SPECTRAL-BASED TESTS FOR PERIODICITIES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Lai Wei, M.S.

* * * * *

The Ohio State University

2008

Dissertation Committee:

Professor Peter F. Craigmile, Adviser
Professor Thomas J. Santner
Professor Wayne M. King

Approved by

Adviser
Graduate Program in
Statistics
ABSTRACT

In this thesis, tests for periodicity are investigated based on a spectral analysis of a time series. Some important fundamental theories of spectral analysis of stationary and harmonic processes are reviewed. A regression model is developed in the frequency domain based on the Fourier transformation. We present some of the periodogram based tests, which are the global test and three local tests, i.e., the hearing test, local $F$ test and Thomson’s multitaper test. We will show that most of the tests can be derived from our regression model with the error term having an approximately diagonal covariance matrix. The distribution of the error term of the spectral regression model is based on the asymptotic distribution of the tapered Fourier transform of the error process. This asymptotic distribution has approximately a diagonal covariance matrix when the sample size is large and the spectral density functions (SDFs) of error processes have small dynamic range. We contrast the $F$ test in the time domain and the local $F$ test in the frequency domain as well as the global and local spectral-based tests. The global test uses the spectral estimates at all sample points, whereas a local test uses a subset of the sample points available. Standard global tests for periodicity are often based on the assumption of a Gaussian IID error process. Using a smoothing spline approach, we extend the global test to the non-IID case. We compare this approach to a number of local tests for periodicity such as the local $F$ test, the test commonly used in hearing sciences, and Thomson’s multitaper $F$
test. Using regression-based F tests, we demonstrate that asymptotic size and power calculations can be made for some of these tests. We compare the size and power at finite sample sizes, under a number of different experimental conditions.

According to the exploratory data analysis, we applied the local F test to hearing data, Distortion Product Otoacoustic Emissions (DPOAEs), collected in the Department of Speech and Hearing Sciences, OSU. The logistic regression model and the noncentral $F$ mixed effects regression models are explored to capture the important features of hearing data. In particular, noncentral-F mixed effects regression models capture within-subject-variability of the distortion products of healthy hearing subjects. This is key to understanding the underlying processes inherent in DPOAE-based hearing tests. The Penalized Quasi-Likelihood method is used to estimate the model parameters and we demonstrate how to do the model selection and diagnosing.
To my family
ACKNOWLEDGMENTS

I would like to express my sincere gratitude toward my advisor, Professor Peter F. Craigmile, for his support and guidance throughout my research. I owed a great deal to him for the knowledge that I have gained in the spectral analysis in time series as well as the experience in it’s applications.

I am also grateful for the DPOAE dataset provided by Professor Wayne King. I want to thank Professor Thomas Santner and Professor Wayne King for taking time to serve in my dissertation committee.

Finally, I would like to thank my parents, my husband Sijin, and my son Kevin for their constant and unselfish love and encouragement.
VITA

1994-1998 ................................. B.S. Electrical Engineering, Electronic Science and Technology of China, Chengdu, China

1998-2000 ................................. Graduate Research Associate, Electronic Science and Technology of China, Chengdu, China

2000-2003 ................................. M.S. Electrical Engineering, The Ohio State University, Columbus, Ohio, USA

2003-2004 ................................. M.S. Statistics, The Ohio State University, Columbus, Ohio, USA

2000-present ............................. Graduate Research/Teaching Associate, The Ohio State University, Columbus, Ohio, USA

FIELDS OF STUDY

Major Field: Statistics

Studies in Spectral-based Tests for Periodicity: Professor Peter F. Craigmile
TABLE OF CONTENTS

Abstract ................................................................. ii
Dedication ............................................................... iv
Acknowledgments ......................................................... v
Vita ........................................................................ vi
List of Tables .............................................................. ix
List of Figures .............................................................. x

Chapters:

1. Introduction ............................................................. 1
   1.1 Spectral-based Tests for Periodicity of Time Series .......... 1
   1.2 Distortion Product Otoacoustic Emissions (DPOAEs) ......... 2
   1.3 Generalized Linear Models ........................................... 4
   1.4 Outline of the Thesis ................................................ 5

2. A Review of The Spectral Analysis of Stationary and Harmonic Processes 7
   2.1 Stationary Processes .................................................. 7
   2.2 Spectral Analysis of Stationary Processes ....................... 11
   2.3 Introduction of Harmonic Processes ............................. 19

3. Spectral-based Tests for Periodicity ................................ 23
   3.1 Analysis of the Tapered Fourier Transform of Stationary Process . 32
   3.2 Regression Model in the Spectral Domain ....................... 41

vii
3.3 Global $F$ test .......................................................... 44
3.3.1 Global $F$ test with one periodicity ($L = 1$) ................. 44
3.3.2 Global $F$ Test with More Than One Periodicities ($L > 1$) . 50
3.4 Local Tests .............................................................. 57
3.4.1 Test from the Hearing Science ................................. 58
3.4.2 Local $F$ Test ......................................................... 58
3.4.3 Thomson’s $F$ Test .................................................. 66

4. Size and Power Comparisons ......................................... 71
4.1 The Size of the Test .................................................. 73
4.2 The Power of the Test ................................................ 82

5. Analysis of Distortion Product Otoacoustic Emissions (DPOAEs) . . . . 96
5.1 Introduction of Distortion Product Otoacoustic Emissions (DPOAEs) 96
5.2 Exploratory Data Analysis ........................................... 99
5.3 Logistic Regression Models ......................................... 102
5.3.1 Review of The Logistic Regression Model .................... 103
5.3.2 Logistic Regression Model for DPOAE-based Healthy Cochlear
      Status ............................................................... 105
5.4 Generalized Linear Mixed Effects Model for the DPOAE Periodicity
      $F$ Statistics ........................................................ 114
5.4.1 Parameter Estimation in the Generalized Linear Mixed Ef-
      fects Model ....................................................... 115
5.4.2 Noncentral-$F$ Linear Mixed Effects Model ................... 120
5.4.3 Generalized Linear Mixed Effects Model for the DPOAE data 125

6. Discussion ..................................................................... 142

Appendices:
A. Noncentral-$F$ distribution .......................................... 146
B. Notation ................................................................. 147

Bibliography .................................................................. 151
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>AIC for different links using the 20% trimmed mean of the times series of all 100 trials.</td>
<td>110</td>
</tr>
<tr>
<td>5.2</td>
<td>AIC for different links using the 10% trimmed mean of the times series of every 20 trials.</td>
<td>113</td>
</tr>
<tr>
<td>5.3</td>
<td>Pearson residual analysis table for independent model</td>
<td>127</td>
</tr>
<tr>
<td>5.4</td>
<td>random effect model</td>
<td>128</td>
</tr>
<tr>
<td>5.5</td>
<td>Cross validation results for the log-Gaussian and noncentral $F$ mixed effects model</td>
<td>140</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>2.1 Plots of rectangular taper and 50% cosine taper in time domain and frequency domain, and the direct spectral estimates of the AR(4) noise process using rectangular taper and 50% cosine taper respectively</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>3.1 Time series plots and ACVS plots of three periodicities, AR(2) noise process, and a time series with three periodicities plus AR(2) noise process.</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>3.2 Plot of diagram of transformation between the time domain and frequency domain</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>3.3 Plot of periodogram of the time series with three periodicities and AR(2) noise process in frequency domain</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>3.4 Plots of variance and correlation for AR(2) with sample size $N = 16$ and $N = 256$, AR(4) with $N = 256$, AR(4) with 20% cosine taper and $N = 256$ respectively. Correlations are calculated between frequency $\eta$ and $\eta' = \eta \pm \frac{1}{N}$ as well as $\eta$ and $\eta' = \eta \pm \frac{2}{N}$.</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>3.5 Plots of absolute correlation for AR(2) with sample size $N = 16$ and $N = 256$, AR(4) with $N = 256$, AR(4) with 20% cosine taper and $N = 256$ respectively.</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>4.1 SDFs of the IID (solid line), AR(2) (dashed line), and AR(4) (dotted line) error processes.</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>4.2 Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test, the Gorga et al. test and the global test for IID and AR(2) error processes with one periodicity ($L = 1$). In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$.</td>
<td>74</td>
<td></td>
</tr>
</tbody>
</table>
4.3 Plots of the estimated size as a function of the nuisance frequency, $\zeta_2$, of the local $F$ test, the Gorga et al. test and the global test for IID and AR(2) error processes with two periodicities ($L = 2$). In each simulation $\zeta_1 = 144/512$. In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$.  

4.4 Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test without taper, the local $F$ tests with 5% cosine taper and the Thomson’s test for AR(4) error processes. In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$. The Thomson’s test uses 7 DPSS tapers with $NW = 7$.  

4.5 Plots of the estimated size as a function of parameter $p$ in the local $F$ test with the 100$p$% cosine taper based on a simulation of 1,000 time series with sample size $N = 64, 128, 256,$ and $512$. The local $F$ tests have test frequency $\zeta_1 = 144/512$ and bin size $B = 4$ for the AR(4), the AR(2) and the IID error processes respectively.  

4.6 Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test with 1 offset, 2 offset, 3 offset, and 4 offset based on a simulation of 1,000 time series with sample size $N = 512$. The local $F$ tests use bin size $B = 4$.  

4.7 A comparison of power curve of the $F$ test in the time domain and the local $F$ test in the frequency domain for AR(2) and AR(4) error processes. Local $F$ test uses bin size $B = 4$.  

4.8 Plots of the estimated power of the local $F$ test, the Gorga et al. test and the global test for IID and AR(2) error processes with two periodicities ($L = 2$, with $\zeta_1 = 144/512$ and $\zeta_2 = 160/512$). In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$.  

4.9 A comparison of the area under the power curve (AUC) and relative AUC of the local $F$ test, the Gorga et al test, Thomson’s $F$ test and the global test for different values of sample size $N$ with an AR(2) and an AR(4) error processes.  

xi
4.10 A comparison of the area under the power curve (AUC) of the local $F$ test, the Gorga et al test and the global test for different number of frequencies used in an neighborhood of test frequency $\zeta_1$ with an AR(2) error process.

4.11 A comparison of power curve of the local $F$ test with approximate and true covariance matrix for IID, AR(2) and AR(4) error processes. We use bin size $B = 4$. And for the tapered cases, a 5% cosine taper was employed.

5.1 Plots of the periodogram of two subjects (one female and one male).

5.2 Plot of observed proportion at each evoking frequency using the female subjects, the male subjects and all the subjects respectively.

5.3 Diagnostic plots for quadratic logistic regression model for females, males, and all subjects, using the 20% trimmed mean of the times series of all 100 trials.

5.4 Diagnostic plots for quadratic logistic regression model for female, males and all subjects using the 10% trimmed mean of the times series of every 20 trials.

5.5 Mean log profile for females and males.

5.6 Residual plots for each subject in the DPOAE dataset in the fixed effects model (5.14) as a function of log $f_2$.

5.7 Residual plots for each subject in the DPOAE dataset in the mixed effects model (5.18) as a function of log $f_2$.

5.8 Residual plots for each subject in the DPOAE dataset in the mixed effects model (5.18) fit to frequencies over 1 kHz as a function of log $f_2$.

5.9 Plots of the fitted log local $F$ statistics for each subject in the DPOAE dataset in the fixed effects model (5.14) as a function of log $f_2$.

5.10 Plots of the fitted log local $F$ statistics for each subject in the DPOAE dataset in the mixed effects model (5.18) as a function of log $f_2$. 

xii
5.11 Residual plots for each subject in the mixed effects model (5.18) as a function of log $f_2$ for the simulated dataset. . . . . . . . . . . . . . . . 137

5.12 Plots of the fitted log local $F$ statistics for each subject in the mixed effects model (5.18) as a function of log $f_2$ for the simulated dataset. . 138
CHAPTER 1

INTRODUCTION

1.1 Spectral-based Tests for Periodicity of Time Series

A time series is a set of observations made sequentially ordered in time. The area of statistics which deals with the modeling of statistical processes observed sequentially in time is commonly referred to as time series analysis. The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions for the data. There exist two main of approaches to time series analysis, the time domain approach and the frequency domain (or harmonic) approach. The general framework for these two approaches for time series analysis have been described in Shumway and Stoffer (2000), Anderson (1994), and Brockwell and Davis (2002). Bloomfield (1976), Priestley (1981a, 1981b), Brillinger (1981), and Percival and Walden (1993) focus especially on the frequency domain approach.

The time domain approach typically focuses on modeling some future value of a time series as a parametric function of the current and past values. A common approach for modeling time series is the autoregressive (AR) model. An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. Another common approach for modeling time series models is the moving average (MA) model. That is, a moving average model is
conceptually a linear regression of the current value of the series against the Gaussian IID noise or random shocks of one or more prior values of the series. Autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models after G. Box and G. M. Jenkins, are typically applied to time series data, which consists of two parts, an AR part and MA part. Autoregressive integrated moving average (ARIMA) models also developed by Box and Jenkins can handle time-correlated modeling and forecasting that may also contain trends.

Conversely, in the frequency domain approach, we are concerned with *spectral analysis*: the exploration of cyclical patterns of data. The purpose of a spectral analysis is to decompose a complex time series with cyclical components into a few underlying sinusoidal (sine and cosine) functions of particular frequencies and amplitudes. In order to determine those periodicities that appear particularly strong or important in the original series, spectral-based tests for periodicities are used. In my research, I develop a regression model in the frequency domain based on the tapered Fourier transform of times series. For a certain design matrix and assumptions, the global and local $F$ tests can be derived from the $F$ test of the regression model. A global test uses the spectral estimates at all sample frequencies in the construction of the test statistic, whereas a local test uses only the points (bins) in the neighborhood of the test frequency. We will compare the global and local tests according to the size and power calculations.

### 1.2 Distortion Product Otoacoustic Emissions (DPOAEs)

Otoacoustic emissions (OAEs), discovered by David Kemp in 1978, are acoustical signals that can be detected in the ear canal. OAEs occur when the tympanum
receives vibrations transmitted backwards through the middle ear from the cochlea. The discovery of OAEs heralded a revolution in thinking about how the auditory system functions and has set the stage for a new appreciation of the nature of hearing impairments caused by damage to the cochlea. There are two types of OAEs, one is transient evoked OAEs (TEOAEs), the other is distortion product otoacoustic emissions (DPOAEs). DPOAEs data will be examined in this dissertation. DPOAEs are produced by intermodulation distortion in the healthy mammalian cochlea by the nonlinear interaction of two simultaneous stimulus tones \( f_1 \) and \( f_2 \) with \( f_2 > f_1 \) (Kemp, 1978). It is said to be distorted because it originates from the cochlea as a tonal signal that is not present in the eliciting stimuli tones. The source of this nonlinear interaction is thought to occur near the characteristic \( f_2 \) place on the basilar membrane. The most robust and frequently measured acoustic intermodulation distortion product in humans is \( 2f_1 - f_2 \), the so-called cubic distortion product (CDP).

Since DPOAEs depend on vulnerable physiologic processes, their presence or absence can be used as an indicator of cochlear status. One widespread application is newborn non-behavioral hearing screening. Because the most common cause of hearing loss in infants is the result of cochlear dysfunction, otoacoustic emissions are an effective screening instrument in this population. In this screening application, DPOAE testing reduces to the detection of the periodicity (CDP), usually using spectral-based tests. My research demonstrates that the spectral-based local \( F \) test in the frequency domain is the best method to detect the CDP (periodicity) for DPOAE data.

Since scientists are interested in both the presence of CDP and the level of CDP at different frequencies across the subjects, we will analyze the DPOAE data in two
ways. First, we model the binary results of the possible presence of a CDP using a logistic regression model. Secondly, a noncentral $F$ linear regression mixed effects models will be used to model the magnitude of the CDP for different subjects across different frequencies.

1.3 Generalized Linear Models

For each individual or experimental unit, the responses are not necessarily normally distributed. Thus, a generalized linear model (GLM) is applied to this situation. In this thesis, two kinds of GLMs are considered. One is for binary responses, which is called logistic linear regression, the other is for noncentral $F$ distributed responses, which is called noncentral $F$ linear regression.

Logistic regression analyzes binomially distributed data. In most simple terms, the response of logistic linear regression can only take one of two possible values, e.g., '0' or '1'. Usually we use '0' and '1' to denote the 'failure' and 'success' respectively. Logistic regression is very different from linear regression because of the underlying mathematics and the computational details. But linear regression plays an important role in the generalized linear model. We use the transformation of a link function that maps the unit interval onto the whole real line $(−\infty, +\infty)$. Then logistic regression applies maximum likelihood estimation (MLE) after transforming the dependent variable into the link function (e.g., logit or probit). In this way, logistic regression estimates the probability of a certain event occurring. In logistic regression, deviance analysis is used to evaluate the model. We will introduce more about logistic regression in Section 5.3.1.
In the noncentral \( F \) linear regression model, the response follows the noncentral \( F \) distribution on \( a \) and \( b \) degrees of freedom with noncentrality parameter \( \eta \). We can model the mean through the log link function. For the longitudinal response individuals measured repeatedly (in this case repetitions of the DPOAE experiment), GLMs can be extended to model the longitudinal data.

In this thesis, random effects model are explored for the DPOAE longitudinal dataset. A generalized Linear Mixed Effects Model (GLMM) is considered because we believe there are both fixed and random effects components. Due to the complication of probability density function (pdf) of the noncentral \( F \) distribution, MLE estimators are difficult to obtain. Then penalized quasi-likelihood (PQL) is used to carry out estimation of the parameters, which will be introduced in detail in Chapter 5. Pearson residuals are considered to diagnose the fit of the noncentral \( F \) mixed effects models.

1.4 Outline of the Thesis

After this introductory presentation, the rest of the thesis is organized as follows. Chapter 2 reviews some important fundamental theories of spectral analysis of stationary and harmonic processes. Chapter 3 develops a regression model in the frequency domain based on the Fourier transform of our time series model. We then present some of the spectral-based tests, which are the global test and three local tests, i.e., the hearing test, local \( F \) test and Thomson’s multitaper test. We will show that most of the tests can be derived from our regression model with the error term having an approximately diagonal covariance matrix. The distribution of the error term of the spectral regression model is based on the asymptotic distribution of the tapered Fourier transform of the error process. We will show that the distribution
has an approximately diagonal covariance matrix when the sample size is large and the spectral density functions (SDFs) of error processes have small dynamic range. Monte Carlo simulations are carried out in Chapter 4 to evaluate the performance of the global and local tests. The size and the power of test are introduced, calculated and compared for the global and local tests under a number different experimental conditions. In Chapter 5 we apply our spectral-based tests for periodicity to datasets from the Department of Speech and Hearing Science. The local $F$ test is chosen to detect the periodicity in Distortion Product Otoacoustic Emissions (DPOAEs) dataset based on the exploratory data analysis. The local $F$ test results are modeled by using two different generalized linear models (GLM). Logistic regression and non-central $F$ regression models are introduced respectively. Logistic regression model is used to model the binary results of the presence of a possible CDP. Noncentral $F$ linear regression model is used to model the level of a CDP. Finally, conclusions and discussions of the thesis are offered and the scope of the future work are suggested in Chapter 6.
Spectral analysis is a technique which is used to explore features of a time series in the frequency domain. For a review of spectral analysis, especially when analyzing periodicity using the periodogram or other spectral estimates, see Shumway and Stoffer (2000), Anderson (1994), Bloomfield (1976), Brockwell and Davis (2002), Priestley (1981a, 1981b), Brillinger (1981), and Percival and Walden (1993). In this chapter, some basic definitions in the spectral analysis of time series are introduced, and some important known statistical properties of spectral estimates are presented.

### 2.1 Stationary Processes

Consider the class of stationary processes. There are two types of stationary processes strictly stationary and weakly stationary. The stochastic process \( \{Z_t : t \in \mathbb{Z}\} \) is \textit{strictly stationary} if the joint distribution of a set of random variables at the different time points is unaffected by time shift \( \tau \). In other words the joint distribution of \( \{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\} \) is as the same as that of \( \{Z_{t_1+\tau}, Z_{t_2+\tau}, \ldots, Z_{t_n+\tau}\} \); i.e.,

\[
F_{t_1,t_2,\ldots,t_n}(a_1, a_2, \ldots, a_n) = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(a_1, a_2, \ldots, a_n).
\]
Unless we build a strictly stationary process out of IID processes, it can be difficult to show that the process is strictly stationary because the above equality needs to be satisfied for every \( \tau \). Thus, we assume a weaker form of stationary process. The process \( \{Z_t\} \) is weakly stationary if \( \{Z_t\} \) has two features: the first moment \( E\{Z_t\} \) does not depend on \( t \) and the second moment \( \text{cov}(Z_t, Z_{t+\tau}) \) only depends on the lag \( \tau \) and not the time \( t \); i.e., for all \( t \) and a constant \( \mu \),

\[
E(Z_t) = \mu \quad \text{and} \quad \text{cov}(Z_t, Z_{t+\tau}) = s_{Z,\tau},
\]

where \( \{s_{Z,\tau} : \tau \in \mathbb{Z}\} \) is called autocovariance sequence (ACVS). For Gaussian processes, strict stationary is equivalent to the weakly stationary, when \( \text{var}(Z_t) \) exists for all \( t \).

The autocovariance sequence, \( \{s_{Z,\tau} : \tau \in \mathbb{Z}\} \), is used to measure the covariance between random variables of a stationary process which are separated by \( \tau \) time units. One of the important properties of ACVS is that the sequence \( \{s_{Z,\tau}\} \) is positive semidefinite; i.e., for all \( n \geq 1 \), for any \( t_1, t_2, \ldots, t_n \) contained in \( \mathbb{Z} \), and for any set of real numbers \( a_1, a_2, \ldots, a_n \),

\[
\sum_{j=1}^{n} \sum_{k=1}^{n} s_{Z,t_j-t_k} a_j a_k \geq 0.
\]

Here are some popular examples of weakly stationary processes: independent and identically-distributed (IID) noise process, the moving average (MA) process, the autoregressive (AR) process, and the autoregressive moving average (ARMA) process.

For the IID noise process, \( \{Z_t\} \) is a sequence of independent random variables with \( E\{Z_t\} = 0 \) and \( \text{var}(Z_t) = \sigma^2 \) for all \( t \). Then \( \{Z_t\} \) is clearly stationary process with ACVS

\[
s_{Z,\tau} = \begin{cases} 
\sigma^2 & \text{if } \tau = 0; \\
0 & \text{if } \tau \neq 0,
\end{cases}
\]

8
which does not depend on $t$.

A process $\{Z_t\}$ is called a $q$th order MA process, denoted by $MA(q)$, if it has the form of

$$Z_t = e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q},$$

where $\theta_q \neq 0$ and $\{e_t\}$ is a IID noise process with zero mean and variance $\sigma^2$. It is straightforward to check that the $MA(q)$ process is a zero mean process with ACVS

$$s_{Z,t} = \left\{ \begin{array}{ll}
\sigma^2 \sum_{j=0}^{q-|\tau|} |\theta_j \theta_{j+|\tau|}| & \text{if } |\tau| \leq q; \\
0 & \text{if } |\tau| > q,
\end{array} \right.$$

where $\theta_0 = 1$ and $\theta_j = 0$ for $j > q$. Since the mean and ACVS of $MA(q)$ process are independent of $t$, the $MA(q)$ process is stationary.

A process $\{Z_t\}$ is called a $p$th order AR process, denoted by $AR(p)$, if it has the form of

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + e_t,$$

where $\phi_1$, $\phi_2$, $\ldots$ and $\phi_p$ are fixed coefficients. Not all of $AR(p)$ processes are stationary. In order to get a stationary $AR(p)$, we need to restrict the values of $\phi_j$ for $j = 1, \ldots, p$ to satisfy

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0, \quad (2.1)$$

for all moduli of $z \in \mathbb{C}$, the set of complex numbers, $|z| = 1$ (Brockwell and Davis (2002)[Chapter 3]). This gives us a general way to check the stationarity other than calculating the mean and ACVS. We can find the roots of $\phi(z)$ first, then show that the roots do not have modulus 1.
A process \( \{Z_t\} \) containing both moving average and autoregressive components is called an ARMA\((p,q)\) process. It is defined by the equation
\[
Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j} = e_t + \sum_{j=1}^{q} \theta_q e_{t-j},
\]
assuming \( \phi_p \neq 0 \) and \( \theta_q \neq 0 \), where \( \phi_j \) are \( \theta_j \) are constants. The MA\((q)\) and AR\((p)\) processes are the special cases of ARMA processes with \( p = 0 \) and \( q = 0 \) respectively; i.e. ARMA\((0,q)\) and ARMA\((p,0)\). The stationarity of ARMA process can be checked by only looking at the AR part and using the same method for the AR\((p)\) process, i.e., a stationary ARMA process exists if and only if \( \phi_j, j = 1, \ldots, p \) satisfy the Equation (2.1) for all \( |z| = 1 \). ARMA process can successfully model a wide range of dependencies in time series. We will use the AR\((2)\) and AR\((4)\) processes in later chapters.

The naive way to calculate the ACVS is to use the definition. But it is hard to calculate the ACVS by definition for ARMA processes with higher order of \( p \) and \( q \). There exists an alternative method to calculate the ACVS more efficiently for causal ARMA process. An ARMA\((p,q)\) process \( \{Z_t\} \) is causal if there exists constants \( \{\psi_j\} \) such that \( \sum_{j=0}^{\infty} |\psi_j| < \infty \) and for all \( t \)
\[
Z_t = \sum_{j=0}^{\infty} \psi_j e_{t-j},
\]
where \( \{e_t\} \) is a IID noise process. The constants \( \{\psi_j\} \) are determined by the relation
\[
\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \text{ or equivalently } \phi(z) \psi(z) = \theta(z).
\]
Then the ACVS can be obtained by using
\[
s_{Z,\tau} = E(Z_t, Z_{t-\tau}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|\tau|}, \quad (2.2)
\]
More details of calculating the constants \( \{\psi_j\} \) can be found from Brockwell and Davis (2002)[Chapter 3].
The ACVS is useful to describe the properties of time series, but it is usually unknown because the time series model is unknown. We can use the estimator of ACVS from a sample to assess the covariances or dependencies present in a time series. The standard estimate of $s_{Z,\tau}$ for $|\tau| < N$, $\hat{s}_{Z,\tau}$, is

$$\hat{s}_{Z,\tau}^{(p)} \equiv \frac{1}{N} \sum_{t=1}^{N-|\tau|} Z_t Z_{t+|\tau|}.$$ 

This estimator of $s_{Z,\tau}$ is biased for $|\tau| > 1$ because $E(\hat{s}_{Z,\tau}^{(p)}) = \left(1 - \frac{|\tau|}{N}\right) s_{Z,\tau}$. The unbiased estimator can be achieved by changing the multiplicative factor in front of the summation, i.e. using $1/(N - |\tau|)$ instead of $1/N$. But statisticians prefer the biased estimator to the unbiased estimator because the sequence of the biased estimator is always positive semidefinite, which satisfies the criteria of ACVS of a stationary process (Percival and Walden (1993)[Chapter 6]). The other reason is that the mean square error (MSE) of the biased estimator is smaller than that of the unbiased estimator for many practical stationary processes. Thus the biased estimator of ACVS is used in this dissertation.

2.2 Spectral Analysis of Stationary Processes

The Fourier transform is the foundation of the spectral analysis. We study only the Fourier transform for discrete-time processes here, which is called the discrete Fourier transform (DFT). It is widely employed in time series analysis and signal processing to analyze the frequencies contained in a time series. Let $f_\Delta = 1/(2\Delta)$ be the Nyquist frequency, where $\Delta$ is the sampling interval of the data. The Fourier transform can be defined for our discrete-time stationary process, $\{Z_t\}$, by

$$Z(f) = \Delta \sum_{t=-\infty}^{\infty} Z_t e^{-i2\pi f t \Delta}, \quad f \in [-f_\Delta, f_\Delta].$$
where \( i \) is the square root of negative one, i.e., \( i = \sqrt{-1} \). Because \( e^{ix} = \cos(x) + i \sin(x) \), \( \{Z(f) : f \in [-f_\Delta, f_\Delta]\} \) is a complex-valued stochastic process. The Fourier transform of a clean periodic signal of infinite length yields a Dirac function at the frequency of the signal, i.e. a line (or peak of zero width) with infinite magnitude.

In the time domain we have investigated the ACVS function of a stationary time series. Alternatively, we can study the spectral density function (SDF) as a function of the frequency. The SDF of a stationary time series \( \{Z_t\} \) is given by the Fourier transform of the ACVS \( \{s_{Z,\tau}\} \),

\[
S_Z(f) = \Delta \sum_{\tau=-\infty}^{\infty} s_{Z,\tau} e^{-i2\pi f \tau \Delta}, \quad f \in [-f_\Delta, f_\Delta]. \tag{2.3}
\]

A naive nonparametric spectral estimator of the SDF is to take the Fourier transform of the biased estimator of ACVS, which is called periodogram, \( \hat{S}_Z^{(p)}(f) \). Then, the periodogram at frequency \( f \in [-f_\Delta, f_\Delta] \) is defined as follows,

\[
\hat{S}_Z^{(p)}(f) = \Delta \sum_{\tau=-(N-1)}^{N-1} \hat{s}_Z^{(p)}_{\tau} e^{-i2\pi f \tau \Delta} = \frac{\Delta}{N} \sum_{\tau=-(N-1)}^{N-1} \sum_{t=1}^{N-|\tau|} Z_t Z_{t+|\tau|} e^{-i2\pi f \tau \Delta} = \frac{\Delta}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} Z_j Z_k e^{-i2\pi f (k-j) \Delta} = \frac{\Delta}{N} \left| \sum_{t=1}^{N} Z_t e^{-i2\pi ft \Delta} \right|^2. \tag{2.4}
\]

When \( \{Z_t\} \) is a Gaussian IID process, the periodogram is an unbiased estimate of the SDF. For non-IID noise at a fixed frequency, the periodogram is unbiased for the true SDF as \( N \) goes to infinity. At finite sample sizes, the periodogram suffers from leakage, in which the spectral estimates at some frequency tend to influence (“leak”) to other frequencies. This phenomenon is most prevalent for SDFs with a
large dynamic range. The *dynamic range*, which is defined as

$$10 \log_{10} \left( \frac{\max_{f \in [-f_{\Delta}, f_{\Delta}]} S_Z(f)}{\min_{f \in [-f_{\Delta}, f_{\Delta}]} S_Z(f)} \right),$$

is used to crudely characterize those stationary processes.

The left panel of Figure 2.1 shows the rectangular taper (no tapering) in time domain, associated spectral window of the rectangular taper, and periodogram of the AR(4) noise process. From the middle plot the sidelobes are prominently displayed for rectangular spectral window, which induces bias in the periodogram of the bottom plot. This bias is called *leakage*.

Tapering (e.g., Percival and Walden, 1993) is an effective mathematical manipulation performed on the time series before the periodogram is calculated to reduce the biases caused by leakage. A data taper \( \{h_t : t = 1, \ldots, N\} \) is a sequence of real-valued constants with \( \sum_t h_t^2 = 1 \). The modulus squared of the Fourier transform of the tapered data, \( \{h_t Z_t : t = 1, \ldots, N\} \), is called the direct spectral estimator,

$$\hat{S}_Z^{(d)}(f) \equiv \Delta \left| \sum_{t=1}^N h_t Z_t e^{-i2\pi f t} \right|^2 = |J_Z(f; h_t)|^2, \quad (2.5)$$

where

$$J_Z(f; h_t) = \sqrt{\Delta} \sum_{t=1}^N h_t Z_t e^{-i2\pi f t}, \quad |f| < 1/2,$$

is the tapered Fourier transform of \( \{Z_t : t = 1, \ldots, N\} \).

The periodogram is an estimator of the SDF in which we use no tapering; we let \( h_t = 1/\sqrt{N} \) for \( 1 \leq t \leq N \). Another commonly used taper is the \( p \times 100\% \) cosine taper, defined by

$$h_t = \begin{cases} \frac{C}{2} \left[ 1 - \cos \left( \frac{2\pi t}{|pN|+1} \right) \right], & 1 \leq t \leq \frac{|pN|}{2}; \\ C, & \frac{|pN|}{2} < t < N + 1 - \frac{|pN|}{2}; \\ \frac{C}{2} \left[ 1 - \cos \left( \frac{2\pi(N+1-t)}{|pN|+1} \right) \right], & N + 1 - \frac{|pN|}{2} \leq t \leq N. \end{cases} \quad (2.6)$$
Figure 2.1: Plots of rectangular taper and 50% cosine taper in time domain and frequency domain, and the direct spectral estimates of the AR(4) noise process using rectangular taper and 50% cosine taper respectively.
Here \( C \) is chosen so that \( \sum t h_t^2 = 1 \). Brillinger (1981)\[Table 3.3.1\] provides some other useful tapers.

The key idea behind tapering is to select \( \{h_t\} \) so that the sidelobes in the spectral estimate can be reduced. The right panel of Figure 2.1 shows 50% cosine taper in time domain, associated spectral window of 50% cosine taper, and the direct spectral estimator of the AR(4) noise process with 50% cosine taper. From the middle panel we can see the sidelobes of 50% cosine taper in the right plot is much smaller than the sidelobes of rectangular taper in the left plot. Thus tapering reduces the sidelobes. But the mainlobes encountered with a 50% cosine taper is larger than that of rectangular taper, which suffers more distortion of the SDF. Then reducing sidelobes (bias) is at the expense of increasing mainlobes (distortion). After we apply the both tapers to the AR(4) noise process, we can see the bias (leakage) of direct spectral estimates is reduced by using the 50% cosine taper in the bottom panel of Figure 2.1.

There are trade-offs involved in using tapers. The effect of tapering depends on both the type of taper and the true SDF. As we show in this dissertation, tapering is particularly important for processes that have an SDF with large dynamic range, but less important for those SDFs with smaller dynamic ranges.

Brillinger (1981) provides the following result for an asymptotic distribution of the direct spectral estimator of the SDF of \( \{Z_t\} \) at the Fourier frequencies. A Fourier frequency is defined as

\[
f_n \equiv \frac{n}{N\Delta}, \quad n = 0, 1, \ldots, \left\lfloor \frac{N}{2} \right\rfloor - 1,
\]

where \( N \) is the sample size.
Proposition 2.2.1 Assume that \( \{Z_t\} \) is a stationary Gaussian process with absolutely summable ACVS; i.e., \( \sum_{\tau} |s_{Z,\tau}| < \infty \). Let \( f_{k,N} = \frac{k(N)}{N\Delta} \) (\( k = 1, \ldots, K \)) be a set of non-zero, non-Nyquist Fourier frequencies that, for each \( k \), tend to some constant \( \xi_k \) as \( N \to \infty \), where \( k(N) \) is a sequence depending on \( N \) such that \( \left| \frac{k(N)}{N\Delta} - \xi_k \right| \to 0 \).

Let \( \Re(z) \) and \( \Im(z) \) denote respectively the real and imaginary parts of any complex number \( z \). Then as \( N \to \infty \) the vector

\[
J_{Z,N} \equiv (\Re(J_Z(f_1,N;h_t)), \ldots, \Re(J_Z(f_K,N;h_t)), \Im(J_Z(f_1,N;h_t)), \ldots, \Im(J_Z(f_K,N;h_t)))^T,
\]

is asymptotically \( 2K \)-variate normal with zero mean and a diagonal covariance matrix with diagonal elements,

\[
(S_Z(\xi_1)/2, \ldots, S_Z(\xi_K)/2, S_Z(\xi_1)/2, \ldots, S_Z(\xi_K)/2)^T.
\]

The theoretical proof can be found in Brillinger (1981, page 94). With the periodogram, this result holds for all \( N \) when \( \{Z_t\} \) is a Gaussian IID process (i.e., an asymptotic is not needed). For a Gaussian non-IID error process, especially the error process with large dynamic range, the independence and structure of covariance matrix might be violated for a small sample size, \( N \), depending upon the taper and the characteristics of the true underlying SDF. We will explore this issue more in this dissertation.

Another way to think of this result is in terms of the direct spectral estimate. The set of direct spectral estimates given by \( \{\hat{S}_Z^{(d)}(f_k,N) = |J_Z(f_k,N;h_t)|^2 : k = 1, \ldots, K\} \) are asymptotically independent \( S_Z(\xi_k)\chi_2^2/2 \) random variables (RVs) as \( N \to \infty \). From this theory, we see that the variance of periodogram and direct spectral estimators does not go to zero as the sample size goes to infinity because \( S_Z(f) \) is typically larger.
than zero. This demonstrates that both the periodogram and the direct spectral estimators are not consistent estimators because the probability that the periodogram and the direct spectral estimators become arbitrarily close to their asymptotic expected value of $S_Z(f)$ is zero. We need to use other spectral estimates to achieve consistency.

Consistency can be achieved by smoothing a direct spectral estimator, or by using multitaper spectral estimates. The traditional smoothing approach is to smooth the direct spectral estimator $S_Z^{(d)}(f)$. There are two smoothed direct spectral estimator: the discretely smoothed direct spectral estimator and the lag window spectral estimator. The discretely smoothed direct spectral estimator is formed by smoothing with a discrete convolution over a countable set of frequencies and the lag window spectral estimator uses a continuous convolution over a continuum of frequencies. See Daniell (1946), Bartlett (1950), Parzen (1957), Whittle (1958), Grenander and Rosenblatt (1957) and Blackman and Tukey (1958) to know more about the lag window spectral estimator. There are other smoothing methods. Wahba (1980) uses smoothing splines (see Section 3.3.1). Moulin (1994) and Percival and Walden (2000) employ wavelet-based shrinkage estimators of the SDF.

Thomson (1982) uses multiple orthogonal tapers (multitapers) to achieve consistency and recover the information lost due to tapering (in the form of an increase in variance). See Percival and Walden (1993) for an in-depth introduction to multitaper spectral estimation. For multitapers a set of approximate uncorrelated estimates of the power spectrum is computed, by pre-multiplying the data by orthogonal tapers which are constructed to minimize the spectral leakage due to the finite length of the series. Let $\{h_{k,t} : k = 1, \ldots, K, t = 1, \ldots, N\}$ denote $K$ orthonormal data tapers
\( (\sum_t h_{k,t}^2 = 1 \text{ and } \sum h_{k,t} h_{k',t} = 0 \text{ for } k \neq k') \). We define the \( k \)th eigenspectrum to be

\[
\widehat{S}_{Z,k}^{(mt)}(f) = \Delta \left| \sum_{t=1}^{N} h_{k,t} z_t e^{-i2\pi ft\Delta} \right|^2.
\] (2.7)

The average of the \( K \) eigenspectra, \( \widehat{S}_{Z}^{(mt)}(f) = K^{-1} \sum_{k=1}^{K} \widehat{S}_{Z,k}^{(mt)}(f) \), is the standard multitaper spectral estimator of the SDF. Under the same asymptotic given above for direct spectral estimates, the multitaper spectral estimator is consistent for the true SDF as long as the number of tapers, \( K \), are increased with the sample size, \( N \).

In terms of the choice of multitaper, the discrete prolate spheroidal sequences (DPSS) tapers and sine tapers are most commonly used. The DPSS tapers are designed to reduce the sidelobes in the spectral estimate. They are the solution to the time-frequency concentration problem in which we find the time limited sequence which has most of its energy concentrated in the interval \([-W, W]\) for some \( W \in \mathbb{R} \), which is less than the Nyquist frequency. In practice there are \( 2NW\Delta - 1 \) “useful” tapers that concentrate the ratio in \([-W, W]\). The choice of the bandwidth and number of tapers \( K \) thus represents the classical trade-off between spectral resolution and the stability or “variance” properties of the spectral estimate. For a fixed resolution bandwidth, \( 2W \), as the number of tapers used to form the estimate increase, the leakage of the estimator (and thus bias) increases and the variance decreases. The second commonly used tapers, the sine tapers introduced by Riedel and Sidorenko (1995), are easily calculated by

\[
h_{k,t} = \left( \frac{2}{N+1} \right)^{1/2} \sin \left( \frac{(k+1)\pi t}{N+1} \right) \quad k = 1, \ldots, K; \quad t = 1, \ldots, N.
\]

They are designed to reduce the smoothing bias, at a compromise to sidelobe reduction. For a given \( K \) the sine tapers are concentrated in the bandwidth \([-W, W]\) for \( W = (K+1)/(2(N+1)\Delta) \). As \( K \) increases we lose resolution but decrease
the variance of the multitaper spectral estimator. Walden (2000) demonstrates other classes of multiple tapers, some which include the smoothed versions of direct spectral estimators above.

### 2.3 Introduction of Harmonic Processes

A common use for spectral analysis of time series, is to examine a periodic signal in the presence of background noise. Suppose that process \( \{X_t : t \in \mathbb{Z}\} \), collected at a sampling interval of \( \Delta \), satisfies the sinusoid plus noise model,

\[
X_t = \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l t \Delta + \phi_l) + Z_t.
\]  

Here the frequencies \( \{\zeta_l : l = 1, \ldots, M\} \) are fixed and known with unknown amplitudes \( \{A_l\} \). We assume that the phases \( \{\phi_l\} \) are independent draws from a uniform\([-\pi, \pi]\) distribution, which are independent of the error process \( \{Z_t : t \in \mathbb{Z}\} \). The amplitudes \( \{A_l\} \) determine the height of the sinusoid functions and the phases \( \{\phi_l\} \) determine the start point of the sinusoid. We let the error process \( \{Z_t\} \) be a mean zero, stationary Gaussian process with autocovariance sequence (ACVS), \( \{s_{Z,\tau} : t \in \mathbb{Z}\} \). The process contains \( L \) cosine waves, which yields \( L \) lines at the frequencies of the periodicities in the frequency domain. Thus we can use the harmonic analysis to determine the line components; i.e., the spikes in the spectrum corresponding to a periodic in terms of their frequency, amplitude, and phase.

The ACVS of process \( \{X_t\} \) can be calculated as following (Exercise [10.3] of Percival and Walden (1993))

\[
s_{X,\tau} = \text{cov}(X_t, X_{t-\tau})
\]
\[ \text{cov} \left( \sum_{k=1}^{L} A_k \cos(2\pi \zeta_k t \Delta + \phi_k), \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l) + Z_t \right) \]

\[ = \text{cov} \left( \sum_{k=1}^{L} A_k \cos(2\pi \zeta_k t \Delta + \phi_k), \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l) \right) \]

\[ + \text{cov} \left( \sum_{k=1}^{L} A_k \cos(2\pi \zeta_k t \Delta + \phi_k), Z_t \right) + \text{cov} \left( \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l), Z_{t - \tau} \right) \]

\[ + \text{cov}(Z_t, Z_{t - \tau}). \]  

(2.9)

Both

\[ \text{cov} \left( \sum_{k=1}^{L} \cos(2\pi \zeta_k t \Delta + \phi_k), Z_t \right) = 0 \]

and

\[ \text{cov} \left( \sum_{l=1}^{L} \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l), Z_{t - \tau} \right) = 0 \]

because we assume that phases \( \{\phi_l\} \) are independent of the error process \( \{Z_t\} \). Then

Equation 2.9 becomes

\[ s_{X,\tau} = \text{cov} \left( \sum_{k=1}^{L} A_k \cos(2\pi \zeta_k t \Delta + \phi_k), \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l) \right) + s_{Z,\tau} \]

\[ = \sum_{k=1}^{L} \sum_{l=1}^{L} A_k A_l \text{cov}(\cos(2\pi \zeta_k t \Delta + \phi_k), \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l)) + s_{Z,\tau} \]

\[ = \sum_{k=1}^{L} \sum_{l=1}^{L} A_k A_l \mathbb{E}(\cos(2\pi \zeta_k t \Delta + \phi_k) \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l)) + s_{Z,\tau}. \]  \( (2.10) \)

The last step uses the fact that \( \{\phi_l\} \) are IID uniform\([-\pi, \pi]\).

\[ E(X_t) = E \left( \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l t \Delta + \phi_l) \right) \]

\[ = \sum_{l=1}^{L} A_l E(\cos(2\pi \zeta_l t \Delta + \phi_l)) \]

\[ = \sum_{l=1}^{L} A_l \int_{-\pi}^{\pi} \cos(2\pi \zeta l t \Delta + \phi_l) \frac{1}{4\pi} d\phi_l = 0. \]

For the double summations, if \( k \neq l \), we have

\[ E(\cos(2\pi \zeta_k t \Delta + \phi_k) \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l)) \]

20
\[ E(\cos(2\pi \zeta_k t \Delta + \phi_k)) E(\cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l)) = 0, \]

because phases \( \{\phi_k\} \) and \( \{\phi_l\} \) are independent. For \( k = l \), the expectation is

\[ E(\cos(2\pi \zeta_l t \Delta + \phi_l) \cos(2\pi \zeta_l (t - \tau) \Delta + \phi_l)) \]

\[ = \frac{1}{4\pi} \int_{-\pi}^{\pi} \cos((4\pi \zeta_l t \Delta - 2\pi \zeta_l \tau) + 2\phi_l) + \cos(2\pi \zeta_l \tau \Delta) \, d\phi_l = \frac{\cos(2\pi \zeta_l \tau \Delta)}{2}. \]

After plugging these results into Equation (2.10), we get

\[ s_{X,\tau} = \sum_{l=1}^{L} A_l^2 \cos(2\pi \zeta_l \tau \Delta)/2 + s_{Z,\tau}. \]

The SDF of a harmonic process \( \{X_t\} \), \( S_X(f) \), is also given by the DFT of the ACVS, \( \{s_{X,\tau}\} \). For a frequency \( f \in [-f_\Delta, f_\Delta] \) it is

\[ S_X(f) = \Delta \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f \tau \Delta}. \]

Then, the periodogram of a harmonic process \( \{X_t\} \) at frequency \( f \in [-f_\Delta, f_\Delta] \) is very similar to the periodogram of the error process \( \{Z_t\} \) in previous section by using \( X_t \) instead of \( Z_t \) in the formula,

\[ \hat{S}_X^{(p)}(f) = \Delta \left| \frac{1}{N} \sum_{t=1}^{N} X_t e^{-i2\pi f t \Delta} \right|^2. \tag{2.11} \]

Comparing the periodogram ordinates of \( X_t \) with that of \( Z_t \), we can see there are \( L \) peaks at the frequencies of the periodicities in the harmonic process \( \{X_t\} \) due to the Fourier transform of \( L \) periodicities.

The direct spectral estimator with a data taper \( \{h_t\} \) has the form

\[ \hat{S}_X^{(d)}(f) \equiv \Delta \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \Delta} \right|^2 = |J_X(f; h_t)|^2, \tag{2.12} \]

where \( J_X(f; h_t) = \sqrt{\Delta} \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \Delta} \) is the tapered Fourier transform of \( \{X_t\} \).

We will analyze the tapered Fourier transform in depth when we study the tests for
periodicity in Chapter 3. All types of the tapers, such as 100\textit{p}% cosine taper, DPSS tapers and sine tapers introduced in Section 2.2, can be applied to harmonic processes to reduce the effect of the leakage.

We have reviewed some important fundamental theories of spectral analysis of stationary and harmonic processes. These theories will be used to develop spectral-based test for periodicity in the next chapter.
CHAPTER 3

SPECTRAL-BASED TESTS FOR PERIODICITY

In this chapter we consider tests for a significant periodicity (i.e., amplitude) based on the spectral analysis of the time series \( \{X_t : t \in \mathbb{Z}\} \), which satisfies the model

\[
X_t = \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l t \Delta + \phi_l) + Z_t. \tag{3.1}
\]

Here the frequencies \( \{\zeta_l : l = 1, \ldots, L\} \) are fixed and known with unknown amplitudes \( \{A_l\} \). We assume the phases \( \{\phi_l\} \) are independent draws from a uniform\([-\pi, \pi]\) distribution, which are independent of the error process \( \{Z_t : t \in \mathbb{Z}\} \). We let the error process \( \{Z_t\} \) be a mean zero, stationary Gaussian process with autocovariance sequence (ACVS), \( \{s_Z, \tau : \tau \in \mathbb{Z}\} \). Without loss of generality, we test whether the first amplitude (periodicity), \( A_1 \), is significantly different from zero; that is,

\[
H_0: A_1 = 0 \quad \text{versus} \quad H_1: A_1 \neq 0.
\]

We will now motivate why it is difficult to test for a significant amplitude (periodicity) based on a time domain (regression) analysis of \( \{X_t\} \) directly. As an example let \( \Delta = 1, L = 3, \zeta_1 = 0.2, \zeta_2 = 0.25, \zeta_3 = 0.3, A_1 = 1, A_2 = 2, A_3 = 3 \) and \( \phi_l = 0 \) for \( l = 1, 2, 3 \). Then the time series model (3.1) becomes

\[
X_t = \sum_{l=1}^{3} A_l \cos(2\pi \zeta_l t) + Z_t = \cos(2(0.2)\pi t) + 2 \cos(2(0.25)\pi t) + 3 \cos(2(0.3)\pi t) + Z_t, \tag{3.2}
\]
Figure 3.1: Time series plots and ACVS plots of three periodicities, AR(2) noise process, and a time series with three periodicities plus AR(2) noise process.
where \( t = 0, \ldots, 255 \). Assume that \( \{Z_t : t \in \mathbb{Z}\} \) is a Gaussian AR(2) noise process, 
\[ Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + e_t, \]
where \( \phi_1 = 0.75, \phi_2 = -0.5, \) and \( \{e_t : t \in \mathbb{Z}\} \) is a set of IID \( \mathcal{N}(0, \sigma^2) \) (with \( \sigma^2 = 1 \)) random variables. The left panel of Figure 3.1 shows times series plots of the three periodicities alone, the AR(2) noise process, and their sum, a realization of (3.2), i.e., the three periodicities plus the AR(2) noise process. The cycles of the periodicity can be seen very clearly from the individual plots or each periodicity, but are harder to identify from a realization of \( \{X_t\} \). Certainly, using the time series plot alone, it is hard to tell if this time series contains a significant periodicity with frequency \( \zeta_1 = 0.2 \).

As we introduced in Chapter 2, the autocovariance sequence (ACVS) can be used to characterize a time series in the time domain. From Shumway and Stoffer (2000)[Chapter 2], the ACVS of a periodicity behaves like a cosine wave with the strongest dependence at multiples of the seasonal lag \( 1/\Delta \zeta \). The ACVS of the noise process also look like a damped down cosine wave. Thus it is hard to identify the periodicity of interest from the ACVS after we add the periodicities and the noise process together. The right panel of Figure 3.1 gives the ACVS of the three periodicities, the AR(2) noise process and the time series with the periodicities plus the AR(2) noise process. The first three plots from the top show the ACVS of the three periodicities respectively. The seasonality of the cosine wave can be seen obviously in these plots, which have the most strongly dependence at the lag \( k/0.2, k/0.25, \) and \( k/0.3 \) for \( k = 1, 2, \ldots, \) respectively (the multiples of the seasonal lag \( 1/\Delta \zeta_l \) for \( \Delta = 1 \) and \( l = 1, 2, 3 \)). From the ACVS plot of the AR(2) noise process, the ACVS curve does look like some damped down cosine wave. But we could not see any clear cosine
pattern and figure out the presence of the periodicity of interest if we only study the ACVS of time series with the three periodicities plus the noise process.

A time-domain-based test for the significance of the first periodicity, $A_1$, is typically based on a regression model. For the general model defined by (3.1) we obtain

$$X_t = \sum_{l=1}^{L} A_l \cos(\phi_l) \cos(2\pi \zeta_l t\Delta) + \sum_{l=1}^{L} A_l \sin(\phi_l) \sin(2\pi \zeta_l t\Delta) + Z_t,$$

for $t = 1, \ldots, N$ where $N$ is the sample size. Letting $X_{td} = (X_1, X_2, \ldots, X_N)$ be an $N$-dimensional vector of observed responses, we can write model (3.3) in matrix form:

$$X_{td} = M_{td} \beta_{td} + Z_{td},$$

where

$$M_{td}^T = \begin{pmatrix}
\cos(2\pi \zeta_1 \Delta) & \cos(4\pi \zeta_1 \Delta) & \cdots & \cos(2N\pi \zeta_1 \Delta) \\
\sin(2\pi \zeta_1 \Delta) & \sin(4\pi \zeta_1 \Delta) & \cdots & \sin(2N\pi \zeta_1 \Delta) \\
\cos(2\pi \zeta_2 \Delta) & \cos(4\pi \zeta_2 \Delta) & \cdots & \cos(2N\pi \zeta_2 \Delta) \\
\sin(2\pi \zeta_2 \Delta) & \sin(4\pi \zeta_2 \Delta) & \cdots & \sin(2N\pi \zeta_2 \Delta) \\
\vdots & \vdots & \ddots & \vdots \\
\cos(2\pi \zeta_L \Delta) & \cos(4\pi \zeta_L \Delta) & \cdots & \cos(2N\pi \zeta_L \Delta) \\
\sin(2\pi \zeta_L \Delta) & \sin(4\pi \zeta_L \Delta) & \cdots & \sin(2N\pi \zeta_L \Delta)
\end{pmatrix}$$

is a $N \times 2L$ design matrix,

$$\beta_{td}^T = (A_1 \cos(\phi_1), -A_1 \sin(\phi_1), A_2 \cos(\phi_2), -A_2 \sin(\phi_2), \ldots, A_L \cos(\phi_L), -A_L \sin(\phi_L)),$$

is a $2L$-dimensional vector of unknown parameters (that depends on the unknown amplitudes and phases), and $Z_{td}$ is a $N$-dimensional random vector of unobserved errors, which has a $N(0, \Sigma_{td})$ distribution. The covariance matrix of the errors depends on the noise process, which is usually unknown and needs to be estimated from the data.

We can use either maximum likelihood or minimize the conditional sum-of-squares to get the estimates of the parameter $\beta_{td}$ and the covariance parameters, $\theta_{td}$, say
in $\Sigma_{td}$ (e.g., for our AR(2) process $\theta_{td} = (\sigma^2, \phi_1, \phi_2)^T$). See Brockwell and Davis (2002)[Chapter 5 and 6]) for further details. Once we get the estimates of the parameters of the model (i.e., we estimate $\beta_{td}$ using $\hat{\beta}_{td}$ and $\theta_{td}$ using $\hat{\theta}_{td}$), we can use a regression $F$ test (e.g., Ravishanker and Dey (2002) [Chapter 4 and 7]) to test if the amplitude $A_1$ is significantly different from zero. Testing $A_1 = 0$ is equivalent to testing $C_{td}^T \beta_{td} = 0$, where

$$C_{td} = \begin{pmatrix} 1 & 0 & [2L - 2 \text{ zeroes}] \\ 0 & 1 & [2L - 2 \text{ zeroes}] \end{pmatrix}^T$$

is a $2L \times 2$ contrast matrix. The test statistic, substituted with estimates of $\beta_{td}$ and $\theta_{td}$ (equivalently $\hat{\Sigma}_{td}$) is

$$T_{\beta_{td} \hat{\Sigma}_{td} \hat{\theta}_{td}} = \frac{(C_{td}^T \hat{\beta}_{td})^T (C_{td}^T (M_{td}^T \hat{\Sigma}_{td}^{-1} M_{td})^{-1} C_{td})^{-1} (C_{td}^T \hat{\beta}_{td})/2} {(X_{td} - M_{td} \hat{\beta}_{td})^T \hat{\Sigma}_{td}^{-1} (X_{td} - M_{td} \hat{\beta}_{td})/N - 2L}.$$ 

Under the null hypothesis of $C^T \beta = 0$ (i.e., $A_1 = 0$), $T_{\beta_{td} \hat{\Sigma}_{td} \hat{\theta}_{td}}$ approximately has a $F$ distribution on 2 and $N - 2L$ degrees of freedom (Ravishanker and Dey (2002) [Chapter 4 and 7]).

Since the noise process is usually unknown, the number of parameters in $\theta_{td}$ is undetermined (e.g., $p$ and $q$ are unknown if we choose to use an ARMA($p$, $q$) process to model the noise process). Thus, it is difficult to estimate $\Sigma_{td}$ (a function of $\theta_{td}$) accurately in the test statistic without knowing the noise process, especially when the noise process exhibits similar periodic patterns in the ACVS to the periodicities. We will compare the time-domain-based test with other spectral-based tests in the frequency domain through the simulations in Chapter 4. We will demonstrate problems with the above time domain testing approach when the noise process is unknown.

In the introduction we showed that the Fourier transform is a bridge used to convert a time series from the time domain to the frequency domain. The spectral
density function (SDF), based on a Fourier analysis of a time series, is used to characterize the time series in frequency domain. Though the SDF is usually unknown, the periodogram or direct spectral estimator can be used to estimate the SDF. The relationships among the ACVS, SDF, as well as the estimates of the ACVS and SDF are shown in Figure 3.2. The left panel of Figure 3.3 shows the periodogram (gray line) of the example time series shown previously with three periodicities plus the AR(2) noise process. The black solid line gives the true SDF of the AR(2) process. The three peaks in the plot occur at the frequencies of the periodicities (i.e., $\zeta_1 = 0.2$, $\zeta_2 = 0.25$ and $\zeta_3 = 0.3$). The different magnitude of peaks increase with the amplitude of each periodicity. Thus the periodicity of frequency $\zeta_3 = 0.3$ has the largest

![Diagram of transformation between the time domain and frequency domain](image)

Figure 3.2: Plot of diagram of transformation between the time domain and frequency domain
amplitude $A_3 = 3$, which corresponds to the highest peak occurring at $\zeta_3 = 0.3$. The majority of the periodogram (gray line) are due to the dependence in the periodogram of the AR(2) noise process.

We divide spectral-based tests for periodicity into two groups: global tests and local tests. A global test uses the spectral estimates at all non-zero, non-Nyquist Fourier frequencies to carry out the spectral-based test ($m = \lceil N/2 \rceil = 128$, as shown in the left panel of Figure 3.3). The Nyquist Fourier frequency cannot be included because there is no way to recover the amplitude spectrum at this frequency. For the AR(2) noise process the SDF is not constant over all the frequencies. The right panel of Figure 3.3 is a zoomed-in plot around a neighborhood of the test frequency, $\zeta_1$. This plot shows the points of spectral estimates used for a local test, which only uses a neighborhood of the test frequency $\zeta_1$. Skipping one Fourier frequency on either side,
this local test uses nine points of the spectral estimates \((m = 9, \text{ four on either sides, plus one at the frequency of interest})\). We can see that it is reasonable to assume that the SDF of error process is locally constant in the neighborhood of \(\zeta_1\) from the figure. Comparing the global test with the local test, we can see that the local test uses much less points of the spectral estimates than the global one. Hence we would expect that the local test is less efficient than a global test.

There is a rich history of spectral-based tests for periodicity in the literature. One of the first examples can be found in Schuster (1898), in which the periodogram, the squared magnitude of the Fourier transform of a time series, is used to test for significant periods. Assuming that the error process, \(\{Z_t\}\), is a Gaussian IID noise process with variance \(\sigma^2\), he considers a periodogram-based test that depends on knowing \(\sigma^2\). Fisher (1929) extends this test, by estimating \(\sigma^2\) using all the periodogram ordinates, and provides the exact distribution of the test statistic in the Gaussian IID noise case. Since Fisher’s global spectral-based test is only powerful for testing a periodicity at only one Fourier frequency, Siegel (1980) derives a test for multiple periodicities based on the extremes of the periodogram rather than just the maximum. All these global tests for periodicity are used in the case of an unknown frequency. Percival (1994) outlines a global \(F\) test for known frequencies. All these global tests are used when the error process, \(\{Z_t\}\), is IID Gaussian. For non-IID (i.e. colored) processes, we will propose a global spectral-based test for periodicity that uses a nonparametric estimate of the spectral density function.

Although a local test statistic is less efficient than a global test because only part of the spectral estimates are used, there are a number of reasons why a local test is appropriate, especially for non-IID noise processes. Firstly, the entire autocovariance
sequence or equivalently, as we have shown in the example, the SDF of the time series does not need to be known; we only need to make assumptions about the process around the frequency of interest. Secondly, local tests of periodicity can be robust to nuisance periodicities that may appear in the time series, as we shall demonstrate.

There have been a number of local spectral-based tests proposed in the literature. When \( \{Z_t\} \) is a Gaussian non-IID process, Thomson (1982) provides a test for a significant periodicity based on the analysis of a complex-valued regression model of multitaper spectral estimates. Thomson’s test is a local because it does not use the information at all frequencies to provide a test for a significant periodicity. It only uses the frequencies in the neighborhood of the test frequency, as defined by the bandwidth, \( W \), of the multiple tapers (the definition of \( W \) is given on P.18). Denison et al. (1999) extends this to test for the significance of a number of periodicities, some of which may be within the bandwidth \( W \). Gorga et al. (1993) provide the standard periodogram-based test employed in the hearing sciences literature, which compares the periodogram at the frequency of interest with the maximum of the average of a number of spectral bins lower and higher of the test frequency. Almasri (2003) defines the local \( F \) test, which is the similar to the Gorga et al. statistic, except the denominator of the statistic is the average of both the lower and higher neighborhoods of the test frequency instead of maximum. Craigmile and King (2004) show via Monte Carlo simulations that the local F test is more powerful than the Gorga et al. test.

In this chapter, the tapered Fourier transform of stationary process is analyzed in depth. Some calculations and simulations are carried out to explore the (asymptotic) distribution of tapered Fourier transform of stationary process. Then we build a regression model in the spectral domain based on the tapered Fourier transform. Global
$F$ tests are derived from the regression $F$ test with or without nuisance periodicities for the IID noise process. We extend the global $F$ test to the non-IID noise process by using a smoothing spline estimate of the spectral density function. The local $F$ test Almasri (2003) and Thomson’s multitaper $F$ test Thomson (1982) are derived in detail from the regression $F$ test under some assumptions. The local $F$ test with tapering is examined for the non-IID noise process with SDFs having a large dynamic range. A standard hearing test Gorga et al. (1993), which can not derived from the regression $F$ test, is also introduced and compared to the local $F$ test and Thomson’s multitaper $F$ test.

3.1 Analysis of the Tapered Fourier Transform of Stationary Process

Before defining the local and global spectral tests, we understand the statistical properties of the tapered spectral estimates under our model for periodicity, Equation (3.1). For some taper $\{h_t : t = 1, \ldots, N\}$, the tapered Fourier transform of $\{X_t : t = 1, \ldots, N\}$ at frequency $\eta \in [-\frac{1}{2\Delta}, \frac{1}{2\Delta}]$ is,

$$J_X(\eta; h_t) = \sqrt{\Delta} \sum_{t=1}^{N} h_t X_t e^{-i2\pi\eta t \Delta}$$

$$= \sqrt{\Delta} \sum_{t=1}^{N} h_t \left( \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l t \Delta + \phi_l) + Z_t \right) e^{-i2\pi\eta t \Delta}$$

$$= \sqrt{\Delta} \sum_{l=1}^{L} \sum_{t=1}^{N} h_t A_l \cos(2\pi \zeta_l t \Delta + \phi_l) e^{-i2\pi\eta t \Delta} + \sqrt{\Delta} \sum_{t=1}^{N} h_t Z_t e^{-i2\pi\eta t \Delta}$$

$$= \sqrt{\Delta} \sum_{l=1}^{L} A_l \sum_{t=1}^{N} h_t \cos(2\pi \zeta_l t \Delta + \phi_l) e^{-i2\pi\eta t \Delta} + J_Z(\eta; h_t)$$

$$= \sqrt{\Delta} \sum_{l=1}^{L} \sum_{t=1}^{N} h_t e^{i\phi_l} \left( e^{i2\pi \zeta_l t \Delta} + e^{-i2\pi \zeta_l t \Delta} \right) e^{-i2\pi\eta t \Delta} + J_Z(\eta; h_t)$$
where \(C_t = A_t e^{i \phi_t}/2\) for each \(l = 1, \ldots, L\). ‘\(^*\)’ denotes the complex conjugation operator, and \(H(\eta) = \Delta \sum_{t=1}^{N} h_t e^{-i 2 \pi \eta t} \) is the transfer function of the taper \(\{h_t\}\). Based on Equation (3.4), if we treat the tapered Fourier transform of the noise process \(\{Z_t\}\), \(J_Z(\eta; h_t)\), as the error term, we can model the tapered Fourier transform of \(\{X_t\}\) as complex-valued linear regression. The distribution of \(J_Z(\eta; h_t)\) is complex Gaussian. Complex Gaussian is defined as follows. Letting \(z = x + iy\), \(z\) is complex Gaussian if the joint distribution of \(x\) and \(y\) is bivariate Gaussian. See Miller (1974)[Sections II.3 and II.4] for further properties of complex Gaussian processes.

We will examine the stationarity of process \(\{J_Z(\eta; h_t)\}\) first. From Chapter 2 we want to calculate the mean and covariance of \(\{J_Z(\eta; h_t)\}\). The mean of \(J_Z(\eta; h_t)\) for frequency \(\eta \in [-\frac{1}{2 \Delta}, \frac{1}{2 \Delta}]\) is

\[
E(J_Z(\eta; h_t)) = E\left(\sqrt{\Delta} \sum_{t=1}^{N} h_t Z_t e^{-i 2 \pi \eta t} \right) = \sqrt{\Delta} \sum_{t=1}^{N} h_t E(Z_t) e^{-i 2 \pi \eta t} = 0,
\]

since \(E(Z_t) = 0\) for each \(t\). For \(\eta, \eta' \in [-\frac{1}{2 \Delta}, \frac{1}{2 \Delta}]\), the covariance is given by

\[
\text{cov}(J_Z(\eta; h_t), J_Z(\eta'; h_{t'})) = E(J_Z(\eta; h_t), J_Z(\eta'; h_{t'}))
\]

\[
= E\left(\Delta \sum_{t=1}^{N} \sum_{t'=1}^{N} h_t h_{t'} Z_t Z_{t'} e^{-i 2 \pi (\eta t - \eta' t')} \Delta \right)
\]

\[
= \Delta \sum_{t=1}^{N} \sum_{t'=1}^{N} h_t h_{t'} \text{cov}(Z_t, Z_{t'}) e^{-i 2 \pi (\eta t - \eta' t')} \Delta
\]

\[
= \Delta \sum_{t=1}^{N} \sum_{t'=1}^{N} h_t h_{t'} s_{Z,t',t} e^{-i 2 \pi (\eta t - \eta' t')} \Delta,
\]

because \(E(J_Z(\eta; h_t)) = 0\) for each \(\eta\) and \(\{Z_t\}\) is a stationary process with ACVS \(\{s_{z,\tau} : \tau = t' - t \in \mathbb{Z}\}\). It is clear that in the frequency domain \(\{J_Z(\eta; h_t)\}\) is
generally not a stationary process on $[-\frac{1}{2\Delta}, \frac{1}{2\Delta}]$ because the covariance depends on the frequency $\eta$. Only when $\{Z_t\}$ is a IID Gaussian process is $\{J_Z(\eta; h_t)\}$ stationary. Now we will study the covariance of $\{J_Z(\eta; h_t)\}$ in depth. From Percival and Walden (1993)[page 197 in Chapter 6] we can write

$$s_{Z,t,t'} = \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} e^{i2\pi u(t-t')} \Delta S_Z(u) du.$$ 

Letting $\eta' = \eta + \tau \eta$ (where $\tau \eta$ is a constant) and plugging the above into Equation (3.5), the covariance becomes

$$\text{cov}(J_Z(\eta; h_t), J^*_Z(\eta'; h_{t'})) = \Delta \sum_{t=1}^{N} \sum_{t'=1}^{N} h_t h_{t'} \left[ \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} e^{i2\pi u(t-t')} \Delta S_Z(u) du \right] e^{-i2\pi(\eta t - \eta' t') \Delta}$$

$$= \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left[ \sum_{t=1}^{N} \sum_{t'=1}^{N} h_t h_{t'} e^{-i2\pi(\eta t - \eta' t') \Delta} e^{i2\pi u(t-t')} \Delta \right] du$$

$$= \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left[ \sum_{t=1}^{N} h_{t'} e^{-i2\pi(\eta t - u) \Delta} \right] \left[ \sum_{t'=1}^{N} h_{t'} e^{-i2\pi(\eta' t' - u) \Delta} \right] du$$

$$= \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left\{ \sum_{t=1}^{N} h_{t'} e^{-i2\pi(\eta - u) \Delta} \right\} \left\{ \sum_{t'=1}^{N} h_{t'} e^{-i2\pi(\eta' - u) \Delta} \right\} du$$

$$= \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) H(\eta - u) H^*(\eta' - u) du$$

$$= \frac{1}{\Delta} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) H(\eta - u) H^*(\eta + \tau \eta - u) du,$$

$$= \frac{1}{\Delta} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) H(\eta - u) H^*((\eta - u) + \tau \eta) du,$$

where $H(\cdot)$ is the transfer function of $\{h_t\}$. If $S_Z(u)$ is constant and the frequency $\eta'$ is close to $\eta$, Equation (3.6) is approximately

$$\text{cov}(J_Z(\eta; h_t), J^*_Z(\eta'; h_{t'})) \approx \frac{S_Z(\eta)}{\Delta} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} |H(\eta - u)|^2 du, = S_Z(\eta),$$

since $\frac{1}{\Delta} \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} |H(f)|^2 df = 1$ (Percival and Walden (1993)[page 208 in Chapter 6]).

If the rectangular taper (corresponding to no tapering) is used, this simplifies to

$$\text{cov}(J_Z(\eta; h_t), J^*_Z(\eta'; h_{t'}))$$
\[
\Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left[ \sum_{t=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi(\eta-u)t\Delta} \right] \left[ \sum_{t'=1}^{N} \frac{1}{\sqrt{N}} e^{-i2\pi(-\eta'+u')t'\Delta} \right] du.
\]

Using Percival and Walden (1993) [Exercise 1.3],

\[
cov(J_Z(\eta; h_t), J_Z^*(\eta'; h_{t'})) = \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left[ \sqrt{N} e^{i(N+1)\pi(\eta-u)\Delta} D_N ((-\eta + u)\Delta) \right] \left[ \sqrt{N} e^{i(N+1)\pi(\eta'-u')\Delta} D_N ((\eta' - u)\Delta) \right] du
\]

\[
= \Delta N \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} S_Z(u) \left[ e^{i(N+1)\pi(\eta'-\eta)\Delta} D_N ((-\eta + u)\Delta) D_N ((\eta' - u)\Delta) \right] du,
\]

(3.7)

where \( D_N(\eta) \equiv \frac{\sin(N\pi\eta)}{N\sin(\pi\eta)} \) is Dirichlet’s kernel.

From the above expressions of the covariance of \( \{J_Z(\eta; h_t)\} \), the stationarity of process \( \{J_Z(\eta; h_t)\} \) depends on the SDF of noise process and the transfer function (or Dirichlet’s kernel if no tapering is used). Thus inspecting Equation (3.6) and (3.7) we see that \( \{J_Z(\eta; h_t)\} \) is stationary only when \( \{Z_t\} \) is an IID process (since \( S_Z(f) \) is constant for all \( f \)). In practice \( \{Z_t\} \) is not IID and thus \( \{J_Z(\eta; h_t)\} \) is not stationary.

In other words, the process \( \{J_Z(\eta; h_t)\} \) is stationary if \( \{Z_t\} \) is a IID Gaussian process; the \( \{J_Z(\eta; h_t)\} \) is approximately stationary if the sample size \( N \) is large, the \( \eta' \) is close to \( \eta \), and the SDF is locally constant in the neighborhood of \( \eta \). In the tests for periodicity that we consider in this dissertation we restrict to the Fourier frequencies. Let \( \{\eta_j : j = 1, \ldots, K\} \), where \( K = \lfloor N/2 \rfloor \), denote the set of non-zero, non-Nyquist Fourier frequencies. Letting

\[
J^c_Z(\eta; h_t) = (J_Z(\eta_1; h_t), \ldots, J_Z(\eta_K; h_t))^T
\]

(3.8)

we have that \( E(J^c_Z) = 0 \) and the covariance matrix is

\[
\Sigma^c = \text{cov}(J^c_Z(\eta; h_t), J^c_Z(\eta'; h_{t'})) = E(J^c_Z(\eta; h_t), J^c_Z^*(\eta'; h_{t'})).
\]

35
We know the distribution of $J^c_Z(\eta; h_t)$ is multivariate complex Gaussian. Except for
Miller (1974), there does not exist a lot of theories about the complex-valued linear
regression. Thus it helps to split Equation (3.4) into its real and imaginary parts.

The real part of $J_X(\eta; h_t)$ is:

$$
\Re(J_X(\eta; h_t)) = \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \Re(H(\eta - \zeta_l)) - \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \Im(H(\eta - \zeta_l))
$$

$$
+ \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \Re(H(\eta + \zeta_l)) + \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \Im(H(\eta + \zeta_l))
$$

$$
+ \Re(J_Z(\eta; h_t))
$$

$$
= \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \{\Re(H(\eta - \zeta_l)) + \Re(H(\eta + \zeta_l))\}
$$

$$
+ \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \{-\Im(H(\eta - \zeta_l)) + \Im(H(\eta + \zeta_l))\} + \Re(J_Z(\eta; h_t)).
$$

Similarly the imaginary part is:

$$
\Im(J_X(\eta; h_t)) = \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \Im(H(\eta - \zeta_l)) + \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \Re(H(\eta - \zeta_l))
$$

$$
+ \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \Im(H(\eta + \zeta_l)) - \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \Re(H(\eta + \zeta_l))
$$

$$
+ \Im(J_Z(\eta; h_t))
$$

$$
= \sum_{l=1}^{L} \frac{A_l \cos \phi_l}{2\sqrt{\Delta}} \{\Im(H(\eta - \zeta_l)) + \Im(H(\eta + \zeta_l))\}
$$

$$
+ \sum_{l=1}^{L} \frac{A_l \sin \phi_l}{2\sqrt{\Delta}} \{-\Re(H(\eta - \zeta_l)) + \Re(H(\eta + \zeta_l))\} + \Im(J_Z(\eta; h_t)).
$$

Let the random variable vector

$$
J_{X,N} \equiv (\Re(J_X(\eta_1; h_t)), \ldots, \Re(J_X(\eta_K; h_t)), \Im(J_X(\eta_1; h_t)), \ldots, \Im(J_X(\eta_K; h_t)))^T
$$

$$
= (\Re J^c_X(\eta; h_t), \Im J^c_X(\eta; h_t))^T,
$$

and

$$
J_{Z,N} \equiv (\Re(J_Z(\eta_1; h_t)), \ldots, \Re(J_Z(\eta_K; h_t)), \Im(J_Z(\eta_1; h_t)), \ldots, \Im(J_Z(\eta_K; h_t)))^T
$$
We can build a linear regression model for $J_{X,N}$ if we know the distribution of $J_{Z,N}$.

The mean of $J_{Z,N} = 0$ because of $J_Z = 0$. The covariance matrix of $J_{Z,N}$ can be calculated by

$$
\Sigma = \text{cov}
\begin{pmatrix}
\Re J^c_Z(\eta; h_t) \\
\Im J^c_Z(\eta; h_t)
\end{pmatrix}
= \frac{1}{2}
\begin{pmatrix}
\Re (\Sigma^c) & -\Im (\Sigma^c) \\
\Im (\Sigma^c) & \Re (\Sigma^c)
\end{pmatrix},
$$

from the results in Chapter 4 of Brillinger (1981). This covariance matrix of $J_{Z,N}$ matches the asymptotic theory presented in Proposition 2.2.1 if $\Re (\Sigma^c)$ is approximately diagonal with diagonal elements, $(S_Z(\eta_1)/2, \ldots, S_Z(\eta_K)/2)^T$, and $\Im (\Sigma^c)$ is approximately zero.

In order to show approximate the asymptotic distribution of $J_{Z,N}$, which is $2K$-variate Gaussian with zero mean and a diagonal covariance matrix with diagonal elements,

$$(S_Z(\eta_1)/2, \ldots, S_Z(\eta_K)/2, S_Z(\eta_1)/2, \ldots, S_Z(\eta_K)/2)^T,$$

we will demonstrate that $\Re (\Sigma^c)$ is approximately diagonal and $\Im (\Sigma^c)$ is approximately zero from simulations. For $\Re (\Sigma^c)$ we will examine the diagonal values as well as the non-diagonal values which should be close to zero. We will assess the non-diagonal values by studying the correlations between frequency of interest, $\eta$, and the neighborhood frequencies of $\eta$, $\eta'$, (e.g., $\eta' = \eta \pm \frac{1}{N}$ and $\eta' = \eta \pm \frac{2}{N}$). This does make sense because we are very interested in local tests which only use the frequencies in the neighborhood of frequency of interest. Examining the correlation between the local frequencies helps us to understand the process around the frequency of interest. For the small dynamic range of AR(2) noise process, the sample size, $N$, is changed to evaluate the distribution of $J_{Z,N}$. For the large dynamic range of AR(4) noise process
we will explore the violation of the theory of asymptotic distribution of $J_{Z,N}$ due to
the leakage. A 20% cosine taper will be used to remedy the asymptotic distribution
of $J_{Z,N}$.

Figure 3.4 gives the plots of variances and correlations between frequency $\eta$ and
$\eta' = \eta \pm \frac{1}{N}$ and $\eta' = \eta \pm \frac{2}{N}$ for the AR(2) process with different sample sizes, as
well as the AR(4) noise process with or without 20% cosine taper. The dashed lines
indicate the theoretical variances for the asymptotic distribution, i.e. $S_Z(\eta)/2$ at the
frequency $\eta$. The first two panels on the top show the variances and correlations
for real part of the Fourier transform of the AR(2) noise process with sample size
$N = 16$ and $N = 256$. The variances are not close to the theoretical values and
correlations are not close to zero for $N = 16$. The correlations between $\eta' = \eta \pm \frac{1}{N}$
and $\eta' = \eta \pm \frac{2}{N}$ depend on the frequency $\eta$ and the SDF of noise process if the
sample size is small. But as the sample size increases to 256, the simulated variances
follow the theoretical values very well over all the frequencies. All the correlations are
close to zero, which indicate the Gaussian random variables are nearly uncorrelated
and hence can be considered to independent in the neighborhood of the frequency
of interest. The two panels on the bottom present the variances and correlations
for the AR(4) noise process with or without a 20% cosine taper. The calculated
variances without the taper are far away from the theoretical values especially at the
higher frequencies because of the leakage (see P. 14 and Figure 2.1 for the definition
of leakage). The random variables are highly dependent because the correlations are
close to 1 at the higher frequencies. When a 20% cosine taper is applied to the AR(4)
process, the calculated and theoretical variances match well. The random variables
Figure 3.4: Plots of variance and correlation for AR(2) with sample size $N = 16$ and $N = 256$, AR(4) with $N = 256$, AR(4) with 20% cosine taper and $N = 256$ respectively. Correlations are calculated between frequency $\eta$ and $\eta' = \eta \pm 1/N$ as well as $\eta$ and $\eta' = \eta \pm 2/N$. 
Figure 3.5: Plots of absolute correlation for AR(2) with sample size $N = 16$ and $N = 256$, AR(4) with $N = 256$, AR(4) with 20% cosine taper and $N = 256$ respectively.

are less correlated than those obtained without tapering, but the independence is still violated around the frequencies corresponding to the peaks in the AR(4) process.

Figure 3.5 gives the boxplots of the correlations for the AR(2) with sample size $N = 16$ and $N = 256$, the AR(4) untapered with $N = 256$, and the AR(4) with a 20% cosine taper and $N = 256$ to assess how close $\Im(\Sigma^c)$ is to being zero. For the AR(2) with sample size $N = 16$ most of correlations are slightly larger than 0 because the sample size is small. But all the correlations are very close to zeros once we increase the sample size to 256. For the AR(4) error process without tapering, the zeros assumption of $\Im(\Sigma^c)$ is violated because of the leakage due to the large dynamic range of the SDF of the AR(4) process. After we use a 20% cosine taper, most correlations are zero.
From these plots we can see the asymptotic distribution of \( J_{Z,N} \), \( 2m \)-variate Gaussian with zero mean and a diagonal covariance matrix, \( \Sigma \), with diagonal elements,

\[
(S_Z(\eta_1)/2, \ldots, S_Z(\eta_K)/2, S_Z(\eta_1)/2, \ldots, S_Z(\eta_K)/2)^T,
\]
is conserved for the noise process with SDFs having a small dynamic range (e.g., the AR(2) process) for large sample sizes, but it is violated for non-IID noise processes if the sample size is small. This is because of the non-stationarity of the tapered Fourier transform process. It is also violated for the process with SDF having a large dynamic range (e.g., AR(4)) even using a large sample size, but a taper can be added to reduce the leakage effect. This asymptotic distribution is extremely violated for the large dynamic range if the sample size is small. Based on this information we know when the asymptotic distribution of \( J_{Z,N} \) is going to fail. Thus, we will derive the regression model for \( J_{X,N} \) in the spectral domain with caution.

### 3.2 Regression Model in the Spectral Domain

Equation (3.4) defines a regression model in the spectral domain. Let the random variable vector of the tapered Fourier transform of time series be

\[
J_X \equiv (\Re(J_X(f_1; h_t)), \ldots, \Re(J_X(f_m; h_t)), \Im(J_X(f_1; h_t)), \ldots, \Im(J_X(f_m; h_t)))^T
\]

and the random variable vector of the tapered Fourier transform of noise process be

\[
J_Z \equiv (\Re(J_Z(f_1; h_t)), \ldots, \Re(J_Z(f_m; h_t)), \Im(J_Z(f_1; h_t)), \ldots, \Im(J_Z(f_m; h_t)))^T,
\]

where \( \{f_k\}, k = 1, \ldots, m \), and \( 1 \leq m \leq \lfloor N/2 \rfloor \), is any set of frequencies. From the previous section we had seen that in certainly situations (i.e., large sample size and small dynamic range) it is reasonable to assume the error term \( J_Z \) is \( 2m \)-variate
normal with zero mean and a diagonal covariance matrix, $\Sigma$, with diagonal elements,

$$(S_Z(f_1)/2, \ldots, S_Z(f_m)/2, S_Z(f_1)/2, \ldots, S_Z(f_m)/2)^T.$$  

We model $J_X$ through a linear regression model.

Though a scaling of $J_X$ we can consider the following regression model in the spectral domain. Let

$$Y_k = \frac{\sqrt{2}}{\sqrt{S_Z(f_k)}} \Re(J_X(f_k; h_t))$$

and

$$Y_{m+k} = \frac{\sqrt{2}}{\sqrt{S_Z(f_k)}} \Im(J_X(f_k; h_t)),$$  \hspace{1cm} (3.10)

for $k = 1, \ldots, m$. With $Y = (Y_1, Y_2, \cdots, Y_{2m-1}, Y_{2m})^T$, regression model is

$$Y = X\beta + \epsilon,$$ \hspace{1cm} (3.11)

where $X$ is a $2m \times 2L$ design matrix depending on the transfer function of the taper and known frequencies, the elements in design matrix $X$ are

$$X_{k,l} = \left\{ \begin{array}{ll}
\frac{1}{\sqrt{2\Delta S_Z(f_k)}} \{ \Re(H(f_k - \zeta_l)) + \Re(H(f_k + \zeta_l)) \} & l = 1, \ldots, L; \\
\frac{1}{\sqrt{2\Delta S_Z(f_k)}} \{ -\Im(H(f_k - \zeta_{L-l})) + \Im(H(f_k + \zeta_{L-l})) \} & l = L+1, \ldots, 2L,
\end{array} \right.$$  \hspace{1cm} (3.12)

$$X_{m+k,l} = \left\{ \begin{array}{ll}
\frac{1}{\sqrt{2\Delta S_Z(f_k)}} \{ \Im(H(f_k - \zeta_l)) + \Im(H(f_k + \zeta_l)) \} & l = 1, \ldots, L; \\
\frac{1}{\sqrt{2\Delta S_Z(f_k)}} \{ \Re(H(f_k - \zeta_{L-l})) - \Re(H(f_k + \zeta_{L-l})) \} & l = L+1, \ldots, 2L,
\end{array} \right.$$  \hspace{1cm} (3.12)

where $k = 1, \ldots, m$ and $L$ is the number of known periodicities; $\beta$ is a $2L$-dimensional vector of unknown parameters defined by

$$\beta_{2L \times 1} = (A_1 \cos \phi_1, A_2 \cos \phi_2, \cdots, A_L \cos \phi_L, A_1 \sin \phi_1, A_2 \sin \phi_2, \cdots, A_L \sin \phi_L)^T,$$ \hspace{1cm} (3.13)
where $A_l$ and $\phi_l$ ($l = 1, \ldots, L$) are the amplitude and phase for the $l$th periodicity, respectively. The $2m$-dimensional random vector of unobserved errors, $\boldsymbol{e}$, is

$$
\boldsymbol{e}_{2m \times 1} = \begin{pmatrix}
\frac{\sqrt{2}}{\sqrt{S_Z(f_1)}} \Re(J_Z(f_1; h_t)) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_m)}} \Re(J_Z(f_m; h_t)) \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_1)}} \Im(J_Z(f_1; h_t)) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_m)}} \Im(J_Z(f_m; h_t))
\end{pmatrix}.
$$

Since $\boldsymbol{e} = \Sigma^{-\frac{1}{2}} \boldsymbol{J}_Z$, From Section 3.1 we can see that $E(\boldsymbol{e}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{e}) \approx \mathbf{I}$. Thus we will assume that $\boldsymbol{e}$ has $N_{2m}(0, \mathbf{I})$ distribution. Using the theory of linear regression models (Ravishanker and Dey (2002) [Chapter 4 and 7]), the ordinary least squares estimator of $\boldsymbol{\beta}$ is denoted by

$$
\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \tag{3.14}
$$

Our goal is to test whether the first amplitude $A_1$ is significantly different from zero, which is equivalent to testing $\mathbf{C}^T \boldsymbol{\beta} = 0$, where

$$
\mathbf{C}_{2L \times 2} = \begin{pmatrix}
1 & [L - 1 \text{ zeroes}] & 0 & [L - 1 \text{ zeroes}] \\
0 & [L - 1 \text{ zeroes}] & 1 & [L - 1 \text{ zeroes}]
\end{pmatrix}^T.
$$

Then the test statistic is

$$
T = \frac{(\mathbf{C}^T \hat{\boldsymbol{\beta}})^T(\mathbf{C}^T(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C})^{-1}(\mathbf{C}^T \hat{\boldsymbol{\beta}})/2}{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})/(2m - 2L)} \sim F(2, 2m - 2L, \eta), \tag{3.15}
$$

where the noncentrality parameter $\lambda = \frac{1}{2}(\mathbf{C}^T \boldsymbol{\beta})^T(\mathbf{C}^T(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C})^{-1}(\mathbf{C}^T \boldsymbol{\beta})$. Under the null hypothesis, $\mathbf{C}^T \boldsymbol{\beta} = 0$, $T$ has a central $F$ distribution.

Using this $F$ test in practice, several different $F$ tests can be derived based on the type of the taper used in Fourier transform of $\{X_t\}$ (e.g., rectangular taper, 100p\%
cosine taper, or DPSS multitaper), the number of known periodicities \( L \) (including periodicity of interest and nuisance periodicities), the number of points of spectral estimate, \( m \), used to carry out the test, and SDF of the error process, \( S_Z(\cdot) \). As we introduced at the beginning of this chapter, the spectral-based tests for periodicity can be divided into two groups, global test and local test, by the number of the points of spectral estimates, \( m \), used to carry out the test. From the Figure 3.3 we have seen a simple example to get a brief idea about the global test and the local test. In the next two sections, we will show in detail that the global and some of the local tests can be derived from the regression model \( F \) test under certain conditions and assumptions.

3.3 Global \( F \) test

A global test evaluates \( H_0: A_1 = 0 \) versus \( H_1: A_1 \neq 0 \), on the basis of an estimate of the SDF across the entire frequency band, i.e., \( m = \lfloor N/2 \rfloor \). For non-IID error process, especially the SDFs of error process with large dynamic range, tapering may be used in a global test to reduce the biases caused by leakage. We will derive the global \( F \) test with or without nuisance periodicities for Gaussian IID and non-IID error process in the following subsections.

3.3.1 Global \( F \) test with one periodicity \((L = 1)\)

The simplest example of the global \( F \) test with one periodicity is the standard \( F \) test for Gaussian IID noise (Percival, 1994, Section 11.2.4). For a sample size \( N \), let \( m = \lfloor N/2 \rfloor \) and \( \{f_{k,N} : k = 1, \ldots, m\} \) be the non-zero, non-Nyquist Fourier frequencies and suppose that the test frequency \( \zeta_1 = f_{k',N} \) for some \( k' \). The test
statistic is

\[
F = \frac{\hat{S}_X^{(p)}(f_{k',N})}{\sum_{k \neq k'} \hat{S}_X^{(p)}(f_{k,N})/(m-1)},
\]

(3.16)

which corresponds to the regression model $F$ test (3.15) assuming that $H(2\zeta_1) = 0$ when we use the periodogram and $\zeta_1$ is a Fourier frequency. Using $h_t = 1/\sqrt{N}$ for all $t$, then $H(f) = \Delta \sqrt{N}$ for $f = 0$ and $H(f) = 0$ otherwise. Evaluating the regression model at the test frequency $\zeta_1 = f_{k',N}$ with $L = 1$ and $m = \lfloor N/2 \rfloor$, we have the parameter vector

\[
\beta_{2 \times 1} = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T,
\]

(3.17)

the design matrix

\[
X_{2m \times 2} = \frac{1}{\sqrt{2\Delta S_Z(f_{k',N})}} \begin{pmatrix}
[k' - 1 \text{ zeros}] & \Delta \sqrt{N} & [m - 1 \text{ zeros}] & 0 & [m - k' \text{ zeros}]
\end{pmatrix}^T
\]

\[
= \frac{\sqrt{\Delta N}}{\sqrt{2\Delta S_Z(f_{k',N})}} \begin{pmatrix}
[k' - 1 \text{ zeros}] & 1 & [m - 1 \text{ zeros}] & 0 & [m - k' \text{ zeros}]
\end{pmatrix}^T,
\]

(3.18)

with

\[
(X^T X)^{-1} = \frac{2S_Z(f_{k',N})}{\Delta N} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.
\]

Then,

\[
\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix} \Re(J_X(f_{k',N})) \\ \Im(J_X(f_{k',N})) \end{pmatrix}.
\]

In order to carry out the test of $C^T \beta = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T = 0$, where $C = I_2$, we get the test statistic by calculating the following terms of the numerator first:

\[
C^T \hat{\beta} = \frac{2}{\sqrt{\Delta N}} I_2 \begin{pmatrix} \Re(J_X(f_{k',N})) \\ \Im(J_X(f_{k',N})) \end{pmatrix}
\]
\[ C^T(X^TX)^{-1}C = \frac{2S_Z(f_{k',N})}{\Delta N} I_2; \]
\[
(C^T \tilde{\beta})^T(C^T(X^TX)^{-1}C)^{-1}(C^T \tilde{\beta}) = \frac{2}{S_Z(f_{k',N})} \left( \Re(J_X(f_{k',N})) \right)^2 + (\Im(J_X(f_{k',N})))^2 \]
\[
= \frac{2}{S_Z(f_{k',N})} \sum_{k \neq k'} S_X(f_{k,N}). \]

Then the denominator is
\[
(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = \frac{2}{S_Z(f_{k',N})} \sum_{k \neq k'} \left( (\Re(J_X(f_{k',N})))^2 + (\Im(J_X(f_{k',N})))^2 \right) \]
\[
= \frac{2}{S_Z(f_{k',N})} \sum_{k \neq k'} S_X(f_{k,N}). \]

since
\[
Y - X\hat{\beta} = \frac{\sqrt{2}}{\sqrt{S_Z(f_{k',N})}} \begin{pmatrix} \Re(J_X(f_{1,N})) \\ \vdots \\ \Re(J_X(f_{k'-1,N})) \\ 0 \\ \Re(J_X(f_{k'+1,N})) \\ \vdots \\ \Re(J_X(f_{m,N})) \\ \Im(J_X(f_{1,N})) \\ \vdots \\ \Im(J_X(f_{k'-1,N})) \\ 0 \\ \Im(J_X(f_{k'+1,N})) \\ \vdots \\ \Im(J_X(f_{m,N})) \end{pmatrix}. \]

Plugging the numerator and denominator into Equation 3.15, we have the F test statistic
\[
F = \frac{(C^T \tilde{\beta})^T(C^T(X^TX)^{-1}C)^{-1}(C^T \tilde{\beta})/2}{(Y - X\hat{\beta})^T(Y - X\beta)/2(m - 1)} \]

46
\begin{align*}
&= \frac{2 \hat{S}_X(f_{k'}, N)}{S_Z(f_{k'}, N)} / 2 \\
&= \frac{2 S_Z(f_{k'}, N)}{\sum_{k \neq k'} S_Z(f_{k}, N) / 2 (m - 1)} \\
&= \frac{\hat{S}_X(f_{k'}, N)}{\sum_{k \neq k'} \hat{S}_X(f_{k}, N) / (m - 1)}.
\end{align*}

This test statistic is the exact global $F$ test statistic (Percival, 1994, Section 11.2.4) for testing one periodicity when \( \{Z_t\} \) is a Gaussian IID process. The $F$ statistic has a central $F$ distribution on 2 and $2(m - 1)$ degrees of freedom under the null hypothesis.

For non-IID Gaussian processes, if we still consider the $F$ test in Equation (3.16), the null $F$ distribution does not apply because the SDF is no longer constant. Let \( \{V_k : k = 1, \ldots, m\} \) be a set of independent $\chi^2_2$ random variables (RVs). We approximate the distribution of the test statistic by a function of weighted chi squared RVs:

\[
\frac{S_Z(f_{k'}, N)V_{k'}}{\sum_{k \neq k'} S_Z(f_{k}, N)V_k / (m - 1)}.
\]

We then use the Satterthwaite approximation to approximate the distribution of $\sum_{k \neq k'} S_Z(f_{k}, N)V_k$ with a chi squared RV with $b$ degrees of freedom multiplied by the constant $a$, where $a$ and $b$ are both unknown. Equating moments (mean and variance), estimates are given by

\[
\hat{a} = \frac{\sum_{k \neq k'} S_Z^2(f_{k}, N)}{\sum_{k \neq k'} S_Z(f_{k}, N)} \quad \text{and} \quad \hat{b} = \frac{2 \left( \sum_{k \neq k'} S_Z(f_{k}, N) \right)^2}{\sum_{k \neq k'} S_Z^2(f_{k}, N)}.
\] (3.19)

It then follows that an approximate distribution for the statistic is an $F$ distribution on 2 and $\hat{b}$ degrees of freedom multiplied by the term $2(m - 1)S_Z(f_{k'}, N)/(\hat{a} \hat{b})$.

To approximate the null distribution for the global test in practice we need an estimate of the underlying SDF of the error process. Since the periodogram or other direct spectral estimates are not consistent estimates, we propose instead to smooth a direct spectral estimator. We employ a smoothing spline approach Wahba (1980)
based on the following regression model for the estimated log SDF:

$$\log \hat{S}^{(d)}(f_{k,N}) + \gamma = \log S_Z(f_{k,N}) + \epsilon_k, \quad k \neq k'.$$

Here the errors \(\{\epsilon_k\}\) are approximately distributed log chi squared (see page 228 of Percival and Walden (1993)) 2 degrees of freedom minus \(\log(2) - \gamma\), and \(\gamma \approx 0.57721\) is the Euler-Mascheroni constant. For each \(k \neq k'\) (i.e., at all Fourier frequencies excluding the test frequency), \(E(\epsilon_k) = 0\) and \(\text{var}(\epsilon_k) = \pi^2/6\). We fit the model using cubic smoothing splines and based on the model predict \(\log(\hat{S}^{(d)}(f_{k',N}))\) which is used in the calculation of the null distribution.

Another way to get the global \(F\) test with Gaussian non-IID processes is to derive the \(F\) test from the regression model. Assuming again that \(H(2\zeta_1) \approx 0\), one approximate distribution for the tapered \(F\) statistic comes from assuming that the errors in the regression model \(J_Z(f_{k,N}; h_t)\), evaluated at the Fourier frequencies are uncorrelated (which we have already shown is an approximation for any fixed \(N\)). Evaluating the regression model at the test frequency \(\zeta_1 = f_{k',N}\), the parameter and the design matrix are the same as Equation (3.17) and (3.18). The estimator of \(\beta\) is

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{2}{\sqrt{\Delta N}} \left( \begin{array}{c} \Re(J_X(f_{k',N})) \\ \Im(J_X(f_{k',N})) \end{array} \right).$$

Letting \(C = I_2\) in \(C^T \beta = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T = 0\), we get the numerator of the test statistic

$$\left( (C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} C^T \hat{\beta} \right) = \frac{2\hat{S}_X(f_{k',N})}{S_Z(f_{k',N})}.$$
and the denominator of the test statistic from

$$ Y - X \hat{\beta} = \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{Z(f_{1,N})}} \Re(J_X(f_{1,N})) \\ \vdots \\ \frac{\sqrt{2}}{\sqrt{Z(f_{k - 1,N})}} \Re(J_X(f_{k - 1,N})) \\ 0 \\ \frac{\sqrt{2}}{\sqrt{Z(f_{k,N})}} \Re(J_X(f_{k,N})) \\ \vdots \\ \frac{\sqrt{2}}{\sqrt{Z(f_{m,N})}} \Re(J_X(f_{m,N})) \\ \frac{\sqrt{2}}{\sqrt{Z(f_{1,N})}} \Im(J_X(f_{1,N})) \\ \vdots \\ \frac{\sqrt{2}}{\sqrt{Z(f_{k - 1,N})}} \Im(J_X(f_{k - 1,N})) \\ 0 \\ \frac{\sqrt{2}}{\sqrt{Z(f_{k,N})}} \Im(J_X(f_{k,N})) \\ \vdots \\ \frac{\sqrt{2}}{\sqrt{Z(f_{m,N})}} \Im(J_X(f_{m,N})) \end{pmatrix} , $$

giving

$$(Y - X \hat{\beta})^T (Y - X \hat{\beta}) = 2 \sum_{k \neq k'}^m \frac{1}{S_Z(f_{k,N})} \{ (\Re(J_X(f_{1,N})))^2 + (\Im(J_X(f_{1,N})))^2 \}$$

$$= 2 \sum_{k \neq k'}^m \frac{1}{S_Z(f_{k,N})} \hat{S}_X(f_{k,N}).$$

Then the $F$ test statistic is

$$ F = \frac{(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta})/2}{(Y - X \hat{\beta})^T (Y - X \hat{\beta})/2(m - 1)}$$

$$= \frac{\hat{S}_X(f_{k',N})}{\sum_{k \neq k'}^m \frac{\hat{S}_X(f_{k,N})}{S_Z(f_{k,N})}} / (m - 1).$$

This $F$ test statistic contains the SDF of error process, $S_Z(\cdot)$, which is not constant for non-IID noise and usually unknown. But we can use the smoothing spline method to estimate them again. For non-IID Global $F$ test, it should be noticed that the
assumption of the stationarity of tapered Fourier transform process is violated because the noise process is not IID and the test uses the whole points. Then the performance of the global test is affected, as we will demonstrate in Chapter 4.

3.3.2 Global $F$ Test with More Than One Periodicities ($L > 1$)

For time series with more than one periodicity, a global test still evaluates $H_0$: $A_1 = 0$ versus $H_1$: $A_1 \neq 0$. For a sample size $N$, let $m = \lfloor N/2 \rfloor$ and $\{f_{k,N} : k = 1, \ldots, m\}$ be the non-zero, non-Nyquist Fourier frequencies. The test frequency of interest is $\zeta_1 = f_{k_1,N}$ and the nuisance frequencies of the periodicities are $\{\zeta_l = f_{k_l,N} : l = 2, \ldots, L\}$, where $1 < k_1, k_2, \ldots, k_L < m$. For a Gaussian IID error process $\{Z_t\}$ with constant SDF, we use rectangular taper, i.e., $h(t) = 1/\sqrt{N}$ for all the $t$. Then $H(0) = \Delta \sqrt{N}$ and $H(\cdot) = 0$ for others. Evaluating the regression model at the test frequency $\zeta_1 = f_{k_1,N}$, we have the parameter

$$
\beta_{2L \times 1} = (A_1 \cos \phi_1, A_2 \cos \phi_2, \ldots, A_L \cos \phi_L, A_1 \sin \phi_1, A_2 \sin \phi_2, \ldots, A_L \sin \phi_L)^T;
$$

and the design matrix

$$
X_{2m \times 2L} = \frac{1}{\sqrt{2\Delta S_Z(f_{k_1,N})}} \times
$$
Then,

\[ (X^T X)^{-1} = \frac{2S_Z(f_{k_1,N})}{\Delta N} I_{2L}, \]

where \( S_Z(f_{1,N}) \) represents the constant SDF of Gaussian IID error process \( \{Z_t\} \). Now the parameter estimator is

\[ \hat{\beta} = (X^T X)^{-1} X^T Y \]
Using

$$C_{2L \times 2} = \begin{pmatrix} 1 & [L-1 \text{ zeroes}] & 0 & [L-1 \text{ zeroes}] \\ 0 & [L-1 \text{ zeroes}] & 1 & [L-1 \text{ zeroes}] \end{pmatrix}^T.$$ 

to test $C^T \beta = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T = 0$, we get the test statistic by calculating the following terms

$$C^T \hat{\beta} = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix} \Re(J_X(f_{k_1,N})) \\ \Im(J_X(f_{k_1,N})) \end{pmatrix};$$

$$C^T (X^T X)^{-1} C = \frac{2S_Z(f_{k_1,N})}{\Delta N} I_2;$$

$$(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta})$$

$$= \frac{2}{S_Z(f_{k_1,N})} \left( \begin{pmatrix} \Re(J_X(f_{k_1,N})) \\ \Im(J_X(f_{k_1,N})) \end{pmatrix}^T I_2 \begin{pmatrix} \Re(J_X(f_{k_1,N})) \\ \Im(J_X(f_{k_1,N})) \end{pmatrix} \right)$$

$$= \frac{2}{S_Z(f_{k_1,N})} \left\{ (\Re(J_X(f_{k_1,N})))^2 + (\Im(J_X(f_{k_1,N})))^2 \right\}$$

$$= \frac{2 \hat{S}_X(f_{k_1,N})}{S_Z(f_{k_1,N})},$$

52
Then the F test statistic is

\[
\hat{Y} - \hat{X} = \frac{\sqrt{2}}{\sqrt{S_Z(f_{k,1})}} \begin{pmatrix}
\Re(J_X(f_{1,N})) \\
\vdots \\
\Re(J_X(f_{k-1,N})) \\
0 \\
\Re(J_X(f_{k+1,N})) \\
\vdots \\
\Re(J_X(f_{2N-1,N})) \\
0 \\
\Re(J_X(f_{2N+1,N})) \\
\vdots \\
\Re(J_X(f_{m,N})) \\
\Im(J_X(f_{1,N})) \\
\Im(J_X(f_{k-1,N})) \\
0 \\
\Im(J_X(f_{k+1,N})) \\
\vdots \\
\Im(J_X(f_{2N-1,N})) \\
0 \\
\Im(J_X(f_{2N+1,N})) \\
\vdots \\
\Im(J_X(f_{m,N}))
\end{pmatrix};
\]

\[
(Y - \hat{X})^T(Y - \hat{X}) = \frac{2}{S_Z(f_{k,1})} \sum_{k \neq \{k_1, k_2, \ldots, k_L\}} \{ (\Re(J_X(f_{k,N})))^2 + (\Im(J_X(f_{k,N})))^2 \}
\]

\[
= \frac{2}{S_Z(f_{k,1})} \sum_{k \neq \{k_1, k_2, \ldots, k_L\}} \hat{S}_X(f_{k,N}).
\]

Then the F test statistic is

\[
F = \frac{(C^T\hat{\beta})^T(C^T(X^TX)^{-1}C)\hat{\beta}}{(Y - \hat{X})^T(Y - \hat{X})/(2(m - L))}
\]

53
When \( \{Z_t\} \) is a Gaussian IID process in the case of L periodicities \((L > 1)\), the test statistic has a \( F \) distribution on 2 and \( 2(m - L) \) degrees of freedom under the null hypothesis.

For Gaussian non-IID processes, the global \( F \) test can be derived by the same steps as before at the test frequency \( \zeta_1 = f_{k1,N} \). The parameter vector is still

\[
\beta_{2L \times 1} = (A_1 \cos \phi_1, A_2 \cos \phi_2, \cdots, A_L \cos \phi_L, A_1 \sin \phi_1, A_2 \sin \phi_2, \cdots, A_L \sin \phi_L)^T,
\]

but the design matrix is changed to

\[
X_{2m \times 2L} = \frac{\sqrt{\Delta N}}{\sqrt{2}} \times
\]

\[
\begin{pmatrix}
0_{k_1-1} & 0_{k_1-1} & \cdots & 0_{k_1-1} & 0_{k_1-1} & 0_{k_1-1} & \cdots & 0_{k_1-1} \\
0_{k_2-k_1-1} & 0_{k_2-k_1-1} & \cdots & 0_{k_2-k_1-1} & 0_{k_2-k_1-1} & 0_{k_2-k_1-1} & \cdots & 0_{k_2-k_1-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & \cdots & 0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & \cdots & 0_{k_L-k_L-1-1} \\
0_{m-k_L+k_1-1} & 0_{m-k_L+k_1-1} & \cdots & 0_{m-k_L+k_1-1} & 0_{m-k_L+k_1-1} & 0_{m-k_L+k_1-1} & \cdots & 0_{m-k_L+k_1-1} \\
0_{k_2-k_1-1} & 0_{k_2-k_1-1} & \cdots & 0_{k_2-k_1-1} & 0_{k_2-k_1-1} & 0_{k_2-k_1-1} & \cdots & 0_{k_2-k_1-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & \cdots & 0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & 0_{k_L-k_L-1-1} & \cdots & 0_{k_L-k_L-1-1} \\
0_{m-k_L} & 0_{m-k_L} & \cdots & 0_{m-k_L} & 0_{m-k_L} & 0_{m-k_L} & \cdots & 0_{m-k_L}
\end{pmatrix}
\]

because of the nonconstant SDF. Then

\[
(X^T X)_{2L \times 2L}^{-1} = \frac{2}{\Delta N} \times
\]
\[
\begin{pmatrix}
S_Z(f_{k_1,N}) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & S_Z(f_{k_2,N}) & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & S_Z(f_{k_L,N}) & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & S_Z(f_{k_1,N}) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & S_Z(f_{k_L,N}) \\
\end{pmatrix}
\]

The parameter estimator is calculated by

\[
\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix}
\Re(J_X(f_{k_1,N})) \\
\Re(J_X(f_{k_2,N})) \\
\vdots \\
\Re(J_X(f_{k_L,N})) \\
\Im(J_X(f_{k_1,N})) \\
\Im(J_X(f_{k_2,N})) \\
\vdots \\
\Im(J_X(f_{k_L,N})) \\
\end{pmatrix}
\]

Letting

\[
C_{2L\times2} = \begin{pmatrix}1 & \left[L - 1 \text{ zeroes}\right] & 0 & \left[L - 1 \text{ zeroes}\right] \\
0 & \left[L - 1 \text{ zeroes}\right] & 1 & \left[L - 1 \text{ zeroes}\right] \end{pmatrix}^T
\]

as before, calculate the numerator and denominator of test statistic respectively,

\[
C^T \hat{\beta} = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix}
\Re(J_X(f_{k_1,N})) \\
\Im(J_X(f_{k_1,N})) \\
\end{pmatrix};
\]

\[
C^T (X^T X)^{-1} C = \frac{2S_Z(f_{k_1,N})}{\Delta N} I_2;
\]

\[
(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta}) = \frac{2\hat{S}_X(f_{k_1,N})}{S_Z(f_{k_1,N})};
\]

55
\[
Y - X\hat{\beta} = \begin{pmatrix}
\frac{\sqrt{2}}{\sqrt{S_Z(f_{1,1})}} \Re(J_X(f_{1,1})) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_1-1,1})}} \Re(J_X(f_{k_1-1,N})) \\
0 \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_1+1,1})}} \Re(J_X(f_{k_1+1,N})) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_2-1,1})}} \Re(J_X(f_{k_2-1,N})) \\
0 \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_2+1,1})}} \Re(J_X(f_{k_2+1,N})) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_L-1,1})}} \Re(J_X(f_{k_L-1,N})) \\
0 \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{k_L+1,1})}} \Re(J_X(f_{k_L+1,N})) \\
\vdots \\
\frac{\sqrt{2}}{\sqrt{S_Z(f_{m,1})}} \Re(J_X(f_{m,N}))
\end{pmatrix};
\]

\[
(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = 2 \sum_{k \neq \{k_1, k_2, \ldots, k_L\}} \frac{S_X(f_{k,N})}{S_Z(f_{k,N})}.
\]
Then

\[
F = \frac{(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta})}{2} \frac{2 S_X(f_{k_1,N})}{S_Z(f_{k_1,N})^2} / 2(m - L) \\
= \frac{2 \sum_{k \neq \{k_1, k_2, ..., k_L\}} \frac{S_X(f_{k,N})}{S_Z(f_{k,N})}}{(m - L)}.
\]

Under the null hypothesis, the \( F \) statistic has a \( F \) distribution on 2 and \( 2(m - L) \) degrees of freedom.

Using this test statistic with nuisance frequencies in practice, again the SDF of the non-IID error process, \( S_Z(\cdot) \), is unknown. We need to use the smoothing spline approach to estimate \( S_Z(\cdot) \). Therefore, all the periodogram ordinates at the nuisance frequencies should be excluded from the denominator of the test statistic (This will affect the null distribution). This causes a problem if we do not know the location of the nuisance frequencies. We will consider the effect of nuisance frequencies in the simulation of Chapter 4. As we seen the global test is very sensitive to the nuisance frequency which gives us motivation to study the local tests.

### 3.4 Local Tests

Local tests of a significant amplitude \( A_1 \) at a frequency \( \zeta_1 \), are based on spectral estimates in a neighborhood of the test frequency. In this section we discuss a local test from the hearing science literature, the local \( F \) test, and Thomson’s multitaper \( F \) test. The local \( F \) test and Thomson’s multitaper \( F \) test can be derived from the regression \( F \) test derived in Section 3.2 by making a number of extra assumptions.
3.4.1 Test from the Hearing Science

The standard test for periodicity used in hearing sciences is calculated using the periodogram Gorga et al. (1993). Assuming that $\zeta_1$ is a Fourier frequency, the test statistic is

$$Q_B(\zeta_1) = \frac{\hat{S}^{(p)}(\zeta_1)}{\max\{L_B(\zeta_1), R_B(\zeta_1)\}},$$

(3.21)

where

$$L_B(\zeta_1) = \frac{1}{B} \sum_{k=1}^{B} \hat{S}^{(p)}(\zeta_1 - \frac{k+1}{N\Delta})$$

and

$$R_B(\zeta_1) = \frac{1}{B} \sum_{k=1}^{B} \hat{S}^{(p)}(\zeta_1 + \frac{k+1}{N\Delta})$$

are respectively, skipping one Fourier frequency on either side, the average of $B$ Fourier frequencies to the left (i.e., lower in frequency) of the frequency of interest and the average of $B$ Fourier frequencies to the right (i.e., higher in frequency). Gorga et al. (1993) reject $H_0$ if $Q_B$ is larger than $10^{(\lambda/10)}$, for some $\lambda$ which has units on the decibel scale. Common values of $\lambda$ include 3, 6, and 9. This test statistic does not relate directly to an F test statistic obtained from the regression model (3.4). Craigmile and King (2004) demonstrated for a fixed value of $N$ and $\Delta$ that the size and power of this test depended on $\lambda$, the number of bins to average, $B$, and the background SDF, $S_Z(\cdot)$. This standard test used in hearing sciences will be compared with the other local tests in Chapter 4.

3.4.2 Local F Test

Almasri (2003) defines the local F test, which is the similar to the Gorga et al. hearing test, except the denominator of the statistic is the average of both the lower and higher neighborhoods of the test frequency, instead of maximum. Craigmile and
King (2004) showed via simulation studies that the local $F$ statistic,

$$F_B(\zeta_1) = \frac{\hat{S}^{(p)}(\zeta_1)}{(L_B(\zeta_1) + R_B(\zeta_1))/2},$$

produces a more powerful test. We shall investigate this in much greater detail, both theoretically and using simulations.

The local $F$ test statistic corresponds to the regression $F$ test for testing that $A_1 = 0$ under model (3.4) when:

1. We restrict the regression model to the frequencies local to $\zeta_1$ given by

$$\left\{ \zeta_1 - \frac{B+1}{N\Delta}, \ldots, \zeta_1 - \frac{2}{N\Delta}, \zeta_1, \zeta_1 + \frac{2}{N\Delta}, \ldots, \zeta_1 + \frac{B+1}{N\Delta} \right\}.$$ 

We assume $\zeta_1$ is a Fourier frequency and that on these frequencies $S_{Z}(\cdot)$ is constant. We skip a frequency on each side of the test frequency because the remaining frequencies in the neighborhood of the test frequency are less likely to be affected by the SDF at $\zeta_1$, as we have seen in Figure 3.3 at the beginning of this chapter.

2. We assume that $H(0) \approx 1$ and $H(2\zeta_1) \approx 0$, and that both $H(\zeta_1 + k/(N\Delta) - \zeta_l) \approx 0$ and $H(\zeta_1 + k/(N\Delta) + \zeta_l) \approx 0$ for each $k = 0, \pm 1, \ldots, \pm (B + 1)$. For large $N$ and decent separation of the frequencies this will be a reasonable assumption because the sidelobes of the taper are close to zero. (The choice of taper can help in this regard – see Chapter 4).

Evaluating the regression model under above conditions at the test frequency $\zeta_1$, we have the parameter vector

$$\beta_{2L \times 1} = (A_1 \cos \phi_1, A_2 \cos \phi_2, \ldots, A_L \cos \phi_L, A_1 \sin \phi_1, A_2 \sin \phi_2, \ldots, A_L \sin \phi_L)^T.$$
and the design matrix

\[
X_{2(2B+1) \times 2L} = \frac{\sqrt{\Delta N}}{\sqrt{2S_Z(\zeta_1)}} \begin{pmatrix}
[B \text{ zeros}] & 1 & [2B \text{ zeros}] & 0 & [B \text{ zeros}] \\
[B \text{ zeros}] & 0 & [2B \text{ zeros}] & 0 & [B \text{ zeros}] \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
[B \text{ zeros}] & 0 & [2B \text{ zeros}] & 0 & [B \text{ zeros}] \\
[B \text{ zeros}] & 0 & [2B \text{ zeros}] & 1 & [B \text{ zeros}] \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
[B \text{ zeros}] & 0 & [2B \text{ zeros}] & 0 & [B \text{ zeros}]
\end{pmatrix}^T,
\]

where \(S_Z(\zeta_1)\) represents the constant SDF of the noise process in the neighborhood \([\zeta_1 - \frac{B+1}{N\Delta}, \zeta_1 + \frac{B+1}{N\Delta}]\). Then the generalized inverse of \(X^TX\) is,

\[
(X^TX)_{2L \times 2L}^{-1} = \frac{2S_Z(\zeta_1)}{\Delta N} \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix},
\]

and the ordinary least squares estimate of \(\beta\) is

\[
\hat{\beta} = (X^TX)^{-1}X^TY = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix}
\Re(J_X(\zeta_1)) \\
0 \\
\vdots \\
0 \\
\Im(J_X(\zeta_1)) \\
0 \\
\vdots \\
0
\end{pmatrix}.
\]

In order to test \(C^T\beta = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T = 0\), let

\[
C_{2L \times 2} = \begin{pmatrix}
1 & [L - 1 \text{ zeroes}] & 0 & [L - 1 \text{ zeroes}] \\
0 & [L - 1 \text{ zeroes}] & 1 & [L - 1 \text{ zeroes}]
\end{pmatrix}^T.
\]

\(C^T\beta\) is testable though \(\hat{\beta}\) is not estimable due to the generalized inverse. We still get the test statistic by calculating the numerator and denominator respectively:

\[
C^T\hat{\beta} = \frac{2}{\sqrt{\Delta N}} \begin{pmatrix}
\Re(J_X(\zeta_1)) \\
\Im(J_X(\zeta_1))
\end{pmatrix};
\]

\[
C^T(X^TX)^{-1}C = \frac{2S_Z(\zeta_1)}{\Delta N} \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}.
\]
\[
(C^T \widehat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \widehat{\beta}) = \frac{2 \widehat{S}_X(\zeta_1)}{S_Z(\zeta_1)};
\]

\[
Y - X\widehat{\beta} = \frac{\sqrt{2}}{\sqrt{S_Z(\zeta_1)}} \left( \begin{array}{c}
\Re(J_X(\zeta_1 - \frac{B+1}{N\Delta}); h_i) \\
\vdots \\
\Re(J_X(\zeta_1 - \frac{2}{N\Delta}); h_i) \\
0 \\
\Re(J_X(\zeta_1 + \frac{2}{N\Delta}); h_i) \\
\vdots \\
\Re(J_X(\zeta_1 + \frac{2}{N\Delta}); h_i) \\
\Re(J_X(\zeta_1 + \frac{B+1}{N\Delta}); h_i) \\
\end{array} \right);
\]

\[
(Y - X\widehat{\beta})^T (Y - X\widehat{\beta}) = \frac{2}{S_Z(\zeta_1)} \left( \sum_{k=1}^{B} \widehat{S}_X(\zeta_1 - \frac{k+1}{\Delta N}) + \sum_{k=1}^{B} \widehat{S}_X(\zeta_1 + \frac{k+1}{\Delta N}) \right).
\]

Then the \( F \) statistic becomes

\[
F = \frac{(C^T \widehat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \widehat{\beta}) / 2}{(Y - X\widehat{\beta})^T (Y - X\widehat{\beta}) / 2(2B)}
\]

\[
= \frac{\widehat{S}_X(\zeta_1)}{ \left( \sum_{k=1}^{B} \widehat{S}_X(\zeta_1 - \frac{k+1}{\Delta N}) + \sum_{k=1}^{B} \widehat{S}_X(\zeta_1 + \frac{k+1}{\Delta N}) \right) / 2B}
\]

\[
= \frac{\widehat{S}_X(\zeta_1)}{ (L_B(\zeta_1) + R_B(\zeta_1)) / 2}.
\]

This test statistic is the \textit{local F test statistic}. Under the null hypothesis \( H_0 : A_1 = 0 \), the local \( F \) statistic approximately has an \( F \) distribution with 2 and \( 4B \) degrees of freedom. Under the alternative hypothesis when \( H_1 : A_1 = A^* \), the test statistic has a noncentral F distribution on 2 and \( 4B \) degrees of freedom with noncentrality parameter

\[
\lambda = \frac{1}{2} (C^T \beta)^T (C^T (X^T X)^{-1} C)^{-1} (C^T \beta)
\]
\[
\Delta N \frac{1}{4S_Z(\xi_1)} \begin{pmatrix} A_1 \cos \phi_1 & A_1 \sin \phi_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \cos \phi_1 \\ A_1 \sin \phi_1 \end{pmatrix} = \Delta N A_1^2 \frac{1}{4S_Z(\xi_1)},
\]

since \( \cos^2 \phi_1 + \sin^2 \phi_1 = 1 \). This second distribution allows us to make theoretical power calculations in Chapter 4.

When there is significant leakage in the spectral estimate (e.g., large dynamic range in the SDF of error process \( \{Z_t\} \)), a taper can be applied to the local \( F \) test.

We need only assume that \( H(\xi_1 + k/(N\Delta) + \xi_l) \approx 0 \) for each \( k = 0, \pm 1, \ldots, \pm (B + 1) \). Evaluating the regression model again at the test frequency \( \xi_1 \), we can use the submodel to evaluate the nonzero part of the matrix, which corresponds the parameter

\[
\beta_{2 \times 1} = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T
\]

and the design matrix

\[
X_{2(2B+1) \times 2} = \frac{1}{\sqrt{2\Delta S_Z(\xi_1)}} \begin{pmatrix}
\Re H(-\frac{B+1}{N\Delta}) & -\Im H(-\frac{B+1}{N\Delta}) \\
\vdots & \vdots \\
\Re H(-\frac{2}{N\Delta}) & -\Im H(-\frac{2}{N\Delta}) \\
\Re H(0) & 0 \\
\Re H(\frac{2}{N\Delta}) & -\Im H(\frac{2}{N\Delta}) \\
\vdots & \vdots \\
\Re H(-\frac{B+1}{N\Delta}) & -\Im H(-\frac{B+1}{N\Delta}) \\
\Im H(-\frac{B+1}{N\Delta}) & \Re H(-\frac{B+1}{N\Delta}) \\
\vdots & \vdots \\
\Im H(-\frac{2}{N\Delta}) & \Re H(-\frac{2}{N\Delta}) \\
0 & \Re H(0) \\
\Im H(\frac{2}{N\Delta}) & \Re H(\frac{2}{N\Delta}) \\
\vdots & \vdots \\
\Im H(\frac{B+1}{N\Delta}) & \Re H(\frac{B+1}{N\Delta})
\end{pmatrix},
\]
where we have $\Re H(-\frac{i+1}{N\Delta}) = \Re H(\frac{i+1}{N\Delta})$ and $\Im H(-\frac{i+1}{N\Delta}) = -\Im H(\frac{i+1}{N\Delta})$ for $i = 1, \ldots, B$.

Then,

$$X^T X = \frac{1}{2\Delta S_Z(\zeta_1)} \times \begin{pmatrix} \Re H(0)^2 + 2 \sum_{i=1}^{B} ((\Re H(\frac{i+1}{N\Delta}))^2 + (\Im H(\frac{i+1}{N\Delta}))^2) & 0 \\ 0 & \Re H(0)^2 + 2 \sum_{i=1}^{B} ((\Re H(\frac{i+1}{N\Delta}))^2 + (\Im H(\frac{i+1}{N\Delta}))^2) \end{pmatrix}$$

$$= \frac{\Re H(0)^2 + 2 \sum_{i=1}^{B} ((\Re H(\frac{i+1}{N\Delta}))^2 + (\Im H(\frac{i+1}{N\Delta}))^2)}{2\Delta S_Z(\zeta_1)} I_2,$$

so that

$$(X^T X)^{-1} = \frac{2\Delta S_Z(\zeta_1)}{\Re H(0)^2 + 2 \sum_{i=1}^{B} ((\Re H(\frac{i+1}{N\Delta}))^2 + (\Im H(\frac{i+1}{N\Delta}))^2)} I_2.$$

Now let $P_1 = (\Re H(0))^2 + 2 \sum_{i=1}^{B} ((\Re H(\frac{i+1}{N\Delta}))^2 + (\Im H(\frac{i+1}{N\Delta}))^2)$. The response vector is

$$Y_{2(2B+1)\times 1} = \begin{pmatrix} \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Re J_X(\zeta_1 - \frac{B+1}{N\Delta}; h_t) \\ \vdots \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Re J_X(\zeta_1 - \frac{1}{N\Delta}; h_t) \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Re J_X(\zeta_1; h_t) \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Re J_X(\zeta_1 + \frac{1}{N\Delta}; h_t) \\ \vdots \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Re J_X(\zeta_1 + \frac{B+1}{N\Delta}; h_t) \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Im J_X(\zeta_1 - \frac{B+1}{N\Delta}; h_t) \\ \vdots \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Im J_X(\zeta_1 - \frac{1}{N\Delta}; h_t) \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Im J_X(\zeta_1; h_t) \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Im J_X(\zeta_1 + \frac{1}{N\Delta}; h_t) \\ \vdots \\ \frac{\sqrt{\pi}}{\sqrt{S_Z(\zeta_1)}} \Im J_X(\zeta_1 + \frac{B+1}{N\Delta}; h_t) \end{pmatrix}.$$
which means that

\[
X^T Y = \frac{1}{\sqrt{\Delta S_Z(\zeta_1)}} \times \left( \begin{array}{c}
\Re H(0) \Re J_X(\zeta_1; h_t) + \\
\sum_{i=1}^{B} (\Re H(\frac{i+1}{N\Delta}) \Re J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) + \Re H(\frac{i+1}{N\Delta}) \Re J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t)) \\
\sum_{i=1}^{B} (\Im H(\frac{i+1}{N\Delta}) \Im J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) + \Im H(\frac{i+1}{N\Delta}) \Im J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t)) \\
\Re H(0) \Im J_X(\zeta_1; h_t) + \\
\sum_{i=1}^{B} (-\Im H(\frac{i+1}{N\Delta}) \Re J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) - \Im H(\frac{i+1}{N\Delta}) \Re J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t)) + \\
\sum_{i=1}^{B} (\Re H(\frac{i+1}{N\Delta}) \Im J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) + \Re H(\frac{i+1}{N\Delta}) \Im J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t))
\end{array} \right)
\]

We use \( P_2 \) and \( P_3 \) to label the first and second part in the vector of \( X^T Y \) respectively.

Now we can estimate \( \beta \) by

\[
\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{2\sqrt{\Delta}}{P_1} \left( \begin{array}{c} P_2 \\ P_3 \end{array} \right).
\]

Letting \( C_{2 \times 2} = I_2 \) again we want to test \( H_0: C^T \beta = (A_1 \cos \phi_1, A_1 \sin \phi_1)^T = 0 \). We calculate the numerator and denominator of the test statistic respectively:

\[
C^T \hat{\beta} = \frac{2\sqrt{\Delta}}{P_1} \left( \begin{array}{c} P_2 \\ P_3 \end{array} \right);
\]

\[
C^T (X^T X)^{-1} C = \frac{2\Delta S_Z(\zeta_1)}{P_1} \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right);
\]

\[
(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta}) = \frac{2}{S_Z(\zeta_1) \cdot P_1} (P_2^2 + P_3^2);
\]

\[
X \hat{\beta} = \frac{\sqrt{2}}{\sqrt{S_Z(\zeta_1) \cdot P_1}} \left( \begin{array}{c}
\Re H(\frac{-B+1}{N\Delta}) \cdot P_2 - \Im H(\frac{-B+1}{N\Delta}) \cdot P_3 \\
\Re H(\frac{-2}{N\Delta}) \cdot P_2 - \Im H(\frac{-2}{N\Delta}) \cdot P_3 \\
\Re H(\frac{2}{N\Delta}) \cdot P_2 - \Im H(\frac{2}{N\Delta}) \cdot P_3 \\
\Re H(\frac{B+1}{N\Delta}) \cdot P_2 - \Im H(\frac{B+1}{N\Delta}) \cdot P_3 \\
\Im H(\frac{-B+1}{N\Delta}) \cdot P_2 + \Re H(\frac{-B+1}{N\Delta}) \cdot P_3 \\
\Im H(\frac{-2}{N\Delta}) \cdot P_2 + \Re H(\frac{-2}{N\Delta}) \cdot P_3 \\
\Im H(\frac{2}{N\Delta}) \cdot P_2 + \Re H(\frac{2}{N\Delta}) \cdot P_3 \\
\Im H(\frac{B+1}{N\Delta}) \cdot P_2 + \Re H(\frac{B+1}{N\Delta}) \cdot P_3
\end{array} \right);
\]

64
and our test statistic for the tapered case is

\[
Y - X\hat{\beta} = \sqrt{2} \times \frac{1}{\sqrt{S_Z(\zeta_1)}} \times \left( \Re J_X(\zeta_1 - \frac{B+1}{N\Delta}; h_t) - (\Re H(-\frac{B+1}{N\Delta}) \cdot P_2 - \Im H(-\frac{B+1}{N\Delta}) \cdot P_3) / P_1 \right)
\]

\[
\vdots
\]

\[
\Im J_X(\zeta_1 - \frac{2}{N\Delta}; h_t) - (\Im H(-\frac{2}{N\Delta}) \cdot P_2 - \Re H(-\frac{2}{N\Delta}) \cdot P_3) / P_1
\]

\[
\Re J_X(\zeta_1 + \frac{2}{N\Delta}; h_t) - (\Re H(\frac{2}{N\Delta}) \cdot P_2 - \Im H(\frac{2}{N\Delta}) \cdot P_3) / P_1
\]

\[
\Im J_X(\zeta_1 - \frac{B+1}{N\Delta}; h_t) - (\Im H(-\frac{B+1}{N\Delta}) \cdot P_2 + \Re H(-\frac{B+1}{N\Delta}) \cdot P_3) / P_1
\]

\[
\Im J_X(\zeta_1 + \frac{B+1}{N\Delta}; h_t) - (\Im H(\frac{B+1}{N\Delta}) \cdot P_2 + \Re H(\frac{B+1}{N\Delta}) \cdot P_3) / P_1
\]

\[
(\Im J_X(\zeta_1 - \frac{2}{N\Delta}; h_t) - (\Im H(-\frac{2}{N\Delta}) \cdot P_2 + \Re H(-\frac{2}{N\Delta}) \cdot P_3) / P_1
\]

\[
\Im J_X(\zeta_1 + \frac{2}{N\Delta}; h_t) - (\Im H(\frac{2}{N\Delta}) \cdot P_2 + \Re H(\frac{2}{N\Delta}) \cdot P_3) / P_1
\]

\[
\Im J_X(\zeta_1 - \frac{B+1}{N\Delta}; h_t) - (\Im H(-\frac{B+1}{N\Delta}) \cdot P_2 + \Re H(-\frac{B+1}{N\Delta}) \cdot P_3) / P_1
\]

In the above expression

\[
P_4 = (\Re J_X(\zeta_1; h_t) - \Re H(0) \cdot P_2 / P_1)^2 + (\Im J_X(\zeta_1; h_t) - \Re H(0) \cdot P_3 / P_1)^2
\]

\[
+ \sum_{i=1}^{B} \left( \Re J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) - \left( \Re H(\frac{i+1}{N\Delta}) \cdot P_2 - \Im H(\frac{i+1}{N\Delta}) \cdot P_3 \right) / P_1 \right)^2
\]

\[
+ \sum_{i=1}^{B} \left( \Im J_X(\zeta_1 + \frac{i+1}{N\Delta}; h_t) - \left( \Im H(\frac{i+1}{N\Delta}) \cdot P_2 + \Re H(\frac{i+1}{N\Delta}) \cdot P_3 \right) / P_1 \right)^2
\]

\[
+ \sum_{i=1}^{B} \left( \Re J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t) - \left( \Re H(-\frac{i+1}{N\Delta}) \cdot P_2 - \Im H(-\frac{i+1}{N\Delta}) \cdot P_3 \right) / P_1 \right)^2
\]

\[
+ \sum_{i=1}^{B} \left( \Im J_X(\zeta_1 - \frac{i+1}{N\Delta}; h_t) - \left( \Im H(-\frac{i+1}{N\Delta}) \cdot P_2 + \Re H(-\frac{i+1}{N\Delta}) \cdot P_3 \right) / P_1 \right)^2,
\]

and our test statistic for the tapered case is

\[
F = \frac{(C^T\hat{\beta})^T(C^T(X^T X)^{-1}C)^{-1}(C^T\hat{\beta})/2}{(Y - X\hat{\beta})^T(Y - X\hat{\beta})/AB}
\]
\[
\text{Again, under the null hypothesis of } H_0 : A_1 = 0, \text{ the local } F \text{ statistic approximately has an } F \text{ distribution with } 2 \text{ and } 4B \text{ degrees of freedom. This local } F \text{ test statistic is very complicated even for one periodicity } (L = 1). \text{ For more than one periodicity, it is not very difficult to calculate it using statistical software (e.g., using R). We only need to specify the design matrix } X \text{ and response vector } Y. \text{ Then the numerical value of } F \text{ test statistic can be obtained after some matrix calculations.}
\]

### 3.4.3 Thomson’s F Test

The final local spectral test we consider is Thomson’s F test, based on the multiple tapers defined in Section 2.2. The multiple tapers test reduces the variance of spectral estimates by using a small set of tapers rather than the unique data taper. A set of independent estimates of the power spectrum is computed, by pre-multiplying the data by orthogonal tapers which are constructed to minimize the spectral leakage due to the finite length of the data set. Percival and Walden (1993)[Chapter 7] demonstrated the Thomson’s multitaper F test related to the integral equation approach.

The Thomson’s F statistic is given by

\[
F_T = \frac{|\hat{C}_1|^2 \sum_k H_k^2(0)/2}{\Delta \sum_k |J_X(\zeta_1; h_{k,t})| - \hat{C}_1 H_k(0)/\sqrt{\Delta}^2/(2K - 2)}.
\]

In this section, we will derive the Thomson’s F test from our regression model (3.4).

For each of the \( k = 1, \ldots, K \) orthonormal multiple tapers, \( \{h_{k,t}\} \), our regression model (3.4) at \( f = \zeta_1 \) becomes

\[
J_X(\zeta_1; h_{k,t}) = \frac{1}{\sqrt{\Delta}} \sum_{l=1}^L C_l H_k(0) + \frac{1}{\sqrt{\Delta}} \sum_{l=1}^L C_l^* H_k(\zeta_1 + \zeta_l) + J_Z(\zeta_1; h_{k,t}). (3.23)
\]
Here $H_k(\cdot)$ is the transfer function of the taper $\{h_{k,t}\}$ for each $k$. Fitting a regression model based on the data, $\{J_X(\zeta_1;h_{k,t}) : k = 1,\ldots,K\}$, we obtain an estimate of $C_1$, $\hat{C}_1$ say, using complex least squares, which corresponds to the regression model $F$ test assuming that

1. The SDF of the noise process, $S_Z(\cdot)$, is locally constant in the region $[\zeta_1 - W,\zeta_1 + W]$, where $W$ is the bandwidth defining the multiple tapers.

2. The frequencies $\{\zeta_l : l \neq 1\}$ are at least $2W$ units away from $\zeta_1$ (the nuisance frequencies are well separated from the test frequencies).

Let the test frequency be $\zeta_1$ for $\{J_X(\zeta_1;h_{k,t}) : k = 1,\ldots,K\}$. The parameter vector is

$$\beta_{2L \times 1} = (A_1 \cos \phi_1, A_1 \sin \phi_1, A_2 \cos \phi_2, A_2 \sin \phi_2, \ldots, A_L \cos \phi_L, A_L \sin \phi_L)^T,$$

the design matrix is

$$X_{2K \times 2L} = (X_T^1, X_T^2, \cdots, X_T^K)^T,$$

where each $X_k, k = 1,\ldots,K$ is a $2 \times 2L$ matrix given by

$$X_k = \frac{1}{\sqrt{2\Delta S_Z(\zeta_1)}} \begin{pmatrix} H_k(0) & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & H_k(0) & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Then

$$X^T X_{2L \times 2L} = \left( \sum_{k=1}^K X_k^T X_k \right)^- = \frac{2\Delta S_Z(\zeta_1)}{\sum_{k=1}^K H_k^2(0)} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}^-, \quad \frac{2\Delta S_Z(\zeta_1)}{\sum_{k=1}^K H_k^2(0)} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$
For the vector of observed responses, we have
\[ Y = (Y_1^T, Y_2^T, \cdots, Y_K^T)^T, \]
and each \( Y_k, k = 1, \cdots, K \) has the form
\[ Y_k = \frac{\sqrt{2}}{\sqrt{S_Z(\zeta_1)}} \left( \begin{array}{c} \Re(J_X(\zeta_1; h_{k,t})) \\ \Im(J_X(\zeta_1; h_{k,t})) \end{array} \right). \]

The estimate of \( \beta \) is
\[
\hat{\beta} = (X^T X)^{-1} X^T Y \\
= (\sum_{k=1}^{K} X_k^T X_k)^{-1} (\sum_{k=1}^{K} X_k^T Y_k) \\
= \frac{2\sqrt{\Delta}}{\sum_{k=1}^{K} H_k^2(0)} \left( \begin{array}{cc} \sum_{k=1}^{K} \Re(J_X(\zeta_1; h_{k,t})H_k(0)) \\ \sum_{k=1}^{K} \Im(J_X(\zeta_1; h_{k,t})H_k(0)) \\ 0 \\ \vdots \\ 0 \end{array} \right).
\]

We know \( C_1 = \frac{A_1 e^{i\phi_1}}{2} = \frac{A_1}{2}(\cos \phi + i \sin \phi) \). Let \( C^T = (\frac{1}{2}, \frac{i}{2}, 0, \cdots, 0)^T \), then \( C_1 = C^T \beta \) is estimable though \( \hat{\beta} \) is not estimable due to the generalized inverse. We have
\[
\hat{C}_1 = \frac{\sqrt{\Delta} \sum_{k=1}^{K} J_X(\zeta_1; h_{k,t})H_k(0)}{\sum_{k=1}^{K} H_k^2(0)}.
\]

In order to test \( C_1 = 0 \), we again calculate the numerator and denominator respectively:
\[
C^T \hat{\beta} = \frac{\sqrt{\Delta} \sum_{k=1}^{K} J_X(\zeta_1; h_{k,t})H_k(0)}{\sum_{k=1}^{K} H_k^2(0)} = \hat{C}_1, \\
C^T (X^T X)^{-1} C = \frac{\Delta S_Z(\zeta_1)}{\sum_{k=1}^{K} H_k^2(0)}; \\
(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta}) = |\hat{C}_1|^2 \frac{\sum_{k=1}^{K} H_k^2(0)}{\Delta S_Z(\zeta_1)}; \\
68
\[ Y - X \hat{\beta} = \left( (Y_1 - X_1 \hat{\beta})^T, (Y_2 - X_2 \hat{\beta})^T, \ldots, (Y_K - X_K \hat{\beta})^T \right)^T; \]

\[ (Y - X \hat{\beta})^T (Y - X \hat{\beta}) = \sum_{k=1}^{K} (Y_K - X_K \hat{\beta})^T (Y_K - X_K \hat{\beta}). \]

In the last expression we have

\[ Y_K - X_K \hat{\beta} = \frac{\sqrt{2}}{\sqrt{S_Z(\zeta_1)}} \left( \Re(J_X(\zeta_1; h_{k,t})) - \Re(\hat{C}_1)H_k(0)/\sqrt{\Delta} \right), \]

so that

\[ (Y - X \hat{\beta})^T (Y - X \hat{\beta}) = \frac{2}{S_Z(\zeta_1)} \sum_{k=1}^{K} \left| J_X(\zeta_1; h_{k,t}) - \hat{C}_1 H_k(0)/\sqrt{\Delta} \right|^2. \]

Then Thomson’s F statistic is

\[ F_T = \frac{(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta})}{(Y - X \hat{\beta})^T (Y - X \hat{\beta})/2(K - 1)} \]

\[ = \frac{2 \sum_{k=1}^{K} |C_1|^2 \sum_{l=1}^{K} H_l^2(0)}{\Delta S_Z(\zeta_1)} \left( \sum_{k=1}^{K} \left| J_X(\zeta_1; h_{k,t}) - \hat{C}_1 H_k(0)/\sqrt{\Delta} \right|^2 /2(K - 1) \right) \]

\[ = \frac{(K - 1) |C_1|^2 \sum_{k} H_k^2(0)}{\Delta \sum_k |J_X(\zeta_1; h_{k,t}) - \hat{C}_1 H_k(0)/\sqrt{\Delta}|^2}. \]

Under the null hypothesis of \( H_0: A_1 = 0, \) Thomson’s F statistic approximately has a \( F_{2,2(K-1)} \) distribution. Denison et al. (1999) adjusted this testing procedure when assumption 2 does not apply (i.e., there is no separation of the periodicities). They demonstrated how to formulate the multitaper test via complex-valued weighted least square for frequencies \( \{ \zeta_l : l \neq 1 \} \) that are within \( 2W \) units from \( \zeta_1, \) given the location of the nuisance frequencies are known.

The global and local spectral-based tests have now been introduced in detail in this chapter. A global test uses all the points of spectral estimates to carry out the spectral-based test, though some assumptions in the theory are violated. A local test statistic is less efficient than a global test because only part of the spectral estimates
are used. The question now is which test is better, the global one or the local one? In order to answer this question, in the next chapter we will carry out Monte Carlo experiments to investigate the size and power properties of global and local tests under a number different experimental conditions.
CHAPTER 4

SIZE AND POWER COMPARISONS

We introduced and derived the time-domain-based test and the spectral-based tests (i.e., the global $F$ test, the Gorga et al. test (the test from the hearing science), the local $F$ test and the Thomson’s $F$ test) in Chapter 3. In order to investigate the performance of the tests, we will compare the size and the power among the spectral-based tests as well as between the time-domain-based test and the spectral-based tests in this chapter. Monte Carlo experiments will be carried out to evaluate the size and the power properties of all the tests under a number different experimental conditions.

To simplify the experimental design we assumed that the sampling interval was $\Delta = 1$ and the phases of the periodicity, $\{ \phi_t \}$ were zero in the following model of the time series $\{X_t : t \in \mathbb{Z}\}$,

$$X_t = \sum_{l=1}^{L} A_l \cos(2\pi \zeta t \Delta + \phi_t) + Z_t.$$ 

In our simulations we used three different error processes:

1. IID noise: $Z_t = e_t$, with $\sigma^2 = 1$,

2. AR(2) process: $Z_t = 0.75Z_{t-1} - 0.5Z_{t-2} + e_t$ with $\sigma^2 = 0.56245$, and
3. AR(4) process: \[ Z_t = 2.7607Z_{t-1} - 3.8106Z_{t-2} + 2.6635Z_{t-3} - 0.9238Z_{t-4} + \epsilon_t \]

with \( \sigma^2 = 0.0013128 \),

where \( \{\epsilon_t\} \) is a set of IID \( N(0,\sigma^2) \) RVs for all three processes. The value of the innovation variance \( \sigma^2 \) guaranteed that \( \text{var}(Z_t) = 1 \) for each process. We choose these two AR processes as the test cases because they have interesting spectral patterns and have been used extensively in the literature. The SDF of these error processes are shown in Figure 4.1. The AR(4) process has the largest dynamic range, followed by the AR(2) process and then the IID process. In our simulations we want to test if the periodicity at frequency \( \zeta_1 \) is significantly different from zero when (i) there are no nuisance periodicities (\( L = 1 \)); and (ii) a nuisance periodicity at \( \zeta_2 \) is observed along with the test periodicity at \( \zeta_1 \) (\( L = 2 \)). We varied the amplitude of interest (\( A_1 \)), sample sizes (\( N \)) and the bin width of neighborhoods used in the local tests.
(B). Then we carried out the Monte Carlo experiment to compare the size and the power of tests between the $F$ test in the time domain and the local $F$ test in the frequency domain, as well as among the spectral-based tests: the global $F$ test, the local $F$ test, the Gorga et al. test, and the Thomson’s multitaper $F$ test.

4.1 The Size of the Test

Let $T$ be the test statistic of interest. The size of the test is

$$P(\text{reject } H_0|H_0 \text{ true}) = P(T > c|A_1 = 0),$$

where $c$ is the critical value so that the test has theoretical size 0.05.

We started from the IID error process and the AR(2) error process with a SDF with a small dynamic range. Firstly, we estimated the size of each test (the global $F$ test, the local $F$ test, and the Gorga et al. test) using 1,000 replicates of simulated realizations of $\{X_t\}$ of length $N = 512$ for the IID Gaussian and AR(2) error process. $B = 4$ is chosen for the local $F$ and the Gorga et al tests because Craigmile and King (2004) have shown that the power performance for both the local $F$ test and the Gorga et al test is not substantial for five bins ($B = 5$) compared to four bins ($B = 4$), but the power performance is better for four bins ($B = 4$) compared to three bins ($B = 3$). Figure 4.2 shows the size of each test as a function of $\zeta_1$. The dotted horizontal lines denote 95% pointwise confidence intervals assuming a test of theoretical size 0.05 (based on a standard binomial calculation). For the IID Gaussian process, regardless of the test used, the size values lie mostly within the confidence limits. For the AR(2) process, the size values for the global test are much higher than the nominal value of 0.05 between 0.1 Hz and 0.2 Hz. This is because the global $F$ test is affected by leakage in the spectral estimates caused by the peaks in the SDF.
Figure 4.2: Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test, the Gorga et al. test and the global test for IID and AR(2) error processes with one periodicity ($L = 1$). In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$. 

Both the local $F$ test and the Gorga et al. test preserve the nominal size for the AR(2) error process.

Secondly, we investigate the size of each test in the presence of a known or unknown nuisance periodicity. We assumed $L = 2$, and fixed the test frequency $\zeta_1 = 144/512$. We varied the nuisance frequency $\zeta_2$, but fixed its amplitude, $A_2 = 1$ (zero on the decibels scale). Figure 4.3 shows the estimated sizes for different values of $\zeta_2$ when, (i) for the purposes of the global $F$ test, we know the location of $\zeta_2$ (dark line) and remove
it before estimating the underlying SDF using a smoothing spline, and (ii) when the location of $\zeta_2$ is unknown (gray line), leading us to leave in the periodogram ordinate that corresponds to $\zeta_2$ in the smoothing spline model. For the local tests, the size of the test is reduced when $\zeta_2$ lies within $B + 1$ Fourier frequencies of the test frequency $\zeta_1$. We can thus see it is important to assume $\zeta_2$ is the nuisance frequency not in a neighborhood of test frequency $\zeta_1$. The size of the global $F$ test is reduced when the location of $\zeta_2$ is unknown. There is weak evidence that the size values of the global test are reduced with the IID process because most values are inside 95% confidence interval, but the size values of the global $F$ test are reduced with the AR(2) process, even adjusting for $\zeta_2$. From the size of test for the IID and AR(2) error processes, the global test is sensitive to the nuisance periodicity and the dynamic range of the error process. As long as the nuisance frequency, $\zeta_2$, is outside the local region, by design the local tests are not sensitive to the nuisance periodicity.

As we introduced in Section 2.2, leakage (bias) in the periodogram is large for the error process with a SDF having a large dynamic range (e.g. AR(4)). The global test is highly affected by the nuisance frequency and leakage in the spectral estimates even for the AR(2) process. Thus, we only explore the size of the local tests (the local $F$ test and the Thomson’s test) for the AR(4) error process. From Figure 2.1 in Chapter 2, we can see tapering reduces the leakage for the error process that have a SDF with a large dynamic range. Thus, tapering is considered to reduce the effect of the leakage for the AR(4) error process in this simulation. In order to guarantee that most values of the size preserve the nominal size of 0.05, a 5% cosine taper is used for the local $F$ test. We can skip a few of frequencies on each side of the test frequency $\zeta_1$ because the remaining frequencies in the neighborhood of the test frequency are less
Figure 4.4: Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test without taper, the local $F$ tests with 5% cosine taper and the Thomson’s test for AR(4) error processes. In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$. The Thomson’s test uses 7 DPSS tapers with $NW = 7$.

likely to be affected by the SDF at $\zeta_1$. The number of frequencies skipped on each side is called *offset*. We let offset = 1 and $N = 512$ in this experiment. Now we will compare the size values for the local $F$ test without a taper, the local $F$ test with a 5% cosine taper, and Thomson’s $F$ test. Figure 4.4 shows the size of the local tests for the AR(4) error process. The left panel of Figure 4.4 displays the size values of local $F$ test without a taper. The size values are highly affected by the large dynamic range of the error process as we expected. This is because the asymptotic distribution of tapered Fourier transform (the local $F$ test is derived from the regression model based on this asymptotic distribution) is violated for large dynamic range error process as we demonstrated in Section 3.1. The middle panel of Figure 4.4 provides the size values of local $F$ test with a 5% (i.e. $p = 0.05$) cosine taper. The dotted horizontal lines denote 95% pointwise confidence intervals assuming a test of theoretical size.
0.05 (based on a standard binomial calculation). This test preserves the nominal size at most test frequencies except the frequencies around the two peaks of the SDF of the AR(4) error process. The size values at the frequencies around the two peaks of the AR(4) error process are either larger than the upper bound or smaller than the lower bound of 95% confidence interval. But it has the same double-peak shape as the SDF of the AR(4) error process because of the leakage effect. The right panel of Figure 4.4 gives the size values of Thomson’s $F$ test with $N = 7$ DPSS tapers and $NW = 7$, where $W$ is the window length. The number of tapers $N$ and the window length $W$ are chosen to match the same effective bandwidth as the direct spectra estimator used in the local $F$ test (see Walden et al. (1995) for equations that can be used to calculate the effective bandwidth for a multitaper spectral estimator).

The size values of Thomson’s $F$ test lie mostly within the confidence limits for the frequencies not around the peaks. The leakage effect around the peaks is reduced a little comparing with the local $F$ test using the cosine taper.

The local $F$ test with a cosine taper is sensitive to the parameter $p$ in the cosine taper. The leakage effect cannot been reduced if a non-appropriate parameter $p$ is picked. The size of the local $F$ test with $100p\%$ cosine taper also depends on the sample size $N$, the type of the error process (dynamic range), and the test frequency. Figure 4.5 provides the estimated size based on a simulation of 1,000 time series with sample size $N = 64, 128, 256, \text{and} 512$ at test frequency $\zeta_1 = 144/512$ using the local $F$ test with a $100p\%$ cosine taper for various values of $p$. In Figure 4.5 the black solid lines give the estimated size of the local $F$ test for AR(4) error processes. For different sample sizes the nominal size of 0.05 is achieved in the local $F$ test with different values of $p$ (e.g., $p = 0.05$ for $N = 512$ and $p = 0.15$ for $N = 256$). The nominal
Figure 4.5: Plots of the estimated size as a function of parameter $p$ in the local $F$ test with the 100$p\%$ cosine taper based on a simulation of 1,000 time series with sample size $N = 64, 128, 256, \text{ and } 512$. The local $F$ tests have test frequency $\zeta_1 = 144/512$ and bin size $B = 4$ for the AR(4), the AR(2) and the IID error processes respectively.
size cannot be achieved for the local $F$ test with small sample sizes (e.g., $N = 64$ and $N = 128$) even using tapering. This is because the asymptotic distribution of the tapered Fourier transform is highly violated for small sample sizes and large dynamic range. Tapering can only reduce the leakage effect if the sample size is not too small.

Tapering is not always helpful especially for the error processes with SDFs having small dynamic range (e.g., IID and AR(2)). The dashed and dot-dashed lines in Figure 4.5 represent the estimated size of the local $F$ test for AR(2) and IID process respectively. Many estimated size values of the local $F$ test with a cosine taper are outside the 95\% confidence interval, especially for IID process. This does agree with what we discussed in Section 3.1. We know the untapered Fourier transform of IID process follows the exact multivariate Gaussian with a diagonal covariance matrix. Then the local $F$ test, exactly derived from the spectral regression model without any approximation, should achieve the nominal size. If we use tapering, correlation in the spectral errors will be present. The tapered Fourier transform of IID process has non-zero values (correlations) for the non-diagonal components of the covariance matrix. This makes the size of the local $F$ test significantly different from the nominal size.

Tapering causes similar problems with achieving the nominal size for error processes with SDFs having small dynamic range. The Fourier transform of the error processes with SDFs having small dynamic range already has a multivariate Gaussian with an approximately diagonal covariance matrix. Tapering might add the correlation to the approximately diagonal covariance matrix. Thus it is reasonable not to use tapering for the error processes with SDFs having small dynamic range.

We also consider the choice of offset (the number of frequencies skipped on each side of the test frequency) used for the local $F$ test. The larger the offset is, the less
Figure 4.6: Plots of the estimated size as a function of the test frequency, $\zeta_1$, of the local $F$ test with 1 offset, 2 offset, 3 offset, and 4 offset based on a simulation of 1,000 time series with sample size $N = 512$. The local $F$ tests use bin size $B = 4$.

likely the remaining frequencies in the neighborhood of the test frequency are to be affected by the estimated SDF at $\zeta_1$. But the locally constant assumption may be weaker as the offset increases since we are making the neighborhood large. Figure 4.6 gives the size plots for the local $F$ test with a 5% cosine taper using different offsets. The local $F$ test uses $B = 4$ and 1,000 replicates of simulated realizations of $\{X_t\}$ of length $N = 512$ for the AR(4) error process. We can see the estimated sizes
are very similar except the values at the frequencies around the two peaks. The size
values around the peaks for the local $F$ test with offset = 1 are smaller than for other
offset values. Combining with the locally constant assumption for the local $F$ test,
offset = 1 is chosen to carry out the rest of simulations.

4.2 The Power of the Test

The power of the test at amplitude $A_1 = a$ for the test frequency $\zeta_1$ is defined by

$$P(\text{reject } H_0| H_0 \text{ false}) = P(T > c|A_1 = a).$$

To summarize the power over a range of amplitudes for the test periodicity of 0 to $A$,
the area under the power curve (AUC) was calculated:

$$\text{AUC} = \int_0^A P(\text{reject } H_0|A_1 = a) \, da,$$

and was approximated with

$$\text{AUC} = \sum_{k=0}^{D-1} w_k \hat{P}(\text{reject } H_0|A_1 = v_k).$$

In the above equation, $\hat{P}$ denotes the estimated power value from the simulations,
and $\{w_k : k = 0, \ldots, D - 1\}$, $\{v_k : k = 0, \ldots, D - 1\}$ respectively denote the weights
and abscissa of a Gaussian quadrature rule. A value of $D = 20$ was chosen to balance
the accuracy of the quadrature rule and the computing time. The AUC is used
to evaluate how well the tests performed. The larger the AUC value is, the better
performance the test has (in terms of aggregate power). The AUC is especially useful
for discriminating between tests if the power curves of the tests seem to be fairly close
together.
In order to compare one particular test of interest (e.g., local $F$ test) with another test, the relative AUC is defined by

$$\text{Relative AUC} = \frac{\text{AUC} - \text{AUC of local } F\text{ test}}{\text{AUC of local } F\text{ test}}.$$

This relative AUC measures the difference of the AUC between one test and the local $F$ test relative to the AUC of the local $F$ test. A value less than zero means the test has worse performance than the local $F$ test (the smaller is the worse), whereas a value greater than zero implies the test has better performance than the local $F$ test (the larger is the better).

Firstly, we examine the power of the $F$ test in the time domain and the local $F$ test in frequency domain. As we introduced at the beginning of Chapter 3, the noise process needs to be estimated for the $F$ test in the time domain. In our simulation, we selected the candidate noise process among different $AR(p)$ processes,
\( p = 0, 1, 2, \ldots, P, \) using the Akaike Information Criterion (AIC). We used \( P = 4 \) for AR(2) process and \( P = 7 \) for AR(4) process by examining the diagnostic plot of ACVS. This estimation procedure for the unknown noise process slows down the computation of the power test a lot. If the noise process is very complicated, we found that it was difficult to determine the range of \( p \) for the AR(\( p \)) process to get an accurate estimated noise process. Figure 4.7 shows the comparison of the \( F \) test in time domain and the local \( F \) test in frequency domain for the AR(2) noise process with sample size \( N = 64 \) and \( N = 512 \) as well as the AR(4) noise process with sample size \( N = 512 \). The maximum pointwise Monte Carlo standard error for each estimate is 0.005 (based on a standard binomial calculation). For the AR(2) noise process with small sample size \( N = 64 \), the test in frequency domain has larger power than the nominal size 0.05 when the amplitude of periodicity is quite small. This indicates the size of the test in time domain is not correct. As the sample size increased, the size is close to nominal size (still slightly larger than 0.05). Though the power of the time-domain-based \( F \) test is also slightly larger than the local \( F \) test, the computational speed of the time domain \( F \) test is much slower because a large dimension of matrix calculation is needed. For the AR(4) noise process the power of the time-domain-based \( F \) test is slightly larger than the local \( F \) test for the existing power values, but it is noticed that there are missing values when the amplitude is larger than \(-26dB\). The missing values are due to the unaccomplished noise process estimation because of nonstationary AR part when the amplitude is large. Thus, the time-domain-based \( F \) test is time consuming and success of the testing procedure can not be guaranteed.
Secondly, we compared the power among the spectral-based tests for different error processes with different sample sizes. The bin width used in the local tests (the local $F$ test and the Gorga et al test) is also explored. As we introduced in Chapter 3, the global $F$ test uses all of the spectral estimates, whereas the local tests only use the spectral estimates in the neighborhood of the test frequency. Then the local tests are less efficient the global test. But they are robust to nuisance periodicities and error processes with SDFs having large dynamic range.

Figure 4.8 shows the estimated power curve based on 1,000 realizations of the IID process (left panel) and AR(2) process (right panel), with $N = 512$. We set $\zeta_1$ at the Fourier frequency $144/512\text{Hz}$, with a nuisance frequency $\zeta_2 = 160/512\text{Hz}$ at an amplitude of $A_2 = 0$ decibels. The amplitude of $\zeta_1$ was varied between $-20$ and 0 decibels (roughly zero and 10 on the original scale). A bin size of $B = 4$ was again used. The maximum pointwise Monte Carlo standard error for each estimate is 0.016 (based on a standard binomial calculation). The power values of global $F$ test are slightly larger than the local tests. This is because the global test uses all the spectral estimates across the entire frequency band for IID and AR(2) error processes. It cannot tell the difference between the local $F$ test and the Gorga et al test from these plots because the power curves are fairly close together. In order to get the better power comparison among the spectral-based tests, we will use AUC to evaluate the power performance.

Figure 4.9 displays the estimated AUC for the global and the local tests using the AR(2) and the AR(4) error processes as a function of sample size. For both error processes, the power of all the tests improves as the sample size increases. The performance of the local tests are closer to that of the global test at large sample sizes.
Figure 4.8: Plots of the estimated power of the local $F$ test, the Gorga et al. test and the global test for IID and AR(2) error processes with two periodicities ($L = 2$, with $\zeta_1 = 144/512$ and $\zeta_2 = 160/512$). In each case, the estimate at each frequency is based on a simulation of 1,000 time series with sample size $N = 512$. The local tests use bin size $B = 4$.

because the approximation made for the spectral regression model, which the local $F$ test is derived from, makes more sense for large sample size. The left plot on the upper panel of Figure 4.9 displays the estimated AUC for the global and local tests using the AR(2) error process. The local $F$ test has more power than the Gorga et al. test. The Thomson’s multitaper test has lower power than the local test without any taper for this AR(2) error process. We conclude that the taper cannot improve the performance in this case. The right plot on the upper panel of Figure 4.9 shows the estimated AUC for the global and local tests using the AR(4) error process as a function of the sample size. Thomson’s multitaper test has much higher power especially for a small sample size of the AR(4) error process. Comparing with the
Figure 4.9: A comparison of the area under the power curve (AUC) and relative AUC of the local $F$ test, the Gorga et al test, Thomson’s $F$ test and the global test for different values of sample size $N$ with an AR(2) and an AR(4) error processes.
AR(2) error process, it is worth using the multitaper test for the AR(4) error process. This highlights that the multitaper test has application to testing periodicity for error processes with SDFs having large dynamic range. From the bottom panel of Figure 4.9, it is easier to compare the global test, the Gorga et al. test, and the Thomson’s multitaper test with the local $F$ test by the relative AUC. For the AR(2) error process the global test has more power than the local tests especially when the sample size is small, but the local $F$ test is as good as the global test if the sample size is large enough. The local $F$ test is better than the Gorga et al. test and the Thomson’s multitaper test. Thus it is not worth tapering for the AR(2) error process with a SDF having small dynamic range. For the AR(4) error process, the Thomson’s multitaper test is much better than other tests especially when the sample size is small. Then, local $F$ test is slightly better than the Gorga et al. test. The global test has smaller power than other tests for the AR(4) error process because the global test is sensitive to the dynamic range of the SDF of the error process.

The left panel of Figure 4.10 displays the estimated AUC at a fixed sample size $N = 512$ as a function of bin width $B$ ($B = 1, 2, \ldots, 5$) used for the AR(2) error process. As expected, the AUC (and hence power) of the local tests increases as the bin width $B$ increases. The local $F$ has more power than the Gorga et al. test. The AUC of the global test dominates the local tests, although the difference is smaller for larger bin sizes, $B$. The right panel of Figure 4.10 gives the relative AUC of local $F$ test comparing to the Gorga et al. test and the global test. We can see the local $F$ is better than the Gorga et al. test. The global test is better than the local tests. The AUC of the local $F$ test is closer to that of the global test as the bin width, $B$, increases. This is what we expected because more information (spectral estimates)
Figure 4.10: A comparison of the area under the power curve (AUC) of the local $F$ test, the Gorga et al test and the global test for different number of frequencies used in a neighborhood of test frequency $\zeta_1$ with an AR(2) error process.

are used to carry out the local $F$ test, though the number of neighborhood frequencies (i.e., $2B + 1$) used by the local $F$ test is still much less than all of the frequencies used by the global test.

From Figure 4.9 and Figure 4.10 we conclude that similar performance (in terms of AUC) can be achieved by the local tests with a reasonable choice of sample size and bin width, using much less information than the global $F$ test.

Finally we compared the power of the local $F$ test used to detect the observed data $Y$, which comes from a multivariate Gaussian distribution with the true covariance matrix $\Sigma$. This matrix, $\Sigma$, is no longer the identity covariance matrix (i.e., $\Sigma \neq I$). We evaluated the local $F$ test under three situations: (i) using the local $F$ test statistic derived from the spectral regression model (3.11) with $\varepsilon \sim (0, I)$ and comparing with
the $F$ critical value from the $F$ distribution; (ii) still using the local $F$ test statistic derived from the regression model, but comparing with the empirical critical value calculated from the repetitive simulations; (iii) Using the true covariance matrix $\Sigma$ to derive the new local $F$ test statistic and comparing with the critical value from the $F$ distribution. If the performance of the local $F$ test under the situation (i) is as powerful as that of the other two situations, especially the situation (iii), we know the local $F$ test works well even when some assumptions are violated.

We generated the errors $\varepsilon$ in the spectral regression model (3.11) from the multivariate Gaussian distribution with the true covariance matrix, $\Sigma$. Supposing that we know the ACVS of the noise process $\{Z_t\}$, $\{s_Z\}$, the true covariance matrix $\Sigma$ can be calculated from Equation (3.5) and (3.9). From Equation (3.14) and (3.15), we have

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

and

$$T = \frac{(C^T \hat{\beta})^T (C^T (X^T X)^{-1} C)^{-1} (C^T \hat{\beta}) / 2}{(Y - X \hat{\beta})^T (Y - X \hat{\beta}) / 4B}.$$ 

Plugging $\hat{\beta}$ into the above equation, we have

$$T = \frac{(C^T (X^T X)^{-1} X^T Y)^T (C^T (X^T X)^{-1} C)^{-1} (C^T (X^T X)^{-1} X^T Y) / 2}{(Y - X (X^T X)^{-1} X^T Y)^T (Y - X (X^T X)^{-1} X^T Y) / 4B} = \frac{Y^T X (X^T X)^{-1} C (X^T X)^{-1} C^T (X^T X)^{-1} X^T Y / 2}{Y^T (I - H)^T (I - H) Y / 4B},$$

where $H = X (X^T X)^{-1} X^T$ and $(I - H)$ is idempotent. Let $M_{\text{den}} = (I - H)$ and $M_{\text{num}} = X (X^T X)^{-1} C (X^T X)^{-1} C^T (X^T X)^{-1} X^T$, the equation is changed to

$$T = \frac{Y^T M_{\text{num}} Y / 2}{Y^T M_{\text{den}} Y / 4B},$$

(4.1)
This test statistic does not follow the $F$ distribution very well if the true covariance matrix is not close to a diagonal one. The power of the local $F$ test at amplitude $A_1 = a$ is calculated by

$$
P(\text{reject } H_0 | H_0 \text{ false})$$

$$= P(T > c | A_1 = a)$$

$$= P\left( \frac{Y^T M_{num} Y / 2}{Y^T M_{den} Y / 4B} > c | A_1 = a \right)$$

$$= P\left( Y^T M_{num} Y / 2 > c(Y^T M_{den} Y / 4B) | A_1 = a \right)$$

$$= P \left( Y^T \left[ \frac{M_{num}}{2} - c \frac{M_{den}}{4B} \right] Y > 0 | A_1 = a \right), \quad (4.2)$$

where $Y \sim N(X\beta, \Sigma)$ and the critical value $c$ is either from the $F$ distribution or an empirical (Monte Carlo) method.

Supposing we know the true covariance matrix $\Sigma$, and using weighted least squares (Ravishanker and Dey (2002) [Chapter 4]), we can develop a general local $F$ test statistic with this true covariance matrix $\Sigma$:

$$T = \frac{(C^T \hat{\beta})^T (C^T (X^T \Sigma^{-1} X)^{-1} C)^{-1} (C^T \hat{\beta}) / 2}{(Y - X\hat{\beta})^T \Sigma^{-1} (Y - X\hat{\beta}) / 4B}, \quad (4.3)$$

where $\hat{\beta}$, the estimator of $\beta$, is now given by the solution to the general least squares problem:

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y.$$ 

Under the null hypothesis $H_0 : C^T \beta = 0$, $T$ has a central $F_{2,4B}$ distribution. This test statistic has no practical value because the true covariance matrix is usually unknown. But we can use this general local $F$ test to assess the performance of the local $F$ test by comparing the power under this case to other approximate cases.
Figure 4.11 gives the estimated power curve based on 1,000 realizations of the IID, the AR(2) and the AR(4) error processes. We set $\zeta_1$ at the Fourier frequency $144/512$Hz. We compared the power curves of (i) the local $F$ test derived from the spectral regression model with the approximate covariance (identity covariance) using the empirical critical value, (ii) the local $F$ test derived from the spectral regression model with the approximate covariance (identity covariance) using the $F$ critical value, and (iii) the general $F$ test derived from the spectral regression model with the true covariance matrix $\Sigma$ using the $F$ critical value, under the different experiment conditions. The upper panel of the figure is for the IID process. Two power curves are overlapped through the whole range, which matches the theory of $\Sigma = I$ for the IID process. The power comparison for AR(2) are shown in the middle panel of the figure with sample size $N = 64$ and $N = 512$ respectively. The general local $F$ test using the true covariance has better power performance for small sample sizes (e.g., $N=64$) than those using the approximately diagonal covariance. This was expected as there is no reason to assume the covariance matrix is diagonal in this case. For the local $F$ test using the approximately diagonal covariance, the power performance is worse if the $F$ critical value is used because the null distribution does not follow a central $F_{2,4B}$ distribution. But the power curves are closer to each other as the sample size increased. This gives a strong support of what we demonstrated in Section 3.1 (the covariance matrix is approximately diagonal for the local $F$ test if the sample size is large enough). Then the true covariance can be approximated by the identity matrix for the local $F$ test if the sample size is large. The bottom panel of Figure 4.11 provides the power comparison of local $F$ test with or without 5% cosine taper for AR(4) error process with sample size $N = 512$ respectively. From these two plots the
power curves are fairly close to each other for the AR(4) process with a large sample size. But the power curve is distorted for the local $F$ test using the approximately covariance matrix if the wrong critical value (the $F$ critical value) is applied. It demonstrates that the asymptotic distribution of the tapered Fourier transform is violated for error processes with SDFs having large dynamic range, but tapering can be used as a remedy for this condition. From the power comparison in Figure 4.11, the $F$ test derived from the spectral regression model with approximate diagonal matrix works well under certain conditions (e.g. large sample size, or tapering).

According to a comparison of size and power from different Monte Carlo simulations in this chapter, we found that the time-domain-based $F$ test is time consuming because of the large dimension matrix calculation, though the performance is slightly better than the local $F$ test for some error processes. Thus spectral-based tests are recommended. Among the spectral-based tests, the global $F$ test has better performance than the local tests for error processes with SDFs having small dynamic range especially when the sample size is small, but the local tests can achieve the similar performance with a reasonable choice of sample size and bin width, using much less information than the global $F$ test. The local tests are robust to the nuisance periodicity and the dynamic range of the error process as long as the nuisance frequency is outside the local region used to form the test statistic. Tapering is very useful strategy that can reduce the leakage effect for error processes with SDFs having large dynamic range. The local $F$ test with a cosine taper and Thomson’s multitaper $F$ test can be used to test the periodicity under such a situation.

When we apply the spectral-based tests for periodicity in practice, we need to explore the underlying properties of the noise process (e.g., dynamic range) and the
periodicities (e.g., unknown nuisance periodicities) as well as the sample size. A certain spectral-based test can be chosen based on the information summarized from the simulation results in this chapter. We will apply our spectral-based tests of periodicity to the field of Speech and Hearing Science in the next chapter.
Figure 4.11: A comparison of power curve of the local $F$ test with approximate and true covariance matrix for IID, AR(2) and AR(4) error processes. We use bin size $B = 4$. And for the tapered cases, a 5% cosine taper was employed.
CHAPTER 5

ANALYSIS OF DISTORTION PRODUCT OTOACOUSTIC EMISSIONS (DPOAES)

The study of time series with periodicities is applicable to many fields, such as Economics, Biology, Psychology, Medicine, Engineering, and Acoustics. We can apply our spectral based tests to detect the periodicities with known frequencies for such time series data. One application we focus on heavily here is to Speech and Hearing Science.

5.1 Introduction of Distortion Product Otoacoustic Emissions (DPOAEs)

Distortion product otoacoustic emissions (DPOAEs) (Kemp, 1978) are one type of otoacoustic emissions (OAEs), which are acoustical signals that can be detected in the ear canal. DPOAEs are generated by intermodulation distortion in the healthy mammalian cochlea by the nonlinear interaction of two simultaneous stimulus tones $f_1$ and $f_2$ with $f_2 > f_1$. It is said to be distorted because it originates from the cochlea as a tonal signal that is not present in the eliciting stimulus tones. The distortion products (DPs) of the form $k f_1 \pm m f_2 (k, m \in \mathbb{Z})$ are observable when they travel back out through the middle ear into the external ear. The most robust and frequently
measured acoustic intermodulation distortion product in humans is $2f_1 - f_2$, the so-called cubic distortion product (CDP). The amplitude of the DPOAEs is dependent on several parameters of the primary stimuli and the overall health of the cochlea. The DPOAE amplitude is usually higher for females than for males due to the different cochlea function (Robinette and Glattke (1997)[Chapter 5]). The DPOAE amplitude is higher for $f_2$ in the region from 2 kHz to 4 kHz. DP levels for $f_2$ below 2 kHz are difficult to measure, especially below 1 kHz, and above 4 kHz, there are measurement issues as well. The frequency separation between the two stimulus tones is crucially important. The relative frequency or the ratio of the higher ($f_2$) to the lower ($f_1$) primary frequency, $f_2/f_1$, defines the separation between the two frequencies. A DPOAE will not be generated if these two primary frequencies are too far apart, i.e., $f_2/f_1$ greater than 1.5, or too close together, i.e., $f_2/f_1$ less than 1.01. It has been demonstrated that the largest amplitudes occur with $f_2/f_1$ ratio around 1.2 (Robinette and Glattke (1997)[Chapter 5]).

Since DPOAEs depend on vulnerable physiologic processes, their presence or absence is routinely used as an indicator of cochlear status. One wide-spread application is newborn hearing screening. Because the most common cause of hearing loss in infants is the result of cochlear dysfunction, Otoacoustic emissions are an effective screening instrument in this population. In this screening applications, DPOAE testing reduces to a binary hypothesis testing problem as shown in Chapter 3. We test if the amplitude of CDP, $A_1$, is significantly different from zero; that is,

$$H_0: A_1 = 0 \text{ versus } H_1: A_1 \neq 0.$$ 

For each $(f_2, f_1)$ pair, the decision (or action) space is normal-abnormal based on the detection of the amplitude of the CDP. Since the CDP generation site is held to be
near the $f_2$ place, varying $f_2$ (and hence $f_1$) in frequency allows for binary hypothesis tests to be performed over a great extent of the cochlea. Tests for hearing, based on otoacoustic emissions, are typically calculated using the periodogram Gorga et al. (1993).

Prof. King in Department of Speech and Hearing Sciences provided us the DPOAE data, which were collected from 8 female and 7 male adult subjects, aged 21-40 years. Subjects had normal hearing sensitivity defined by audiometric thresholds at or below 15 dB hearing level (HL) at the standard audiometric test frequencies (250 Hz - 8 kHz). The subjects had no air-bone gap and normal tympanograms, indicating normal middle ear function. Otoscopic examination revealed clear external auditory canals on the day of testing. All human testing was carried out with the approval of the Ohio State University Institutional Review Board.

All DPOAE testing was performed in a sound-attenuated room. Stimulus presentation and recording was done using a Tucker-Davis System II software and hardware (Tucker Davis System Technologies) coupled to an Etymotic Research ER10B+ research microphone with an ER10B pre-amplifier providing +40 decibels of gain. The ratio of the evoking tones $f_2/f_1$ was fixed at 1.2 to guarantee that the CDP has maximum amplitude for easier detecting. The overall levels were held at 65 dB and 55 dB sound pressure level (SPL) respectively. These parameters (evoking tone frequency ratio and amplitudes) have been shown to evoke the most robust emissions in humans. The higher frequency evoking tone, $f_2$ was varied from 500 to 900 in 100 Hz steps and then from 1 kHz to 8 kHz in 1 kHz steps for a total of 13 ($f_2, f_1$) pairs. For each stimulus frequency pair one hundred acquisitions of $N = 8192$ points were sampled at 50 kHz. The $f_2 = 500$ Hz data was missing for one male subject and the $f_2 = 5$
kHz data was missing for one female subject. We will explore this time series dataset throughout this chapter.

5.2 Exploratory Data Analysis

Using the harmonic process model,

\[ X_t = \sum_{l=1}^{L} A_l \cos(2\pi \zeta_l t \Delta + \phi_l) + Z_t, \quad t = 1, \ldots, 8192, \]

introduced in Chapter 2, the DPOAEs can be represented by this model with \( L = 3 \) using \( \zeta_1 = 2f_1 - f_2, \zeta_2 = f_1, \) and \( \zeta_3 = f_2. \) We want to detect if the amplitude of the CDP, \( A_1, \) is significantly different from zero. The spectral-based tests for periodicity introduced in Chapter 3 can be applied to the DPOAE dataset. As we learned from the simulations, the global \( F \) test usually has larger power than the local tests because it uses all the sample points (information), but the global test is very sensitive to the nuisance periodicity and the dynamic range of the error process comparing to the local tests. The local tests can have similar power performance as the global one with a reasonable choice of sample size and bin width, but using much less information than the global \( F \) test. Among the local tests, the local \( F \) test has more power than the Gorga et al. hearing test, and Thomson’s multitaper test has better power performance than other local tests only for the small sample size and the error process with SDFs having large dynamic range.

For an illustration of the testing methodology, two subjects are randomly picked from DPOAE dataset: one is female, the other is male. The periodogram of the DPOAE data at \( f_2 = 2 \) kHz and sample size \( N = 8192 \) for randomly picked one trial (acquisition) is examined, which is given in Figure 5.1. The three vertical dashed lines indicate the three frequencies (from the left to the right), CDP (left), \( f_1 \) (middle), and
Figure 5.1: Plots of the periodogram of two subjects (one female and one male).

It is noticed that there are also some spikes at other unknown frequencies, which highlights the presence of nuisance frequencies (indicating $L > 3$). As we have shown, this will make it difficult to use the global $F$ test (it will be hard estimate the underlying SDF of the non-IID error process using smoothing splines). We can see the dynamic range of the periodogram is not very large (about 40-50dB). Thus Thomson’s multitaper test might not improve the performance. The Local $F$ test is a better choice than the Gorga et al. hearing test since the local $F$ test has larger power to detect CDPs. We choose to apply the local $F$ test to detect the periodicity of CDP. We will use the bin width $B = 4$ for the local $F$ test as we have shown in Chapter 4 that the power performance for the local $F$ test is not substantial for five bins ($B = 5$) compared to four bins ($B = 4$), but the power performance is better for four bins ($B = 4$) compared to three bins ($B = 3$).
Applying the local $F$ test with bin width $B = 4$ to this randomly selected one trial from these two subjects respectively, we obtain $F(\zeta_1) = 888.97$ for the female signal and $F(\zeta_1) = 124.54$ for the male signal. Comparing these $F$ test statistics to the approximate critical value of $F_{2,4,\alpha} = F_{2,16,0.05} = 3.6337$, both $F$ test statistics for the female and male subjects are larger than the critical value. Thus we can conclude that the cochlear status of both female and male subjects are normal in this case, as we expected. The test statistic of the female subject is larger than that of the male subject because the DPOAE amplitude is higher for female than for male as we can see from the periodogram.

Based on the local $F$ test results, there are several models we can use to explore the DPOAE dataset. After we apply the local $F$ test, we have a binary response: either evidence of normal or abnormal hearing. We will use “1” to represent the rejection of the null hypothesis, i.e., the amplitude at the frequency of CDP is significant different from 0 (the normal cochlear function). “0” is used when we fail to reject of null hypothesis, i.e., the amplitude at the frequency of CDP is not significantly different from 0 (the abnormal cochlear function). These binary normal-abnormal responses can be modeled by a logistic regression model in terms of the evoking frequency $f_2$ and gender, which we will examine in detail in the next section.

Secondly, we want to model the random effects presented in the signal-to-noise ratio once we account for a number of known factors such as the evoking frequency, $f_2$, and the gender of each subject (remember that the other evoked frequency $f_1 = f_2/1.2$ is a function of $f_2$). This leads us to explore the local $F$ test statistics of all the subjects at the different $f_2$ frequencies for all the 100 replicates collected for each $(f_2, f_1)$ pair. We know from Chapter 3 that, to within a good approximation, the local
\textit{F} statistics have a noncentral \textit{F} distribution. Then a generalized linear mixed effect model (GLMM) is used to model the local \textit{F} test statistics to capture the variation among subjects across the frequencies. The GLMM will be introduced in Section 5.4. As we mentioned in Section 5.1, the amplitude of the DPOAEs dependents on the frequencies of the primary stimuli (i.e., \( f_1 \) and \( f_2 \)) and the overall health of the cochlea (e.g., gender). Thus we will consider to use the evoking frequency \( f_2 \) (since \( f_1 \) is a function of \( f_2 \)) and gender as our covariates.

### 5.3 Logistic Regression Models

Applying the local \textit{F} test to DPOAE dataset, the local \textit{F} test statistics can be compared with the \textit{F}-based critical value. If test statistic is larger than the critical value, the null hypothesis of \( A_1 = 0 \) is rejected (indicating evidence that the cochlear status is normal). We use “1” to denote this situation. Otherwise the null hypothesis is not rejected (there is evidence that cochlear status is abnormal). We use “0” is used to represent this case. Let this binary response of the local \textit{F} test result be \( Y_{jkl} \), which \( Y_{jkl} \) denotes the \( l \)th replicate \( l = 1, \ldots, L \) (\( L = 100 \) here) for subject \( j = 1, \ldots, J \) (\( J = 15 \) here) at the different evoking frequencies \( f_{2,k} : k = 1, \ldots, K \) (\( K = 13 \) here). The distribution of \( Y_{jkl} \) is \( Y_{jkl} \sim \text{Bern}(p_k) \), where \( p_k \) is the probability of the healthy cochlear status at the evoking frequency \( f_{2,k} \) for the subject \( j \). Assume independence of the random variables, we will model this binary data using binomial generalized linear model (GLM) with logit link (a logistic model). Before we apply a logistic model to the DPOAE dataset, the logistic regression model is reviewed first.
5.3.1 Review of The Logistic Regression Model

Logistic regression model is a special case of the generalized linear model (GLM) with binary data. For a review of GLM, see McCullagh and Nelder (1999). A GLM has the following components: a random component, for a set of independent random variables \( \{Y_i, i = 1, \ldots, n\} \), \( Y_i \sim f(y_i|\theta_i, \phi) \), where \( f(\cdot) \) is a pdf/pmf for an exponential family and the observed data \( \{y_i\} \) is a realization of \( \{Y_i\} \), a link function \( g(\cdot) \) which satisfies \( \eta_i = g(\mu_i) \), where \( \mu_i = E(Y_i) \), and a systematic or linear predictor component, \( \eta_i = x_i^T \beta \). Let \( f(y_i|\theta_i, \phi) \) have the well known Aitkin form:

\[
f(y_i|\theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right),
\]

where \( \theta_i \) is the canonical parameter and \( \phi \) is the scale parameter. The log-likelihood function is

\[
l(\theta_i|y_i) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi).
\]

The score function, \( U_j(\beta) \), can be attained by taking derivatives of the log-likelihood with respect to \( j \)th parameter, \( \beta_j, j = 1, \ldots, p \),

\[
U_j(\beta) = \frac{\partial l(\beta)}{\partial \beta_j}.
\]

The parameter \( \beta \) can be estimated by solving the score equations, but since there is no closed form expression, we can use iteratively weighted least squares (IWLS) (see McCullagh and Nelder (1999)[Section 2.5]).

The deviance is often used to measure the goodness of fit of a GLM model, which is defined by

\[
D(y, \hat{\mu}) = 2a(\phi)\{l(\hat{\theta}) - l(\tilde{\theta})\}
\]

\[
= 2 \sum_{i=1}^{n} \{y_i(\hat{\theta}_i - \tilde{\theta}_i) - (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}
\]
\[ \sum_{i=1}^{n} D_i, \]

where \( \tilde{\theta}_i \) is the estimate of \( \theta_i \) when we fit the full model (when we estimate \( \mu_i \) using \( \tilde{\mu}_i = y_i \)), \( \hat{\theta}_i \) denotes the estimate of \( \theta_i \) when we fit the reduced model, and \( \hat{\mu}_i \) is the associated estimate of \( \mu_i \) under the reduced model. The deviance residual is then for \( i = 1, \ldots, n \)

\[ (r_D)_i = \text{sign}(y_i - \hat{\mu}_i) \sqrt{D_i}. \]

We can then create an analysis of deviance table and deviance residual plots to evaluate the generalized linear model (McCullagh and Nelder (1999)[Section 2.4.3 and 4.4.3]).

For independently distributed \( \{Y_i : i = 1, \ldots, n\} \) with \( Y_i \sim \text{Bin}(m_i, p_i) \) for each \( i \), using the logit link function, \( \eta_i = \theta_i = \log \left( \frac{p_i}{1-p_i} \right) \), we have

\[
f(y_i | p_i) = \binom{m_i}{y_i} p_i^{y_i} (1 - p_i)^{m_i-y_i} = \exp \left( \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right),
\]

where \( a(\phi) = \phi, \phi = 1, b(\theta_i) = m_i \log(1 + e^{\theta_i}), \) and \( c(y_i, \phi) = \log \left( \frac{m_i}{y_i} \right) \). Taking derivatives of the log-likelihood with respect to \( \beta \) we obtain the score function. Then \( \beta \) can be estimated by solving the score equation with IWLS. The deviance can be calculated for this binary data, which is

\[
D(y, \hat{\mu}) = 2 \sum_{i=1}^{n} \left\{ y_i \log \left( \frac{y_i}{\hat{\mu}} \right) + (m_i - y_i) \log \left( \frac{m_i - y_i}{m_i - \hat{\mu}} \right) \right\}.
\]

Next, we will apply the binomial GLM (logistic regression model) to the DPOAE dataset.
5.3.2 Logistic Regression Model for DPOAE-based Healthy Cochlear Status

In order to get a binary response of normal or abnormal hearing, we apply the local $F$ test to the DPOAE data. Then we compare the local $F$ test statistics value with the $F$ based critical value, $F_{2.4b,\alpha}$, from the $F$ distribution (e.g., using the critical value $F_{2.16,0.05} = 3.6337$ for $B = 4$ and $\alpha = 0.05$). Let “1” denote the case of when the local $F$ test statistic exceeds the critical value, which rejects the null hypothesis $A_1 = 0$ (i.e., there is evidence of healthy cochlear function). Then “0” denotes the local $F$ test statistic is less than the critical value, which fails to reject the null hypothesis (i.e., there is no evidence of healthy cochlear function). Let $y_{jkl}$ denote the observed binary response of the local $F$ test result for the subject $j = 1, \ldots, J$ with the $l$th replicate $(l = 1, \ldots, L)$ at the different evoking frequencies $f_{2,k}: k = 1, \ldots, K$. If we are interested in the population probability of independent healthy cochlear status at each evoking frequency, then $\{y_{jkl}\}$ is a realization of independent random variables $\{Y_{jkl}\}$, where $Y_{jkl} \sim \text{Bern}(p_k)$ and $p_k$ is the population probability at the evoking frequency $f_{2,k}$. If a characteristic of each subject $j$ (e.g., gender) also matters, $\{y_{jkl}\}$ is a realization of independent random variables $\{Y_{jkl}\}$, where $Y_{jkl} \sim \text{Bern}(p_{jk})$ and $p_{jk}$ is the population probability at the evoking frequency $f_{2,k}$. In our experiment we have $K = 15$ and $f_{2,k} = 500$ Hz, $600$ Hz, $\ldots$, $900$ Hz, $1$ kHz, $2$ kHz, $\ldots$, $8$ kHz. If we consider the females, the males, as well as the females and males respectively, we have a different number of subjects, $J$: $J = 8$ (females), $J = 7$ (males) and $J = 15$ (both females and males). Since we are interested in modeling the population probability of the normal cochlear status at each evoking frequency, we average the time series of several trials for each subject at each evoking frequency (see Jones (August 2003)).
Figure 5.2: Plot of observed proportion at each evoking frequency using the female subjects, the male subjects and all the subjects respectively.

Then the local $F$ test statistic is calculated from the periodogram of the average of these time series. A trimmed mean is considered to reduce the effect of the variability of the recording system because some trials are quite noisy and prone to outliers. Here we will average the times series of 100 trials from each subject at each frequency in two different ways. Firstly, we will use a 20% trimmed mean of the time series of all 100 trials to calculate the local $F$ test statistic for each subject at each evoking frequency (i.e., $L = 1$ (no replicates)). Secondly, we will apply 10% trimmed mean to the times series of every 20 trials. Thus we have 5 values of the local $F$ test statistic, and thus 5 binary responses (i.e., $L = 5$) for each subject at each evoking frequency.

Although we know what covariates we are going to use, we still want to get a rough idea of what kind of logistic model to fit. We use log 10 of the frequency $f_2$,
log \( f_2 \), as the explanatory variable. Then we draw the estimated proportion of healthy cochlear status at each evoking frequency using the observed data from the female subjects, the male subjects and all the subjects respectively. From Figure 5.2 we can see that there might exist a certain quadratic (or maybe cubic) pattern in terms of log 10 of the evoking frequency. The DPOAE is easier to detect (higher probabilities) for \( f_2 \) in the region from 1 kHz to 5 kHz. This is because the DP level for \( f_2 \) below 1 kHz and above 5 kHz are difficult to measure. Thus we will try quadratic and cubic logistic models for the females, the males, and all subjects respectively. We also fit the linear logistic regression model to evaluate those quadratic and cubic models.

For the first case, we apply the local \( F \) test to the 20\% trimmed mean of the times series of all 100 trials for each subject at each frequency (i.e., \( L = 1 \)). We will examine the logistic models for only female subjects, for only males subjects and then for the female and male subjects together respectively. For the females, the estimated quadratic model for the logit of the probability of healthy cochlear status for females is

\[
\log \left( \frac{\hat{p}_k}{1 - \hat{p}_k} \right) = \hat{\beta}_0 + \hat{\beta}_1 (\log f_2)_k + \hat{\beta}_2 (\log f_2)_k^2, \tag{5.1}
\]

where \( \hat{\beta}_0 = -111.09 \) with standard error s.e. = 35.66, \( \hat{\beta}_1 = 68.65 \) with s.e. = 22.06 and \( \hat{\beta}_2 = -10.21 \) with s.e. = 3.33. All the terms are very significant because the corresponding marginal p-values are less than 0.003. From the upper panel of the diagnostic plots showed in Figure 5.3, we can see the model fits most of the data well except some data at lower or higher frequency (i.e., 700, 7k and 8k). But the normality is not very good from both of the Q-Q plot and the histogram. The cubic term of log \( f_2 \) was added into the logistic model, but it was not significant.
Figure 5.3: Diagnostic plots for quadratic logistic regression model for females, males, and all subjects, using the 20% trimmed mean of the times series of all 100 trials.
For the males, the cubic and quadratic terms are not significant in the logistic regression model. Then the estimated logit of the probability of healthy cochlear status for males is

\[
\log\left(\frac{\widehat{p}_k}{1 - \widehat{p}_k}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 (\log f_2)_k,
\]

where \(\widehat{\beta}_0 = -13.92\) (s.e. = 4.16) and \(\widehat{\beta}_1 = 4.95\) (s.e. = 1.42) with p-values less than 0.001. The middle panel of Figure 5.3 gives the diagnostic plots for this model. We can see the model fits the probability of healthy cochlear status fairly good only at higher frequency, which is different from the females (the fit of the model for the females is good at the middle range of the frequencies). This is because two male subjects fail to reject the null hypothesis at the evoking frequency \(f_2 = 3\) kHz. The normality is still violated from the Q-Q plot and the histogram.

For all the subjects, the quadratic model (5.1) with \(\widehat{\beta}_0 = -82.93\) (s.e. = 23.18), \(\widehat{\beta}_1 = 50.02\) (s.e. = 14.33) and \(\widehat{\beta}_2 = -7.23\) (s.e. = 2.17) is a good candidate. All the p-values are very small. From the diagnostic plots in the bottom panel of Figure 5.3, we can see the model fits the most of proportions well except some higher frequencies (i.e. 3k, 7k and 8k). If we add a gender explanatory variable (1 denotes female, 0 denotes male) into the model, we find the gender is not significant in either a logit linear model or a logit quadratic model.

Moreover, we can compare the binomial GLM with the logit link, probit link, and complementary log-log link using the Akaike Information Criterion (AIC). A smaller AIC value indicates a better fitting model and a possible choice of link function. Table 5.1 lists all the AIC values based on the candidate models for the females, the males and all the subjects respectively. We can see there is no big difference in using different links. The logit link gives the smallest AIC value for both the males and all
Table 5.1: AIC for different links using the 20% trimmed mean of the times series of all 100 trials.

<table>
<thead>
<tr>
<th>links</th>
<th>logit</th>
<th>probit</th>
<th>c log-log</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC(female)</td>
<td>72.00</td>
<td>71.29</td>
<td>70.45</td>
</tr>
<tr>
<td>AIC(male)</td>
<td>67.73</td>
<td>68.30</td>
<td>69.31</td>
</tr>
<tr>
<td>AIC(all)</td>
<td>144.04</td>
<td>144.72</td>
<td>146.29</td>
</tr>
</tbody>
</table>

the subjects, but the complementary log-log (c log-log) link gives the smallest AIC value for females. Comparing the diagnostic figures with different links, there is no big improvement. Thus we stick to the simplest to interpret link, the logit link.

Secondly, we calculate the local $F$ test statistic based on the 10% trimmed mean of the time series, every 20 trials, for each subject at each frequency. In this way, we can get more data to analyze the probability of healthy cochlear status at each evoking frequency. Since the 100 trials for each subject at each frequency are correlated, using the 10% trimmed mean of the time series of every 20 trials can reduce the correlation among the replicates. Here we have 5 replicates ($L = 5$) at each frequency for each subject. Thus, we use the same strategy as before to fit the binomial GLM models.

For the females, we still have quadratic logistic regression model (5.1) with $\hat{\beta}_0 = -119.53$ (s.e. = 13.98), $\hat{\beta}_1 = 72.05$ (s.e. = 8.57) and $\hat{\beta}_2 = -10.66$ (s.e. = 1.29) as the good candidate. All the terms in the model are very significant since all the p-values are quite small. From the diagnostic plots provided in the upper panel of Figure 5.4, we can see the model still fits most of the data well except some data at higher frequency (i.e. 7k and 8k). Comparing the histogram with the corresponding histogram in Figure 5.3, we can see the data is less left skewed than before. The model seems to fit better.
Figure 5.4: Diagnostic plots for quadratic logistic regression model for female, males and all subjects using the 10% trimmed mean of the times series of every 20 trials.
For the males with $L = 5$ replicates for each evoking frequency, the quadratic model (5.1) is selected instead of the linear logistic model (the model for the 1 replicate case). The parameter estimates are $\hat{\beta}_0 = -67.40$ (s.e. = 14.06), $\hat{\beta}_1 = 38.20$ (s.e. = 8.60) and $\hat{\beta}_2 = -5.22$ (s.e. = 1.30), which are marginally very significant due to the small p-values. The middle panel of Figure 5.4 gives the diagnostic plots for this model. We can see the model fits the proportion fairly good at higher frequencies, but for the middle range of the frequencies, there still exists some variability. This is still due to the same two male subjects as the first case, who frequently fail to reject the null hypothesis at these evoking frequencies.

Following the same procedure to fit all the subjects together by using the logistic regression model of $\log f_2$, we find all the terms ($\hat{\beta}_0 = -89.15$ (s.e. = 9.36), $\hat{\beta}_1 = 52.81$ (s.e. = 5.73), and $\hat{\beta}_2 = -7.58$ (s.e. = 0.86)) in the quadratic model (5.1) are marginally significant. Since we found the quadratic shape of the females is different from that of the males, the new explanatory variable gender is added into the model. The new model is

$$\log \left( \frac{\hat{p}_k}{1 - \hat{p}_k} \right) = \hat{\beta}_0 + \hat{\beta}_1 (\log f_2)_k + \hat{\beta}_2 (\log f_2)_k^2 + \hat{\beta}_3 (\text{gender})_{jk} + \hat{\beta}_4 ((\log f_2)_k \ast (\text{gender})_{jk}) + \hat{\beta}_5 ((\log f_2)_k^2 \ast (\text{gender})_{jk}), \quad (5.2)$$

where $\hat{\beta}_0 = -67.40$ (s.e. = 14.06), $\hat{\beta}_1 = 38.20$ (s.e. = 8.60), $\hat{\beta}_2 = -5.22$ (s.e. = 1.30), $\hat{\beta}_3 = -52.13$ (s.e. = 19.83), $\hat{\beta}_4 = 34.25$ (s.e. = 12.14), and $\hat{\beta}_5 = -5.44$ (s.e. = 1.83). All the terms (including the gender variable and the interactions between the evoking frequency and the gender) are significant in this new logistic quadratic model (all the p-values are less than 0.003). Adding the interaction term matches what we saw from the exploratory plot (Figure 5.2): the quadratic shape of the females is different from that of the males, but they are not parallel. From the diagnostic plots in the
Table 5.2: AIC for different links using the 10% trimmed mean of the times series of every 20 trials.

<table>
<thead>
<tr>
<th>links</th>
<th>logit</th>
<th>probit</th>
<th>c log-log</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC(female)</td>
<td>436.84</td>
<td>436.04</td>
<td>437.38</td>
</tr>
<tr>
<td>AIC(male)</td>
<td>392.84</td>
<td>393.3</td>
<td>402.25</td>
</tr>
<tr>
<td>AIC(all)</td>
<td>829.67</td>
<td>829.34</td>
<td>839.64</td>
</tr>
</tbody>
</table>

bottom panel of Figure 5.4, we can see the model fits the most of proportions well. The normality is still bad by looking at the Q-Q plot and the histogram. The fitted proportions of healthy cochlear status for females are larger than those for males at most of the evoking frequencies except the higher ones (i.e., 6 kHz, 7 kHz and 8 kHz). This does make sense because females have larger local $F$ test statistics at these evoking frequencies in this DPOAE dataset as we saw from Section 5.2, which makes the null hypothesis be rejected more easily.

Again, we compare the binomial GLM with the logit link, probit link, and complementary log-log link using AIC. From Table 5.1, we can see there is no big difference by using different links. Then we will use the simplest logit link.

We can see all the logistic regression models fit the most of proportions well except some higher frequencies (i.e. 7 kHz and 8 kHz) because it is difficult to measure the DPOAE at that range. The logit link is good choice for DPOAE data in the logistic regression model, but the normality for the deviance residuals is violated, which is not surprising as we had a smaller number of trials at each evoking frequency. In next chapter we will model the local $F$ test statistics directly instead of the decisions made based on the local $F$ test.
5.4 Generalized Linear Mixed Effects Model for the DPOAE Periodicity $F$ Statistics

Researchers in Speech and Hearing Sciences are interested not only in testing for the presence of DPOAEs, but also in estimating the magnitude of DPOAEs. We know from Section 3.4.2 that the local $F$ test statistic has a noncentral $F$ distribution with noncentrality parameter $\lambda = \frac{\Delta N A_1^2}{4S_Z(\zeta_1)}$, which is the function of the amplitude of CDP, $A_1$. We use signal-to-noise ratio (SNR), the power ratio between a signal (meaningful information) and the background noise, which is

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}}, \quad (5.3)$$

where $P$ is the power. We can see this noncentrality parameter is also a function of the SNR because $A_1^2$ is the function of the power of the CDP and $S_Z(\zeta_1)$ is a function of the power of the background noise. In this section, a generalized linear model (GLM) with the noncentral $F$ distribution is applied to model the local $F$ test statistics obtained using our DPOAE dataset. For longitudinal data analysis used for different subjects, mixed effects models (with random and fixed effects) have broader use than fixed effects model, and allow us to capture the many subject and frequency specific effects that we expect to observe in DPOAE data. In this section we consider a generalized linear mixed effects model (GLMM) (e.g., McCullagh and Nelder, 1999; Diggle et al., 2002), for $F$ statistics derived from the DPOAE dataset to capture the variation within the subjects across the different evoking frequencies of interest.
5.4.1 Parameter Estimation in the Generalized Linear Mixed Effects Model

For GLMs, the most common way to estimate the parameters is using the frequentist method. A likelihood function can be constructed to do the statistical inference (e.g. Binomial GLM in Section 5.3). But there might have insufficient information to construct a likelihood function or the likelihood function is difficult to calculate. Then the other way of estimating the parameters is to use the applicable methodology, quasi-likelihood (QL), which has been first used by Wedderburn (1974). He showed that the log likelihood function is identical to the QL if and only if this family is an exponential family. Instead of full distributional assumptions, Wedderburn gives the definition of QL function which only relies on weaker assumptions about the means and variances/covariances. The precision of the maximum QL estimates is also explored and the Gauss-Newton method is applied to calculate the estimates. McCullagh and Nelder (1999, Chapter 9) also give a comprehensive introduction of the quasi-likelihood function. The QL method cannot directly apply to the GLMM. For mixed effects, the traditional maximum likelihood can be considered, even though the likelihood is difficult to evaluate due to the complicated numerical integration. Bayesian approaches can be used by taking repeated samples from the posterior distribution instead of doing the numerical integration. Karim and Zeger (1992) used the Monte Carlo method, the Gibbs sampler, to overcome the difficulties of high-dimensional numerical integration. They applied the Gibbs sampler to the data set which has random effects. Instead of using the likelihood function, in this dissertation we examine how to adapt quasi-likelihood methods to estimate the parameters of the GLMM. But the relationship between mean and variance in the quasi-likelihood

115
function is not sufficient to estimate the variance-covariance structure of the GLMM. Green (1987) suggested to add a penalty function to the quasi-likelihood, which is called penalized quasi-likelihood (PQL). Semi-parametric regression models have also been explored by using the PQL method. Schall (1991) proposed the PQL as an approximate Bayes procedure. Our interest is in the work of Breslow and Clayton (1993) who applied the PQL criterion to the GLMM. We present the outline of the computational procedure. The standard weighted least squares (IWLS) can be simply applied to solve the score equations when the PQL is maximized. More details about the QL and the PQL method now follow.

For generalized linear models (GLMs), the quasi-likelihood model relies on us stating the following about our random variables of interest:

1. Whether or not the random variables are independent;

2. The mean of the random variables;

3. The variance of the random variables.

No further distribution assumptions are made. For a generalized linear fixed effects model we would assume that the $n$ components of the response variable $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ are independent with mean $\mathbf{\mu} = (\mu_1, \ldots, \mu_n)^T$ and variance $\phi V(\mu_i)$ for each $i = 1, \ldots, n$, where $V(\cdot)$ is a known variance function and $\phi$ is a dispersion parameter. We model the mean structure through a link function $g(\cdot)$, with

$$g(\mathbf{\mu}) = \mathbf{X} \mathbf{\beta}, \quad (5.4)$$
where $X$ is the $n \times p$ design matrix and $\beta$ is a vector of $p$ fixed effects parameters. Then the quasi-likelihood function for an observation $Y_i$ ($i = 1, \ldots, n$) is defined by

$$Q_i = \int_{Y_i}^{\mu_i} \frac{Y_i - t}{\phi V(t)} \, dt; \quad (5.5)$$

and the associated quasi-score function is

$$U_i = \frac{Y_i - \mu_i}{\phi V(\mu_i)}. \quad (5.6)$$

The quasi-likelihood for the complete data is the sum of the individual contributions, namely $Q = \sum_{i=1}^{n} Q_i$ (similarly for the quasi-score for the complete data).

For a generalized linear mixed effects model (GLMM), we extend (5.4) to incorporate random effects. We have

$$g(\mu) = X\beta + Wb,$$

where $\beta$ is a $p \times 1$ fixed effects vector, $b$ is a $q \times 1$ random effects vectors, and $X$ and $W$ are, respectively, the $n \times p$ and $n \times q$ explanatory variable matrices associated with the fixed and random effects. In this model we define $\mu = E(Y|b)$ and $\text{var}(Y|b) = \phi V(\mu)$ to be the mean and variance conditional on the random effects $b$, where $\phi > 0$ is a dispersion parameter, and $V(\cdot)$ is a $n \times n$ matrix of specified variance functions. The conditional mean is related to the linear predictor $\psi = X\beta + Wb$ through the link function $\psi = g(\mu)$, where $g(\cdot)$ is assumed to be differentiable. We assume $b$ has a multivariate normal distribution with mean 0 and covariance matrix $D_\omega$, where $\omega$ is a vector of covariance parameters.

The full likelihood for $\beta$ and $\omega$ is

$$L(\beta, \omega|Y) = \int_{\mathbb{R}^q} \left[ \prod_{i=1}^{n} f(y_i|\beta, b) \right] f(b|\omega) \, db, \quad (5.7)$$
where \( f(b|\omega) \propto \exp\left(-\frac{1}{2}b^T D_\omega^{-1}b\right) \) denotes the multivariate normal density for the random effects, that depends on \( \omega \). Maximizing (5.7) with respect to \( \beta \) and \( \omega \), we obtain the maximum likelihood estimates. For most cases, this integral does not have a closed form. Thus, the maximum likelihood estimates are not directly available; instead some computational method is required to maximize the likelihood, or we can use a quasi-likelihood-based approach.

Adapting the quasi-likelihood method to estimate the parameters of the GLMM, a penalty function is added to the quasi-likelihood (Green, 1987). This is called the penalized quasi-likelihood (PQL) and is defined by

\[
PQL = \sum_{i=1}^{n} Q_i - \frac{1}{2} b^T D_\omega^{-1} b,
\]

where \( Q_i \) is the quasi-likelihood for individual \( i \) that is given by Equation (5.5). We maximize this PQL by differentiating with respect to \( \beta \) and \( b \). Then the two score equations are

\[
X^T \Omega (Y - \mu) = 0,
\]

\[
W^T \Omega (Y - \mu) - D_\omega^{-1} b = 0,
\]

where \( \Omega \) is a diagonal matrix with diagonal terms \( \{\phi V(\mu_i) g'(\mu_i)\}^{-1} \) \( (i = 1, \ldots, n) \) and \( g'(\cdot) \) is the first derivative of the link function \( g(\cdot) \). Hence \( \beta \) and \( b \) can be estimated by solving these two equations. Breslow and Clayton (1993) solved these two equations by reexpressing the Fisher scoring algorithm developed by Green (1987) as an IWLS problem. They defined the working vector \( a = \psi + (Y - \mu)g'(\mu) \) and estimated the parameter \( \beta \) and \( b \) in the associated normal theory model \( a = X\beta + Wb + e \), where \( e \sim N(0, G^{-1}) \), \( b \sim N(0, D_\omega) \), and \( G \) is the diagonal matrix with diagonal terms \( G_i = \{\phi V(\mu_i)[g'(\mu_i)]^2\}^{-1} \). Here, \( e \) and \( b \) are independent. From
the usual normal theory,

\[ \hat{\beta} = (X^T R^{-1} X)^{-1} X R^{-1} a, \quad (5.8) \]

and

\[ \hat{b} = D\omega W^T R^{-1}(a - X\hat{\beta}), \quad (5.9) \]

where \( R = G^{-1} + WD\omega W^T \).

The integrated quasi-likelihood function for \((\beta, \omega)\), defined in Breslow and Clayton (1993), is

\[ L(\beta, \omega) = e^{l(\beta; \omega)} \]

\[ \propto |D\omega|^{-1/2} \int \exp \left\{ \sum_{i=1}^{n} Q_i(\beta; b) - \frac{1}{2} b^T D\omega^{-1} b \right\} \, db, \quad (5.10) \]

where for each \( i = 1, \ldots, n \),

\[ Q_i(\beta; b) \propto \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} \, dt. \]

We examine the special case with the observations \( y_i \)'s are drawn from an exponential family with variance function \( V(\cdot) \). It is well known that \( Q_i(\beta; b) = [f(y_i; \mu_i, \phi) - f(y_i; y_i, \phi)] \) for exponential families, where \( f(y_i; \mu_i, \phi) \) is a conditional exponential family likelihood of \( y_i \) given its mean \( \mu_i \) (McCullagh and Nelder (1999)[page 327]). Then the integrated quasi-likelihood \( L(\beta, \omega) \) in the Equation (5.10) is the same as the likelihood in Equation (5.7). We now evaluate the integral in Equation (5.10) by using a first-order Laplace approximation of the form \( \int e^{-\kappa(b)} \, db \) as Breslow and Clayton (1993) did. Then,

\[ l(\beta, \omega) \approx -\frac{1}{2} \log |I + W^T \Omega WD\omega| + \sum_{i=1}^{n} Q_i(\beta; \hat{b}) - \frac{1}{2} \hat{b}^T D\omega^{-1} \hat{b}, \]

119
where \( \hat{b} \) denotes the solution to \(-\kappa'(b) = W^T\Omega(Y - \mu) - D^{-1}\omega b = 0\). Assuming the first term changes slowly with \( \beta \) for fixed \( \omega \), Breslow and Clayton (1993) shows that

\[
 l(\beta, \omega) \approx \sum_{i=1}^{n} Q_i(\beta; \hat{b}) - \frac{1}{2} \hat{b}^T D^{-1}\omega \hat{b},
\]

which is the PQL evaluated at \( b = \hat{b} \). Therefore, we can see the PQL is a good approximation to the likelihood based on this exponential family case. In other words, the PQL is based on a Laplace approximation to the marginal likelihood.

### 5.4.2 Noncentral-\( F \) Linear Mixed Effects Model

For the DPOAE data we want to model the random effects present in the signal-to-noise ratio once we account for a number of known factors such as the evoking frequency, \( f_2 \), and the gender of each subject. Let \( f_{2,k} : k = 1, \ldots, K \) denote the different evoking frequency values. For a local \( F \) statistic with \( B \) Fourier bins to the left and right of the DP frequency of interest, let \( Y_{jkl} \) denote the \( l \)th replicate \((l = 1, \ldots, L)\) of the local \( F \) statistic for subject \( j = 1, \ldots, J \), at frequency \( f_2 = f_{2,k} \). (In our experiment \( J = 13, K = 15, L = 100 \) and \( f_{2,k} = 500 \text{ Hz}, 600 \text{ Hz}, \ldots, 900 \text{ Hz}, 1 \text{ kHz}, 2 \text{ kHz}, \ldots, 8 \text{ kHz} \)).

Based on the result in Section 3.4.2, we assume \( Y_{jkl} \) has a noncentral-\( F \) distribution with 2 and \( 4B \) degrees of freedom, and noncentrality parameter \( \lambda_{jkl} = \frac{\Delta N A^2}{4S_Z(z_i)} \) (a proportion to the SNR of the CDP and the background noise). Rather than modeling the signal-to-noise ratio equation (Equation 5.3) directly, we choose to model the mean of the random variables. This is commonly done for generalized linear mixed models. It is highly interpretable, and can allow for eased fitting of the model as we will demonstrate. Letting \( \mu_{jkl} = EY_{jkl} \), we have that the mean of the local \( F \) statistic
is related to the signal-to-noise ratio through the relationship,

$$
\mu_{jkl} = \frac{B(\lambda_{jkl} + 2)}{2B - 1}.
$$

We then model this mean through a log link, with

$$
\log \mu_{jkl} = x_{jkl}^T \beta + w_{jkl}^T b.
$$

(5.11)

Here $\beta$ is a vector of length $p$ of fixed effects parameters, associated with the design vector of length $p$, $x_{jkl}$, for replicate $l$ of subject $j$, at frequency $f_{2,k}$, and the design vector of length $q$, $w_{jkl}$ is associated with the vector of length $q$ of random effects, $b$. We assume that the random variables $\{Y_{jkl}\}$ are independent, conditional on these random effects. We finish the model by assuming some distributional form for $b$: assume that $b \sim N_q(0, D_\omega)$, where $N_q$ denotes the $q$-variate normal distribution and $D_\omega$ is a $q \times q$ positive definite covariance matrix, where $\omega$ is a vector of covariance parameters. For later, let $X$ denote the design matrix consisting of all the fixed effects design vectors, $x_{jkl}$, and $W$ be the design matrix consisting of all the random effect design vectors, $w_{jkl}$.

Conditional on $b$, the probability density function (pdf) for $\{Y_{jkl}\}$ is

$$
f(y_{jkl}|\beta, b) = \sum_{r=0}^{\infty} \left\{ \frac{\exp(-(2 - B^{-1})\mu_{jkl} - 2)) ((2 - B^{-1})\mu_{jkl} - 2)^r}{r!} \times \frac{\Gamma(2 + 4B + r) 2^{2+r} (4B)^{4B} y^{2+r-1}}{\Gamma(2 + r) \Gamma(4B) (2 + 4By)^{2+4B+r}} \right\}.
$$

(5.12)

where $\mu_{jkl}$ satisfies (5.11) (Properties of the noncentral $F$ distribution are listed in Appendix A). Letting $Y = (Y_{jkl})$ denote the vector of all the random variables for each replicate $l$ of subject $j$ at frequency $f_{2,k}$, the full likelihood for $\beta$ and $\omega$ is (in the complete data case)

$$
L(\beta, \omega|Y) = \int_{\mathbb{R}^q} \left[ \prod_{j=1}^{J} \prod_{k=1}^{K_j} \prod_{l=1}^{L_{jk}} f(y_{jkl}|\beta, b) \right] f(b|\omega) \, db.
$$

(5.13)
where \( f(b|\omega) \) denotes the multivariate normal density function for the random effects, that depends on \( \omega \). Maximizing (5.13) with respect to \( \beta \) and \( \omega \), we obtain the maximum likelihood estimates. To calculate this likelihood, we need to evaluate an infinite sum in the noncentral-F density given by (5.12), and then we need to integrate that expression over the density for the random effects. Thus, the maximum likelihood estimates are not directly available; instead some computational method is required to maximize the likelihood. It is possible to truncate the infinite sum to approximate the pdf for the noncentral-F random variables. For the integration, when \( q \) is of small dimension, one can use quadrature methods to numerically approximate the likelihood. When \( q \) is larger, Monte Carlo techniques are needed.

We apply the PQL method instead for our DPOAE data. Conditional on the random effects \( b \), the quasi-likelihood function and quasi-score function for an observation \( Y_{jkl} \) with mean \( \mu_{jkl} = B(\lambda_{jkl} + 2)/(2B - 1) \) and variance \( V(\mu_{jkl}) = B^2[8B(1 + \lambda_{jkl}) + \lambda_{jkl}^2]/(2(2B - 1)^2(B - 1)) \) (see Appendix A) are given by

\[
Q_{jkl} = \int_{Y_{jkl}}^{\mu_{jkl}} \frac{Y_{jkl} - t}{\phi V(t)} \, dt \quad \text{and} \quad U_{jkl} = \frac{Y_{jkl} - \mu_{jkl}}{\phi V(\mu_{jkl})},
\]

respectively. Expressed in terms of the mean, the variance of \( Y_{jkl} \) is \( c_0 + c_1 \mu_{jkl} + c_2 \mu_{jkl}^2 \), where \( c_0 = -2B^2/(2B - 1)(B - 1) \), \( c_1 = 2B/(B - 1) \), and \( c_2 = 1/(2(B - 1)) \). Then the variance \( V(\mu_{jkl}) = c_2(\mu_{jkl} + r_1)(\mu_{jkl} + r_2) \), where \( r_1 = (-c_1 + \sqrt{c_1^2 - 4c_0c_2})/2c_2 \) and \( r_2 = (-c_1 - \sqrt{c_1^2 - 4c_0c_2})/2c_2 \). The quasi-likelihood function is given by

\[
Q_{jkl} = \int_{Y_{jkl}}^{\mu_{jkl}} \frac{Y_{jkl} - t}{\phi V(t)} \, dt = \frac{1}{\phi c_2} \int_{Y_{jkl}}^{\mu_{jkl}} \frac{Y_{jkl} - t}{(t + r_1)(t + r_2)} \, dt
\]

\[
= \frac{1}{\phi c_2(r_2 - r_1)} \int_{Y_{jkl}}^{\mu_{jkl}} \frac{Y_{jkl} - t}{r_2 - r_1} \left( \frac{1}{t + r_1} - \frac{1}{t + r_2} \right) \, dt
\]
which was introduced in Section 5.4.1. The algorithm is followed, and thus $
abla = 1 + \frac{1}{\phi c_2(r_2 - r_1)} \left( \int_{Y_{jkl}}^\mu \frac{Y_{jkl} - t}{t + r_1} dt - \int_{Y_{jkl}}^\mu \frac{Y_{jkl} - t}{t + r_2} dt \right) - 
abla = 1 + \frac{1}{\phi c_2(r_2 - r_1)} \left( \int_{Y_{jkl}}^\mu \frac{Y_{jkl}}{t + r_1} dt + \int_{Y_{jkl}}^\mu \frac{t}{t + r_2} dt \right) - 
abla = 1 + \frac{1}{\phi c_2(r_2 - r_1)} \left( \int_{Y_{jkl}}^\mu \frac{Y_{jkl}}{t + r_1} dt - \int_{Y_{jkl}}^\mu \frac{1 - r_1}{t + r_1} dt \right) - 
abla = 1 + \frac{1}{\phi c_2(r_2 - r_1)} \left( \int_{Y_{jkl}}^\mu \frac{Y_{jkl}}{t + r_1} dt + \int_{Y_{jkl}}^\mu \frac{1 - r_2}{t + r_2} dt \right) - 
abla = 1 + \phi c_2(r_2 - r_1) \left( Y_{jkl} \log(t + r_1) - t + r_1 \log(t + r_1) - Y_{jkl} \log(t + r_2) \right) \left| Y_{jkl} + \frac{1}{\phi c_2(r_2 - r_1)} \left( t - r_2 \log(t + r_2) \right) \right| Y_{jkl} + \frac{1}{\phi c_2(r_2 - r_1)} \left( Y_{jkl} \log \left( \frac{t + r_1}{t + r_2} \right) \right) 
abla = 1 + \phi c_2(r_2 - r_1) \left( \log \left( \frac{(t + r_1)(Y_{jkl} + r_1)}{(t + r_2)(Y_{jkl} + r_2)} \right) \right)^{\mu_{jkl}} \nabla = 1 + \phi c_2(r_2 - r_1) \left( \log \left( \frac{(\mu_{jkl} + r_1)(Y_{jkl} + r_1)}{(\mu_{jkl} + r_2)(Y_{jkl} + r_2)} \right) \right) - \log \left( \frac{(Y_{jkl} + r_1)(Y_{jkl} + r_1)}{(Y_{jkl} + r_2)(Y_{jkl} + r_2)} \right) 
abla = 1 + \phi c_2(r_2 - r_1) \left( Y_{jkl} + r_1 \right) \log \frac{Y_{jkl} + r_1}{Y_{jkl} + r_1} - \left( Y_{jkl} + r_2 \right) \log \left( \frac{Y_{jkl} + r_2}{Y_{jkl} + r_2} \right) + \frac{1}{\phi c_2(r_2 - r_1)} \left( Y_{jkl} + r_1 \right) \log \left( \frac{\mu_{jkl} + r_1}{Y_{jkl} + r_1} \right) - \left( Y_{jkl} + r_2 \right) \log \left( \frac{\mu_{jkl} + r_2}{Y_{jkl} + r_2} \right) + \frac{1}{\phi c_2(r_2 - r_1)} \left( Y_{jkl} + r_1 \right) \log \left( \frac{\mu_{jkl} + r_1}{Y_{jkl} + r_1} \right) - \left( Y_{jkl} + r_2 \right) \log \left( \frac{\mu_{jkl} + r_2}{Y_{jkl} + r_2} \right).

We evaluate the quasi-likelihood function at $B = 4$ which yields $c_0 = \frac{32}{27}$, $c_1 = \frac{8}{3}$, $c_2 = \frac{1}{6}$, $r_1 = 8 + 16\sqrt{2}$ and $r_2 = 8 - 16\sqrt{2}$ for our DPOAE data.

For our noncentral-F distributed mixed-effects model, we alternately estimate $\beta$ and $b$ via IWLS first (Breslow and Clayton (1993)), and then the parameters $\phi$, $\omega$ and thus $D_\omega$ are estimated by fitting a linear mixed effects model on the $\psi$ scale, which was introduced in Section 5.4.1. The algorithm is followed,

- Specify the initial values $\beta^{(0)}$, $b^{(0)}$, $D_\omega^{(0)}$ and $\phi^{(0)}$, and a tolerance error.
- For $t = 0, 1, 2, \ldots$

123
– Let \( \psi^{(t+1)} = X\beta^{(t)} + Wb^{(t)} \) and \( \mu^{(t+1)} = e^{\psi^{(t+1)}} \).

– Compute \( V = c_0 + c_1\mu^{(t+1)} + c_2(\mu^{(t+1)})^2 \) where \( c_0 = -B^2/((B-1)(4B-2)) \), \( c_1 = 4B/(B-1) \), and \( c_2 = (2B-2)^{-1} \).

– For each \( i \) compute:
  \[
  G_i = \left( \frac{\phi^{(t+1)}V(\mu_i^{(t+1)})}{(\mu_i^{(t+1)})^2} \right)^{-1};
  a_i = \psi_i^{(t+1)} + \frac{(y_i-\mu_i^{(t+1)})}{\mu_i^{(t+1)}}.
  \]

– Build the diagonal matrix \( G \) with the diagonal elements \( \{G_i\} \) and the vector \( a \) with \( \{a_i\} \).

– Compute \( R = G^{-1} + WD^{(t+1)}W' \).

– Compute the fixed effects parameters \( \beta^{(t+1)} = (X'R^{-1}X)^{-1}X'R^{-1}a \).

– Compute the random effects parameters \( b^{(t+1)} = D^{(t+1)}W'R^{-1}(a-X\beta^{(t+1)}) \).

– Estimate \( \phi^{(t+1)} \) and \( D^{(t+1)} \) by fitting a normal linear mixed effects model
  \( a = X\beta^{(t+1)} + Wb^{(t+1)} + e. \)

Repeat while \( \| \beta^{(t+1)} - \beta^{(t)} \| < \) tolerance error.

This algorithm is applied to the mixed effects models of our DPOAE data using the software package R. The matrix computation can be simplified due to the block diagonal features of the variance-covariance matrix, \( R \). In order to examine how well the mixed effects model works, the fixed effects model is also considered and then compared.

Residuals are very useful to assess the adequacy of fit of a model. For generalized linear models, some extended definition of residuals are used, such as Pearson residual, deviance residual and Anscomble residual. We had used deviance residual for the
binomial GLM in Section 5.3.2. Instead we consider Pearson residuals for analysis and diagnostics because some of the deviance residuals are not available. The Pearson residual, defined by

\[(r_P)_{jkl} = \frac{(Y_{jkl} - \hat{\mu}_{jkl})}{\sqrt{V(\hat{\mu}_{jkl})}},\]

is the raw residual scaled by the estimated standard deviation of \(Y_{jkl}\), and the Pearson \(X^2\) statistic is defined as

\[X^2 = \sum (r_P)_{jkl}^2,\]

which can be used as a goodness-of-fit statistic to do the model selection.

5.4.3 Generalized Linear Mixed Effects Model for the DPOAE data

In our experiment, let the observed data \(Y_{jkl}\) denote the \(l\)th replicate (\(l = 1, \ldots, 100\)) of the local \(F\) statistic for subject \(j = 1, \ldots, 15\), at evoking frequency \(f_2 = f_{2,k}\), \(k = 1, \ldots, 15\) (\(f_{2,k} = 500\) Hz, \(600\) Hz, \(\ldots\), \(900\) Hz, \(1\) kHz, \(2\) kHz, \(\ldots\), \(8\) kHz). Before studying the whole DPOAE dataset, we randomly selected one female subject and one male subject. The mean of the local \(F\) test statistic from 100 trials are calculated for these two subjects at each evoking frequency \(f_2\). From Figure 5.5, we can see that there still exists a certain quadratic pattern of the evoking frequency for both subjects, but the quadratic pattern is different between the female subject and the male subject. The DPOAE is difficult to detect for \(f_2\) at higher or lower frequencies. Thus we will try to fit a quadratic GLMM with the covariates, \(\log\) (to base 10) of \(f_2\) and gender.

In order to evaluate the noncentral \(F\) GLMM, we will compare the candidate GLMMs to the corresponding generalized fixed effects model for the DPOAE data.
first. Then we generate some noncentral $F$ distributed local $F$ test statistic data based on the estimated noncentral parameters from the DPOAEs to understand how well this methodology applies to the modeling of local $F$ statistic. Finally, we will evaluate the noncentral $F$ GLMM by comparing it with easy to fit Gaussian mixed effects model, using cross validation mean square errors.

**Comparison of Generalized Linear Mixed Effects Model and Generalized Linear Fixed Effects for the DPOAE Data**

Firstly, we find a good candidate of the fixed effects model. Our model for the estimated mean is

$$\log(\hat{\mu}_{jkl}) = \hat{\beta}_0 + \hat{\beta}_1 (\log f_2)_k + \hat{\beta}_2 (\log f_2)_k^2 + \hat{\beta}_3 (\text{gender})_j$$

(5.14)
Table 5.3: Pearson residual analysis table for independent model

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Difference of Pearson statistic</th>
<th>Df</th>
<th>Pearson statistic</th>
<th>adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>19299</td>
<td>17514</td>
<td>(log $f_2$)</td>
<td>19298</td>
<td>15194</td>
</tr>
<tr>
<td>(log $f_2$)$^2$</td>
<td>1</td>
<td>1837</td>
<td>19297</td>
<td>13357</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>760</td>
<td>19296</td>
<td>12597</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

where $\hat{\beta}_0 = -35.530$ (s.e. = 0.60), $\hat{\beta}_1 = 21.445$ (s.e. = 0.36), $\hat{\beta}_2 = -3.081$ (s.e. = 0.05) and $\hat{\beta}_3 = 0.307$ (s.e. = 0.01). The interaction term of $(\log f_2) \ast (\text{gender})$ and $(\log f_2)^2 \ast (\text{gender})$ appeared to be not jointly significant according to the Pearson $X^2$ statistic. Adjusted p-values were calculated considering the underdispersion parameter of $\hat{\phi} = 0.6528$ (significantly different from 1 with p-value is less than < 0.0001 ). From Table 5.3, it can be seen that all the terms are significant. Thus we deem that the model (5.14) is the best fixed effects model we have.

From the exploratory analysis, we know there exists variability in the local $F$ statistics that depends on gender, and it is different for different subjects at different frequencies. Thus a mixed effects model needs to be considered. Several mixed effects models are fitted, and the Pearson statistics are compared. All these models are based on the noncentral $F$ fixed effects model (5.14) with the log link and the quadratic term of log $f_2$, but they have different random effect terms. Here are the models we considered:

$$\log (\mu_{jkl}) = (\beta_0 + U_{jk}) + \beta_1 \log f_{2k} + \beta_2 (\log f_2)_{k}^2 + \beta_3 (\text{gender})_j,$$  \hspace{1cm} (5.15)

$$\log (\mu_{jkl}) = \beta_0 + (\beta_1 + U_{jk}) (\log f_2)_k + (\beta_2 + V_{jk}) (\log f_2)_k^2 + \beta_3 (\text{gender})_j,$$  \hspace{1cm} (5.16)
Table 5.4: random effect model

\[
\log(\mu_{jkl}) = (\beta_0 + U_{jk}) + (\beta_1 + V_{jk})(\log f_2)_k + (\beta_2 + W_{jk})(\log f_2)_k^2 \\
+ \beta_3(\text{gender})_j,
\]

(5.17)

\[
\log(\mu_{jkl}) = (\beta_0 + U_j + V_{jk}) + \beta_1(\log f_2)_k + \beta_2(\log f_2)_k^2 \\
+ (\beta_3 + W_j)(\text{gender})_j.
\]

(5.18)

We assume \( U_j \sim N(0, \sigma_U^2) \) and \( W_j \sim N(0, \sigma_W^2) \) for each \( j \), and \( U_{jk} \sim N(0, \tau_U^2) \), \( V_{jk} \sim N(0, \tau_V^2) \) and \( W_{jk} \sim N(0, \tau_W^2) \) for each \( k \) within \( j \). Here \( j = 1 \ldots 15 \) is the subject index, \( k = 1 \ldots 13 \) represents the frequency index, and \( l = 1 \ldots 100 \) is the replicate index. All these models tell us that the DPOAE amplitudes are different for the different subjects at the different frequencies, which matches what we found from the exploratory analysis. We can see that Model (5.15) has only a random intercept, which tells that the DPOAE amplitudes have the similar parallel pattern for the different subjects across the frequencies. Model (5.16) has a random slope for the different subjects at each evoking frequency. Model (5.17) has both a random intercept and a random slope for the evoking frequencies for each subject. Model (5.18) has a random intercept plus a random slope for females for each subject at each evoking frequency. We compare these four mixed effects models using the Pearson \( X^2 \) statistic.
From Table 5.4, Model (5.18) with random effects in the intercept and the gender terms has the smallest Pearson statistic and give us clear scientific meaning. The estimated parameters are $\hat{\beta}_0 = -36.67$ (s.e. $= 3.95$), $\hat{\beta}_1 = 22.07$ (s.e. $= 2.42$), $\hat{\beta}_2 = -3.18$ (s.e. $= 0.37$), $\hat{\beta}_3 = 0.28$ (s.e. $= 0.14$), $\hat{\sigma}^2_U = 0.024$, $\hat{\sigma}^2_W = 0.049$, and $\hat{\tau}^2_V = 0.322$. This is consistent with Robinette and Glattke (1997)[Chapter 5] who stated there exists a difference of the average DPOAE amplitudes in gender and frequency. This model also extends the conclusions of Robinette and Glattke and tells us there exist random intercepts for the different subjects at the different frequencies and that the variability of the random effects are modulated by gender (females are more variable). Hence it is reasonable to have the random intercept at the different frequencies and the random slope for the gender in the mixed effects model like Model (5.18). They also mentioned that the DPOAEs at the frequencies less than 1 kHz are difficult to measure. This might cause some unexpected variability at the lower frequencies. To explore this, we removed the data with frequencies less than 1 kHz and refit the mixed effects model. We explore this mixed effects model with diagnostic plots and compare it with the fixed effects model.

We now use residual plots as a straightforward graphical tool to check the fit of the model. There are a bunch of types of residual plots that could be applied. For example, we could show the residuals against the explanatory variables. In order to see if the model capture the variation among the subjects, it is convenient to draw the residuals against the evoking frequency $f_2$ for each subject respectively. This is also a good way to compare the fixed effects model with the mixed effects model because the advantage of the mixed effects model is to capture within-subject-variability of DPOAEs. Only using the fixed effects in the model, we estimate the average of the
DPOAE periodicity $F$ statistics for the subjects at each evoking frequency. If there exists the variation within the subjects across the evoking frequencies, the fixed effects part cannot capture this variation. The estimated values will be far away from the observed values, and the residuals will show systematic differences from zero. On the other hand, the random effects part should capture the variation within each subject across the evoking frequencies. We expect residual plots for the random effects model to be more spread evenly about zero.

Figure 5.6, 5.7, and 5.8 give the residual plots of the residuals against the evoking frequency $f_2$ for each subject under the fixed effects model (5.14), the mixed effects model (5.18) and the mixed effects model (5.18) fit to only the higher $f_2$ frequencies (1 kHz $\sim$ 8 kHz) respectively. In each figure, the solid line represents the median of the residuals of the 100 replicates, the dashed lines show the 25% and 75% of the residuals of the 100 replicates, and the dotted lines give the 5% and 95% of the residuals of the 100 replicates. Comparing Figure 5.6 with Figure 5.7, the smaller Pearson residuals are obtained from the mixed effects model. It can be seen that the pattern of the residuals are changed a lot (systematically different from zero) for each subject under the fixed effects model. The median of the residuals are close to zero under the mixed effects model. These tell us that the fixed effects model cannot capture the within-subject-variability, whereas the mixed effect model can. As expected the residuals follow the assumptions of the model better when we restrict to the higher evoking frequencies (Figure 5.8). Figure 5.9 and 5.10 show plots of the fitted values against log of the evoking frequency $f_2$ for each subject under the fixed effects model and the mixed effects model. In each figure, the solid line represents the median of the local $F$ test statistics of the 100 replicates, the dashed lines show the 25% and 75% of the
Figure 5.6: Residual plots for each subject in the DPOAE dataset in the fixed effects model (5.14) as a function of $\log f_2$. 

131
Figure 5.7: Residual plots for each subject in the DPOAE dataset in the mixed effects model (5.18) as a function of $\log f_2$. 
Figure 5.8: Residual plots for each subject in the DPOAE dataset in the mixed effects model (5.18) fit to frequencies over 1 kHz as a function of $\log f_2$. 

133
Figure 5.9: Plots of the fitted log local $F$ statistics for each subject in the DPOAE dataset in the fixed effects model (5.14) as a function of log $f_2$. 
Figure 5.10: Plots of the fitted log local $F$ statistics for each subject in the DPOAE dataset in the mixed effects model (5.18) as a function of log $f_2$. 
local $F$ test statistics of the 100 replicates, the dotted lines give the 5% and 95% of
the local $F$ statistics of the 100 replicates, and the solid line demonstrates the fitted
local $F$ test statistic by using the corresponding models. We can find that the fitted
local $F$ test statistics under the fixed effects model do not match the observed local $F$
statistics well at certain frequencies for some subjects. On the other hand, the fitted
local $F$ test statistics under the mixed effects model are very close to the median of
the observed local $F$ test statistics, which tells us the mixed effects model can capture
the variability in the gender, the subjects and the frequencies of the DPOAE data.

The Generalized Linear Mixed Effects Model for the Simulated Data

We also simulate some noncentral $F$ distributed local $F$ test statistics based on the
estimated parameters in the model, and apply the PQL method to fit the mixed effects
model. In this way, we know exactly that this simulated dataset has the variation
within each subject across the evoking frequencies. We want to see if the mixed
effects model (5.18) can capture the variation in this simulated dataset as expected
and also to simply verify the validity of the PQL method for this data. Figure 5.11,
and 5.12 give the residual plots and the fitted plots respectively. Again the solid line
represents the median line, the dashed lines show the 25% and 75% lines, the dotted
lines give the 5% and 95% lines, and the grey solid line demonstrates the fitted local
$F$ test statistics. We can see the residuals plots are fairly good, and the fitted local
$F$ test statistics match the simulated local $F$ statistics very well. Thus, this mixed
effects model capture the variability in the gender, the subjects and the frequencies
of the simulated data.
Figure 5.11: Residual plots for each subject in the mixed effects model (5.18) as a function of $\log f_2$ for the simulated dataset.
Figure 5.12: Plots of the fitted log local $F$ statistics for each subject in the mixed effects model (5.18) as a function of $\log f_2$ for the simulated dataset.
Comparison of the Noncentral $F$ Linear Mixed Effects Model and the log-Gaussian Linear Mixed Effects for the DPOAE Data

As far as we know, there is no existing software package that can fit the mixed effects model with a noncentral $F$ distribution. However, the log-Gaussian mixed effects model can be fitted easily in many statistical packages. Thus, we ask if the log-Gaussian model works well as a model for the DPOAE-based $F$ statistics as compared to the more complicated noncentral $F$-based mixed effects model. To answer this question we need to select a method to evaluate these two types of models. Motivated by how well the fitted model would perform at the new points, the cross validation method is considered to compare the log-Gaussian model with the noncentral $F$ model. Cross validation method involves building a new dataset by removing a single observation from the original data, fitting the model based on this new dataset, and yielding a prediction of the removed observation. The goodness of the match can be measured by the squared error between the removed observed value and the predicted value from the model in which the observation is removed. For the DPOAE data we remove one replicate randomly from each subject at each frequency. Then we fit the models and calculate the prediction errors for both the log-Gaussian and the noncentral $F$ mixed effects models. After repeating this procedure 1000 times, we calculate the mean and the variance of the mean square error (MSE) for both the log-Gaussian and the noncentral $F$ mixed effects models. We can see from Table 5.5 that the noncentral $F$ GLMM has consistently smaller cross validation mean squared errors than with the log-Gaussian mixed effects model. Thus, it is worth to fitting the noncentral $F$ mixed effects model.
Table 5.5: Cross validation results for the log-Gaussian and noncentral $F$ mixed effects model

<table>
<thead>
<tr>
<th>GLMM</th>
<th>$\mu_{mse}$</th>
<th>$\sigma^2_{mse}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncentral $F$</td>
<td>1.975</td>
<td>0.086</td>
</tr>
<tr>
<td>log-Gaussian</td>
<td>16.456</td>
<td>0.550</td>
</tr>
</tbody>
</table>

Discussion of the Noncentral $F$ Mixed Effects Model for the DPOAE Data

We should indicate that there exist some problems when using the noncentral $F$ GLMM with the PQL method for the DPOAE data. First of all, we are using the Pearson residual to do the model selection and diagnosing, though we know it is often skewed. We can calculate the deviance residual instead since we have already evaluated the quasi-likelihood function in this specific situation. However, it should be noticed that the quasi-likelihood exists if only if all the $Y_{jkl}$ are greater than $-r_2 = -8 + 16\sqrt{\frac{2}{7}} = 0.552$ in Equation (5.5). Some of our local $F$ test statistics from the DPOAE data are less than 0.552, which makes the deviance residuals unavailable. Hence, we need to find a way to adjust the $Y_{jkl}$. There might have two ways to consider. One is adding a constant to $Y_{jkl}$ to guarantee all the values are greater than 0.552. The other way is using a method used to adjust the deviance in Gamma-distributed observations (McCullagh and Nelder (1999)[Chapter 8]), which uses an arbitrary bounded function $C(Y)$, the function of $Y$, instead of $(Y_{jkl} + r_2) \log(Y_{jkl} + r_2)$ in the deviance function.

A second problem with our noncentral $F$ GLMM is that we assume the variances of the random effects are identical to all the subjects and the frequencies in the variance-covariance matrix $D_\omega$. But we can see from Figure 5.7, this is not quite
true. If we assume the unequal variance here, the normal linear mixed effects model does not work well to evaluate the variance-covariance matrix. This requires writing software to estimate the variance-covariance matrix.

A third issue in our approach is the bias of the standard error for the fixed effect and the random effect parameters. From Equation (5.8) and (5.9), we can estimate the standard errors for the fixed effects parameters $\hat{\beta}$ and the random effects parameters $\hat{b}$ by using the exact values from the normal linear model (following the idea of Breslow and Clayton (1993)) in our algorithm, which assumes that $\omega$ is known. Then both sets of the standard errors are biased because we ignore a certain variability without estimating the $\omega$. We know the PQL algorithm greatly underestimates the random effects standard derivation relative to the numerical ML estimators. Then we might consider EM method instead, or maybe the Bayesian method.
CHAPTER 6

DISCUSSION

In this thesis, our research work focuses on spectral-based tests for periodicities. We have explored the time-domain-based $F$ test and found that it is time consuming especially for the large sample size due to the large dimensional matrix calculation, and success of the testing procedure can not be guaranteed because of estimating the unknown error process. Then spectral-based tests have been studied in depth. The global $F$ test, the local $F$ test and Thomson’s multitaper test can be derived from the spectral regression model with the error term having an approximately diagonal covariance matrix. The distribution of the error term of the spectral regression model is based on the asymptotic distribution of the tapered Fourier transform of the error process. This asymptotic distribution has the approximately diagonal covariance matrix when the sample size is large and the SDFs of error processes have small dynamic range (closer to a constant). This does match the simulation result that the spectral-based tests have better size and power performance for the large sample sizes and the error processes with SDFs having small dynamic range (e.g., IID and AR(2) error processes). For the non-IID error processes, the local tests are more robust to the dynamic range of the SDFs because we use only the frequencies in the neighborhood of the test frequency. Then it is reasonable to assume they are
locally constant. For the non-IID error processes with SDFs having large dynamic range, the asymptotic distribution of the tapered Fourier transform does not have an approximately diagonal matrix due to leakage in the spectral estimates. A cosine taper can be applied or Thomson’s multitaper test can be used to reduce the effect of leakage. In general the spectral-based tests derived from the spectral regression model work well if the covariance of the spectral regression model is approximately diagonal.

A local test uses only part of the spectral estimates, whereas a global test uses all the information. Then the local test is notionally less efficient than the global test. The performance of the global test, the local $F$ test, the Gorga et al. test and Thomson’s $F$ test are investigated under both the IID and the non-IID error processes. We found that the local tests can achieve the similar performance with a reasonable choice of sample size and bin width, using much less information than the global $F$ test. The local tests are robust to the nuisance periodicity and the dynamic range of the error process as long as the nuisance frequency is outside the local region used in the test. For the non-IID error processes, the smoothing spline is used to estimate the background SDF before carrying out the global $F$ test, which is not stable and accurate especially for the unknown nuisance frequencies. Thus, a local test is preferred – the local $F$ test has slightly better performance than the Gorga et al. test for this case. Thomson’s multitaper test is less powerful than the other spectral-based tests (no tapering) when the SDFs of error processes have small dynamic range. These can be treated as the criteria to select the spectral-based test for periodicity.
In our research, all the derivations and simulations are based on the fact that the periodicities to test for along with the nuisance periodicities are at Fourier frequencies. The performance of these spectral-based tests will be affected when the frequencies do not line up with Fourier frequencies. More work could be done if the test frequency and the nuisance frequencies are not Fourier frequencies. We also assume the case that the phase is zero in our simulations. We could explore the performance of the spectral-based tests when the phase is different from zero.

We applied our spectral-based tests to the field of Speech and Hearing Science. We used the logistic regression models to model the binary normal-abnormal responses and the noncentral $F$ GLMMs to model the local $F$ test statistic in terms of the evoking frequency $f_2$ and gender. From the logistic regression model, we can see there exists the downward pattern in all the deviance plots and the normality assumption for deviance residuals is violated. We could use the random effects logistic model (or adapt the covariance matrix of the error term) to capture the many subject and frequency specific effects. For the noncentral $F$ GLMM, there still exist some problems and this needs to be improved. Firstly, the Pearson residual is not a good choice to do the model selection and diagnostics because it may be skewed. Also the deviance residual is not available since the deviance could be negative, which makes no sense to the square root of the deviance. We might adjust either the observed data or the deviance to guarantee that the deviance is nonnegative (e.g., McCullagh and Nelder (1999)[Section 8.3.2]). Secondly, it is not quite true that we assume the variances of the random effects are identical to all the subjects and frequencies in the variance-covariance matrix. We could use more powerful software or write a function to estimate the variance-covariance matrix. Finally the bias of
the standard error exists for the fixed effect and random effect parameters because
the PQL algorithm greatly underestimates the random effects standard derivation
relative to the numerical ML estimators. Then we might consider an EM method
instead, or maybe a Bayesian method.

In our research we only apply the spectral-based tests for periodicity to the
DPOAE dataset in the field of Speech and Hearing Science. The spectral-based tests
for periodicities are also interesting to many other fields, such as Economics, Biology,
Psychology, Medicine, Engineering, and Acoustics. When we apply the spectral-based
tests for periodicity in practice, we need to explore the underlying properties of the
noise process (e.g., dynamic range) and the periodicities (e.g., unknown nuisance peri-
odicities) as well as the sample size. Then a certain spectral-based test can be chosen
based on the information summarized in this thesis.
APPENDIX A

NONCENTRAL-F DISTRIBUTION

Here we provide, without proof, some elementary properties of the noncentral-F distribution. For a further summary of results regarding this distribution, see Johnson and Kotz (1970)[Chapter 26]. Let $U$ have a noncentral chisquare distribution with $a$ degrees of freedom (df) and noncentrality parameter $\eta$ (we use the notation $U \sim \chi^2_a(\eta)$), which is independent of a central chisquare distribution $V$ with $b$ df. Then the random variable defined by

$$Y = \frac{U/a}{V/b}$$

is said to the noncentral-F distribution with $a$ and $b$ df and noncentrality parameter $\eta$. The density of $Y$ is

$$f(y|\eta) = \sum_{r=0}^{\infty} e^{-\eta} \frac{\eta^r}{r!} \frac{\Gamma(a+b+r)}{\Gamma(a+r)\Gamma(b)} \frac{a^{a+r} b^b y^{a+r-1}}{(a+by)^{a+b+r}}.$$

The mean and variance of $Y$ are

$$\mu = E(Y) = \frac{b(\eta + a)}{a(b-2) \text{ and } \text{var}(Y) = \frac{2b^2[(a + 2\eta)(a + b - 2) + \eta^2]}{a^2(b-2)^2(b-4)},}$$

respectively. Expressed in terms of the mean, the variance of $Y$ is, $c_0 + c_1\mu + c_2\mu^2$, where $c_0 = -2b^2/(a(b-2)(b-4))$, $c_1 = 4b/(a(b-4))$, and $c_2 = 2/(b-4)$. 

146
APPENDIX B

NOTATION

\(a\) constant
\(\hat{a}\) estimate of \(a\)
\(a\) working vector
\(\{A_t\}\) unknown amplitudes of periodicities
\(b\) degrees of freedom for a chi squared random variable
\(\hat{b}\) estimate of \(b\)
\(b\) random effects vector
\(\hat{b}\) estimate of random effects vector
\(B\) bin width
\(\text{cov}(\cdot)\) covariance
\(C(\cdot)\) arbitrary bounded function
\(C\) matrix for hypothesis testing
\(\mathbb{C}\) the set of complex numbers
\(C_{td}\) matrix of hypothesis testing in the time domain
\(D\) deviance
\(D_N(\cdot)\) Dirichlet’s kernel
\(D_\omega\) covariance matrix
\(\{e_t\}\) a set of IID Gaussian processes
\(e\) error vector
\(E(\cdot)\) mean
\(f\) Fourier frequency
\(f(\cdot)\) a pdf/pmf for an exponential family
\(f_1\) lower evoking frequency in DPOAE
\(f_2\) higher evoking frequency in DPOAE
\(f_\Delta\) Nyquist frequency
\(F_t(\cdot)\) cumulative probability distribution function
\(F\) \(F\) test statistic
\(F_T\) Thomson’s \(F\) statistic
\(g(\cdot)\) link function
\(g'(\cdot)\) first derivative of \(g(\cdot)\)
\(G\) diagonal matrix with diagonal terms \(G_i = \{\phi V(\mu_i)[g'(\mu_i)]^2\}^{-1}\)
\{G_i\} diagonal elements of \( G \)
\{h_t\} taper
\( H \) hat matrix
\( H(\cdot) \) transfer function of taper
\( i \) square root of negative one
\( I \) identity matrix
\( \Im(\cdot) \) imaginary part of a complex value
\( j \) index
\( J_X(\cdot) \) tapered Fourier transform of \( X_t \)
\( J_Z(\cdot) \) tapered Fourier transform of \( Z_t \)
\( \mathbf{J}_Z \) vector of tapered Fourier transform of \( Z_t \)
\( \mathbf{J}_X \) vector of real and imaginary parts of tapered Fourier transform of \( X_t \)
\( \mathbf{J}_{Z,N} \) vector of real and imaginary parts of tapered Fourier transform of \( Z_t \)
\( k \) index
\( k(N) \) a sequence depending on \( N \)
\( L \) number of periodicities
\( l \) index
\( l(\cdot) \) log-likelihood function
\( L_B(\cdot) \) average of \( B \) Fourier frequencies to the left of the test frequency
\( m \) number of points of spectral estimate used to carry out the test
\( M_{td} \) design matrix in the time domain
\( N \) sample size
\( p \) probability of binomial distribution
\( P \) power
\( Q \) quasi-likelihood function
\( r_D \) deviance residual
\( r_P \) Pearson residual
\( \mathbb{R} \) the set of real numbers
\( \mathbf{R} \) \( \mathbf{R} = \mathbf{G}^{-1} + \mathbf{W} \mathbf{D} \omega \mathbf{W}^T \)
\( R_B(\cdot) \) average of \( B \) Fourier frequencies to the right of the test frequency
\( \Re(\cdot) \) real part of a complex value
\( \{s_{Z,T}\} \) ACVS
\( \{\hat{s}_{Z,T}\} \) standard estimates of ACVS
\( S(\cdot) \) spectral density function
\( \hat{S}^{(p)}(\cdot) \) periodogram
\( \hat{S}^{(d)}(\cdot) \) direct spectral estimator
\( \hat{S}^{(mt)}(\cdot) \) multitaper spectral estimator
\( t \) time index
\( T \) test statistic
\( U(\cdot) \) score function
$U_t$ random effect
$V(\cdot)$ variance function
$V_t$ random effect
$\mathbf{V}(\cdot)$ variance function matrix
$\text{var}(\cdot)$ variance
$W$ bandwidth
$W_t$ random effect
$\mathbf{W}$ explanatory variables matrix associated with the random effects
$X^2$ Pearson test statistic
$\mathbf{X}$ design matrix, explanatory variables matrix associated with the fixed effects
$\{X_t\}$ harmonic process
$\mathbf{X}_{td}$ vector of observed responses in the time domain
$\{y_t\}$ observed data
$\{Y_i\}$ random variables
$\mathbf{Y}$ vector of response variables
$\mathbb{Z}$ the set of integers
$\{Z_f\}$ Fourier transform of $\{Z_t\}$
$\{Z_t\}$ real-valued discrete parameter stochastic process
$\mathbf{Z}_{td}$ vector of errors in the time domain
$\alpha$ level of a test
$\mathbf{\beta}$ vector of parameters, fixed effects vector
$\hat{\mathbf{\beta}}$ estimate of vector of parameters, estimate of fixed effects vector
$\mathbf{\beta}_{td}$ vector of parameters in the time domain
$\hat{\mathbf{\beta}}_{td}$ estimate of vector of parameters in the time domain
$\Delta$ sampling interval
$\varepsilon$ error term
$\eta$ Fourier frequency
$\eta'$ Fourier frequency
$\gamma$ Euler-Mascheroni constant
$\lambda$ noncentrality parameter of noncentral $F$ distribution
$\mu$ mean
$\mu$ mean vector
$\mu$ estimate of $\mu$ under the full model
$\hat{\mu}$ estimate of $\mu$ under the reduced model
$\mathbf{\Omega}$ diagonal matrix with diagonal terms $\{\phi V(\mu_i) g'(\mu_i)\}^{-1}$ ($i = 1, \ldots, n$)
$\phi$ dispersion parameter
$\hat{\phi}$ estimate of dispersion parameter
$\{\phi_j\}$ AR coefficients
$\{\phi_l\}$ unknown phases
$\{\psi_j\}$ constants for causality of ARMA model
$\phi$ linear predictor
$\sigma^2$ variance
$\Sigma$ true covariance matrix
$\Sigma^c$ covariance matrix of $J_f$
$\Sigma_{td}$ covariance matrix in the time domain
$\hat{\Sigma}_{td}$ estimate of covariance matrix in the time domain
$\tau$ lag value in the time domain
$\tau_{\eta}$ lag value in the frequency domain
$\{\theta_j\}$ MA coefficients
$\theta_{td}$ covariance parameters in the time domain
$\hat{\theta}_{td}$ estimate of covariance parameters in the time domain
$\{\xi_l\}$ fixed known frequencies of periodicities
BIBLIOGRAPHY


