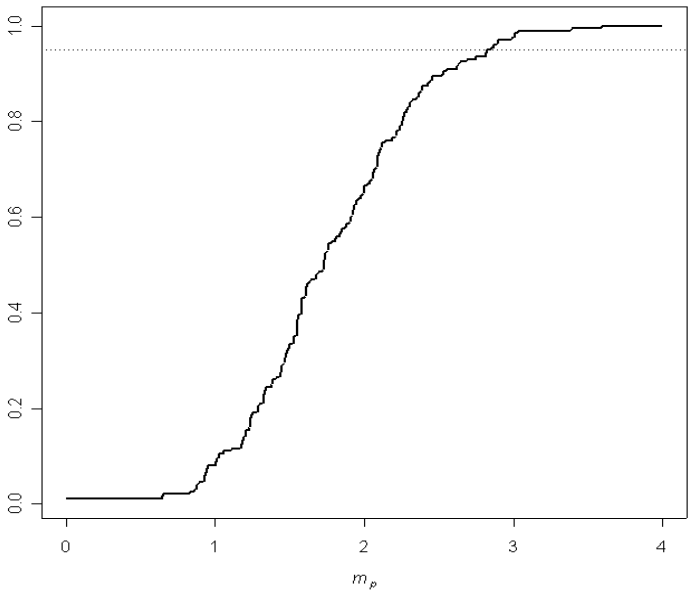
Note that using frequency response function (pp 220, [28]), $A_{\epsilon x}(\omega) = \sum_{t=-\infty}^{t=\infty} a_t e^{-2\pi i\omega t}$,

$$f_\epsilon(\omega) = |A_{\epsilon x}(\omega)|^2 f_x(\omega), \qquad \text{(B.2)}$$

where $A_{\epsilon x}(\omega) = 1 - 0.9e^{-2\pi i\omega} - 0.05e^{-4\pi i\omega}$ because $a_1 = 1, a_2 = -0.9, a_3 = -0.05$, and $a_r = 0$ otherwise. We can simply find the same $f_x(\omega)$.

APPENDIX C

EQUATION 4.5 WHEN GROUP 1 = AR(0.9, 0.05) AND GROUP 2 = AR(-0.9, 0.05)



APPENDIX D

R FUNCTIONS

D.1 EXAMPLE : TIME SERIES

Figure 1.1

```
han=scan("c:\\eq+exp.dat")  ## Read 34816 items
earth1=han[1:2048]
earth2=han[2049:4096]
earth3=han[4097:6144]
earth4=han[6145:8192]
earth5=han[8193:10240]
earth6=han[10241:12288]
earth7=han[12289:14336]
earth8=han[14337:16384]
exp1=han[16385:18432]
exp2=han[18433:20480]
exp3=han[20481:22528]
exp4=han[22529:24576]
exp5=han[24577:26624]
exp6=han[26625:28672]
exp7=han[28673:30720]
exp8=han[30721:32768]
NZ=han[32769:34816]

ear1=earth1[1025:2048]
ear2=earth2[1025:2048]
ear3=earth3[1025:2048]
ear4=earth4[1025:2048]
ear5=earth5[1025:2048]
ear6=earth6[1025:2048]
ear7=earth7[1025:2048]
ear8=earth8[1025:2048]
ex1=exp1[1025:2048]
ex2=exp2[1025:2048]
ex3=exp3[1025:2048]
```

```
ex4=exp4[1025:2048]
ex5=exp5[1025:2048]
ex6=exp6[1025:2048]
ex7=exp7[1025:2048]
ex8=exp8[1025:2048]

par(mfrow=c(2,1)) ## 5th and 6th series.
ts.plot(ear5,ylab="", main="Earthquake")
ts.plot(ex6,ylab="", main="Explosion")
```

D.2 EXAMPLE : WAHBA'S ESTIMATION

Figure 2.1

```
## After we get Figure 1.1 ##
nearth1=(earth1-mean(earth1))/sd(earth1)
nearth2=(earth2-mean(earth2))/sd(earth2)
nearth3=(earth3-mean(earth3))/sd(earth3)
nearth4=(earth4-mean(earth4))/sd(earth4)
nearth5=(earth5-mean(earth5))/sd(earth5)
nearth6=(earth6-mean(earth6))/sd(earth6)
nearth7=(earth7-mean(earth7))/sd(earth7)
nearth8=(earth8-mean(earth8))/sd(earth8)
nexp1=(exp1-mean(exp1))/sd(exp1)
nexp2=(exp2-mean(exp2))/sd(exp2)
nexp3=(exp3-mean(exp3))/sd(exp3)
nexp4=(exp4-mean(exp4))/sd(exp4)
nexp5=(exp5-mean(exp5))/sd(exp5)
nexp6=(exp6-mean(exp6))/sd(exp6)
nexp7=(exp7-mean(exp7))/sd(exp7)
nexp8=(exp8-mean(exp8))/sd(exp8)

##  For earthquakes  ##

snearth1=nearth1[1025:2048]
snearth2=nearth2[1025:2048]
snearth3=nearth3[1025:2048]
snearth4=nearth4[1025:2048]
snearth5=nearth5[1025:2048]
snearth6=nearth6[1025:2048]
snearth7=nearth7[1025:2048]
snearth8=nearth8[1025:2048]

## For explosions  ##

snexp1=nexp1[1025:2048]
snexp2=nexp2[1025:2048]
snexp3=nexp3[1025:2048]
snexp4=nexp4[1025:2048]
snexp5=nexp5[1025:2048]
```

```
snexp6=nexp6[1025:2048]
snexp7=nexp7[1025:2048]
snexp8=nexp8[1025:2048]


par(mfrow=c(4,4))
searth1.per=spec.pgram(snearth1,taper=0,log="no")
searth2.per=spec.pgram(snearth2,taper=0,log="no")
searth3.per=spec.pgram(snearth3,taper=0,log="no")
searth4.per=spec.pgram(snearth4,taper=0,log="no")
searth5.per=spec.pgram(snearth5,taper=0,log="no")
searth6.per=spec.pgram(snearth6,taper=0,log="no")
searth7.per=spec.pgram(snearth7,taper=0,log="no")
searth8.per=spec.pgram(snearth8,taper=0,log="no")
sexp1.per=spec.pgram(snexp1,taper=0,log="no")
sexp2.per=spec.pgram(snexp2,taper=0,log="no")
sexp3.per=spec.pgram(snexp3,taper=0,log="no")
sexp4.per=spec.pgram(snexp4,taper=0,log="no")
sexp5.per=spec.pgram(snexp5,taper=0,log="no")
sexp6.per=spec.pgram(snexp6,taper=0,log="no")
sexp7.per=spec.pgram(snexp7,taper=0,log="no")
sexp8.per=spec.pgram(snexp8,taper=0,log="no")

## we get spectra ##

s1=searth1.per$spec
s2=searth2.per$spec
s3=searth3.per$spec
s4=searth4.per$spec
s5=searth5.per$spec
s6=searth6.per$spec
s7=searth7.per$spec
s8=searth8.per$spec
s9=sexp1.per$spec
s10=sexp2.per$spec
s11=sexp3.per$spec
s12=sexp4.per$spec
s13=sexp5.per$spec
s14=sexp6.per$spec
s15=sexp7.per$spec
s16=sexp8.per$spec


smatH=cbind(s1,s2,s3,s4,s5,s6,s7,s8)
smatL=cbind(s9,s10,s11,s12,s13,s14,s15,s16)
smeanH=apply(smatH,1,mean)
smeanL=apply(smatL,1,mean)

## main for Whaba graph ##

meanH=apply(smatH,1,mean)
meanL=apply(smatL,1,mean)

I1=meanH
```

```
freq=searth1.per$freq
freq2=sort(-freq)
frequency=c(freq2,freq)

Y11=log(I1)+0.0637944
Y12=Y11[512:1]
Y1=c(Y12,Y11)             ## logI[1]=logI[960]=-0.6110726  ##

c0=(log(2)+0.0637944)/pi
Y1[1024]=c0+(Y1[1024]-0.0637944)
Y1[1]=c0+(Y1[1]-0.0637944)


I2=meanL
Y21=log(I2)+0.0637944
Y22=Y21[512:1]
Y2=c(Y22,Y21)

d0=(log(2)+0.0637944)/pi
Y2[1024]=d0+(Y2[1024]-0.0637944)
Y2[1]=d0+(Y2[1]-0.0637944)


gv1=rep(1,1024)
gv2=rep(1,1024)

v=-511:512
k=-511:512
for(i in 1:1024) { gv1[i]=(1/1024)*sum(Y1*exp(-2*pi*1i*v[i]*k/1024))}
for(j in 1:1024) { gv2[j]=(1/1024)*sum(Y2*exp(-2*pi*1i*v[j]*k/1024))}

lamda1=10**(-10)   ## test with different parameters ##
mm1=4

lamda2=10**(-5)
mm2=2

lamda3=10**(-14)
mm3=4

lamda4=10**(-7)
mm4=2

w=seq(-511/1024,512/1024,1/1024)

est1=rep(1,1024)
for(i in 1:1024) {est1[i]=sum((gv1/(1+lamda3*(2*pi*v)**(2*mm3))) *exp(2*pi*1i*v*w[i]))}

est2=rep(1,1024)
for(j in 1:1024) {est2[j]=sum((gv2/(1+lamda3*(2*pi*v)**(2*mm3))) *exp(2*pi*1i*v*w[j]))}

par(mfrow=c(2,1))  ## Graph ##
plot(frequency,Y1,lty="dotted",type="l",ylim=c(-8,7),xlab=expression(italic(frequency))
     ,ylab="",main="S components of earthquake")
```

```
lines(frequency,est1,type="l")
legend(0.27,6.5,legend=c("estimate","log I(w) + C"),lty=c(1,3))
plot(frequency,Y2,lty="dotted",type="l",ylim=c(-8,7),xlab=expression(italic(frequency))
     ,ylab="",main="S components of explosion")
lines(frequency,est2,type="l")
```

D.3 EXAMPLE : SAMPLE COEFFICIENT VARIATION

Figure 2.2

```
## After we get Figure 1.1 and Figure 2.1 ##
s=apply(smatH,1,sd)
I=apply(smatH,1,mean)
plot(freq,s/I,type="l",ylab=expression(italic(CV)),xlab=expression(italic(frequency)))
abline(h=1,lty="dotted")
```

D.4 EXAMPLE : SPECTRAL DENSITIES

Figure 4.1

```
w=seq(0.02,0.5,0.02)
f1=log(1/(1.8125-1.71*cos(2*pi*w)-0.1*cos(4*pi*w)))
f2=log(1/(2.81-3.8*cos(2*pi*w)+1.8*cos(4*pi*w)))
f3=log(1/(1.8125+1.71*cos(2*pi*w)-0.1*cos(4*pi*w)))
matplot(w,cbind(f1,f2),xlab=expression(bold(italic(frequency))),ylim=c(-2.5,6.5),
        ylab=expression(italic(logf)~(w)),lwd=1,type="l",col=c(1,1))
lines(w,f3,lty=4)

text(locator(1),"AR(0.9,0.05)")
text(locator(1),"AR(1,-0.9)")
text(locator(1),"AR(-0.9,0.05)")
```

D.5 SIMULATION : SPECTRAL DENSITY WITH FITTING CURVE

Figure 4.2

```
y1=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y2=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y3=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y4=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y5=arima.sim(list(order=c(2,0,0),ar=c(1,-0.9)),n=500)
y6=arima.sim(list(order=c(2,0,0),ar=c(1,-0.9)),n=500)
```

```
y7=arima.sim(list(order=c(2,0,0),ar=c(1,-0.9)),n=500)
y8=arima.sim(list(order=c(2,0,0),ar=c(1,-0.9)),n=500)
z1=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z2=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z3=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z4=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)

par(mfrow=c(4,3))
b1.per=spec.pgram(y1,taper=0,log="no")
b2.per=spec.pgram(y2,taper=0,log="no")
b3.per=spec.pgram(y3,taper=0,log="no")
b4.per=spec.pgram(y4,taper=0,log="no")
b5.per=spec.pgram(y5,taper=0,log="no")
b6.per=spec.pgram(y6,taper=0,log="no")
b7.per=spec.pgram(y7,taper=0,log="no")
b8.per=spec.pgram(y8,taper=0,log="no")
z1.per=spec.pgram(z1,taper=0,log="no")
z2.per=spec.pgram(z2,taper=0,log="no")
z3.per=spec.pgram(z3,taper=0,log="no")
z4.per=spec.pgram(z4,taper=0,log="no")

YY1=log(b1.per$spec)+0.57721
YY2=log(b2.per$spec)+0.57721
YY3=log(b3.per$spec)+0.57721
YY4=log(b4.per$spec)+0.57721
YY5=log(b5.per$spec)+0.57721
YY6=log(b6.per$spec)+0.57721
YY7=log(b7.per$spec)+0.57721
YY8=log(b8.per$spec)+0.57721
ZZ1=log(z1.per$spec)+0.57721
ZZ2=log(z2.per$spec)+0.57721
ZZ3=log(z3.per$spec)+0.57721
ZZ4=log(z4.per$spec)+0.57721

Y1=c(YY1,YY2,YY3,YY4)
Y2=c(YY5,YY6,YY7,YY8)
Z1=c(ZZ1,ZZ2,ZZ3,ZZ4)

## 1st ##
t1=rep(1,1000)
t2=rep(b1.per$freq,4)
T=cbind(t1,t2)

a100=matrix(b1.per$freq,250,nrow=250)
ss=function(a)
{
c=ncol(a)
r=nrow(a)
mat=matrix(1,r,ncol=c)
for(i in 1:c) {
for(j in 1:r) { mat[i,j]=1*((abs(i-j)*0.02)^4-2*(abs(i-j)*0.02)^3+(abs(i-j)*0.02)^2
               -(1/30))/(-24)}
}
return(mat)
```

```
}

m=ss(a100)
q=rbind(m,m,m,m)
Q=cbind(q,q,q,q)
W=diag(1,1000)
n=1000
lamda=0.001
M=W%*%Q%*%t(W) + diag(n*lamda,1000)
D=solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%t(W)%*%solve(M)%*%Y1
C=W%*%solve(M)%*%(diag(1,1000)-W%*%T%*%solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%
  t(W)%*%solve(M))%*%Y1
beta1=T%*%D+Q%*%C
Yhat1=W%*%beta1
fit1=Yhat1[1:250]

##  2nd ##
D=solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%t(W)%*%solve(M)%*%Y2
C=W%*%solve(M)%*%(diag(1,1000)-W%*%T%*%solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%
  t(W)%*%solve(M))%*%Y2
beta2=T%*%D+Q%*%C
Yhat2=W%*%beta2
fit2=Yhat2[1:250]

##  3rd ##
D=solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%t(W)%*%solve(M)%*%Z1
C=W%*%solve(M)%*%(diag(1,1000)-W%*%T%*%solve(t(T)%*%t(W)%*%solve(M)%*%W%*%T)%*%t(T)%*%
  t(W)%*%solve(M))%*%Z1
beta3=T%*%D+Q%*%C
Yhat3=W%*%beta3
fit3=Yhat3[1:250]

##  3 fitted log-density ##
par(mfrow=c(4,3))
matplot(b1.per$freq,cbind(fit1,YY1),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(0.9,0.05))))
matplot(b1.per$freq,cbind(fit2,YY5),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(1,-0.9))))
matplot(b1.per$freq,cbind(fit3,ZZ1),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(-0.9,0.05))))


matplot(b1.per$freq,cbind(fit1,YY2),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(0.9,0.05))))
matplot(b1.per$freq,cbind(fit2,YY6),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(1,-0.9))))
matplot(b1.per$freq,cbind(fit3,ZZ2),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(-0.9,0.05))))
```

```
matplot(b1.per$freq,cbind(fit1,YY3),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(0.9,0.05))))
matplot(b1.per$freq,cbind(fit2,YY7),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(1,-0.9))))
matplot(b1.per$freq,cbind(fit3,ZZ3),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(-0.9,0.05))))

matplot(b1.per$freq,cbind(fit1,YY4),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(0.9,0.05))))
matplot(b1.per$freq,cbind(fit2,YY8),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(1,-0.9))))
matplot(b1.per$freq,cbind(fit3,ZZ4),xlab=expression(italic(frequency)),ylab="",lwd=1,
        type="l",col=c(1,1))
title(expression(italic(AR(-0.9,0.05))))
```

D.6 SIMULATION : COMPARING AR(0.9, 0.05) WITH AR(1, -0.9) WITH CONFIDENCE INTERVALS

Figure 4.5

```
## After we get Figure 4.5 ##

beta1hat=fit1-fit2  ## length = 250 ##

### 1st ####
booth=function(){
B=matrix(0,4,200)
for(i in 1:200) {
p1=sample(c(1:4),4,replace=TRUE)
B[,i]=sort(c(p1))
}
return(B)
}
sam=booth()

YMAT=cbind(YY1,YY2,YY3,YY4)
YYY=matrix(0,1000,200)
for (i in 1:200)
YYY[,i]=c(YMAT[,sam[1,i]],YMAT[,sam[2,i]],YMAT[,sam[3,i]],YMAT[,sam[4,i]])

ahn=function(YYY){
temp=matrix(0,250,200)
for ( k in 1:200){
Y=YYY[,k]
```

```
t1=rep(1,1000)
t2=rep(b1.per$freq,4)
T=cbind(t1,t2)
a100=matrix(b1.per$freq,250,nrow=250)
ss=function(a)
{
c=ncol(a)
r=nrow(a)
mat=matrix(1,r,ncol=c)
for(i in 1:c) {
for(j in 1:r) { mat[i,j]=((abs(i-j)*0.02)^4-2*(abs(i-j)*0.02)^3+(abs(i-j)*0.02)^2
-(1/30))/(-24)}
}
return(mat)
}
m=ss(a100)
q=rbind(m,m,m,m)
Q=cbind(q,q,q,q)

W1=diag(1,1000)
n=1000
lamda=0.001
M=W1%*%Q%*%t(W1) + diag(n*lamda,1000)
D=solve(t(T)%*%t(W1)%*%solve(M)%*%W1%*%T)%*%t(T)%*%t(W1)%*%solve(M)%*%Y
C=W1%*%solve(M)%*%(diag(1,1000)-W1%*%T%*%solve(t(T)%*%t(W1)%*%solve(M)%*%W1%*%T)%*%
t(T)%*%t(W1)%*%solve(M))%*%Y
FIT1=(W1%*%T%*%D+Q%*%C)[1:250]
temp[,k]=FIT1
}
return(temp)
}
CILIT1=ahn(YYY)
v1=apply(CILIT1,1,var)

## 2nd ##
booth=function(){
B=matrix(0,4,200)
for(i in 1:200) {
p1=sample(c(1:4),4,replace=TRUE)
B[,i]=sort(c(p1))
}
return(B)
}
sam=booth()
MAT=cbind(YY1,YY2,YY3,YY4)
ZZZ=matrix(0,1000,200)
for (i in 1:200)
ZZZ[,i]=c(ZMAT[,sam[1,i]],ZMAT[,sam[2,i]],ZMAT[,sam[3,i]],ZMAT[,sam[4,i]])

ahn=function(ZZZ){
temp=matrix(0,250,200)
for ( k in 1:200){
Z=ZZZ[,k]
t1=rep(1,1000)
```

```
t2=rep(b1.per$freq,4)
T=cbind(t1,t2)
a100=matrix(z1.per$freq,250,nrow=250)
ss=function(a)
{
c=ncol(a)
r=nrow(a)
mat=matrix(1,r,ncol=c)
for(i in 1:c) {
for(j in 1:r) { mat[i,j]=((abs(i-j)*0.02)^4-2*(abs(i-j)*0.02)^3+(abs(i-j)*0.02)^2
-(1/30))/(-24)}
}
return(mat)
}

m=ss(a100)
q=rbind(m,m,m,m)
Q=cbind(q,q,q,q)
W1=diag(1,1000)
n=1000
lamda=0.001
M=W1%*%Q%*%t(W1) + diag(n*lamda,1000)
D=solve(t(T)%*%t(W1)%*%solve(M)%*%W1%*%T)%*%t(T)%*%t(W1)%*%solve(M)%*%Z
C=W1%*%solve(M)%*%(diag(1,1000)-W1%*%T%*%solve(t(T)%*%t(W1)%*%solve(M)%*%W1%*%T)%*%
t(T)%*%t(W1)%*%solve(M))%*%Z

FIT2=(W1%*%T%*%D+Q%*%C)[1:250]
temp[,k]=FIT2
}
return(temp)
}
CILIT2=ahn(ZZZ)
v2=apply(CILIT2,1,var)

## Theoretical value ##
w=b1.per$freq
logf1=log(1/(1.8125-1.71*cos(2*pi*w)-0.1*cos(4*pi*w)))
logf0=log(1/(2.81-3.8*cos(2*pi*w)+1.8*cos(4*pi*w)))
Beta0=logf0
Beta1=logf1-logf0

## Confidence intervl ##
z=qnorm(1/2+1/2*(0.95)^(1/250))  ## z=3.712 ##
m=2.85

sigma=sqrt(v1+v2)
fhat=CILIT1-CILIT2

upperz=beta1hat+sigma*z
lowz=beta1hat-sigma*z
upperm=beta1hat+sigma*m
lowm=beta1hat-sigma*m

par(mfrow=c(1,2))  ## graphs ##
```

```
matplot(w[1:249],cbind(Beta1[1:249],upperz[1:249]),ylim=c(-6,6),
xlab=expression(italic(frequency)),
ylab="",lwd=1,type="l",col=c(2,4))
lines(w[1:249],lowz[1:249],lty=2,col=4)
lines(w[1:249],beta1hat[1:249],lty="dotted",col=4)
title("Independent Simultaneous CI")
matplot(w[1:249],cbind(Beta1[1:249],upperm[1:249]),ylim=c(-6,6),
xlab=expression(italic(frequency)),
ylab="",lwd=1,type="l",col=c(2,4))
lines(w[1:249],lowm[1:249],lty=2,col=4)
lines(w[1:249],beta1hat[1:249],lty="dotted",col=4)
title("Bootstrapping CI")
```

D.7 EXAMPLE : GRAPH OF EQUATION (4.5) WHEN GROUP 1 = AR(0.9, 0.05) AND GROUP 2 = AR(1, -0.9)

Figure 4.4

```
## After we get Figure 4.5 ##

mx=rep(0,200)
for (i in 1:200){
mx[i]=max(abs((fhat[,i]-beta1hat)/sigma))}

m20=cbind(sort(mx[1:20]),sort(mx[21:40]),sort(mx[41:60]),sort(mx[61:80]),
sort(mx[81:100]),sort(mx[101:120]),sort(mx[121:140]),sort(mx[141:160]),
sort(mx[161:180]),sort(mx[181:200]))
sort(m20)
m=rep(0,401)
for (j in 1:401){
m[j]=mean(c(1/20*length(m20[,1][m20[,1]<=0.01*j]),1/20*length(m20[,2][m20[,2]<=0.01*j]),
1/20*length(m20[,3][m20[,3]<=0.01*j]),1/20*length(m20[,4][m20[,4]<=0.01*j]),
1/20*length(m20[,5][m20[,5]<=0.01*j]),1/20*length(m20[,6][m20[,6]<=0.01*j]),
1/20*length(m20[,7][m20[,7]<=0.01*j]),1/20*length(m20[,8][m20[,8]<=0.01*j]),
1/20*length(m20[,9][m20[,9]<=0.01*j]),
1/20*length(m20[,10][m20[,10]<=0.01*j])))
}

plot(seq(0,4,0.01),m,type="l",lwd=2,xlab=expression(italic(m[p])),ylab="")
abline(h=0.95,lty="dotted")
```

D.8 SIMULATION : COMPARING AR(0.9, 0.05) WITH AR(-0.9, 0.05) WITH CONFIDENCE INTERVALS

Figure 4.6

```
## Everything is the same as Figure 4.8 except belows ##

beta1hat=fit1-fit3
MAT=cbind(ZZ1,ZZ2,ZZ3,ZZ4)
m=2.82

y1=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y2=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y3=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
y4=arima.sim(list(order=c(2,0,0),ar=c(0.9,0.05)),n=500)
z1=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z2=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z3=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
z4=arima.sim(list(order=c(2,0,0),ar=c(-0.9,0.05)),n=500)
```

BIBLIOGRAPHY

[1] Adams, R. (1975). *Sobolev Space*. Academic Press, New York.

[2] Aronszajn, N. (1950). Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, **68**, 337-404.

[3] Bateman, H. (1954). *Tables of Integral Transforms*. New York: McGraw-Hill Book Co.

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

[5] Blandford, R. R. (1993). *Discrimination of Earthquakes and Explosions.* AFTAC-TR-93-044 HQ, Air Force Technical Applications Center, Patrick Air Force Base, FL.

[6] Babu, J. and Singh, K. (1983). Inference on Means Using Bootstrap. *The Annals of Statistics,* Vol. 11, **3**, 999-1003.

[7] Cogburn, R. and Davis, H. T. (1974). Periodic Splines and Spectral Estimation. *The Annals of Statistics*, **2**, 1108-1126.

[8] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25**, 1-37.

[9] Diggle, P. and Al Wasel, I. (1997). Spectral analysis of replicated biomedical time series. *Appl. Statist.*, **46**, 31-71.

[10] Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. CRC press.

[11] Green, P. J. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall, London.

[12] Gu, C. (2002). *Smoothing Spline ANOVA Models*. New York: Springer-Verlag.

[13] Gu, C. and Wahba, G. (1993a). Semiparametric Analysis of Variance With Tensor Product Thin Plate Spline. *Journal of the Royal Statistical Society*, Ser. B, **55**, 353-368.

[14] Gu, C. and Wahba, G. (1993b). Smoothing Spline ANOVA With Component-Wise Bayesian Confidence Intervals. *Journal of the Computational and Graphical Statistics*, **2**. 97-117.

[15] Guo, W., Dai, M., Ombao, H., and von Sachs, R. (2003). smoothing spline ANOVA for time-dependent spectral analysis. *Journal of the American Statistical Association,* **98**, 643-652.

[16] Hastie, T., Tibshirani, R and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag.

[17] Little, R. J. & Rubin, D. B. (2002). *Statistical Analysis with missing data*. New Jersey: John Wiely & Sons.

[18] Nychka, D. (1988). Bayesian Confidence Intervals for Smoothing Splines. *Journal of the American Statistical Association*, **83**, 1134-1143.

[19] Ombao, H., Raz, J., Sachs, R. and Malow, B. (2001). Automatic Statistical Analysis of Bivariate Nonstationary Time Series. *Journal of the American Statistical Association*, **25**, 38-76.

[20] Ombao, H., Raz, J., Sachs,R., and Guo, W. (2002). The SLEX Model of a Nonstationary Time Series. *Annals of the Institute of Statistical Mathematics,* **54**, 171-200.

[21] Olshen, R., Biden, E., Wyatt, M. and Sutherland, D. (1989). Gait Analysis and the Bootstrap. *The Annals of Statistics,* Vol. 17, **4**, 1419-1440.

[22] Parzen, E. (1961). Mathematical Considerations in the Estimation of Spectra. *Technometrics*, **3**, 167-190.

[23] Pawitan, Y. and O'Sullivan, F (1994). Nonparametric Spectral Density Estimation Using Penalized Whittle Likelihood. *Journal of the American Statistical Association* **89**, 600-610.

[24] Qin, L., Guo, W. and Litt, B. (2006). Time-Frequency Functional Model. *Technical report*.

[25] Ramsay, J.O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York: Springer-Verlag.

[26] Ruppert, D, Wand, M.P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

[27] Shumway, R. H. (1970). Applied Regression and Analysis of Variance for Stationary Time Series. *Journal of the American Statistical Association*, **65**, 1527-1546.

[28] Shumway, R. H. & Stoffer, D. S. (2006). *Time Series Analysis and Its Application with R Examples*. New York: Springer-Verlag.

[29] Wahba, G. (1980). Automatic Smoothing of the log periodogram. *Journal of the American Statistical Association* **75**, 122-132.

[30] Wahba, G. (1983). Bayesian confidence Interval for the Cross-Validated Smoothing Spline. *Journal of the Royal Statistical Society*, Ser. B, **45**, 133-150.

[31] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameters in the generalized spline smoothing problem. *The Annals of Statistics*, **13**, 1378 - 1402.

[32] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.