Bayesian Tapering Test for Comparing Two Spectral Densities with Application to EEG Data

Chenyi Pan

M.S., University of Virginia, United States, 2012

B.S., Shanghai University of Finance and Economics, P.R. China, 2011

A Dissertation Presented to the Graduate Falculty

of the University of Virginia in Candidacy for the Degree of

Doctor of Philosophy

Department of Statistics

University of Virginia

May, 2016

Abstract

Human brain is still a mystery and finding the possible relationships between different parts of the brain is an important and unsolved topic. Our research provides a new perspective to explore this mystery by looking at the Electroencephalography (EEG) data recorded when the subject is performing cognitive task or gambling task. EEG is the multichannel recordings of the electrical activities generated by collections of neurons within the brain. Different channels reflect the activity within different brain regions. Motivated by the EEG dataset from Reward Two-back Continuous Performance Task, a novel methodology of Bayesian tapering test and corresponding Bayesian multiple testing procedures for assessing whether two independent stationary time series have same spectral density has been proposed. Tapering test based on the raw log-periodograms, Fourier transformation of the log-periodograms, and "optimal" kernel smoothed log-periodograms as estimates for log-spectrum are explored. The tapering testing procedures when the replication exists is further investigated. The formal setup of asymptotic performance of models are deduced based on "rate-of-testing" theory, a framework to find the rate at which the power is retained under geometric smoothness constraints. As a result, the "optimal" kernel smoothing model and corresponding tapering test procedures are recommended wether the replication exists or not. Bayesian testing procedures are further explored due to its practical advantages. And the Bayesian multiple testing procedure based on the closed form of Bayesian tapering test is explored. Furthermore, the empirical power of the newly proposed tapering test are demonstrated through a comprehensive simulation study. The proposed "optimal" tapering test is found to be generally more powerful than the existing tests. Finally, the operation of the proposed Bayesian tapering test and Bayesian multiple testing procedure is demonstrated on the motivating scalp EEG (EEG) and the Stereotactic EEG (SEEG) dataset. Clinical meaningful results are found, which help us to gain more insights into the brain activities.

KEY WORDS: electroencephalography (EEG); spectral density; tapering test; kernel smoothing; bandwidth; rate of testing; Bayesian testing; Bayesian multiple testing.

Acknowledgements

It is a genuine pleasure to press my deep sense of thanks and gratitude to all people who helped me and have contributed in making this dissertation research possible.

Firstly, I want to thank my advisor Prof. Spitzner. He is so nice, patient and smart. He always think of his students and try to help his students both in research and in life. He can always provide me suggestions of which direction I should go whenever I am confused. He can always answer my questions whenever I get stuck. He always encourage me and give me confidence whenever I am upset. Every time I felt too stressful and wanted to give up, I could gain confidence and courage from him. He is such a good advisor that he can always find the potential of students and encourage them to achieve their best. Under his supervision, I finally understand the true meaning of the training in PHD program. I am so proud that I can be one of his students.

I also want to thank my co-advisor Prof. Keenan. He provided a lot of very important inputs in our real data application part. I kept bothering him with a lot of different questions. He is so patient that he always answer my questions and share with me his ideas. I really appreciate his time and efforts in trying to help me.

I really appreciate Prof. Jianhui Zhou's help in his courses I have taken and being my committee member. He has solid background in both theoretical and practical parts of statistics. I was benefited a lot from the classes he taught including probability and theory, longitudinal data analysis, categorical data analysis etc. He also provided a lot of suggestions and helps in this dissertation research.

I am especially grateful to Prof. Guofen Yan in the Department of Biostatistics in School of Medicine. I started to work as the research assistant under her supervision from my second year of PHD program. She is more like a friend or a family member to me. She taught me a lot of things not only in how to apply statistical methods to help people in real world, how to make myself more outstanding in job market, but also in how to manage my life. She played an important role in my PHD life and I benefitted so much from the collaborative work with her. She also provided a lot of very valuable suggestions to my dissertation research and encouraged me a lot.

I also got a lot of helps from many faculties in the Department of Statistics. I want to give

the sincere acknowledgement to Prof. Karen Kafadar, Prof. Tingting Zhang, Prof. Chao Du, Prof. Jeff Holt. They provided many suggestions to my dissertation research and helped me a lot during my PHD program. I also want to express my special thanks to Karen Dalton, our administrative. She organized the whole department more like a family; she dealt with every trivial or big thing regarding to our undergraduate and graduate program; she provided sounded supports for every student in our department, etc.

Finally, I want to thank my family and all of my friends, who always be there for me. I have so many friends that they always provide me sounded supports no matter what happens. Especially, I want to thank my parents. They give me their best love and provide me the best study opportunity. Even I am their only child, even I didn't go back home to visit them for more than two years, even they miss me so much that they cannot go to sleep every night, they still ask me to do not worry about them and encourage me to do my best in my PHD program. The luckiest thing in my life is to have my parents.

Contents

Chapter 1

Introduction

1.1 Description of the Testing Problem

In the research work presented here, we developed a new non-parametric testing approach for comparing the spectral density of two independent stationary time series, or equivalently, their auto-covariance functions. Firstly let me introduce our testing problem in details.

Let $n_1, n_2 \in \mathbb{Z}^+$ with $n_1 \leq n_2$ and consider two independent stationary real-valued time series, $\{X_{1,t_1},\, t_1 = 1, \cdots, n_1\}$ and $\{X_{2,t_2},\, t_2 = 1, \cdots, n_2\}$. We are interested in testing the hypothesis of two time series process $\{X_{1,t_1}\}$ and $\{X_{2,t_2}\}$ have the same spectral density function can be written as

$$\begin{aligned} H_0 :&\quad f_1(\omega) = f_2(\omega), &&\text{for all } \omega \in (0, \pi) \\ H_1 :&\quad f_1(\omega) \neq f_2(\omega), &&\text{for some } \omega \in (0, \pi) \end{aligned} \tag{1.1}$$

where $f_1(\omega)$ and $f_2(\omega)$ are spectral density functions of these two time series process.

If $\{X_t\}$ is a stationary time series with absolutely summable auto-covariance function $\gamma(.)$, $\{X_t\}$ has a continuous spectral density (power spectrum) given by

$$f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}, \quad \omega \in [-\pi, \pi],$$

where ω is the frequency, $\gamma(h)$ is the auto-covariance function, and $e^{-ih\omega}$ denotes complex exponentiation. And note that the definition of spectral density indicates that it is the Fourier transform

of the auto-covariance sequence.

Given a stationary stochastic process $\{X_t\}$, the auto-covariance is a function that gives the covariance of the process with itself at pairs of time points, which is given by

$$\gamma(h) = E\left[\left(X_t - \mu_t\right)\left(X_{t+h} - \mu_{t+h}\right)\right],$$

where $\mu_t = E(X_t)$, $\mu_{t+h} = E(X_{t+h})$.

Function f is the spectral density of a real-valued stationary process if and only if the following conditions are satisfied,

(1) $f(\omega) = f(-\omega)$

(2) $f(\omega) \geq 0$

(3) $\int_{-\pi}^{\pi} f(\omega) d\omega < \infty$

Since spectral density is unknown, we need certain methods to estimate it. The most convenient way to estimate it is to use periodogram. Suppose we have finite number of observations $\{x_t, t = 1, \cdots, n\}$ of the stochastic process $\{X_t\}$. The periodogram of $\{x_t, t = 1, \cdots, n\}$ is defined at the Fourier frequencies $\omega_j = 2\pi j/n, \quad \omega_j \in [-\pi, \pi]$, by

$$\begin{aligned} I(\omega_j) &= \frac{1}{n}\left|\sum_{t=1}^{n} e^{-it\omega_j} X_t\right|^2 \\ &= \sum_{|h|<n} \widehat{\gamma}(h) e^{-ih\omega}, \quad \omega \neq 0 \end{aligned}$$

And the estimated auto-covariance function $\widehat{\gamma}(h)$ is defined by

$$\widehat{\gamma}_k(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_{k,t} - \overline{x}_k\right)\left(x_{k,t+h} - \overline{x}_k\right)),$$

where $\overline{x}_k = \frac{1}{n}\sum_{t=1}^{n} x_{k,t}$ is the sample mean.

From the definition, we can find that the periodogram is simply the discrete Fourier transform of the biased estimator of the auto-covariance sequence, that's why periodogram can be regarded as an estimator of spectral density. Actually, periodogram is the unbiased but non-consistent estimate of spectral density, see Priestley (1981) [20], Brockwell and Davis (1991) [6], and Brillinger (2001) [5].

Additionally, periodogram is a symmetric and periodic function of ω with period 2π, and thus we only need to look at $\omega \in (0, \pi]$. And by Theorem 10.3.2 in Brockwell and Davis (1991) [6],

the random variables $I(\lambda_l)/(\pi f(\lambda_l))$ are **approximately independent** and distributed as chi-squared with 2 degree of freedom among $l = 1, \cdots, m$ if $0 < \lambda_1 < \cdots < \lambda_m < \pi$. Note that $\lambda = \pi$ is excluded since its sampling distribution is proportional to $\chi^2(1)$ not $\chi^2(2)$.

For simplicity assume n_1 and n_2 are even, let's focus on $\omega_{k,j} \in (0, \pi)$ i.e. $j = 1, 2, \cdots, \lfloor \frac{n_k}{2} \rfloor$ and define $p_n = \frac{n_1}{2} - 1$ as the dimensionality parameter. Therefore, for $j = 1, \cdots, p_n$,

$$I_k(\omega_{1,j}) \xrightarrow{D} \pi f_k(\omega_{1,j}) \chi^2(2)$$

where $I_2(\omega_{1,j}) = \frac{1}{n_2} \left| \sum_{t_2=1}^{n_2} e^{-it_2\omega_{1,j}} X_{2,t_2} \right|^2$.

1.2 Overview of the Methodological Development

The comparison and clustering of different time series is an important topic in statistical data analysis and has various application. Our method to identify similarities or dissimilarities between two process is to compare the entire auto-covariance structure of two time series, which can be effectively done in the frequency domain by comparing their spectral characteristics. In this context frequency domain methods are more appealing and related procedures have found considerable interest in the literature. Such problems were posed by Coates and Diggle (1986) [8], who proposed periodogram based tests and studied the homogeneity of a single wheat price series over time and compared the wall thicknesses of a gas pipe at two different locations. This paper included two nonparamtric tests. However, one test, is based on the range of peridogram ratios with the test statistic

$$R := \max \left\{ \log \frac{I_1(\omega_j)}{I_2(\omega_j)} \right\} - \min \left\{ \log \frac{I_1(\omega_j)}{I_2(\omega_j)} \right\},$$

is weaker whereas the other, based on cumulative sums of transformed periodogram ratios depends on the arbitrary labelling of the two series $\{X_{1,t}\}$ and $\{X_{2,t}\}$. That is, the achieved significance level does not remain the same if the series $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are interchanged. Note that $I_1(\omega_j)$ and $I_2(\omega_j)$ are periodograms for these two time series and ω_j is the Fourier frequency, and we will define those concepts latter on. In this paper, they also proposed a semiparametric procedure based on model that the underlying spectral densities $f_1(\omega)$ and $f_2(\omega)$ are related via the equation

$$f_2(\omega) = f_1(\omega) exp(\alpha + \beta\omega + \gamma\omega^2)$$

However, in general, a semiparametric approach is likely to give a more powerful test when the parametric assumptions are approximately valid and vice versa. We believe that there is merit in using a nonparametric procedure at least in the exploratory phase of the data analysis.

Diggle and Fisher (1991) [10] use graphical devices to compare cumulative periodograms and apply Kolmogorov-Smirnov or Cramer-von Mises type test statistics based on empirical spectral distributions, i.e.

$$D_m = \sup \mid F_1(\omega) - F_2(\omega) \mid$$
$$F_1(\omega_j) = \sum_{i=1}^{j} I_1(\omega_i) \,/\, \sum_{i=1}^{p_n} I_1(\omega_i)$$
$$F_2(\omega_j) = \sum_{i=1}^{j} I_2(\omega_i) \,/\, \sum_{i=1}^{p_n} I_2(\omega_i).$$

They applied the methods to test whether the frequency characteristics of the pulsatile release pattern for luteinizing hormone (LH) concentrations in blood are the same in two phases of the subject's menstrual cycle. But both for Coates and Diggle (1986) [8] and Diggle and Fisher (1991) [10], their interests are in detecting shape rather than scale differences between the two underlying spectra. Thus, the hypothesis are

$$H_0: \qquad f_1(\omega) = f_2(\omega), \qquad 0 < \omega \leq \pi$$
$$H_1: \qquad f_1(\omega) = \kappa f_2(\omega), \qquad 0 < \omega \leq \pi$$

Lund et al. (2009) [18] considered both time domain and frequency domain tests to assess whether two stationary and independent time series have the same auto-covariances at all lags. They reviewed frequency domain tests based on the average of log ratio of periodograms as well as the likelihood principles. and also devises a time domain approach similar to Caiado et al. (2006) [7], based on the asymptotic distribution of the estimated auto-covariance function. But time-domain tests have less satisfactory performance and merit further exploration. The frequency domain test that we will compare with in our simulation study is defined through the test statistic

$$\overline{D} := \frac{1}{(n/2-1)} \sum_{j=1}^{n/2-1} \left| \log \frac{I_1(\omega_j)}{I_2(\omega_j)} \right|.$$

They applied these testing procedures to the analysis of temperatures and precipitations in Atlanta and Athens in Georgia in order to identify a good climatological reference series for given stations. Multivariate version of these tests are also considered in this paper.

Recently Lu and Li (2013) [17] considered tests for assessing whether two stationary and independent time series have the same spectral densities (or same auto-covariance functions). Both frequency domain and time domain test statistics for this purpose are reviewed. The adaptive Neyman test (ANT) introduced by Fan (1996) [11] are then applied to the log-ratio periodogram test and time domain test in Lund et al. (2009) [18] and their performances are investigated. The frequency domain test that we will compare with in the simulation study is defined through the test statistic

$$\begin{aligned}
\overline{LL}^{\star} &:= \max_{1\le k\le n_m} \frac{1}{\sqrt{k\hat{\sigma}_2^2}} \sum\nolimits_{i=1}^{k} \left(\left(\bar{D}^{\star}_{m,i}\right)^2 - \hat{\sigma}_1^2 \right) \\
\overline{D}_{m,i} &= ln \frac{m^{-1} \sum\limits_{k=(i-1)m+1}^{im} I_1(\omega_k)}{m^{-1} \sum\limits_{k=(i-1)m+1}^{im} I_2(\omega_k)} \quad i = 1, \cdots, N_m \\
\overline{D}^{\star}_{m,i} &= \frac{1}{N_m} \sum_{i=1}^{N_m} \overline{D}_{m,i} \psi_{j,i} \\
\widehat{\sigma}_1^2 &= \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}^{\star}_{m,i}\right)^2 - \left\{ \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \overline{D}^{\star}_{m,i} \right\}^2 \\
\widehat{\sigma}_2^2 &= \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}^{\star}_{m,i}\right)^4 - \left\{ \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}^{\star}_{m,i}\right)^2 \right\}^2 ,
\end{aligned}$$

where $\psi_{j,i}$ is the Fourier basis functions as defined before, $N_m = \left[\frac{n}{m}\right]$, $I_{N_m} = \left[\frac{N_m}{4}\right]$.

However, their definition of tests and simulation study are problematic and lack of theoretical derivation especially for the time domain test, we will discuss more about this in Chapter 2.

There are a number of tests in the literature on this topic with various applications, for instance, Kakizawa et al. (1998) [15] studied whether seismological series were more likely earthquakes or nuclear tests based on discrimination and clustering; Shumway (2003) [25] considered similar classification problems related to functional magnetic resonance image series; Quinn (2006) [21] tested signal equality by fitting autoregressive models to the series and then comparing model coefficients. The following authors have considered spectral-based extensions to locally stationary and non-stationary series: Huang et al. (2004) [12], Shumway and Stoffer(2013) [26], and Bengtsson and Cavanaugh (2008) [3]. Our work concentrates on the stationary case as there is much to be done even in this setting.

However, those existing methods in literatures either suffer from low power or cannot be directly apply to our context. Therefore, the main objective of our research is to propose a powerful test for comparing two estimated spectral densities that can easily fit into our context of EEG dataset. **Note that this objective is twofold, primarily, we want powerful testing procedure; secondly, we want a procedure which is very convenient to implement in practice.** The proposed approach is particularly inspired by techniques described in Lu and Li (2013) [17], in which a test for comparing spectral densities is developed from the ANT of Fan (1996) [11]. The original context of the ANT is goodness-of-fit testing, nonparametric regression, and functional data analysis, which is the same context in which Spitzners (2008) [28] tapering test is developed. In that context, Spitzner compares the operating characteristics of various procedures, including the ANT, and identifies broad, practically relevant parameter configurations under which tapering is expected to exhibit superior performance. Our original interest in comparing spectra thus stemmed from curiosity in how well Spitzners tapering test would adapt to the time-series context.

1.3 Overview of Spectral Density Estimation Methods

Inspired by the method proposed by Coates and Diggle (1986) [8], tests proposed in our research were built on the sequence of

$$log\left(\widehat{f}_1(\omega_j)\right) - log\left(\widehat{f}_2(\omega_j)\right) \tag{1.2}$$

Then, the challenge comes as to find proper estimate of the spectral density $\widehat{f}(\omega_j)$.

A natural estimate of $f(\omega_j)$ for $\omega_j \neq 0$ is the periodogram $I(\omega_j)$, which is unbiased estimator of $f(\omega)$ for $\omega \neq 0$ and $\omega \in [-\pi, \pi]$, as defined in Section 1.1. However, the raw periodogram as the estimate of power spectrum suffers from several problems. First of all is the inefficiency. $I(\omega_j)$ is not a consistent estimator of $f(\omega)$. If we add more samples to the signal, or increase the sampling rate, the estimate are not improved since this will only increase the frequency domain resolution of estimate. Additionally, it suffers from high variance. See for example Priestley (1981)[20], Brockwell and Davis (1991) [6], and Brillinger (2001) [5]. There are three different ways to gain a consistent estimate: **local average smoothing**; lag window smoothing and periodogram averaging (Welch's-Bartlett spectral estimation method). But the Welch's-Bartlett method lead to the spectral leakage and bias. The most popular approach is to directly smooth on the periodogram via a local average,

which is equivalent to the kernel smoothing via discrete compact kernel form. For considering of simplicity and the computation time, only local average smoothed periodogram is considered in this paper as the estimates for spectral density. The periodograms are smoothed by averaging in a small neighborhood of ω. Since the number of Fourier frequencies in a given interval increases approximately linearly with n, we can indeed construct consistent spectral density estimators by averaging over a suitably increasing number of frequencies in a neighborhood.

The procedure of local average smoothing can provide consistent estimates, however, it increases the bias. When we have enough data, we may obtain the spectral estimate with satisfactory bias and variance. When the data are limited, a tradeoff between bias and variance is inevitable. Perhaps the first non-parametric method to address simultaneously the issues of bias and variance in an optimal fashion was the multitaper spectral estimation technique developed by David Thomson (1982) [31]. Instead of local average smoothing, the Multitaper Method (MTM) reduces estimation bias by obtaining multiple independent estimates from the same sample. Assume $\mathbf{X}(t), t = 0, \pm 1, \cdots$ is a stationary r vector-valued series with mean vector μ_X. Components of $\mathbf{X}(t)$, i.e. $X_j(t), j = 1, \cdots, r$ have all moments exist, and have absolutely summable auto-covariance functions, $\sum |\gamma_{ab}| < \infty, a, b = 1, \cdots, r$. Define the orthonormal tapers $h_k(t), t = 1, \cdots, n-1$ which is bounded , of bounded variation and vanishes for $t > T$. And the Discrete Fourier Transform (DFT) of the tapered series $h_k(t)\mathbf{X}(t), t = 1, \cdots, n-1$ can be constructed as,

$$d_a^k(\lambda) = \sum_{t=0}^{n-1} h_k(t) X_a(t) exp(-i\lambda t), \quad -\infty < \lambda < \infty, \quad \text{for} \quad a = 1, \cdots, r.$$

Each data taper is multiplied element-wise by the signal to provide a windowed trial from which one estimates the power at each component frequency. As each taper is pairwise orthogonal to all other tapers, the windowed signals provide statistically independent estimates of the underlying spectrum. Let r=2, define the eigen-spectra for component series $X_a(t), a = 1, \cdots, r$ of $\mathbf{X}(t)$ based on the taper $h_k(t)$, as

$$\widehat{f}_a^k(\omega_j) = \frac{1}{2\pi n} d_a(\omega_j) \overline{d_a(\omega_j)}, \quad a = 1, 2.$$

Then the multitaper spectral density estimate is defined as the weighted average of eigen-spectra,

$$\widehat{f}(\omega_j) = \frac{\sum_{k=1}^K u_k \widehat{f}_k(\omega_j)}{\sum_{k=1}^K u_k},$$

where u_k are weights corresponding to each eigen-spectra. The final spectrum is obtained by averaging over all the tapered spectra. Averaging over this (small) ensemble of spectra yields

a better and more stable estimate, i.e., one with lower variance than do single taper methods. Thomson chose the Slepian or discrete prolate spheroidal sequences as tapers since these vectors are mutually orthogonal and possess desirable spectral concentration properties. Note that the case with only one set of tapers reduces to the trivial Blakman-Tukey case of a single tapered DFT. MTM offers the appeal of being nonparametric, in that it does not prescribe an a priori (e.g., autoregressive) model for the process generating the time series under analysis. Because the windowing functions or eigentapers are the specific solution to an appropriate variational problem, this method is less heuristic than traditional nonparametric techniques.

To derive a powerful testing procedure, in Section 2.1, we will firstly introduce the basis model based on the raw periodogram (log-periodogram) as the estimate of power spectrum (log-spectrum). However, this model suffers from issues since periodogram or log-periodogram is not a consistent estimator. Then we will explore the Fourier transform of the log-periodogram as estimates for log-spectrum and proposed the Model 2 in Section 2.2, inspired by Lu and Li's (2013) [17] paper. By Fourier transformation, we achieved a vanishing variance, which makes it possible to even think about the asymptotic performance based on rate of testing criteria. However, Fourier transform of log-periodogram as estimates for log-spectrum doesn't have a very meaningful interpretation in time series context, instead, kernel smoothing of the periodogram (log-periodogram) is more acceptable. See for example Priestley (1981)[20], Brockwell and Davis (1991) [6], and Brillinger (2001) [5]. There are three different ways to smooth the periodogram (log-periodogram): local average smoothing, lag window smoothing and periodogram averaging (Welch's-Bartlett spectral estimation method). But the Welch's-Bartlett method lead to the spectral leakage and bias. Considering that the objective of our research is to develop a powerful testing procedure which is convenient to use even for people with little statistical background, the simple moving averaged log-periodogram as the estimate for log-spectrum will be used, which is equivalent to do kernel smoothing with the Uniform Kernel. Since the number of Fourier frequencies in a given interval increases approximately linearly with n, we can indeed construct consistent log-spectrum estimators by averaging over a suitably increasing number of frequencies in a neighborhood.

1.4 Spitzner's Tapering Test

Spitzner (2008) [28] proposed the test based on tapering for use in testing a global linear hypothesis under a functional linear model. The test statistic is constructed as a weighted sum of squared linear combinations of Fourier coefficients, a tapered quadratic form, in which higher Fourier frequencies are down-weighted so as to emphasize the smooth attributes of the model.

After some initial pre-processsing, functional data may typically presented by a discrete, high-dimensional model,

$$Y_{n,j} = \theta_j + n^{-1/2} e_{n,j} \tag{1.3}$$

for $j = 1, \cdots, p_n$, ignoring any replication, where p_n represents some (high) maximum number of dimensions to be accounted for at a given n. $e_{n,j}$ are zero-mean, unit-variance error-vectors that the $e_{n,j}$ are independent across j. (However, small correlations among the $e_{n,j}$ are possible.) Then the functional linear hypothesis translates to

$$\begin{aligned} H_0 : &\quad \theta_j = 0 \ \ for\, j = 1, \cdots, p_n \\ H_1 : &\quad not\, H_0 \end{aligned} \tag{1.4}$$

Based on the model (1.3), the test statistic based on the quadratic form are proposed

$$Q_n^{OPT} = n \sum_{j=1}^{p_n} j^{-1/2} Y_{n,j}^2 \tag{1.5}$$

This test incorporates the smoothness assumption that $(\theta_1, ..., \theta_{p_n})$ belongs to Sobolev class. In this article, the "rate of testing" theory and "adaptive rate of testing" theory is discussed and the asymptotic optimality of the test based on the weight $j^{-1/2}$ among tests based on tapering are proved, under further assumptions that the fourth moment of $e_{n,j}$ is bounded and the covariance of $Y_{n,j}$ dies out as the sample size increases.

Simulation studies on Gaussian case to test the null hypothesis versus spiked alternatives as well as smoothed alternatives are conducted in this article. The performance of tapering test and other testing procedures including adaptive Neyman test, wavelet thresholding test are compared. Results suggest that any test based on tapering is asymptotically suboptimal to tests based on truncation or thresholding; however, Q_n^{OPT} exhibits better empirical power under the class of smoothed alternatives that would be of natural interest in many applications.

While this article argues in support of the tapering test procedure on the basis of the power properties, another reason for choosing tapering test is that the test statistic in Equation (1.5) is easy to be written through a formal Bayesian construction, as a monotone transformation of the posterior null probability. While in literature, even there exists Bayesian construction of the thresholding mechanism for use in estimation, which suggests that the test statistic might be viewed as a summary metric based on a Bayesian estimator. This fails to represent test statistic as a monotone transformation of a posterior null probability.

1.5 Rate of Testing Theory

The asymptotic performance criteria used in Spitzner(2008) [28] and in our research is from "rate of testing" theory, a framework to find the rate at which the power is retained under geometric smoothness constraints.

Rate of Testing Criteria:

We wish to test the null hypothesis

$$H_0 : \theta_j = 0 \ \ for\, j = 1, \cdots, p_n$$

against the composite nonparametric alternative that the mean vector $\{\theta_j, j = 1, \cdots, p_n\}$ is separated away from zero in L_2 norm, $\| \boldsymbol{\theta} \| \geq \delta$ and also $\{\theta_j, j = 1, \cdots, p_n\}$ possesses some smoothness properties. The problem is to describe the minimal (optimal) rate $\delta_n \to 0$ for which testing with prescribed error probabilities is still possible. The "rate" $\delta_n \to 0$ in rate of testing theory characterize test performance. Then the rate of testing theory is to look at the asymptotic hypothesis testing problem as the sample size $n \to \infty$, and evaluate the optimal (fastest) rate of decay to zero of the δ_n as a function of n.

A key issue in testing is how to impose the smoothness assumption, so as not to waste statistical power attempting to distinguish the rougher aspects of the model. Rate of testing criteria restrict the mean vector of the discrete model $Y_{n,j} = \theta_j + n^{-1/2} e_{n,j}$ to smooth-function class. In the most general settings, this would be a Besov class $B_{s,p,q}(M)$ with $p < 2$, but here it is taken to be a Sobolev class as a special class when $p = q = 2$, which is appropriate when working with Fourier decompositions. Suppose $\boldsymbol{\theta}$ is an element of a Sobolev ellipsoid in continuous space, the restriction

can be expressed as

$$\boldsymbol{\theta} = (\theta_j, j = 1, 2, \cdots) \in \mathcal{B}_{s,M}$$
$$\mathcal{B}_{s,M} = \left\{ (\theta_1, \theta_2, \cdots) : \sqrt{\sum_{j=1}^{\infty} j^{2s}\theta_j^2} \leq M \right\}$$

, where $\mathcal{B}_{s,M}$ is a Sobolev ellipsoid of radius M in infinite-dimensional discrete space, and $M > 0$, $s > 1/2$ are fixed constants, larger s makes the restriction stronger.

Fix $M > 0$, $s > 1/2$, and for each n, let $\phi_n = \phi_n(Y_{n,1}, \cdots, Y_{n,p_n})$ be the test such that $\lim_n P_{\mathbf{0}}[\phi_n = 1] \leq \alpha$ for a fixed level $\alpha \in (0, 1)$. $P_{\boldsymbol{\theta}}$ denotes the probability under the model $Y_{n,j} = \theta_{n,j} + n^{-1/2}e_{n,j}$ for a specific $\boldsymbol{\theta}$ and fixed n. The rate of testing theory are formulated from sequance $\delta_n \to 0$ satisfying

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} P_{\boldsymbol{\theta}}[\phi_n = 1] \to 1, \forall \delta_n^\star \to 0 \tag{1.6}$$
$$H_1(\delta; s, M) = \left\{ \boldsymbol{\theta} \in \mathcal{B}_{s,M} : \sqrt{\sum_{j=1}^{\infty} \theta_j^2} \geq \delta \right\}$$

Remark 1.5.1. *The rate of testing criterion describes the rate at which a gap may shrink between the null hypothesis and a class of distinguishable alternatives, those the test would be able to detect with high power, asymptotically. The better performing tests allow this gap to shrink faster. If for some $\delta_n \to 0$ the criterion is satisfied for one test, but not another, the former test is preferred.*

Minimax Rate and Adaptive Minimax Rate:

Ingster (1993) [13] established the optimal rate of testing:

$$\widehat{\delta}_n^M(s) = n^{-2s/(4s+1)}, \tag{1.7}$$

which is commonly expressed as the **minimax rate** for geometry $\mathcal{B}_{s,M}$ at specific s. In the paper, he proved that no test does any have the rate δ_n converge to zero faster than $\widehat{\delta}_n^M(s)$ that satisfy the criteria (1.6). And he proposed the rate-optimal test based on tapering with the "rate" $\widehat{\delta}_n^M(s)$ that satisfies the criteria (1.6). However, the practical applications of this test that proposed by Ingster meet the crucial problem that it depends on the smoothness parameter s, which are typically unknown.

To address this problem with Ingster's "rate of testing" theory, Spokoiny (1996) [30] proposed "adaptive rate of testing" theory, by restricting the unknown smoothness parameter s to the interval

$(s_\star, s^\star)$ and $\frac{1}{2} < s_\star < s^\star$ are two known constants. Spokoinys showed that the adaptive (assumption free) testing with the same rate is impossible with regarding to the problem that the smoothness parameter$s$ is unknown, and he defined the optimal adaptive rate of testing:

$$\widehat{\delta}_n^{AM}(s) = \left[n^2 \left(loglogn\right)^{-1}\right]^{4s/(4s+1)} \qquad across \quad s_\star < s < s^\star. \tag{1.8}$$

The rate $\widehat{\delta}_n^{AM}(s)$ is commonly expressed as the **adaptive minimax rate** for geometry $\mathcal{B}_{s,M}$ across $s_\star < s < s^\star$. Note that comparing $\widehat{\delta}_n^{AM}(s)$ with $\widehat{\delta}_n^{M}(s)$, the adaptive minimax rate add an adaptive factor $\left(loglog\sigma_n^{-2}\right)^{1/4}$ in $\widehat{\delta}_n^{M}(s)$ to make it adaptive, and thus is slower than Ingster's minimax rate. Spokoiny in this paper proved that no test does any with the rate δ_n converge to zero faster than $\widehat{\delta}_n^{AM}(s)$ that satisfy the criteria (1.6) uniformly across $s_\star < s < s^\star$, and derived the rate-optimal adaptive test based on thresholding with the "rate" $\widehat{\delta}_n^{AM}(s)$ that satisfies the criteria (1.6).

Spitzner (2008) [28] proved that no tapering test with rate δ_n goest to 0 faster than

$$\widehat{\delta}_n^{Q}(s) = \left[n^2 \left(\log n\right)^{-1}\right]^{-s/(4s+1)} \qquad across \quad s_\star < s < s^\star$$

will satisfy the rate of testing criteria, and proposed the optimal tapering test Q_n^{OPT} in (1.5) with rate $\widehat{\delta}_n^{Q}$ satisfies the rate of testing criteria.

Remark 1.5.2. *Note that compared with Ingster's minimax rate in (1.7) and Spokoiny's adaptive minimax rate in (1.8), the rate $\widehat{\delta}_n^{Q}$ is slower than both of them. But $\widehat{\delta}_n^{Q}$ already represents an adaptively optimal configuration among tests based on tapering. Spitzner in his paper reiterates that an empirical investigation in a non-asymptotic context has demonstrated that the test Q_n^{OPT} has superior power against a class of alternatives. Another important reason that Spitzner recommended this test is that the test statistic in (1.5) may arise through a formal Bayesian construction, as a monotone transformation of a posterior null probability, which has been shown in Spitzner (2008) [27].*

1.6 Applied Motivation

As our research is actually motivated by solving real world problem regarding to EEG dataset. Let me introduce this motivating example in details.

The human brain is the most complex organ in the body. Composed of 50 to 100 billion neurons, the human brain remains one of the world's greatest unsolved mysteries. Our research will provide

a way to take a closer look at the spectral characteristics of scalp Electroencephalography (EEG) recordings to discover more about the relationship between lobes during cognitive tasks. And this is achieved by developing more powerful testing procedures in spectral densities of EEG signals from the Reward Two Back Study.

The research in this article was driven by solving real world problem regarding to the dataset from the Swartz Center for Computational Neuroscience (SCCN) of the University of California, San Diego. SCCN is founded by U.S. National Institutes of Health grants. Studies at SCCN focus on how EEG data, alone or combined with functional hymodynamic imaging data, can be used to observe, model and test new theories about how different parts of the brain interact dynamically to support human awareness and behavior.

EEG is the multichannel recording of the electrical activity generated by collections of neurons within the brain. Different channels reflect the activity within different brain regions. When the EEG is measured using non-invasive electrodes arrayed on an individual's scalp, it is referred to as scalp EEG. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a short period of time, as recorded from multiple electrodes placed on the scalp. Diagnostic applications generally focus on the spectral content of the EEG, that is, the type of neural oscillations that can be observed in EEG signals.

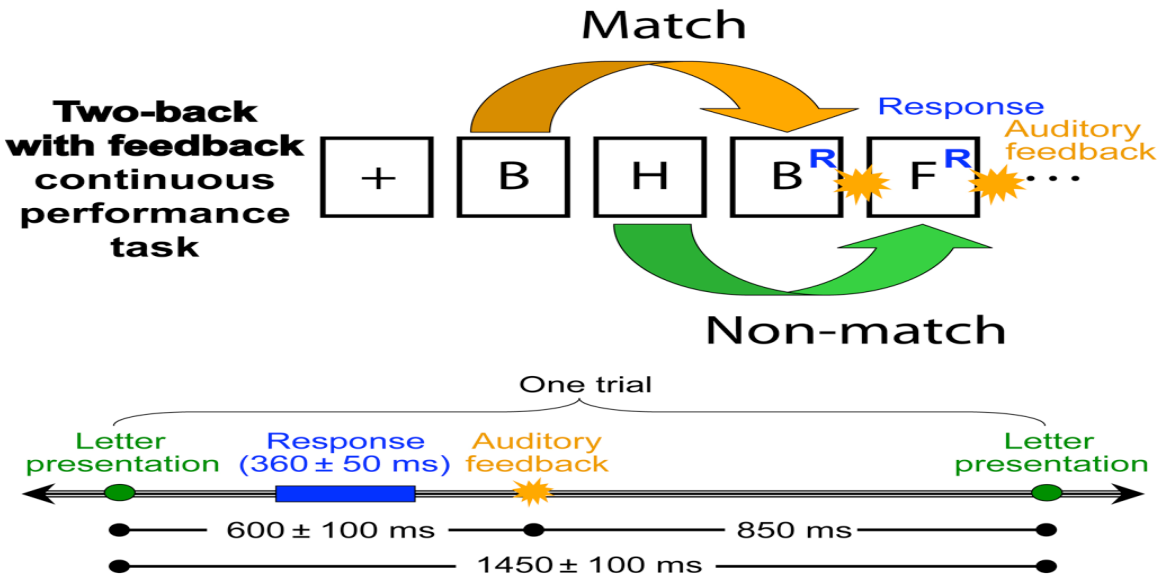
The scalp EEG is non-invasive measure of the electrical potentials generated by the activity of tens of millions of neurons within the brain. An EEG signal, or channel is formed by taking the difference between potentials measured at two electrodes. Each EEG channel summarizes activity localized within a region of the brain. The neurons that contribute the most to the scalp EEG are those closest to the scalp surface. Scalp EEG activity is modulated by the state of vigilance of an individual. In particular, the dominant frequency and spatial distribution of EEG activity during the awake state is different than that during sleep. EEG signals arriving at each electrode are the sum of activities in all such EEG source areas, as well as electrical artifacts from muscles, eyes, electrodes, movements, and the electrical environment.

EEG signals are stochastic, i.e. they can be represented as a sequence of related random variables. Statistical spectral analysis treats EEGs as time series data generated by stationary stochastic (random) processes. Recently more and more advanced mathematical tools may be brought to bear on the problem of understanding the link between brain activity, as seen in the

EEG, and function as determined via the implementation of experimental protocols. The task oriented brain activity analysis and classification is a prime issue in EEG signal processing these days. The similar attempt has been done here to estimate the brain activity on the basis of power spectrum analysis. The accurate testing procedures of electrical activity for a particular state of human brain helps in neurological diagnosis and also for establishing standards for instrumentation development. This procedure also helps in the brain computer interfacing which has been gaining wide attraction in the research industry.

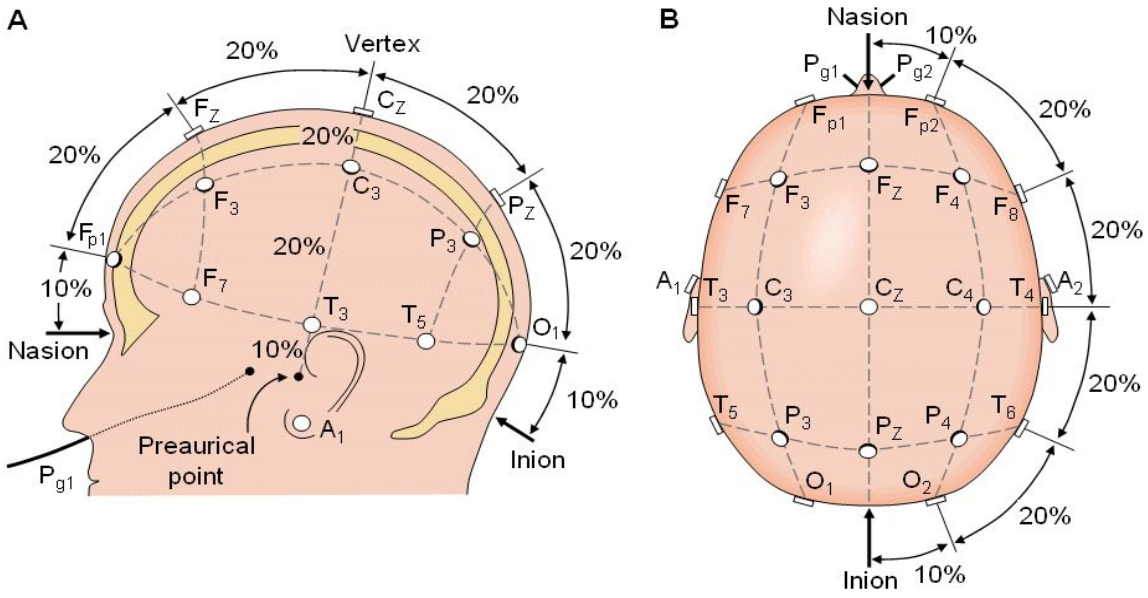
Our research is driven by the data of EEG recordings from "Reward Two-back Continuous Performance Task Study". The aim of this study is to explore the cognitive effects of positive and negative reinforcement. This study utilizes the Two-back continuous Performance Task (CPT) with auditory feedback and reward/punishment. Subjects were seated in front of computer monitor with a response pad in the laps that was held with both hands. During the task, subject were presented sequential single letters. Beginning with the third letter, subject respond to each letter, specifying with a right or left thumb press whether the current letter was the same as the one presented two before. An auditory feedback signal at letter offset informed the subject of whether their answer was correct or wrong. After 850 milliseconds, the next letter was presented. Correct responses add 1 point ($ in their pocket), and incorrect or failures to respond deducted 1 cent from the subject's performance reward ($ out of their pocket). Visualization of the trial can be found in Figure 1.1.

Figure 1.1: *Visualization of the Reward Two-back Continuous Performance Task Study.*



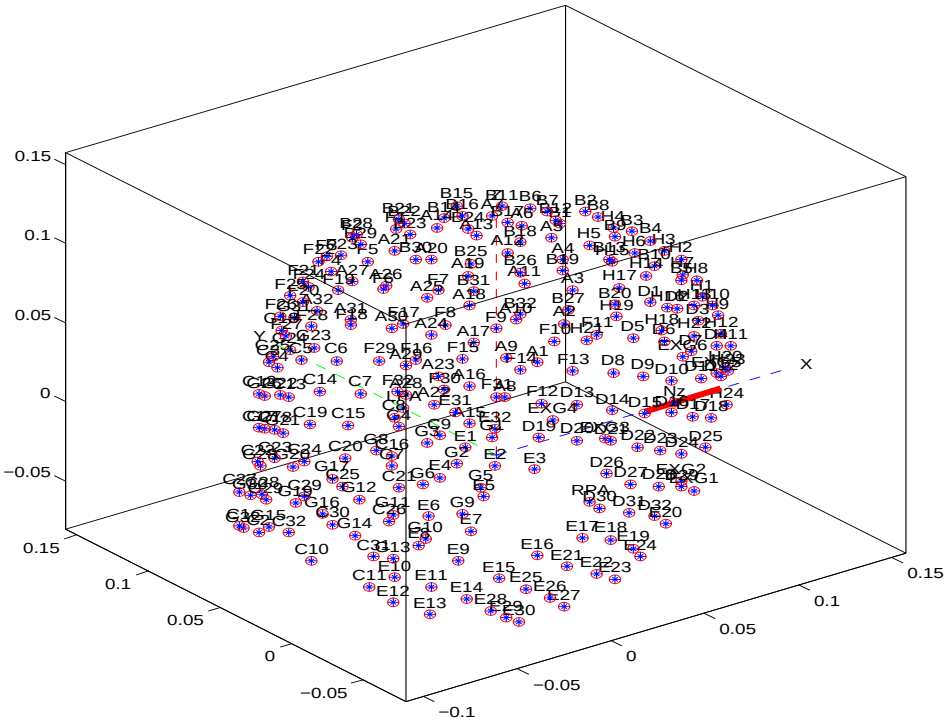
During the study, 256 electrodes were put on the subject's scalp and EEG recordings are collected from these subjects performing the task. To facilitate repetition, a standard is used for electrode positions known as the international 10-20 system. This includes guidelines on position and an associated system for labeling those positions. The Figure 1.2 shows how electrodes are located on scalp as for this system.

Figure 1.2: *The international 10-20 system seen from (A) left and (B) above the head. A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital.*



We are interested in just one trial from letter shown up on the screen until the next letter shown up as in 1.1. The sampling rate is 256 Hz, which means that we have 256 samples every second. Thus for one tail in Figure 1.1, we will have 340 discrete EEG observations. EEG signals are recorded for a normal 22 years old, right-handed young man with only electrode D27 unstable, otherwise clean EEG recordings is studied. The 3D plot of channel locations are shown in Figure 5.4. Data were read and preprocessed using MATLAB toolbox EEGLAB.

Figure 1.3: *3D channel locations.*

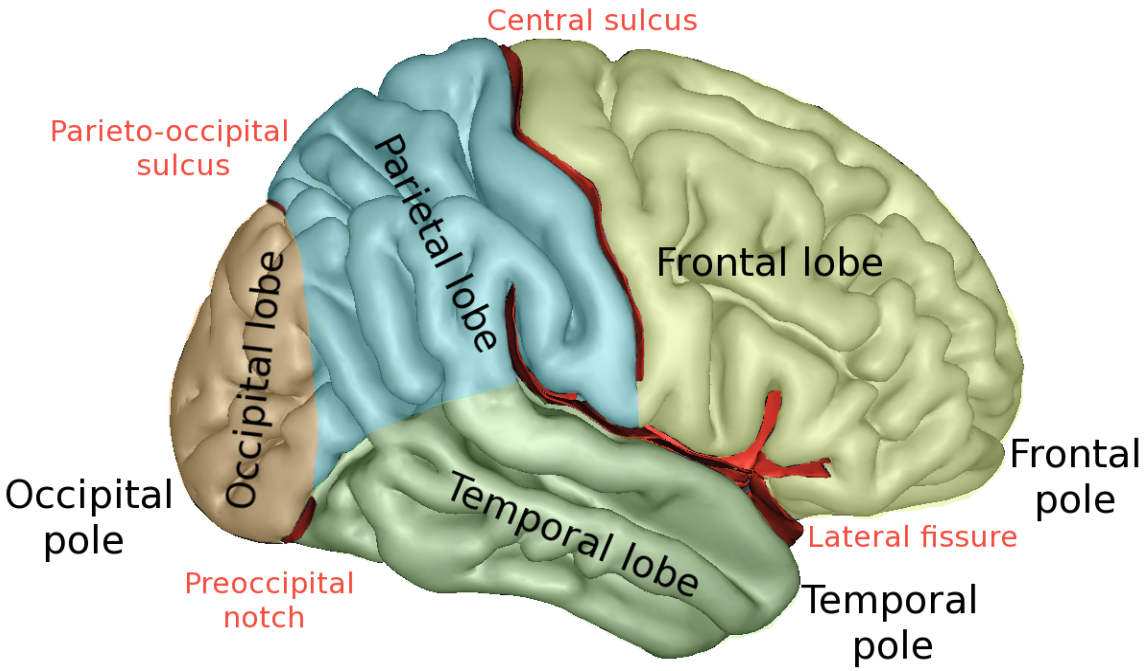


The event related EEG recordings are assumed to be realizations of a specific type of non-stationary random process, namely, a locally stationary process. Thus, within short segments of time, the process exhibits stationary, even though the global characteristics vary throughout the trial. For non-parametric of the spectral density or the cross-spectral density of the EEG recordings, a trade-off has to be made regarding the length of the data segment for analysis. The segment should not be too long to satisfy the assumption of the local stationarity but must provide adequate spectral resolution. One trial with the during of around 1.5 sec segment was chosen as being short enough to ensure the approximate stationarity, but at the same time give a proper spectral resolution.

There are four main different lobes on the brain, Frontal Lobe, Temporal Lobe, Occipital Lobe and Parietal Lobe, as shown in the Figure 5.8. The frontal lobe has many functions most of

which center on cognition, problem solving and reasoning. For the temporal lobe, there are two temporal lobes located on both sides of the brain that are in close proximity to the ears. The primary function of the temporal lobes is to processing auditory sounds. The occipital lobe, the smallest of the four lobes, is located near the posterior region of the cerebral cortex, near the back of the skull. The occipital lobe is the primary visual processing center of the brain. The parietal lobe is responsible for processing sensory information from various parts of the body. Thus, for the right-handed healthy young man in the Reward Two-back Study, the Frontal, Temporal Left and Occipital Lobe will be most active. We will only focus on these three lobes for our testing of potential relationships.

Figure 1.4: *Four different lobes in the brain.*



The electroencephalogram (EEG) is a complex signal and an important brain state indicator (e.g. waking, sleep, seizure). Crucial aspects of the signal might not be recognized by visual inspection of the EEG. Therefore, additional quantitative analysis is fundamental to investigate the EEG in more detail. Modern brain research is intimately linked to the feasibility to record the EEG and to its quantitative analysis. EEG spectral analysis (decomposing a signal into its

constituent frequency components) is an important method to investigate brain activity. Spectral analysis is a mathematical approach to quantify the EEG. It have been used for a long time in the analysis of EEG signals, which referred to as frequency domain analysis or spectral density estimation. Its purpose is the decomposition of signals such as the EEG, into its constituting frequency components. The fast Fourier transform is a widely applied method for obtaining the EEG spectrum. The power density spectrum or power spectrum displays the distribution of power or variance over the frequency components of a signal. It is defined as the Fourier transform of the autocorrelation function. Spectral content reveals neural oscillation. And characterizing such neural oscillatory behavior gives insight into the underlying dynamics of brain systems.

In this research, we will look at the spectral contents of these three lobes and apply our powerful testing procedures to find possible links and similar patterns between them.

Chapter 2

Models and Asymptotics

Our main goal in this chapter is to develop a new powerful test based on "rate of testing" theory that is very convenient for implement in practice, even for some people have no background in statistics. Although Spitzner's tapering test was originally proposed in functional data analysis framework, we can still borrow the idea of tapering and try to identify suitable definitions of those quantities for time series problems. Thus, our testing procedures will be derived from Spitzner's tapering test based on certain discrete models. The model is constructed based on different methods of estimating spectral densities as discussed in Section 1.3. We will mainly explore three different models based on the raw periodogram, Fourier transformation and kernel smoothing. Under each model, the optimality of the tapering test is derived based on "rate of testing" theory.

2.1 Model 1 (Basic Model Based on Raw Periodogram)

As introduced in Section 1.3, a natural and most convenient estimate of spectral density is the periodogram. We will firstly try to derive the model and testing procedure based on the model derived from periodogram as the estimate of spectral density. For now, let's suppose the length of these two time series $\{X_{1,t_1},\, t_1 = 1, \cdots, n_1\}$ and $\{X_{2,t_2},\, t_2 = 1, \cdots, n_2\}$ are equal, i.e. $n_1 = n_2 = n$.

Define

$$Y_j = \log I_1(\omega_j) - \log I_2(\omega_j),$$

which relfects the context of our problem. And define $\theta_j = \log f_1(\omega_j) - \log f_2(\omega_j)$ for $j = 1, \cdots, p_n$. p_n is the number of frequencies and $p_n = n/2$ if n is even, $p_n = n/2 - 1$ if n is odd. Then it is easy to prove that $Y_j \xrightarrow{D} N\left(\theta_j, \sigma^2\right)$ asymptotically independent across j, where $\sigma^2 = \frac{\pi^2}{3}$.

We are interested in testing the hypothesis of whether these two time series have the same spectral density. Note that by the definition of θ_j, the hypothesis is equivalent to

$$H_0 : \theta_j = 0 \text{ for all } j = 1, \cdots, p_n \quad \text{versus} \quad H_1 : \theta_j \neq 0 \text{ for some } j = 1, \cdots, p_n. \tag{2.1}$$

Since the natural and most convenient estimate of spectral density is the periodogram, the basic model for tapering test is derived firstly based on the asymptotic properties of periodogram:

$$Y_j = \theta_j + \sigma\epsilon_j \quad \text{for } j = 1, \cdots, p_n,$$

where $\epsilon_j = \frac{Y_j - \theta_j}{\sigma}$ are asymptotic independent standard normal random variables.

The tapering test statistic is constructed in the form of a tapered sum of squared summary coefficients Y_js,

$$Q_n = \sum_{j=1}^{p_n} w_{n,j} Y_j^2,$$

whose wights $w_{n,j}$ are presumed to taper to zero in the sense that $w_{n,j} \to 0$ as $j \to \infty$. For such asymptotic statements to make sense, we think conceptually that p_n is arbitrarily large, however, in practice p_n is finite and determined by data configuration.

Remark 2.1.1. *As defined before the "rate" in rate of testing theory, δ_n, defines the gap between the null hypothesis and the class of "distinguishable" alternatives where the test can detect with high power, asymptotically. The class of distinguishable alternatives, $\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)$ are defined based on the smoothness assumption, where*

$$H_1(\delta; s, M) = \left\{ \boldsymbol{\theta} \in \mathcal{B}_{s,M} : \sqrt{\sum_{j=1}^{\infty} \theta_j^2} \geq \delta \right\},$$

and the smoothness restriction on $\boldsymbol{\theta}$ is defined as

$$\mathcal{B}_{s,M} = \left\{ (\theta_1, \theta_2, \cdots) : \sqrt{\sum_{j=1}^{\infty} j^{2s} \theta_j^2} \leq M \right\},$$

a Sobolev geometry ellipsoid of radius M in infinite-dimensional discrete space, where $M > 0$ and $s > 1/2$ are fixed constants, with larger s making the restriction stronger.

Thus δ_n characterized the test performance and the optimal testing procedures should satisfy the rate of testing criteria with the fastest rate of δ_n among the same testing mechanism.

Then we will explain step by step as how do we get assess the asymptotic performance of the proposed test based on "rate of testing" theory.

Lets firstly define some items that we will use latter in the proofs. Define

$$S_n(p) = \sum_{j=1}^{p} w_{n,j}^2$$

$$W_n(p) = min\left\{w_{n,j}^2, j \leq p\right\}$$

$$U_n(p,q) = qW_n(q)/S_n(p)$$

$$U_n(p) = U_n(p,p),$$

where the weights $0 < w_{n,j} \leq 1$ and the dimension here is $p_n = n/2 - 1$. When the weights $w_{n,j}$ dose not depend on n, then $S_n(p)$ becomes $S(p)$, $W_n(p)$ becomes $W(p)$, $U_n(p,q)$ becomes $U(p,q)$, and $U_n(p)$ becomes $U(p)$.

The results of the following lemma will be used in the proof of Lemma 2.1.2 to derive the mean of variance of our test statistic Q_n.

Lemma 2.1.1. *Let X_n and Y_n are two asymptotically independent sequence satisfying*

$$\lim_{n\to\infty} Cov(X_n, Y_n) = 0$$

, then

$$\lim_{n\to\infty} Cov\left\{g\left(X_n\right), g\left(Y_n\right)\right\} = 0$$

for any suitably integrable function g.

Proof.

$$\begin{aligned}
&\lim_{n\to\infty} Cov(X_n, Y_n) = 0 \quad \Rightarrow \\
&\lim_{n\to\infty} \left[E(X_nY_n) - E(X_n)E(Y_n)\right] = 0 \quad \Rightarrow \\
&\lim_{n\to\infty} \int\int f(x,y)dxdy = \lim_{n\to\infty} \int\int f_X(x)f_Y(y)dxdy \quad \Rightarrow \\
&\lim_{n\to\infty} f(x,y) = \lim_{n\to\infty} f_X(x)f_Y(y)
\end{aligned}$$

. Then

$$\begin{aligned}
&\lim_{n\to\infty} Cov\left\{g\left(X_n\right), g\left(Y_n\right)\right\} \quad = \\
&\lim_{n\to\infty} \left\{E\left[g(X_n)g(Y_n)\right] - E\left[g\left(X_n\right)\right] E\left[g\left(Y_n\right)\right]\right\} \quad = \\
&\lim_{n\to\infty} \left\{\int\int g(x)g(y)f(x,y)dxdy - \int\int g(x)g(y)f_X(x)f_Y(y)dxdy\right\} \quad = \\
&\int\int g(x)g(y) \lim_{n\to\infty} \left\{f(x,y) - f_X(x)f_Y(y)\right\} dxdy = 0
\end{aligned}$$

□

The key of the "rate of testing" criteria is to look at the asymptotic power of our test Q_n, to calculate the power, we firstly need to get the asymptotic mean and variance of Q_n. The following lemma provides the formula of $E\left(Q_n\right)$ and $Var\left(Q_n\right)$ when $n \to \infty$.

Lemma 2.1.2. *Based on the model of Y_j as discussed before, to test the hypothesis of equal spectral densities, the test is specified through the tapering test statistic $Q_n = \sum_{j=1}^{p_n} w_{n,j} Y_j^2$, which rejects null hypothesis when Q_n is large. we have*

$$\begin{aligned}
\lim_{n\to\infty} E\left(Q_n\right) &= \sigma^2 R_n(p_n) + \theta_{1,n}^2(p_n) \\
\lim_{n\to\infty} Var\left(Q_n\right) &= A_n S_n(p_n) + 4\sigma^2\theta_{2,n}^2(p_n)
\end{aligned}$$

, where $R_n(p) = \sum_{j=1}^{p} w_{n,j}$, $\theta_{m,n}^2(p) = \sum_{j=1}^{p} w_{n,j}^m \theta_j^2$, and $A_n \asymp 1$.

Proof. Firstly, let's look at the asymptotic expectation of the test statistics $\lim_{n\to\infty} E\left(Q_n\right)$,

$$\begin{aligned}
E\left(Q_n\right) &= E(\sum_{j=1}^{p_n} w_{n,j} Y_j^2) = \sum_{j=1}^{p_n} w_{n,j} E(Y_j^2) \\
&= \sum_{j=1}^{p_n} w_{n,j} \left[Var(Y_j) + (EY_j)^2\right] \\
&\Rightarrow \\
\lim_{n\to\infty} E\left(Q_n\right) &= \sigma^2 \sum_{j=1}^{p_n} w_{n,j} + \sum_{j=1}^{p_n} w_{n,j}\theta_j^2 \\
&= \sigma^2 R_n(p_n) + \theta_{1,n}^2(p_n)
\end{aligned}$$

Secondly, let's look at the asymptotic variance of the test statistics $\lim_{n\to\infty} Var\left(Q_n\right)$,

$$Var\left(Q_n\right) = Var(\sum_{j=1}^{p_n} w_{n,j} Y_j^2) = \sum_{j=1}^{p_n} w_{n,j}^2 Var(Y_j^2) + 2 \sum_{1\le j\le k\le p_n} w_{n,j} w_{n,k} Cov(Y_j^2, Y_k^2)$$

. Let's firstly look at the covariance part $Cov(Y_j^2, Y_k^2)$.

$$\begin{aligned} Cov\{Y_j, Y_k\} &= Cov\left\{log\frac{\widehat{f}_1(\omega_j)}{\widehat{f}_2(\omega_j)}, log\frac{\widehat{f}_1(\omega_k)}{\widehat{f}_2(\omega_k)}\right\} \\ &= Cov\left\{log\widehat{f}_1(\omega_j), log\widehat{f}_1(\omega_k)\right\} + Cov\left\{log\widehat{f}_2(\omega_j), log\widehat{f}_2(\omega_k)\right\} \end{aligned}$$

, since these two time series are independent.

For periodograms, we have $\lim_{n\to\infty} Cov\left\{log\widehat{f}_1(\omega_j), log\widehat{f}_1(\omega_k)\right\} \to 0$, and $\lim_{n\to\infty} Cov\left\{log\widehat{f}_2(\omega_j), log\widehat{f}_2(\omega_k)\right\} \to 0$. Thus we will have

$$\begin{aligned} &\lim_{n\to\infty} Cov\{Y_j, Y_k\} = 0 \quad \Rightarrow \\ &\lim_{n\to\infty} Cov\left\{Y_j^2, Y_k^2\right\} = 0 \quad \text{(Lemma 2.1.1)} \end{aligned}$$

Thus, $\exists$ constant M and integer n_0 s.t. for $n > n_0$,

$$\left|\frac{2\sum_{1\le j\le k\le p_n} w_{n,j}w_{n,k}Cov(Y_j^2, Y_k^2)}{S_n(p_n)}\right| \le M$$

, that is

$$2\sum_{1\le j\le k\le p_n} w_{n,j}w_{n,k}Cov(Y_j^2, Y_k^2) = O\left(S_n(p_n)\right)$$

Let $2\sum_{1\le j\le k\le p_n} w_{n,j}w_{n,k}Cov(Y_j^2, Y_k^2) = a_nS_n(p_n)$, where $a_n \asymp 1$. Then

$$Var\left(Q_n\right) = \sum_{j=1}^{p_n} w_{n,j}^2 Var(Y_j^2) + a_nS_n(p_n)$$

Then let's look at the first part of $Var\left(Q_n\right)$. We have $Var(Y_j^2) = E(Y_j^4) - \left[E(Y_j^2)\right]^2$.

$$\begin{aligned} \left[E(Y_j^2)\right]^2 &= \left[Var(Y_j) + (EY_j)^2\right]^2 \quad \Rightarrow \\ \lim_{n\to\infty}\left[E(Y_j^2)\right]^2 &= \sigma^4 + \theta_j^4 + 2\theta_j^2\sigma^2. \end{aligned}$$

For $E(Y_j^4)$, which is the fourth moment of Y_j, since the asymptotic distribution for Y_j is normal with mean θ_j and variance σ_n^2,

$$\lim_{n\to\infty} E(Y_j^4) = \theta_j^4 + 6\theta_j^2\sigma^2 + 3\sigma^4.$$

Thus

$$\begin{aligned} \lim_{n\to\infty} Var(Y_j^2) &= \lim_{n\to\infty} E(Y_j^4) - \lim_{n\to\infty}\left[E(Y_j^2)\right]^2 \\ &= 4\theta_j^2\sigma^2 + 2\sigma^4. \end{aligned}$$

Thus

$$\begin{aligned} Var\left(Q_n\right) &= \sum_{j=1}^{p_n} w_{n,j}^2 Var(Y_j^2) + a_n S_n(p_n) \\ &= 4\sigma^2 \sum_{j=1}^{p_n} w_{n,j}^2 \sigma^4 \theta_j^2 + 2\sum_{j=1}^{p_n} w_{n,j}^2 + a_n S_n(p_n) \\ &= 4\sigma^2 \theta_{2,n}^2(p_n) + 2\sigma^4 S_n(p_n) + a_n S_n(p_n) \\ &= A_n S_n(p_n) + 4\sigma^2 \theta_{2,n}^2(p_n) \end{aligned}$$

, where $A_n \asymp 1$. □

To take a closer look at the "rate of testing" criteria, let's firstly write it into two equivalent conditions involving the asymptotic mean and variance of Q_n.

Lemma 2.1.3. *For fixed smoothness parameter $s > 1/2$ and $M > 0$, and some positive sequence $\delta_n \to 0$, the rate of testing criteria holds if and only if*

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_0}(Q_n)} \to \infty \quad \textit{and} \tag{2.2}$$

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_{1-0}}(Q_n)} \to \infty, \tag{2.3}$$

where

$$\begin{aligned} E_{H_{1-0}}(Q_n) &= E_{H_1}(Q_n) - E_{H_0}(Q_n) \\ Var_{H_{1-0}}(Q_n) &= Var_{H_1}(Q_n) - Var_{H_0}(Q_n) \end{aligned}$$

Proof. In the rate of testing criteria, to look at the power of the test under the smoothness constraint of the alternative, let's first derive the forms of the critical value for the size-α test based on Q_n. Define the critical value be C, that is, $P_{H_0}(Q_n > C) = \alpha$. Then,

$$\begin{aligned} \alpha &= P\left(\frac{Q_n - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} > \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} \right) \\ &= P\left(\frac{Q_n - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} > \widetilde{C}_{H_0} \right), \end{aligned}$$

where $\widetilde{C}_{H_0} = \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}}$. Then by Chebychev's Inequality, we have

$$\begin{aligned}
&\alpha \leq \frac{1}{1+\widetilde{C}_{H_0}} \\
\Longrightarrow \quad &\frac{1}{\alpha} - 1 \geq \widetilde{C}_{H_0} = \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} \\
\Longrightarrow \quad &C \leq \left\{ E_{H_0}(Q_n) + C\sqrt{Var_{H_0}(Q_n)} \right\} \{1 + o(1)\},
\end{aligned}$$

where $C = \frac{1}{\alpha} - 1$ is a constant. Thus the critical value for size α test that rejects for large Q_n is $E_{H_0}(Q_n) + C\sqrt{Var_{H_0}(Q_n)}$.

Then the power can be written as

$$\begin{aligned}
&P_{H_1(\delta_n/\delta_n^\star;s,M)} \left(Q_n > E_{H_0}(Q_n) + C\sqrt{Var_{H_0}(Q_n)} \right) \\
= \quad &P\left(\widehat{Q}_n = \frac{Q_n - E_{H_1}(Q_n)}{\sqrt{Var_{H_1}(Q_n)}} > \frac{E_{H_0}(Q_n) + C\sqrt{Var_{H_0}(Q_n)} - E_{H_1}(Q_n)}{\sqrt{Var_{H_1}(Q_n)}} \right) \\
= \quad &P\left(\widehat{Q}_n > \frac{-E_{H_{1-0}}(Q_n) + C\sqrt{Var_{H_0}(Q_n)}}{\sqrt{Var_{H_{1-0}}(Q_n) + Var_{H_0}(Q_n)}} \right) \\
= \quad &P\left(\widehat{Q}_n > \frac{-E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_0}(Q_n)} + C}{\sqrt{1 + Var_{H_{1-0}}(Q_n)/Var_{H_0}(Q_n)}} \{1 + o(1)\} \right),
\end{aligned}$$

where $\widehat{Q}_n$ is the standardized test statistic under alternative hypothesis. Since

$$\begin{aligned}
\lim_{n\to\infty} E(\widehat{Q}_n) &= 0 \\
\lim_{n\to\infty} Var(\widehat{Q}_n) &= 1,
\end{aligned}$$

the probability coverges to 1 if the right term inside diverge to $-\infty$ as $n \to \infty$. Thus the rate of testing criteria $P_{H_1(\delta_n/\delta_n^\star;s,M)} \left(Q_n > E_{H_0}(Q_n) + C\sqrt{Var_{H_0}(Q_n)} \right) \to 1$ is eqivalent to

$$\frac{-E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_0}(Q_n)} + C}{\sqrt{1 + Var_{H_{1-0}}(Q_n)/Var_{H_0}(Q_n)}} \to -\infty \text{ under the constraint } \boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M) \tag{2.4}$$

Since the asymptotic distribution for e_j is normal, then $P(e_j \leq -t) > 0$ for each $t > 0$. Thus $P_{H_1(\delta_n/\delta_n^\star;s,M)} \left(\widehat{Q}_n \leq -t \right) > 0$ for each $t > 0$. Thus such divergence is also a necessary condition. Finally it is easy to verify that the condition (2.4) is equivalent to

$$\begin{aligned}
&\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_0}(Q_n)} \to \infty \quad \text{and} \\
&\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} E_{H_{1-0}}(Q_n)/\sqrt{Var_{H_{1-0}}(Q_n)} \to \infty
\end{aligned}$$

□

Then we can plug the results of Lemma 2.1.2 into Lemma 2.3.1, and after some manipulations we can get the equivalent conditions of "rate of testing" criteria as in the following theorem.

Theorem 2.1.1. *Based on the model of Y_j, to test the hypothesis of equal spectral densities, the test is specified through the tapering test statistic $Q_n = \sum_{j=1}^{p_n} w_{n,j} Y_j^2$, which rejects null hypothesis when Q_n is large. For fixed smoothness parameter $s > 1/2$ and $M > 0$, and some positive sequence $\delta_n \to 0$, the rate of testing criteria holds if and only if*

$$\limsup_{n\to\infty} \; U_n(p_n) p_n^{-(4s+1)} < \infty \tag{2.5}$$

$$\liminf_{n\to\infty} \; U_n(p_n, q_n) q_n^{-(4s+1)} > 0, \tag{2.6}$$

where $q_n = \left\{\frac{\delta_n}{M}\right\}^{-1/s}$.

Proof. According to Lemma 2.3.1 and Lemma 2.1.2, the rate of testing criteria is equivalent to

$$\begin{aligned} \inf_{\theta \in H_1(\delta_n/\delta_n^\star; s, M)} \theta_{1,n}^2(p_n)/\sqrt{S_n(p_n)} \to \infty \quad \text{and} \\ \inf_{\theta \in H_1(\delta_n/\delta_n^\star; s, M)} \theta_{1,n}^2(p_n)/\sqrt{\theta_{2,n}^2(p_n)} \to \infty \end{aligned} \tag{2.7}$$

, where $\theta_{1,n}^2(p_n)$ and $\theta_{2,n}^2(p_n)$ are defined as $\theta_{m,n}^2(p) = \sum_{j=1}^{p} w_{n,j}^m \theta_j^2$ in Lemma 2.1.2.

Furthermore,

$$\begin{aligned} &0 < w_{n,j} \le 1 \Rightarrow \theta_{1,n}^2(p_n) \ge \theta_{2,n}^2(p_n) \\ &\Rightarrow \theta_{1,n}^2(p_n)/\sqrt{\theta_{2,n}^2(p_n)} \ge \theta_{1,n}^2(p_n)/\sqrt{\theta_{1,n}^2(p_n)} = \theta_{1,n}(p_n). \end{aligned}$$

And since $S_n(p_n) \to \infty$, the first condition in Equation (2.7) implies $\theta_{1,n}^2(p_n) \to \infty$ and also implies the second condition in Equation (2.7). Thus to develop equivalent conditions for rate of testing criteria is equivalent to develop equivalent conditions for Equation (2.7), and is equivalent to develop equivalent conditions for

$$\inf_{\theta \in H_1(\delta_n/\delta_n^\star; s, M)} \theta_{1,n}^2(p_n)/\sqrt{S_n(p_n)} \to \infty \tag{2.8}$$

To look at Equation (2.8), firstly, we have $\theta_{1,n}^2(p_n) = \theta_{1,n}^2(\infty) - \sum_{j=p_n+1}^{\infty} w_{n,j}\theta_j^2$. Additionally, for $\theta \in \mathcal{B}_{s,M}$,

$$\begin{aligned} \sum_{j=p_n+1}^{\infty} w_{n,j}\theta_j^2 &= \sum_{j=p_n+1}^{\infty} w_{n,j} j^{-2s} j^{2s} \theta_j^2 \\ &\le w_{n,\widetilde{p}_n} p_n^{-2s} \sum_{j=p_n+1}^{\infty} j^{2s}\theta_j^2 \le w_{n,\widetilde{p}_n} p_n^{-2s} M^2 \end{aligned}$$

, where $\widetilde{p}_n$ is the index between 1 and p_n that minimize $w_{n,j}$ for $j = 1, \cdots, p_n$.

Using Lagrange multipliers method, we can get that the vector θ that achieves the lower bound of $\theta^2_{1,n}(\infty)$ subject to constraints $\theta \in H_1(\delta_n/\delta_n^\star; s, M)$ is that

$$\theta_j = \begin{cases} \frac{\delta_n}{\delta_n^\star} & j = \widetilde{J} \\ 0 & o.w. \end{cases}$$

, where $\widetilde{J}$ is the index between 1 and p_n that satisfy that $\widetilde{J} \leq \left\{\frac{\delta_n}{\delta_n^\star M}\right\}^{-1/s}$ and $w_{n,\widetilde{J}}$ minimize $w_{n,1}, \cdots, w_{n,p_n}$ Then $\sum_{j=1}^{\infty} j^{2s}\theta_j^2 = \widetilde{J}^{2s}\delta^2 \leq M^2$, i.e. $\theta \in \mathcal{B}_{s,M}$ satisfied. And $\sum_{j=1}^{\infty}\theta_j^2 = \delta^2$, i.e., $\theta \in H_1(\delta_n/\delta_n^\star; s, M)$ satisfied.

Thus,

$$\begin{aligned}
&\theta^2_{1,n}(\infty) = \sum_{j=1}^{\infty} w_{n,j}\theta_j^2 \geq w_{n,\widetilde{J}}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 \text{ under } \quad \theta \in H_1(\delta_n/\delta_n^\star; s, M) \\
&\Rightarrow \theta^2_{1,n}(p_n) \geq w_{n,\widetilde{J}}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 - w_{n,\widetilde{p}_n}p_n^{-2s}M^2 \\
&\Rightarrow \frac{1}{\sigma_n^2}\theta^2_{1,n}(p_n)/\sqrt{S_n(p_n)} \geq \frac{1}{\sigma_n^2}\left\{w_{n,\widetilde{J}}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 - w_{n,\widetilde{p}_n}p_n^{-2s}M^2\right\}/\sqrt{S_n(p_n)}.
\end{aligned}$$

Then Condition (2.8) holds for every $\delta_n^\star \to 0$ is equivalent to

$$\begin{aligned}
&\limsup_{n\to\infty}\ w_{n,\widetilde{p}_n}p_n^{-2s}/\sqrt{S_n(p_n)} < \infty \quad \text{and} \\
&\lim_{n\to\infty} w_{n,\widetilde{J}}\left(\frac{\delta_n}{\delta_n^\star}\right)^2/\sqrt{S_n(p_n)} = \infty,
\end{aligned} \tag{2.9}$$

where the first condition is needed because the rate of the first one converge to ∞ can be slowed down arbitrarily by $\forall\delta_n^\star \to 0$.

Let's firstly look at the first condition in Equation (2.9),

$$\begin{aligned}
&\limsup_{n\to\infty}\ w_{n,\widetilde{p}_n}p_n^{-2s}/\sqrt{S_n(p_n)} < \infty \quad \Leftrightarrow \\
&\limsup_{n\to\infty}\ w^2_{n,\widetilde{p}_n}p_n^{-4s}/S_n(p_n) < \infty \quad \Leftrightarrow \\
&\limsup_{n\to\infty}\ U_n(p_n)p_n^{-(4s+1)} < \infty.
\end{aligned}$$

Then let's firstly look at the second condition in Equation (2.9),

$$\lim_{n\to\infty} \; w_{n,\tilde{J}} \left(\frac{\delta_n}{\delta_n^\star}\right)^2 / \sqrt{S_n(p_n)} = \infty \quad \Leftrightarrow$$

$$\lim_{n\to\infty} \left\{ \frac{w^2_{n,\tilde{J}}}{\delta_n^{\star 4} W_n\left(\left\{\frac{\delta_n}{M}\right\}^{-1/s}\right)} \right\} U_n\left(p_n, \left\{\frac{\delta_n}{M}\right\}^{-1/s}\right) \delta_n^{(4s+1)/s} = \infty \quad \Leftrightarrow$$

Take the square, and since $w^2_{n,\tilde{J}} = min\left\{ w^2_{n,j}, j \le \left(\frac{\delta_n}{\delta_n^\star M}\right)^{-1/s} \right\} = W_n\left(\left\{\frac{\delta_n}{\delta_n^\star M}\right\}^{-1/s}\right)$

$$\lim_{n\to\infty} \left\{ \frac{W_n\left(\left\{\frac{\delta_n}{\delta_n^\star M}\right\}^{-1/s}\right)}{\delta_n^{\star 4} W_n\left(\left\{\frac{\delta_n}{M}\right\}^{-1/s}\right)} \right\} U_n\left(p_n, q_n\right) \left\{\frac{\delta_n}{M}\right\}^{(4s+1)/s} = \infty \quad \Leftrightarrow$$

Take the square, and since $W_n\left(\left\{\frac{\delta_n}{\delta_n^\star M}\right\}^{-1/s}\right) \ge W_n\left(\left\{\frac{\delta_n}{M}\right\}^{-1/s}\right) \quad for \quad \forall \delta_n^\star \to 0$

$$\liminf_{n\to\infty} \; U_n(p_n, q_n) q_n^{-(4s+1)} > 0$$

□

Finally let's take a closer look at those equivalent condition of "rate of testing" criteria".

Remark 2.1.2. *Condition (2.6) will never be satisfied, since*

$$U_n(p_n, q_n) q_n^{-(4s+1)} = \frac{q_n^{1-r_1-(4s+1)} \left(\log q_n\right)^{r_2(q_n)}}{S(p_n)} \to 0$$

$$since \; 4s+1 > 3, r_1 \ge 0, r_2(q_n) \le 0, q_n \to \infty \; and \; p_n \to \infty \; as \; n \to \infty.$$

Thus rate of testing criteria will never be satisfied for tapering testing based on raw periodogram.

Then main reason that our Model 1 based on raw periodograms cannot satisfy "rate of testing" criteria is that the periodogram is not a consistent estimator of spectral density and thus the σ in Model 1 is a constant and does not go to zero as n goes to infinity. To further explore impact of the non-consistent estimator, we compared the asymptotic power of our test based on consistent estimators with that based on non-consistent estimators and have the following theorem.

Theorem 2.1.2. *Based on the model of Y_j, to test the hypothesis of equal spectral densities, the tapering tests based on the model with constant σ to construct test statistic Q_n will always asymptotically have lower power than tests based on the model with σ in the model goes to 0 as $n \to \infty$.*

Proof. To construct the power of tapering test, let's first look at the critical value. Let the critical value be C, that is, $P_{H_0}(Q_n > C) = \alpha$. Then,

$$\begin{aligned}\alpha &= P\left(\frac{Q_n - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} > \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}}\right) \\ &= P\left(\frac{Q_n - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} > \widehat{C}_{H_0}\right)\end{aligned}$$

, where $\widehat{C}_{H_0} = \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}}$. Then by Chebychev's Inequality, we have

$$\begin{aligned}&\alpha \leq \frac{1}{1+\widehat{C}_{H_0}} \quad \Rightarrow \\ &\frac{1}{\alpha} - 1 \geq \widehat{C}_{H_0} = \frac{C - E_{H_0}(Q_n)}{\sqrt{Var_{H_0}(Q_n)}} \quad \Rightarrow \\ &C \leq \left\{\sigma^2 R_n(p_n) + \left(\frac{1}{\alpha} - 1\right)\sqrt{A_n S_n(p_n)}\right\}\{1 + o(1)\} \quad \text{(Lemma 2.1.2)} \quad \Rightarrow \\ &C \leq \left\{\sigma^2 R_n(p_n) + C_n\sqrt{A_n S_n(p_n)}\right\}\{1 + o(1)\},\end{aligned}$$

where $C_n = \frac{1}{\alpha} - 1 \asymp 1$ and $R_n(p_n)$, $S_n(p_n)$, and $A_n \asymp 1$ are notations defined before. Thus the critical value for size α test that rejects for large Q_n is $\sigma^2 R_n(p_n) + C_n\sqrt{A_n S_n(p_n)}$ for $A_n \asymp 1$ and $C_n \asymp 1$.

Then the power can be written as

$$\begin{aligned}&P_{H_1}\left(Q_n > \sigma^2 R_n(p_n) + C_n\sqrt{A_n S_n(p_n)}\right) \quad = \\ &P\left(\widehat{Q}_{nH_1} = \frac{Q_n - E_{H_1}(Q_n)}{\sqrt{Var_{H_1}(Q_n)}} > \frac{\sigma^2 R_n(p_n) + C_n\sqrt{A_n S_n(p_n)} - E_{H_1}(Q_n)}{\sqrt{Var_{H_1}(Q_n)}}\right) \quad = \\ &P\left(\widehat{Q}_{nH_1} > \frac{-\theta_{1,n}^2(p_n)/\sqrt{A_n S_n(p_n)} + C_n}{\sqrt{1 + 4\sigma^2\theta_{2,n}^2(p_n)/\sqrt{A_n S_n(p_n)}}}\{1 + o(1)\}\right)\end{aligned}$$

, where $\widehat{Q}_{nH_1}$ is the standardized test statistic under alternative hypothesis.

Define $CV = \frac{-\theta_{1,n}^2(p_n)/\sqrt{A_n S_n(p_n)} + C_n}{\sqrt{1+4\sigma^2\theta_{2,n}^2(p_n)/\sqrt{A_n S_n(p_n)}}}$. Then let's look at the value of CV for constant σ as well as $\sigma \to 0$ as $n \to \infty$. The difference of CV is mainly on the σ in denominator of CV. Thus, when $\sigma \to 0$ as $n \to \infty$, CV have smaller value and thus the power is higher. □

Remark 2.1.3. *By Remark 2.1.2, the tapering test based on the model of raw log-periodograms as estimates for log-spectrums will never satisfy the rate of testing criteria and according to Theorem 2.1.2, the tapering test based on non-consistent estimates of log-spectrum will always sub-optimal to the test based on model based on consistent estimates with* $\sigma \to 0$ *when* $n \to \infty$*. However, even we cannot based on Model 1 to derive the "optimal" tapering test, it still can serve as the basis for making transformations and constructing more complex model. The following sections will explore the way of doing Fourier transformation and kernel smoothing to the basis model and therefore explore the asymptotic optimality of tapering test.*

2.2 Model 2 (Fourier Transform)

The raw periodogram as the estimate of power spectrum suffers from several problems since periodogram is not a consistent estimator of spectral density. Consequently, the biggest problem with Model 1 is that σ does not goes to zero as n goes to infinity. Some transformation on Y_js and θ_js are needed to satisfy the "rate of testing" criteria or "adaptive rate of testing" criteria. In Lu and Li (2013) [17] paper, they proposed the testing scheme based on the Adaptive Neyman test in Fan (1996) [11]. The test statistic is calculated on the Fourier transform of block-average-periodograms. Model 2 here is an improvement of their idea. We summarize the periodogram locally to reduce variability and thus the "rate of testing criteria" can be satisfied. For now, let's suppose the length of these two time series $\{X_{1,t_1},\, t_1 = 1, \cdots, n_1\}$ and $\{X_{2,t_2},\, t_2 = 1, \cdots, n_2\}$ are equal, i.e. $n_1 = n_2 = n$.

Define the Fourier coefficients of Y_js and θ_js:

$$Y_j^\star = \frac{1}{p_n} \sum_{k=1}^{p_n} Y_k \psi_{j,k}$$
$$\theta_j^\star = \frac{1}{p_n} \sum_{k=1}^{p_n} \theta_k \psi_{j,k},$$

where $\psi_{j,k}$ is the Fourier basis function defined as below:

p_n is odd	p_n is even
$\psi_{1,k} = 1$ for $j = 1, \cdots, \frac{p_n - 1}{2}$ $\psi_{2j,k} = \cos\left(\frac{2\pi jk}{p_n}\right)$ $\psi_{2j+1,k} = \sin\left(\frac{2\pi jk}{p_n}\right)$	$\psi_{1,k} = 1$ for $j = 1, \cdots, \left(\frac{p_n}{2} - 1\right)$ $\psi_{2j,k} = \cos\left(\frac{2\pi jk}{p_n}\right)$ $\psi_{2j+1,k} = \sin\left(\frac{2\pi jk}{p_n}\right)$ and $\psi_{p_n,k} = \cos(k\pi)$

Let's firstly try to derive the asymptotic distribution of those newly defined $Y_j^\star$s.

Lemma 2.2.1. *For $\forall j \neq i$,*

$$\sum_{k=1}^{p_n} \cos\left(\frac{2\pi jk}{p_n}\right) \cos\left(\frac{2\pi ik}{p_n}\right) \tag{2.10}$$

$$= \sum_{k=1}^{p_n} \sin\left(\frac{2\pi jk}{p_n}\right) \sin\left(\frac{2\pi ik}{p_n}\right) \tag{2.11}$$

$$= \sum_{k=1}^{p_n} \cos\left(\frac{2\pi jk}{p_n}\right) \sin\left(\frac{2\pi ik}{p_n}\right) \tag{2.12}$$

$$= 0$$

Proof. For $\forall j \neq i$,

$$1 \leq i, j \leq \frac{p_n - 1}{2} (p_n \text{ is odd }) \text{ or } \frac{p_n}{2} - 1 (p_n \text{ is even })$$

$$\Longrightarrow 1 \leq |i - j| \leq \frac{p_n - 1}{2} - 1 (p_n \text{ is odd }) \text{ or } \frac{p_n}{2} - 2 (p_n \text{ is even })$$

$$2 \leq i + j \leq p_n - 2 (p_n \text{ is odd }) \text{ or } p_n - 3 (p_n \text{ is even })$$

And we have

$$\text{summation (2.10)} = \frac{1}{2}\sum_{j=1}^{p_n}\left[\cos\left(\frac{2\pi(j-i)}{p_n}k\right)+\cos\left(\frac{2\pi(j+i)}{p_n}k\right)\right]$$

$$\text{summation (2.11)} = \frac{1}{2}\sum_{j=1}^{p_n}\left[\cos\left(\frac{2\pi(j-i)}{p_n}k\right)-\cos\left(\frac{2\pi(j+i)}{p_n}k\right)\right]$$

$$\text{summation (2.12)} = \frac{1}{2}\sum_{j=1}^{p_n}\left[\sin\left(\frac{2\pi(j+i)}{p_n}k\right)-\sin\left(\frac{2\pi(j-i)}{p_n}k\right)\right]$$

Define $A=\frac{2\pi(j-i)}{p_n}$. Since $\sin\frac{A}{2}=0$ only when $\frac{j-i}{p_n}$ is integer which is not possible as $|i-j|<p_n-1$, then $\sin\frac{A}{2}\neq 0$. And thus

$$\begin{aligned}\sum_{j=1}^{p_n}\cos\left(\frac{2\pi(j-i)}{p_n}k\right) &= \cos A+\cos 2A+\cdots+\cos p_nA\\ &= \frac{\sin\frac{A}{2}\left(\cos A+\cos 2A+\cdots+\cos p_nA\right)}{\sin\frac{A}{2}}\\ &= \frac{\frac{1}{2}\left[\sin\left((\frac{1}{2}+p_n)A\right)-\sin\frac{A}{2}\right]}{\sin\frac{A}{2}}\\ &= 0\\ &\left(\text{Since }\sin\left((\frac{1}{2}+p_n)A\right)=\sin\left(\frac{A}{2}+2\pi(j-i)\right)=\sin\frac{A}{2}\right)\end{aligned}$$

$$\begin{aligned}\sum_{j=1}^{p_n}\sin\left(\frac{2\pi(j-i)}{p_n}k\right) &= \frac{\sin\frac{A}{2}\left(\sin A+\sin 2A+\cdots+\sin p_nA\right)}{\sin\frac{A}{2}}\\ &= \frac{\frac{1}{2}\left[\cos\frac{A}{2}-\cos\left((\frac{1}{2}+p_n)A\right)\right]}{\sin\frac{A}{2}}\\ &= 0\\ &\left(\text{Since }\cos\left((\frac{1}{2}+p_n)A\right)=\cos\left(\frac{A}{2}+2\pi(j-i)\right)=\cos\frac{A}{2}\right).\end{aligned}$$

Similarly, if we set $A=\frac{2\pi(j+i)}{p_n}$, we will get $\cos\left(\frac{2\pi(j+i)}{p_n}k\right)=0$ and $\sin\left(\frac{2\pi(j+i)}{p_n}k\right)=0$.

Thus, summation (2.10)=summation (2.11)=summation (2.12)=0. □

Lemma 2.2.2.

$$\sum_{k=1}^{p_n}\cos^2\left(\frac{2\pi jk}{p_n}\right)=\sum_{k=1}^{p_n}\sin^2\left(\frac{2\pi jk}{p_n}\right)=\frac{p_n}{2}$$

Proof. $1 \leq j \leq \frac{p_n-1}{2}(p_n \text{ is odd }) \text{ or } \frac{p_n}{2}-1(p_n \text{ is even })$

$$\sum_{k=1}^{p_n} \cos^2\left(\frac{2\pi jk}{p_n}\right) = \sum_{k=1}^{p_n}\left[\frac{1+\cos\left(\frac{4\pi jk}{p_n}\right)}{2}\right]$$
$$\sum_{k=1}^{p_n} \sin^2\left(\frac{2\pi jk}{p_n}\right) = \sum_{k=1}^{p_n}\left[\frac{1-\cos\left(\frac{4\pi jk}{p_n}\right)}{2}\right]$$

Then let's look at $\sum_{k=1}^{p_n} \cos\left(\frac{4\pi jk}{p_n}\right)$, define $B = \frac{4\pi j}{p_n}$. Since $\sin\frac{B}{2} = 0$ only when $\frac{2j}{p_n}$ is integer which is not possible as $1 \leq j \leq \frac{p_n-1}{2}(p_n \text{ is odd }) \text{ or } \frac{p_n}{2}-1(p_n \text{ is even })$, then $\sin\frac{B}{2} \neq 0$. And thus

$$\begin{aligned}\sum_{k=1}^{p_n} \cos\left(\frac{4\pi jk}{p_n}\right) &= \frac{\sin\frac{B}{2}\left(\cos B + \cos 2B + \cdots + \cos p_n Ba\right)}{\sin\frac{B}{2}} \\ &= \frac{\frac{1}{2}\left[\sin\left((\frac{1}{2}+p_n)B\right) - \sin\frac{B}{2}\right]}{\sin\frac{B}{2}} \\ &= 0 \\ &\text{(Since } \sin\left((\frac{1}{2}+p_n)B\right) = \sin\left(\frac{B}{2}+4\pi j\right) = \sin\frac{B}{2})\end{aligned}$$

, Thus summation $\sum_{k=1}^{p_n} \cos^2\left(\frac{2\pi jk}{p_n}\right) = \sum_{k=1}^{p_n} \sin^2\left(\frac{2\pi jk}{p_n}\right) = \frac{p_n}{2}$. □

Theorem 2.2.1. *If Y_j is defined before as $Y_j \xrightarrow{D} N\left(\theta_j, \sigma^2\right)$, asymptotically independent across j for $j = 1, \cdots p_n$, and $Y_j^\star$ is the Fourier coefficients of Y_j defined before. Then*

$$Y_j^\star \xrightarrow{D} N\left(\theta_j^\star, \sigma_n^2\right)$$
$$Cov(Y_j^\star, Y_i^\star) \to 0, \forall j \neq i,$$

where $\sigma_n^2 = \frac{\sigma^2}{2p_n}$.

Proof. The Fourier coefficients $Y_j^\star$ is the linear combination of Y_js, and since the distribution of Y_j is asymptotic normal, then the distribution of $Y_j^\star$ is also asymptotic normal. For the mean and variance,

$$E(Y_j^\star) = \frac{1}{p_n}\sum_{k=1}^{p_n} \psi_{j,k} E(Y_k) \to \frac{1}{p_n}\sum_{k=1}^{p_n} \psi_{j,k}\theta_k = \theta_j^\star$$
$$Var(Y_j^\star) = \frac{1}{p_n^2}\sum_{k=1}^{p_n} \psi_{j,k}^2 Var(Y_k) \to \frac{\sigma^2}{p_n^2}\sum_{k=1}^{p_n} \psi_{j,k}^2.$$

And for $\sum_{k=1}^{p_n} \psi_{j,k}^2$, by Lemma 2.2.2,

$$\sum_{k=1}^{p_n} \psi_{j,k}^2 = \sum_{k=1}^{p_n} \cos^2\left(\frac{2\pi jk}{p_n}\right) = \frac{p_n}{2}, \text{ or}$$
$$\sum_{k=1}^{p_n} \psi_{j,k}^2 = \sum_{k=1}^{p_n} \sin^2\left(\frac{2\pi jk}{p_n}\right) = \frac{p_n}{2}$$
$$\implies \quad Var(Y_j^\star) \longrightarrow \frac{\sigma^2}{p_n^2} \sum_{k=1}^{p_n} \psi_{j,k}^2 = \frac{\sigma^2}{2p_n}.$$

Then $Y_j^\star \xrightarrow{D} N\left(\theta_j^\star, \sigma_n^2\right)$, where $\sigma_n^2 = \frac{\sigma^2}{2p_n}$.

And for $j \neq i$,

$$\begin{aligned} Cov\left(Y_j^\star, Y_i^\star\right) &= Cov\left(\frac{1}{p_n}\sum_{k=1}^{p_n} Y_k\psi_{j,k}, \frac{1}{p_n}\sum_{l=1}^{p_n} Y_l\psi_{i,l}\right) \\ &= \frac{1}{p_n^2}\sum_{k=1}^{p_n}\sum_{l=1}^{p_n} \psi_{j,k}\psi_{i,l} Cov\left(Y_k, Y_l\right) \\ &= \frac{1}{p_n^2}\sum_{k=1}^{p_n} \psi_{j,k}\psi_{i,k} Cov\left(Y_k, Y_k\right) + \frac{1}{p_n^2}\sum_{k\neq l}\sum \psi_{j,k}\psi_{i,l} Cov\left(Y_k, Y_l\right) \\ &\to 0 \end{aligned}$$

(since $\frac{1}{p_n^2}\sum_{k=1}^{p_n} \psi_{j,k}\psi_{i,k} Cov\left(Y_k, Y_k\right) \to \frac{\sigma^2}{2p_n}\sum_{k=1}^{p_n} \psi_{j,k}\psi_{i,k} = 0$ for $\forall j \neq i$(Lemma 2.2.1)

$\frac{1}{p_n^2}\sum_{k\neq l}\sum \psi_{j,k}\psi_{i,l} Cov\left(Y_k, Y_l\right) \to 0$ since $Cov\left(Y_k, Y_l\right) \to 0$ for $k \neq l$)

□

Theorem 2.2.1 gives the asymptotic distribution of those $Y_j^\star$s and their asymptotic covariance. Then we can start to construct our model and deduce our tapering test statistic.

Model 2 based on Fourier transformations is constructed as follows:

$$\begin{aligned} &Y_j^\star = \theta_j^\star + \sigma_n e_j^\star \quad \text{ for } j = 1, \cdots, p_n \\ &Y_j^\star \xrightarrow{D} N\left(\theta_j^\star, \sigma_n^2\right), \text{ asymptotically independent across j ,} \end{aligned} \tag{2.13}$$

where

$Y_j^\star = \frac{1}{p_n}\sum_{k=1}^{p_n}\psi_{j,k}\log I_1(\omega_k) - \frac{1}{p_n}\sum_{k=1}^{p_n}\psi_{j,k}\log I_2(\omega_k)$ defined based on Fourier coefficients of log-periodograms

$\theta_j^\star = \frac{1}{p_n}\sum_{k=1}^{p_n}\psi_{j,k}\log f_1(\omega_k) - \frac{1}{p_n}\sum_{k=1}^{p_n}\psi_{j,k}\log f_2(\omega_k)$ defined based on Fourier coefficients of log-spectrum

$e_j^\star = \frac{Y_j^\star - \theta_j^\star}{\sigma_n}$ are asymptotic i.i.d. standard normal random variables

$\sigma_n^2 = \frac{\sigma^2}{2p_n} \to 0$ as $n \to \infty$.

Y_j, θ_j and σ are defined the same as in Model 1.

Note that now $\sigma_n^2 \to 0$ as $n \to \infty$. We have achieved a vanishing variance. And this is what makes it possible to even think about rates of testing criteria.

Now the hypothesis we want to test becomes

$$H_0 : \theta_j^\star = 0 \text{ for all } j = 1, \cdots, p_n$$

$$H_1 : \theta_j^\star \neq 0 \text{ for some } j = 1, \cdots, p_n$$

Then tapering test statistic is constructed in the form of a tapered sum of squared summary coefficients $Y_j^\star$s,

$$Q_n^\star = \sum_{j=1}^{p_n} w_{n,j} Y_j^{\star 2},$$

whose wights $w_{n,j}$ are presumed to taper to zero in the sense that $w_{n,j} \to 0$ as $j \to \infty$. For such asymptotic statements to make sense, we think conceptually that p_n is arbitrarily large, however, in practice p_n is finite and determined by data configuration.

Remark 2.2.1. *As defined before, the "rate" in rate of testing theory, δ_n, defines the gap between the null hypothesis and the class of distinguishable alternatives, where the test can detect with high power, asymptotically. The class of distinguishable alternatives, $\boldsymbol{\theta}^\star \in H_1(\delta_n/\delta_n^\star; s, M)$ are defined based on the smoothness assumption, where*

$$H_1(\delta; s, M) = \left\{ \boldsymbol{\theta} \in \mathcal{B}_{s,M} : \sqrt{\sum_{j=1}^{\infty} \theta_j^{\star 2}} \geq \delta \right\}.$$

Thus δ_n characterized the test performance and the optimal testing procedures should achieve the fastest rate of δ_n among the same testing mechanism.

Here, smoothness constrains are defined as a restriction on the parameter $\boldsymbol{\theta}$ to a smooth-function class, $\boldsymbol{\theta} \in \mathcal{B}_{s,M}$.

Note that the smoothness constraint is based on θ_js, however, the "distinguishable region" $H_1(\delta; s, M)$ is defined on $\theta_j^\star$s. The following lemma will link these two definitions and provide the basis for future proofs for "rate of testing" criteria. This lemma represents one of the key innovations, and original contributions, of our work.

Lemma 2.2.3. *The smoothness constraint on $\boldsymbol{\theta}$ and the smoothness constraint on $\boldsymbol{\theta}^\star$ have the following relationship:*

$$\boldsymbol{\theta} \in \mathcal{B}_{s,M} \Rightarrow \boldsymbol{\theta}^\star \in \mathcal{B}_{s,M^\star}.$$

Proof.

$$\begin{aligned}
& \boldsymbol{\theta}^\star \in \mathcal{B}_{s,M^\star} \\
\iff & \sum_{j=1}^{\infty} j^{2s}\theta_j^{\star 2} \leq M^{\star 2} \\
\iff & \boldsymbol{\theta}^{\star T}\widetilde{J}^\star\boldsymbol{\theta}^\star \leq M^{\star 2} \\
\iff & tr\left(\boldsymbol{\theta}^{\star T}\widetilde{J}^\star\boldsymbol{\theta}^\star\right) \leq M^{\star 2} \\
\iff & tr\left(\boldsymbol{\theta}^T F^T \widetilde{J}^\star F\boldsymbol{\theta}\right) = tr\left(\boldsymbol{\theta}^T \widetilde{J}^\star\boldsymbol{\theta}F^T F\right) \leq M^{\star 2},
\end{aligned}$$

where $\widetilde{J}^\star$ is the $p_n \times p_n$ diagonal matrix

$$\widetilde{J}^\star = \begin{pmatrix} 1^{2s} & & & \\ & 2^{2s} & & \\ & & \ddots & \\ & & & p_n^{2s} \end{pmatrix}$$

and F is the $p_n \times p_n$ matrix of Fourier basis functions

$$F = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,p_n} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,p_n} \\ \vdots & \vdots & & \vdots \\ \psi_{p_n,1} & \psi_{p_n,2} & \cdots & \psi_{p_n,p_n} \end{pmatrix}$$

Since

$$tr\left(\boldsymbol{\theta}^T \widetilde{J}^\star\boldsymbol{\theta}F^T F\right) \leq tr\left(\boldsymbol{\theta}^T \widetilde{J}^\star\boldsymbol{\theta}\right) tr\left(F^T F\right),$$

and

$$\begin{aligned} tr\left(F^T F\right) &= \sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\psi_{ij}^2 \\ &\leq \left(\frac{p_n}{2}+1\right)p_n, \end{aligned}$$

Then

$$tr\left(\boldsymbol{\theta}^T\widetilde{J}^{\star}\boldsymbol{\theta}F^T F\right) \leq tr\left(\boldsymbol{\theta}^T\widetilde{J}^{\star}\boldsymbol{\theta}\right)\left(\frac{p_n}{2}+1\right)p_n.$$

And since

$$\begin{aligned} & \boldsymbol{\theta} \in \mathcal{B}_{s,M} \\ \iff & \sum_{j=1}^{\infty} j^{2s}\theta_j^2 \leq M^2 \\ \iff & \boldsymbol{\theta}^T\widetilde{J}^{\star}\boldsymbol{\theta} \leq M^2 \\ \iff & tr\left(\boldsymbol{\theta}^T\widetilde{J}^{\star}\boldsymbol{\theta}\right) \leq M^2, \end{aligned}$$

Define the fixed constant $M^{\star 2} = \left(\frac{p_n}{2}+1\right)p_n M^2$, then our smoothness constraint

$$\boldsymbol{\theta} \in \mathcal{B}_{s,M} \Rightarrow \boldsymbol{\theta}^{\star} \in \mathcal{B}_{s,M^{\star}}.$$

□

From our specification of model (2.13), our definition of σ_n^2, Remark 2.2.1 and Lemma 2.2.3, we can find that our current context is quite similar to the context in Spitzner (2008) [28] and our Model 2 is similar to the model (1.3) as in this paper. So, similar asymptotic results can be derived and the corresponding theorems and proofs can be found in Spitzner (2008b) [29].

Remark 2.2.2. *Based on the model of $Y_j^\star$, the test is specified through the tapering test statistic $Q_n^\star = \sum_{j=1}^{p_n} w_{n,j} Y_j^{\star 2}$, which rejects null hypothesis when $Q_n^\star$ is large. The Fourier coefficients are specified as before. w_j is the regular tapering scheme. Define known constants $1/2 < s_\star < s^\star$ and*

$$\widehat{\delta}_n^F(s) = \left\{ p_n^2 \left(\log p_n\right)^{-1} \right\}^{-s/(4s+1)},$$

then no tapering test will satisfy the "rate of testing" criteria with rate $\delta_n = o\left(\widehat{\delta}_n^F(s)\right)$, i.e. $\widehat{\delta}_n^F(s)$ is the optimal adaptive rate of testing for the tapering mechanism. Additionally, when

$$w_{n,j} = j^{-1/2},$$

then the test $Q_n^\star$ will satisfy the rate of testing criteria for $\delta_n = \widehat{\delta}_n^F(s)$ across $s_\star < s < s^\star$.

2.3 Model 3 (Kernel Smoothing)

In Section 2.1, we introduced the basis model based on the raw periodogram (log-periodogram) as the estimate of power spectrum (log-spectrum). Since Model 1 will never satisfy the "rate of testing" criteria, we explored the Fourier transform in basis model and proposed the Model 2 in Section 2.2, inspired by Lu and Li (2013) [17] paper. Based on Model 2, we investigated the asymptotic optimality property and derived the "optimal" tapering test from the "optimal weights" $w_{n,j} = j^{-1/2}$ in the definition of tapering test. However, Fourier transform of log-periodogram as estimates for log-spectrum doesn't have a very meaningful interpretation in time series context, instead, kernel smoothing of the periodogram (log-periodogram) is more acceptable. See for example Priestley (1981)[20], Brockwell and Davis (1991) [6], and Brillinger (2001) [5]. There are three different ways to smooth the periodogram (log-periodogram): local average smoothing, lag window smoothing and periodogram averaging (Welch's-Bartlett spectral estimation method). But the Welch's-Bartlett method lead to the spectral leakage and bias. Considering that the objective of our research is to develop a powerful testing procedure which is convenient to use even for people with little statistical background, the simple moving averaged log-periodogram as the estimate for log-spectrum will be used, which is equivalent to do kernel smoothing with the Uniform Kernel. Since the number of Fourier frequencies in a given interval increases approximately linearly with n, we can indeed construct consistent log-spectrum estimators by averaging over a suitably increasing

number of frequencies in a neighborhood. Firstly, let's define our problem based on the kernel smoothing.

The advantages of kernel smoothing is twofold: firstly, a smoothed periodogram, when suitably constructed, is an asymptotically consistent estimate of its underlying spectral density (provided that density is also smooth), which admits consideration of asymptotic performance concepts used in rates of testing theory; secondly, a smoothed periodogram may be calculated at any frequency, ω, which provides a means of extension to the case of unequal time series lengths, $n_1 \neq n_2$.

Let's firstly define the kernel smoothing function as

$$C_{i,j,k_i} = K\left(\frac{\omega_{ik_i} - \omega_j}{b_i}\right) \Bigg/ \sum_{k_i=1}^{p_i} K\left(\frac{\omega_{ik_i} - \omega_j}{b_i}\right),$$

for frequencies ω_j spread on an even grid between 0 and π. And $w_{i,k_i} = 2\pi k_i/n_i$, n_1 and n_2 are lengths of these two time seires, $K(.)$ is a specified kernel function, satisfying $K(t) \to 0$ as $t \to \pm\infty$. In addition, b_i are bandwidth parameters, and note that for these two time series, the bandwidth parameter can be different. In the present investigation, the grid is $w_j = 2\pi j/n$ across $j = 1, \cdots, p$, where n is the minimum of n_1 and n_2 and p is minimum of p_1 and p_2. Then define the kernel smoothed log-periodogram:

$$\log \widetilde{I}_i(\omega_j) = \sum_{k_i=1}^{p_i} C_{i,j,k_i} \log I_i(\omega_j)$$

In the proposed methodology, the Y_j are defined according to

$$\widetilde{Y}_j = \log \widetilde{I}_1(\omega_j) - \log \widetilde{I}_2(\omega_j),$$

for frequencies ω_j spread on an even grid between 0 and π. In the present investigation, the grid is $\omega_j = 2\pi j/n$ across $j = 1, \cdots, p_n$, where n is the minimum of n_1 and n_2 and p_n is the minimum of $p_{n,1}$ and $p_{n,2}$.

By Brockwell and Davis (1991) [6], it is easy to prove that $\widetilde{Y}_j \xrightarrow{D} N\left(\theta_j, \sigma_n'^2\right)$ asymptotically independent across j, where $\sigma_n'^2 = Var(\widetilde{Y}_j) = \frac{\sigma^2}{2}\left\{\sum_{k_1=1}^{p_1} C_{1,j,k_1}^2 + \sum_{k_2=1}^{p_2} C_{2,j,k_2}^2\right\}$.

To reflect the asymptotic framework, the subsequent notation is updated to indicate a dependency of relevant quantities on n, so that, e.g., $b_i = b_{n,i}$, $\widetilde{Y}_j = \widetilde{Y}_{n,j}$, etc. Limiting behavior is defined with respect to $n \to \infty$, assuming that both n_1/n_2 and n_2/n_1 are bounded.

The bandwidth parameter in the kernel function, $b_{n,i}$ is a positive integer depends on n_i, which is also denoted as "bandwidth" parameter in kernel smoothing. And the way we deduce the optimality of tapering test is to derive the optimal rate for the bandwidth parameter $b_{n,i}$. In order for this estimate of the spectral density to be consistent, we impose the following conditions on b and the kernel function:

$$b_{n,i} \to \infty \quad and \quad b_{n,i}/n_i \to 0 \quad as\, n_i \to \infty$$
$$\sum_{k_i=1}^{p_{n,i}} C_{i,j,k_i}^2 \to 0 \quad as\, n_i \to \infty.$$

The above conditions ensure that the discrete spectral average estimator converges to true value and its covariance converges to zero. That is,

$$\begin{aligned} Cov\left(\log \widetilde{I}_1(\omega_j), \log \widetilde{I}_2(\omega_k)\right) \to 0, \;\; if \;\; j \neq k \\ \Rightarrow \quad Cov\left(\widetilde{Y}_{n,j}, \widetilde{Y}_{n,k}\right) \to 0, \;\; if \;\; j \neq k \\ \Rightarrow \quad Cov\left(\widetilde{Y}_{n,j}^2, \widetilde{Y}_{n,k}^2\right) \to 0, \;\; if \;\; j \neq k \end{aligned}$$

as n_i sufficiently large. Note that the rate of vanishing moments is important here, because in our mathematical results we will work with sums of coefficients, whose number of terms increase with p_n, and thus are affected by aggregate covariance, which could grow out of control if individual covariances do not vanish quickly enough. See Brockwell and Davis (1991) [6], and Brillinger (2001) [5].

Remark 2.3.1. *Since the form of the kernel is not our focus, instead, we are interested in deriving the optimal tuning parameter in the kernel smoothing so as to do the tapering test based on Model 2. More importantly, our target is to propose a test with simple form but good performance regarding to both asymptotic and empirical powers. So we would rather use the simplest kernel which is easiest to understand as well as implement in practice and try to optimize the tapering test based on this kernel. In the following theorems and proofs, we will take the kernel defined in before as general case and the Uniform discrete compact kernel as the special case.*

Uniform Kernel *means* $K(t_l) = 1_{\{t_l \leq 1\}}$*, which is equivalent to simple local moving average, as a common technique to smooth the periodogram (log-periodogram), see Priestley (1981)[20], Brockwell and Davis (1991) [6], and Brillinger (2001) [5].*

Model 3 based on the kernel smoothing of log-periodogram as estimates for log-spectrum is

defined as follows:

$$\widetilde{Y}_{n,j} = \theta_j + \sigma'_n \widetilde{\epsilon}_j \quad \text{for } j = 1, \cdots, p_n \tag{2.14}$$

$$\widetilde{Y}_{n,j} \xrightarrow{D} N\left(\theta_j, \sigma'^2_n\right), \text{ asymptotically independent across j ,}$$

where

$$\begin{aligned}
\widetilde{Y}_{n,j} &= \log \widetilde{I}_1(\omega_j) - \log \widetilde{I}_2(\omega_j) \\
\theta_j &= \log f_1(\omega_j) - \log f_2(\omega_j) \\
\widetilde{\epsilon}_j &= \frac{\widetilde{Y}_{n,j} - \theta_j}{\sigma'_n} \text{ are asymptotic standard normal random variables} \\
\sigma'^2_n &= \frac{\sigma^2}{2} \left\{ \sum_{k_1=1}^{p_1} C^2_{1,j,k_1} + \sum_{k_2=1}^{p_2} C^2_{2,j,k_2} \right\} \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Note that now $\sigma'^2_n \to 0$ as $n \to \infty$. We have achieved a vanishing variance. And this is what makes it possible to even think about rates of testing criteria.

Similarly to the setting in Model 1, to test the hypothesis of equal spectral densities is equivalent to test

$$H_0 : \theta_j = 0 \text{ for all } j = 1, \cdots, p_n$$

$$H_1 : \theta_j \neq 0 \text{ for some } j = 1, \cdots, p_n$$

Finally, the corresponding tapering test statistic is constructed as

$$\widetilde{Q}_n = \sum_{j=1}^{p_n} j^{-1/2} \widetilde{Y}^2_{n,j}.$$

Additionally, in order for this estimate of the spectral density to be consistent, we impose the following conditions on $b_{n,i}$ and the kernel function:

$$\begin{aligned}
&b_{n,i} \to \infty \quad and \quad b_{n,i}/n_i \to 0 \quad as\, n_i \to \infty \\
&\sum_{k_i=1}^{p_{n,i}} C^2_{i,j,k_i} \to 0 \quad as\, n_i \to \infty
\end{aligned}$$

The above conditions ensure that the kernel smoothed spectral estimator converges to true value

and its covariance converges to zero. And by Lemma 2.1.1,

$$\begin{aligned} & Cov\left(\log \widetilde{I}_1(\omega_j), \log \widetilde{I}_2(\omega_k)\right) \to 0, \ \ if \ \ j \neq k \\ \Rightarrow \quad & Cov\left(Y_{n,j}, Y_{n,k}\right) \to 0, \ \ if \ \ j \neq k \\ \Rightarrow \quad & Cov\left(Y_{n,j}^2, Y_{n,k}^2\right) \to 0, \ \ if \ \ j \neq k \end{aligned}$$

as n_i sufficiently large. See Brockwell and Davis (1991) [6], and Brillinger (2001) [5].

To look the "rate of testing" criteria, we need to investigate the asymptotic power of our test. Thus the following remark is very important as it defines the "distinguishable region" of our test, i.e. $H_1(\delta_n/\delta_n^\star; s, M)$.

Remark 2.3.2. *The "rate" in rate of testing theory, δ_n, defines the range of the indistinguishable region where the test cannot detect with high power, asymptotically. The class of "distinguishable" alternatives, $H_1 : \boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)$ are defined based on the smoothness assumption, where*

$$H_1(\delta; s, M) = \left\{ \boldsymbol{\theta} \in \mathcal{B}_{s,M} : \sqrt{\sum_{j=1}^{\infty} \theta_j^2} \geq \delta \right\}.$$

Thus δ_n characterized the test performance and the optimal testing procedures should achieve the fastest rate of $\delta_n \to 0$ among the same testing mechanism.

To take a closer look at the "rate of testing" criteria, let's firstly write it into two equivalent conditions involving the asymptotic mean and variance of Q_n. Similarly to the proof of Lemma 2.3.1, we can prove the following lemma.

Lemma 2.3.1. *For fixed smoothness parameter $s > 1/2$ and $M > 0$, and some positive sequence $\delta_n \to 0$, the rate of testing criteria holds for the test statistic $Q_n = \sum_{j=1}^{p_n} j^{-1/2} Y_j^2$ if and only if*

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} E_{H_{1-0}}(Q_n) / \sqrt{Var_{H_0}(Q_n)} \to \infty \quad and \tag{2.15}$$

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} E_{H_{1-0}}(Q_n) / \sqrt{Var_{H_{1-0}}(Q_n)} \to \infty, \tag{2.16}$$

where $E_{H_{1-0}}(Q_n) = E_{H_1}(Q_n) - E_{H_0}(Q_n)$ and $Var_{H_{1-0}}(Q_n) = Var_{H_1}(Q_n) - Var_{H_0}(Q_n)$.

The following lemma will be used in the proof of our main results, Theorem 2.3.1

Lemma 2.3.2.

$$\sum_{j=1}^{p_n} j^{-1} \asymp \log p_n$$

Proof. For $\sum_{j=1}^{p_n} j^{-1}$, consider the area of rectangles with the curve $y = 1/x$, we will get $\sum_{j=1}^{p_n} j^{-1} > \int_1^{p_n+1} \frac{1}{x} dx = \log(p_n + 1)$. And since the partial sums of the series have logarithmic growth, in particular $\sum_{j=1}^{p_n} j^{-1} < \log(p_n) + 1$. Thus we will have $1 < \frac{\log(p_n+1)}{\log p_n} < \frac{\sum_{j=1}^{p_n} j^{-1}}{\log p_n} < \frac{\log(p_n)+1}{\log p_n} = 1 + \frac{1}{\log p_n} \leq 1 + \frac{1}{\log 2}$, thus $\sum_{j=1}^{p_n} j^{-1} \asymp \log p_n$. □

The following theorem serves as the main results and the key contributions of our research. We write the "rate of testing" criteria into two equivalent conditions involving the unknown parameters of our interest and do some manipulations to finally get the "optimal" bandwidth parameter and "optimal" rate.

Theorem 2.3.1. *Based on the model of $\widetilde{Y}_j$, the test is specified through the tapering test statistic $\widetilde{Q}_n$, which rejects null hypothesis when $\widetilde{Q}_n$ is large. Define $\widehat{b}_{n,i}$ to be the setting of the bandwidth parameter $b_{n,i}$, such that*

$$\min_{i,j} C_{i,j,k_i} = \left\{ p_n^{4s+1} \log p_n \right\}^{-1/4}, \tag{2.17}$$

where $\min_{i,j} C_{i,j,k_i}$ is the minimal non-zero value of C_{i,j,k_i} across $i = 1, 2$, and $j = 1, \cdots, p_n$. Fix $b_{n,i} = \widehat{b}_{n,i}$, at $n = 1, 2$ and define known constants $1/2 < s_\star < s^\star$ and the sequence

$$\widehat{\delta}_n(s) = \left\{ \sigma_n'^4 \log p_n \right\}^{s/(4s+1)}. \tag{2.18}$$

then no tapering test will satisfy the "rate of testing" criteria with rate $\delta_n = o\left(\widehat{\delta}_n(s)\right)$, i.e. $\widehat{\delta}_n(s)$ is the optimal adaptive rate of testing for the tapering mechanism. Additionally, when $b_{n,i} = \widehat{b}_{n,i}$ for $i = 1, 2$, then the test $\widetilde{Q}_n$ will satisfy the rate of testing criteria for $\widehat{\delta}_n(s)$ across $s_\star < s < s^\star$.

Proof. By Lemma 2.3.1, to find the equivalent conditions for the adaptive rate of testing criteria, we need to look at condition (2.15) and (2.16). Let's look at the mean and variance of the test

statistic $\widetilde{Q}_n$.

$$\begin{aligned}
E\left(\widetilde{Q}_n\right) = &\quad E(\sum_{j=1}^{p_n} j^{-1/2} Y_{n,j}^2) \\
= &\quad \sum_{j=1}^{p_n} j^{-1/2} E(Y_{n,j}^2) \\
= &\quad \sum_{j=1}^{p_n} j^{-1/2} \left\{ Var(Y_{n,j}) + [E(Y_{n,j})]^2 \right\} \\
\longrightarrow &\quad \sum_{j=1}^{p_n} j^{-1/2} \left\{ \sigma_n'^2 + \theta_j^2 \right\} \\
= &\quad \sum_{j=1}^{p_n} j^{-1/2} \sigma_n'^2 + \sum_{j=1}^{p_n} j^{-1/2} \theta_j^2 \\
Var\left(\widetilde{Q}_n\right) = &\quad Var(\sum_{j=1}^{p_n} j^{-1/2} Y_{n,j}) \\
= &\quad \sum_{j=1}^{p_n} j^{-1} Var(Y_{n,j}^2) + \sum\sum_{j \neq k} j^{-1/2} k^{-1/2} Cov(Y_{n,j}^2, Y_{n,k}^2) \\
\longrightarrow &\quad \sum_{j=1}^{p_n} j^{-1} Var(Y_{n,j}^2) \quad \text{(Lemma 2.2.1)} \\
= &\quad \sum_{j=1}^{p_n} j^{-1} \left\{ E(Y_{n,j}^4) + \left[E(Y_{n,j}^2)\right]^2 \right\} \\
\longrightarrow &\quad \sum_{j=1}^{p_n} j^{-1} \left\{ 4\theta_j^2 \sigma_n'^2 + 2\sigma_n'^4 \right\} \\
= &\; 4\sigma_n'^2 \sum_{j=1}^{p_n} j^{-1} \theta_j^2 + 2\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}
\end{aligned}$$

Then condition (2.15) and (2.16) becomes

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} \frac{\sum_{j=1}^{p_n} j^{-1/2} \theta_j^2}{\sqrt{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}}} \to \infty \quad \text{and} \tag{2.19}$$

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} \frac{\sum_{j=1}^{p_n} j^{-1/2} \theta_j^2}{\sqrt{\sigma_n'^2 \sum_{j=1}^{p_n} j^{-1} \theta_j^2}} \to \infty \tag{2.20}$$

Afterwards, let's look at the condition (2.20). Since

$$\begin{aligned}
&\sum_{j=1}^{p_n} j^{-1}\theta_j^2 < \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 \\
\Longrightarrow \quad &\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 \Big/ \sqrt{\sigma_n'^2 \sum_{j=1}^{p_n} j^{-1}\theta_j^2} > \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 \Big/ \sqrt{\sigma_n'^2 \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2} \\
\Longrightarrow \quad &\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 \Big/ \sqrt{\sigma_n'^2 \sum_{j=1}^{p_n} j^{-1}\theta_j^2} > \sqrt{\frac{\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2}{\sigma_n'^2}} = A_n.
\end{aligned}$$

And if condition (2.19) is satisfied, since $\sum_{j=1}^{p_n} j^{-1} \to \infty$ as $n \to \infty$, then $A_n \to \infty$ as $n \to \infty$. And thus condition (2.20) is satisfied. Thus condition (2.19) implies condition (2.20).

Therefore, we only need to look at condition (2.19). Now, for condition (2.19), since the denominator is not related with θ_js, let's firstly look at the lower bound of $\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2$ under the constraint $\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)$. Since

$$\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 = \sum_{j=1}^{\infty} j^{-1/2}\theta_j^2 - \sum_{j=p_n+1}^{\infty} j^{-1/2}\theta_j^2.$$

Let's firstly look at $\sum_{j=1}^{\infty} j^{-1/2}\theta_j^2$. By Langrange multiplier, we can conclude that the lower bound of $\sum_{j=1}^{\infty} j^{-1/2}\theta_j^2$ subject to the constraint $\boldsymbol{\theta} \in \mathcal{B}_{s,M}$ is achieved by

$$\theta_j = \begin{cases} \frac{\delta_n}{\delta_n^\star} & j = \widehat{j} \\ 0 & o.w. \end{cases},$$

where $\widehat{j}$ is the largest index between 1 and ∞ that satisfies $\widehat{j} \le \left(\frac{\delta_n^\star M}{\delta_n}\right)^{1/s}$. Then the lower bound of $\sum_{j=1}^{\infty} j^{-1/2}\theta_j^2$ is $\widehat{j}^{-1/2}\left(\frac{\delta_n}{\delta_n^\star}\right)^2$. And since

$$\begin{aligned}
\sum_{j=p_n+1}^{\infty} j^{-1/2}\theta_j^2 &= \sum_{j=p_n+1}^{\infty} j^{-1/2} j^{-2s}\left(j^{2s}\theta_j^2\right) \\
&\le p_n^{-(2s+1/2)} \sum_{j=p_n+1}^{\infty} j^{2s}\theta_j^2 \\
&\le p_n^{-(2s+1/2)} M^2 \\
&\quad \text{(since } \boldsymbol{\theta} \in \mathcal{B}_{s,M} \quad i.e. \quad \sqrt{\sum_{j=1}^{\infty} j^{2s}\theta_j^2} \le M^2\text{)}.
\end{aligned}$$

Additionally, when θ_j is specified as before and $\widehat{j} = \left(\frac{\delta_n^\star M}{\delta_n}\right)^{1/s} > p_n$ then $\sum_{j=p_n+1}^{\infty} j^{2s}\theta_j^2 = M^2$, and $\sum_{j=p_n+1}^{\infty} j^{-1/2}\theta_j^2$ achieves the upper bound as $p_n^{-(2s+1/2)}M^2$.

Thus the lower bound of $\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2$ subject to the constraint $\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)$ is

$$\begin{aligned}\inf_{\boldsymbol{\theta}\in H_1(\delta_n/\delta_n^\star;s,M)} \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 &= \widehat{j}^{-1/2}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 - p_n^{-(2s+1/2)}M^2 \\ &\geq \left(\frac{\delta_n}{\delta_n^\star}\right)^{\frac{4s+1}{2s}} M^{-\frac{1}{2s}} - p_n^{-\frac{4s+1}{2}}M^2,\end{aligned}$$

which is achieved by

$$\theta_j = \begin{cases} \frac{\delta_n}{\delta_n^\star} & j = \left(\frac{\delta_n^\star M}{\delta_n}\right)^{1/s} > p_n \\ 0 & o.w. \end{cases}$$

Thus condition (2.19) becomes

$$\frac{\left(\frac{\delta_n}{\delta_n^\star}\right)^{\frac{4s+1}{2s}} - p_n^{-\frac{4s+1}{2}}}{\sqrt{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}}} \to \infty$$

$\Leftrightarrow$

$$\lim_{n\to\infty} \frac{\left(\frac{\delta_n}{\delta_n^\star}\right)^{\frac{4s+1}{2s}}}{\sqrt{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}}} = \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{p_n^{-\frac{4s+1}{2}}}{\sqrt{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}}} < \infty$$

(the second condition is needed because the rate of the first condition $\to \infty$ can be arbitrarily slowed down by $\forall \delta_n^\star \to 0$)

And then the first condition take the square is equivalent to

$$\lim_{n\to\infty} \left\{ \frac{1}{\delta_n^{\star\frac{4s+1}{s}}} \right\} \frac{\delta_n^{\frac{4s+1}{s}}}{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}} = \infty.$$

Since this condition is valid for $\forall \delta_n^\star \to 0$, then the item in braces tends to infinity but with arbitrarily slow rate. Then the condition is equivalent to

$$\liminf_{n\to\infty} \frac{\delta_n^{(4s+1)/s}}{\sigma_n'^4 \sum_{j=1}^{p_n} j^{-1}} > 0.$$

And by Lemma 2.3.2, this condiiton is equivalent to

$$\liminf_{n\to\infty} \frac{\delta_n^{(4s+1)/s}}{\sigma_n'^4 \log p_n} > 0. \tag{2.21}$$

For the second condition, take the square, by Lemma 2.3.2, and based on the definition of $\min_{i,j} C_{i,j,k_i}$, this condition is equivalent to

$$\limsup_{n\to\infty} \frac{p_n^{-(4s+1)}}{\min_{i,j} C_{i,j,k_i}^4 \log p_n} < \infty. \tag{2.22}$$

For $\underset{i,j}{min}C_{i,j,k_i}$, it is defined as *minimally allowed*, if it is the fastest rate goes to zero that satisfies condition (2.22).

For the positive sequence $\delta_n \to 0$, it is defined as *minimally allowed*, if it is the fastest rate decaying to zero that satisfies condition (2.21). And if the condition (2.22) is also satisfied, then δ_n is the *minimax rate* we are looking for.

Then let's prove that $\underset{i,j}{min}C_{i,j,k_i}$ is minimally allowed, if and only if

$$\frac{p_n^{-(4s+1)}}{\underset{i,j}{min}C_{i,j,k_i}^4 \log p_n} \asymp 1. \tag{2.23}$$

It is easy to prove that condition (2.23) is a necessary condition. To prove it is a sufficient condition, if exits another sequence $\underset{i,j}{min}C_{i,j,k_i}^\star \to 0$ satisfies condition (2.23) and $\underset{i,j}{min}C_{i,j,k_i}^\star = o(\underset{i,j}{min}C_{i,j,k_i})$. Then

$$\frac{p_n^{-(4s+1)}}{\underset{i,j}{min}C_{i,j,k_i}^{\star 4} \log p_n} = \frac{p_n^{-(4s+1)}}{\underset{i,j}{min}C_{i,j,k_i}^4 \log p_n} \left(\frac{\underset{i,j}{min}C_{i,j,k_i}}{\underset{i,j}{min}C_{i,j,k_i}^\star} \right)^4 \to \infty,$$

and thus condition (2.22) dose not hold. Therefore, condition (2.23) is equivalent to condition (2.22).

Then let's prove that the sequence δ_n is minimally allowed, if and only if

$$\frac{\delta_n^{(4s+1)/s}}{\sigma_n'^4 \log p_n} \asymp 1. \tag{2.24}$$

It is easy to prove that condition (2.23) is a necessary condition. To prove it is a sufficient condition, if exits another sequence $\delta_n^\star \to 0$ satisfies condition (2.24 and $\delta_n^\star = o(\delta_n)$. Then

$$\frac{\delta_n^{\star(4s+1)/s}}{\sigma_n'^4 \log p_n} = \frac{\delta_n^{(4s+1)/s}}{\sigma_n'^4 \log p_n} \left(\frac{\delta_n^\star}{\delta_n} \right)^{(4s+1)/s} \to 0$$

and thus condition (2.21) dose not hold. Therefore, condition (2.24) is equivalent to condition (2.21).

Then the "rate of testing" criteria is equivalent to condition (2.23) and condition (2.24). From condition (2.23), we can get the optimal bandwidth in equation (2.17) and minimax rate in equation (2.18). □

Note that this is the theorem for the general kernel functions and these the log-periodogram can be smoothed with different bandwidth parameter. By plugging the specific form of the kernel

function into Equation (2.17), we can derive the corresponding "optimal" bandwidth $\widehat{b}_{n,i}$ for that kernel. Additionally, when we need to smooth differently for these two time series, the results in Equation (2.17) becomes the relationship of the "optimal" bandwidth parameter of these two time series, i.e., $\widehat{b}_{n,1}$ and $\widehat{b}_{n,2}$.

Remark 2.3.3. *In Theorem 2.3.1, we have developed the tapering test procedure for the general case when n_1 is not necessary equal to n_2 and/or $b_{n,1}$ is not necessary equal to $b_{n,2}$. However, in our empirical work, driven by the needs of the motivating SEEG dataset, let's consider the scenario that the length of these two time series are equal and the bandwidth for these two time series are equal, i.e. $n_1 = n_2 = n$, $b_{n,1} = b_{n,2} = b_n$.*

Remark 2.3.4. *Assume the notation and conditions of Theorem 2.3.1, for the Uniform Kernel (i.e. simple moving average), the "optimal bandwidth" is*

$$\widehat{b}_n = B\left\lfloor \frac{1}{2} p_n^{3/4} \left(\log p_n\right)^{1/4} - \frac{1}{2} \right\rfloor, \tag{2.25}$$

where the lower bound of smoothness parameter s is used to construct the rate, $B > 0$ is a leading constant and $\lfloor a \rfloor$ denotes the integer part of a number. And the "optimal adaptive rate of testing" is

$$\widehat{\delta}_n(s) = \left\{ p_n^{4s+1} \left(\log p_n\right)^{-1} \right\}^{-s/(2(4s+1))}. \tag{2.26}$$

Remark 2.3.5. *However, in Equation (2.25), $\left\lfloor \frac{1}{2} p_n^{3/4} \left(\log p_n\right)^{1/4} - \frac{1}{2} \right\rfloor$ is just the "optimal" rate of bandwidth for n goes to infinity. To get the value of bandwidth in practice for a specific value of n, we add the leading constant B into $\widehat{b}_n$. The "optimal" leading constant is defined as the value of B that results in optimal power of the tapering test across different value of p_n, since our main consideration of performance is the power. However, it's challenging to think about the approach that can address this question asymptotically. Here for simplicity, in our empirical work, we will set $B = 1$ as it is often used by convention, see Ma and He (2016) [19] for similar specifications. Additionally, in the following chapter, we will show that setting $B = 1$, our proposed tapering test $\widetilde{Q}_n$ exhibits very strong empirical performance across wide ranges of value of p_n.*

Remark 2.3.6. *Note that the "optimal bandwidth" and "optimal adaptive rate of testing" can be*

derived by plugging any kernel function into Equation (2.17) and (2.18). For the simplicity of our testing procedure, Uniform kernel is used and the corresponding "optimal bandwidth" $\widehat{b}_n$ *and "optimal adaptive rate of testing"* $\widehat{\widehat{\delta}}_n$ *are deduced as in Corollary 2.3.4. Other forms of kernel can also be plugged into these two conditions in Theorem 2.3.1.*

For example, for Epanechnikov kernel with

$$K(t_l) = \frac{3}{4}(1 - t_l^2)1_{\{t_l \leq 1\}}$$

the optimal bandwidth is

$$\widehat{b}_n = B\left\lfloor p_n^{3/8} \left(\log p_n\right)^{1/8} \right\rfloor$$

and the "optimal adaptive rate of testing" is

$$\widehat{\delta}_n(s) = \left\{ p_n^{4s+1} \left(\log p_n\right)^{-3} \right\}^{-s/(4(4s+1))}.$$

2.4 Connection Between Fourier Transform and Kernel Smoothing

In Section 2.2 we introduced the model based on Fourier transform of the log periodogram as the consistent estimate of log spectral density, and in Section 2.3 based on "optimal" kernel smoothing of the log periodogram as the consistent estimate of log spectral density. As sometimes Fourier transformation is equivalent to kernel smoothing under certain form of the kernel, we will now try to look at whether this connection is applicable in our case.

For any unknown ***probability density function*** $f(x)$ of a direction x, it is ***periodic with period*** 2π. We have i.i.d. random sample $\{x_k\}$ from f, $k = 1, \cdots, n$. $\phi_j = \int_{-\pi}^{\pi} f(x) e^{-ijx} d_x = E\left(e^{-ijx}\right)$ for each j are called ***Fourier coefficients*** of f. The collection of ϕ_j for all j is called the ***Fourier transform*** of f. Since ϕ_js are unknown, we have estimates $\widehat{\phi}_j = E_n\left(e^{-ijx}\right) = \frac{1}{n}\sum_{k=1}^{n} e^{-ijx_k}$. If we reconstruct f from its Fourier transform at any point x where it is continuous, we will get the ***Fourier series density estimate*** g. And after doing some transformations, we

have

$$\begin{aligned} g(x) &= \frac{1}{2\pi}\sum_{j=-m}^{m}\widehat{\phi}_j e^{ijx} \\ &= \frac{1}{2\pi}\sum_{j=-m}^{m}\Big(\frac{1}{n}\sum_{k=1}^{n}e^{-ijx_k}\Big)e^{ijx} \\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(\frac{1}{2\pi}\sum_{j=-m}^{m}e^{ij(x-x_k)}\Big) \\ &= \frac{1}{n}\sum_{k=1}^{n}K_m\big(x-x_k\big), \end{aligned}$$

with $K_m\big(y\big) = \frac{1}{2\pi}\sum_{j=-m}^{m}e^{ijy} = \frac{1}{2\pi}\frac{\sin\big[\big(m+\frac{1}{2}\big)y\big]}{\sin\big(\frac{1}{2}y\big)}$ is called a ***Poisson Kernel***, which is symmetric about 0 and $\int_{-\pi}^{\pi}K_m(y)d_y = 1$ and is more concentrated near zero as m grows. m is the ***smoothing parameter***. The larger m, the more wiggly complexity is possible to g. g actually converges to f, and is thus on longer just an approximation, at points where f is continuous. Additionally, since ***kernel density estimates*** are defined as estimates of the form $g(x) = \frac{1}{n}\sum_{k=1}^{n}K_m\big(x-x_k\big)$, Fourier series density estimate is equivalent to kernel density estimate with Poisson kernel.

In our case, our goal is not to get density estimates based on a sample from this density, instead, we are trying to estimate spectral density using periodograms, i.e. Model 1. However, Model 1 will never satisfy rate of testing criteria. The reason is that the σ in the model is a constant, it dose not goes to zero as n goes to infinity. And we have proved that model with constant σ will always sub-optimal to the model with σ goes with zero, with regarding to the asymptotic power. Thus we want to apply some transformation to Model 1 and thus get the new model with σ goes to zero. Fourier transformation and kernel smoothing are chosen to transform the Model 1. So our goal here is not to reconstruct the density from the Fourier transformation of the sample from this density, we mainly use Fourier transformation to get a new model with σ goes to zero as n goes to infinity, which satisfies rate of testing theory and thus have better asymptotic power. Consequently, our Fourier transformed model is not equivalent to kernel smoothed model.

2.5 Model 4 (Basic Model with Replications)

Our Model 1, Model 2 and Model 3 are based on one single realization $\{x_{1,1},\cdots,x_{1,n_1}\}$ and $\{x_{2,1},\cdots,x_{2,n_2}\}$ from time series process $\{X_{1,t_1}\}$ and $\{X_{2,t_2}\}$. How about we have m realiza-

tions of $\{X_{1,t_1}\}$ and $\{X_{2,t_2}\}$? Do we still need kernel smoothing? The answer is yes and I will explain step by step through the evaluation of empirical performance based on "rate of testing" criteria.

Suppose there are m replicated samples for each of $\{X_{k,t_k}\}$ and m can be either a constant or $m \to \infty$ and $\frac{m}{n} \to 0$ as $n \to \infty$, the tapering test procedure is still applicable and similar testing framework can be derived. For the $i = 1, \cdots, m^{th}$ sample, define

$$Y_j^{(i)} = \log I_1^{(i)}(\omega_{1,j}) - \log I_2^{(i)}(\omega_{1,j}),$$

$j = 1, \cdots, p_n$ and the average over m replicated samples,

$$Y_{m,j} = \frac{1}{m} \sum\nolimits_{i=1}^{m} Y_j^{(i)}.$$

Then $Y_{m,j} \xrightarrow{D} N\left(\theta_j, \frac{\sigma^2}{m}\right)$. The basic Model 1 based on raw periodogram now becomes

$$Y_{m,j} = \theta_j + \frac{\sigma}{\sqrt{m}} \epsilon_{m,j} \quad \text{for } j = 1, \cdots, p_n, \tag{2.27}$$

where $\epsilon_{m,j} = \frac{Y_{m,j} - \theta_j}{\sigma/\sqrt{m}}$ are asymptotic independent standard normal random variables. And the test statistic becomes

$$Q_{m,n} = \sum_{j=1}^{p_n} j^{-1/2} Y_{m,j}^2.$$

Now let's consider the model with m replications of each time series. Suppose $m \to \infty$ and$\frac{m}{n} \to 0$ as $n \to \infty$, then $\sigma/\sqrt{m} \to 0$ in the basic model (2.27). Thus the question comes as whether kernel smoothing is still needed. The following theorem provides the "optimal adaptive rate of testing" under basic model with replications.

Theorem 2.5.1. *Based on the basic model with replication (2.27), the goal is to test the hypothesis (2.1) through the tapering test statistic $Q_{m,n}$, which rejects null hypothesis when $Q_{m,n}$ is large. Set*

$$\widehat{\delta}_{m,n}(s) = \left\{ p_n^{4s+1} \left(\log p_n\right)^{-1} \right\}^{-s/(2(4s+1))} \tag{2.28}$$

and suppose $1/2 < s_\star < s^\star$. For no test $Q_{m,n}$ is "rate of testing criteria" satisfied with $\delta_{m,n}(s) = o\left(\widehat{\delta}_{m,n}(s)\right)$ across $s_\star < s < s^\star$. Moreover, if the number of replications m is set to be

$$\widehat{m}_n = B \left\lfloor p_n^{(4s+1)/2} \log p_n^{1/2} \right\rfloor, \tag{2.29}$$

where $B > 0$ is a leading constant and $\lfloor a \rfloor$ denotes the integer part of a number. Then test $Q_{m,n}$ satisfies rate of testing criteria with "optimal adaptive rate of testing" $\widehat{\delta}_{m,n}(s)$ across $s_\star < s < s^\star$.

Proof. The same results can be derived for basic model with replication as in Lemma 2.3.1, i.e. the "rate of testing criteria" is equivalent to

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} E_{H_{1-0}}(Q_{m,n}) / \sqrt{Var_{H_0}(Q_{m,n})} \to \infty$$
$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} E_{H_{1-0}}(Q_{m,n}) / \sqrt{Var_{H_{1-0}}(Q_{m,n})} \to \infty.$$

Additionally, since $Y_{m,j} \xrightarrow{D} N\left(\theta_j, \frac{\sigma^2}{m}\right)$ and $Q_{m,n} = \sum_{j=1}^{p_n} j^{-1/2} Y_{m,j}^2$, then

$$\lim_{n\to\infty} E\left(Q_{m,n}\right) = \frac{\sigma^2}{m} \sum_{j=1}^{p_n} j^{-1/2} + \sum_{j=1}^{p_n} j^{-1/2} \theta_j^2$$
$$\lim_{n\to\infty} Var\left(Q_{m,n}\right) = \frac{2\sigma^4}{m^2} \sum_{j=1}^{p_n} j^{-1} + \frac{4\sigma^2}{m} \sum_{j=1}^{p_n} j^{-1} \theta_j^2.$$

Thus, those two conditions becomes

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} \frac{\sum_{j=1}^{p_n} j^{-1/2} \theta_j^2}{\sqrt{\sum_{j=1}^{p_n} j^{-1}/m^2}} \to \infty \quad \text{and} \quad \inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} \frac{\sum_{j=1}^{p_n} j^{-1/2} \theta_j^2}{\sqrt{\sum_{j=1}^{p_n} j^{-1} \theta_j^2/m}} \to \infty. \tag{2.30}$$

In the first condition of (2.30), since $\sum_{j=1}^{p_n} j^{-1/2} \to \infty$, then

$$\inf_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star;s,M)} \frac{\sum_{j=1}^{p_n} j^{-1/2} \theta_j^2}{\sqrt{\sum_{j=1}^{p_n} j^{-1}/m^2}} \to \infty \Rightarrow \sum_{j=1}^{p_n} j^{-1/2} \theta_j^2 m \to \infty.$$

And since

$$\frac{\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2}{\sqrt{\sum_{j=1}^{p_n} j^{-1}\theta_j^2/m}} \geq \frac{\sum_{j=1}^{p_n} j^{-1}\theta_j^2}{\sqrt{\sum_{j=1}^{p_n} j^{-1}\theta_j^2/m}} = \sqrt{\sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 m} \to \infty,$$

the first condition in (2.30) $\Rightarrow$ the second condition in (2.30). Then we only need to look at the first condition in (2.30).

Firstly, let's look at $\underset{\boldsymbol{\theta}\in H_1(\delta_n/\delta_n^\star;s,M)}{inf} \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 \to \infty$. Similar to proof of Theorem 2.3.1, we can get

$$\begin{aligned}\underset{\boldsymbol{\theta}\in H_1(\delta_n/\delta_n^\star;s,M)}{inf} \sum_{j=1}^{p_n} j^{-1/2}\theta_j^2 &= \widehat{j}^{-1/2}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 - p_n^{-(2s+1/2)}\widetilde{M}^2 \\ &= \left(\frac{\delta_n}{\delta_n^\star}\right)^{\frac{4s+1}{2s}} \widetilde{M}^{-\frac{1}{2s}} - p_n^{-\frac{4s+1}{2}}\widetilde{M}^2,\end{aligned}$$

which is achieved by

$$\theta_j = \begin{cases} \frac{\delta_n}{\delta_n^\star} & j = \left(\frac{\delta_n^\star \widetilde{M}}{\delta_n}\right)^{1/s} > p_n \\ 0 & o.w. \end{cases},$$

where $\widehat{j}$ is the largest index between 1 and ∞ that satisfies $\widehat{j} \leq \left(\frac{\delta_n^\star \widetilde{M}}{\delta_n}\right)^{1/s}$. Then the condition becomes

$$\lim_{n\to\infty} \frac{\widehat{j}^{-1/2}\left(\frac{\delta_n}{\delta_n^\star}\right)^2 m}{\sqrt{\sum_{j=1}^{p_n} j^{-1}}} = \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{p_n^{-2s-1/2} m}{\sqrt{\sum_{j=1}^{p_n} j^{-1}}} < \infty. \tag{2.31}$$

The second condition is needed because the rate of the first condition can be arbitrarily slowed down by $\forall \delta_n^\star \to 0$. For the first condition in (2.31), take the square,

$$\frac{\widehat{j}^{-1}\left(\frac{\delta_n}{\delta_n^\star}\right)^4 m^2}{\sum_{j=1}^{p_n} j^{-1}} = \left\{\frac{\widehat{j}^{-1}}{\delta_n^{\star 4}\widetilde{j}^{-1}}\right\} \frac{\widetilde{j}^{-1}\delta_n^4 m^2}{\sum_{j=1}^{p_n} j^{-1}} \to \infty,$$

where $\widetilde{j}$ is defined as $\left(\frac{M}{\delta_n}\right)^{1/s}$. Then $\widehat{j} \leq \widetilde{j} \Rightarrow \widehat{j}^{-1/2} \geq \widetilde{j}^{-1/2}$, and thus the item in braces $\to \infty$ for $\forall \delta_n^\star \to 0$. Therefore, the first condition in (2.31) is equivalent to

$$\liminf_{n\to\infty} \frac{\widetilde{j}^{-1}\delta_n^4 m^2}{\sum_{j=1}^{p_n} j^{-1}} > 0 \quad \text{i.e.} \quad \liminf_{n\to\infty} \frac{\widetilde{j}^{-1}\delta_n^4 m^2}{\log p_n} > 0, \tag{2.32}$$

based on Lemma 2.3.2. For the second condition in (2.31), take the square, then it is equivalent to

$$\limsup_{n\to\infty} \frac{p_n^{-4s-1} m^2}{\sum_{j=1}^{p_n} j^{-1}} < \infty \quad \text{i.e.} \quad \limsup_{n\to\infty} \frac{p_n^{-4s-1} m^2}{\log p_n} < \infty, \tag{2.33}$$

based on Lemma 2.3.2.

Similarly to the definition in Theorem 2.3.1, for the number of replications m, where $m \to \infty$ and $\frac{m}{n} \to 0$ as $n \to \infty$, it is defined as **minimally allowed**, if it is the fastest rate $\to \infty$ satisfying the first condition in (2.33). Define δ_n as minimally allowed if it is the fastest rate $\to 0$ that satisfies the first condition of (2.33) and if the second condition of (2.33) is also satisfied, δ_n is the minimax rate we are looking for.

Similarly to proofs of Theorem 2.3.1, we can prove that the number of replications m is minimally allowed if and only if

$$\frac{\widetilde{j}^{-1}\delta_n^4 m^2}{\log p_n} \asymp 1$$
$$\Leftrightarrow \frac{m\delta_n^{(4s+1)/s}}{\log p_n} \asymp 1$$

the sequence δ_n is minimally allowed if and only if

$$\frac{p_n^{-(4s+1)} m^2}{\log p_n} \asymp 1.$$

From the second condition, the optimal number of replications can be derived as

$$\widehat{m}_n = \left\{p_n^{4s+1} \log p_n\right\}^{1/2}$$

and if we plug $\widehat{m}$ into the first condition, the optimal adaptive rate of testing can be derived as

$$\widehat{\delta}_{m,n}(s) = \left\{p_n^{4s+1} \left(\log p_n\right)^{-1}\right\}^{-s/(2(4s+1))}.$$

□

Remark 2.5.1. *From results of Theorem 2.5.1, in order to achieve the "rate of testing" criteria, the number of replications needs to be as large as $\widehat{m}_n$. And from Equation (2.29), we can easily find that $\frac{\widehat{m}}{n} \to \infty$ as $n \to \infty$. Thus, although the "rate of testing criteria" can be satisfied under model (2.27) with "optimal rate" $\delta_{m,n}(s)$, the number of replications to make that happen is too large, which is contrary to our assumption of $\frac{m}{n} \to 0$ and is not meaningful in practice. Thus kernel smoothing model is recommended over replicated basic model.*

2.6 Model 5 (Kernel Smoothing Model with Replications)

Since the basic model with replications still dose not satisfy the "rate of testing" criteria, the kernel smoothing Model 3 is further extended to the current context with replications.

Suppose m is treated as constant as $n \to \infty$, which is the common case and is more meaningful in practice. Similar to the replicated basic model, define

$$\widetilde{Y}_j^{(l)} = \left(\log \widetilde{I}_1^{(l)}(\omega_j)\right) - \left(\log \widetilde{I}_2^{(l)}(\omega_j)\right),$$

$j = 1, \cdots, p_n$ and the average over m replicated samples,

$$\widetilde{\bar{Y}}_{m,j} = \frac{1}{m} \sum\nolimits_{l=1}^{m} \widetilde{Y}_j^{(l)}.$$

Then $\widetilde{\bar{Y}}_{m,j} \xrightarrow{D} N\left(\theta_j, \frac{\sigma_n'^2}{m}\right)$. The Model 3 based on kernel smoothing now becomes

$$\widetilde{\bar{Y}}_{m,j} = \theta_j + \frac{\sigma_n'}{\sqrt{m}} \widetilde{\epsilon}_{m,j} \quad \text{for } j = 1, \cdots, p_n \tag{2.34}$$

where $\widetilde{\epsilon}_{m,j} = \frac{\widetilde{\bar{Y}}_{m,j} - \theta_j}{\sigma_n'/\sqrt{m}}$ are asymptotic independent standard normal random variables. And the test statistic becomes

$$\widetilde{Q}_{m,n} = \sum_{j=1}^{p_n} j^{-1/2} \widetilde{\bar{Y}}_{m,j}^2.$$

Remark 2.6.1. *We can easily prove that the same theoretical results can be derived for kernel smoothing model as in Theorem 2.3.1, which means that kernel smoothing model with "optimal bandwidth" is still recommended over basic model.*

2.7 Summary

In this chapter, we are trying to develop the tapering test under time series context. Firstly, we build Model 1, the basic model based on the periodogram (log-periodogram) as the estimate for spectrum (log-spectrum). However, this model suffers from in-efficiency problem, mainly because of the in-consistency of periodogram (log-periodogram). And we prove that the model based on in-consistent estimate will always be sub-optimal to the model based on consistent estimate. Thus

we are trying to overcome this problem by Fourier transform of the basic model as Model 2 which is inspired by Lu and Li (2013) [17] paper. After Fourier transformation, the variance of the new $Y_j^\star$s go to zero as n goes to infinity. The corresponding "optimal weights" and "optimal adaptive rate of testing" are derived. However, Fourier transform is not generally used in time series to estimate periodogram (log-periodogram). Thus we propose the Model 3 based on kernel smoothing of log-periodogram as the estimate for log-spectrum. If the bandwidth of the kernel satisfy certain conditions, the kernel smoothed estimator is consistent and thus the "rate of testing criteria" can be satisfied. The corresponding "optimal bandwidth" and "optimal adaptive rate of testing" are derived. Therefore, the "optimal tapering test" based on kernel smoothing model using the "optimal bandwidth" is derived.

We also explores the relationship between Fourier transform and kernel smoothing to see whether the Fourier transform in Model 2 can be written into kernel smoothing as in Model 3. But we find out that these two models are actually not related.

Additionally, since our current "optimal tapering test" are based on single sample from each of the time series. But in practice, we will encounter with the scenario where there are multiple samples for each time series. We future look at the replicated model. And since the replicated basic model does not have the problem of in-consistency anymore, the question is whether kernel smoothing is still needed. We explore the "rate of testing theory" under basic model. The "optimal number of replications" and "optimal adaptive rate of testing" are derived. However, the number of replications needed to satisfy rate of testing criteria is too large, which is contrary to our assumption and is meaningless in practice. Thus, kernel smoothing is sill recommended. The replicated kernel smooth model is built based on the reasonable assumption that the number of replication m is constant as n goes to infinity. The same "optimal bandwidth" and "optimal adaptive rate of testing" are derived as Model 3.

In summary, no matter whether replication exists, kernel smoothing model is recommended for its optimal asymptotic power.

Chapter 3

Bayesian Tapering Test and Bayesian Multiple Testing

3.1 Bayesian Tapering Test

Since the tests under kernel smoothing model, i.e. Model 3 without replication and Model 5 with replications, the Bayesian testing framework explored in this chapter is based on Model 3 and Model 5. The reasons we want to write our test into Bayesian framework are

- firstly, to prepare for our future work in Bayesian multiple testing;
- secondly, to actually implement the tests in real data, the asymptotic critical value performs not good as the empirical critical value, however, it is too hard to get the empirical critical value for two time series with unknown relationship of their spectral densities;
- finally, we will explain latter that the Bayes factor can be written as the monotone transformation of the tapering test statistic, which makes it easier to implement and understand the Bayes tapering test framework.

3.1.1 Bayesian Setup

As discussed in Section 1.4, another important advantage of Spitzner's tapering test is that the test statistic in Equation (1.5) is easy to be written through a formal Bayesian construction. Spitzner (2008) [27] studied the Bayesian point-null testing problem under a high-dimensional normal-means model. A non-informative prior structure is proposed for general problems, and then refined for the specialized contexts of goodness-of-fit testing and functional data analysis. Here we further explore our testing procedures in Bayesian framework following the idea of Spitzner (2008) [27].

For ease of explanation, we will only look at Model 3 based on kernel smoothing as in Section 2.3, and the Bayesian framework for models based on Fourier transformation is similar. The asymptotic framework under consideration is defined according to

$$\widetilde{Y}_j = \theta_j + \sigma_n' \widetilde{\epsilon}_j \quad \text{for } j = 1, \cdots, p_n$$

$$\widetilde{Y}_j \xrightarrow{D} N\left(\theta_j, \sigma_n'^2\right), \text{ asymptotically independent across j}$$

The objective is to test the hypothesis:

$$H_0 : \theta_j = 0 \text{ for all } j = 1, \cdots, p_n$$

$$H_1 : \theta_j \neq 0 \text{ for some } j = 1, \cdots, p_n$$

Special attention is given to the context where smoothness assumption $\boldsymbol{\theta} \in \mathcal{B}_{s,M}$ is incorporated into the Bayesian framework.

Following the idea of Spitzner (2008) [27] by building the standard Bayesian setup as in Berger (1985) [4] and Robert (2001) [23],

- a prior mass $\rho_{0,n} \in (0,1)$ is placed on the null hypothesis H_0.
- a continuous distribution $(1-\rho_{0,n})\,\pi\left(\boldsymbol{\theta}|H_1\right)$ is placed on the alternative hypothesis H_1.
- the prior on $\boldsymbol{\theta}$ is specified according to $\boldsymbol{\theta}|H_1 \sim N\left(\boldsymbol{\xi}, \tau_n^2 \mathbf{W}_n \Sigma_n\right)$,
  where $\mathbf{W}_n = \text{ diag}\left(w_{n,1}, \cdots, w_{n,p_n}\right)$ is the diagonal matrix of weights, $\boldsymbol{\Sigma}_n = \sigma_n' I = \frac{\sigma^2}{2b_n+1} I$ is the covariance matrix depending on the unknown tuning parameter b.
- the likelihood is $P(\widetilde{\mathbf{Y}}|\boldsymbol{\theta}) = N(\boldsymbol{\theta}, \Sigma_n)$.

Thus the marginal likelihood under null hypothesis is $\widetilde{\mathbf{Y}}|H_0 \sim N(\mathbf{0}, \Sigma_n)$, and the marginal likelihood of the data under alternative hypothesis is $\widetilde{\mathbf{Y}}|H_1 \sim N\left(\boldsymbol{\xi}, \Sigma_n + \tau_n^2 \mathbf{W}_n \Sigma_n\right)$.

Our procedure will focus on $\boldsymbol{\xi} = \mathbf{0}$, which means that $\boldsymbol{\theta}|H_1 \sim N\left(\mathbf{0}, \tau_n^2 \mathbf{W}_n \Sigma_n\right)$. And define $\upsilon_{n,j} = \left\{1 + \frac{1}{\tau_n^2 w_{n,j}}\right\}^{-1}$. Then $\tau_n^2 w_{n,j} = \left(\upsilon_{n,j}^{-1} - 1\right)^{-1}$, and thus $1 + \tau_n^2 w_{n,j} = (1 - \upsilon_{n,j})^{-1}$.

To determine whether to reject null hypothesis in the Bayesian testing framework, we need the Bayes factor and/or the posterior null probability. Bayes factor is defined as the marginal likelihood of the data under null hypothesis divided by the marginal likelihood of the data under alternative hypothesis:

$$
\begin{aligned}
BF_0(\widetilde{\mathbf{Y}}) &= \frac{m\left(\widetilde{\mathbf{Y}}|H_0\right)}{m\left(\widetilde{\mathbf{Y}}|H_1\right)} \\
&= \frac{N(\mathbf{0}, \Sigma_n)}{N\left(\mathbf{0}, \Sigma_n + \tau_n^2 \mathbf{W}_n \Sigma_n\right)} \\
&= \frac{\prod_{j=1}^{p_n} \left\{\sigma_n'^{-1} exp\left(-\frac{1}{2}\widetilde{Y}_j^2 / \sigma_n'^2\right)\right\}}{\prod_{j=1}^{p_n} \left\{\left(1 + \tau_n^2 w_{n,j}\right)^{-\frac{1}{2}} \sigma_n'^{-1} exp\left(-\frac{1}{2}\widetilde{Y}_j^2 / \left[\left(1 + \tau_n^2 w_{n,j}\right)\sigma_n'^2\right]\right)\right\}} \\
&= \prod_{j=1}^{p_n} \left\{(1 - \upsilon_{n,j})^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_n'^2}\left(\widetilde{Y}_j^2 - (1 - \upsilon_{n,j})\widetilde{Y}_j^2\right)\right]\right\} \\
&= \prod_{j=1}^{p_n} \exp\left\{-\frac{1}{2\sigma_n'^2}\upsilon_{n,j}\widetilde{Y}_j^2 - \frac{1}{2}\log(1 - \upsilon_{n,j})\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_n'^2}\sum_{j=1}^{p_n} \upsilon_{n,j}\widetilde{Y}_j^2 - \frac{1}{2}\sum_{j=1}^{p_n} log(1 - \upsilon_{n,j})\right\}
\end{aligned}
\tag{3.1}
$$

The decision is made based on the value of of $2\log BF$, which provides us information that how much confident we have for the H_0 based on the data we have. As provided by Kass and Raftery (1995) [16], the value of $2\log BF$ from 0 to 2 means "Not worth more than a bare mention"; those from 2 to 6 means "positive" evidence for H_0; those greater than 6 indicate "strong" evidence for H_0; and those greater than 10 means "very strong" evidence for H_0. While the negative values indicate the strength of evidence for H_1.

Posterior null probability can also be derived from Bayes factor:

$$
P\left(H_0|\widetilde{\mathbf{Y}}\right) = \left\{1 + \frac{\rho_{0,n}^{-1} - 1}{BF_0(\widetilde{\mathbf{Y}})}\right\}^{-1} \tag{3.2}
$$

Thus, to make decisions, we need to firstly specify these unknown parameters $\rho_{0,n}$, τ_n and $\mathbf{W}_n$.

3.1.2 Specification of the Prior Null Probability $\rho_{0,n}$

However, the Bayesian testing problem is challenging even in more basic scenarios, especially if prior information is vague or absent, and further challenges arise if the dimensionality is high. Because the priors used in testing place mass on a point-null hypothesis, the standard techniques used in estimation to construct non-informative priors lead to test procedures that are not sensible. For instance, at, improper priors, which are commonly used in estimation, lead to test procedures that are sensitive to arbitrary normalizing constants.

To address this issue, Spitzner (2008) [27] considered different procedures and decided to adapted Robert (1993) [22]'s idea to impose a dependency structure among the prior parameters. But instead of invoking an equiponderance device as in Robert (1993) [22], Spitzner (2008) [27] assumes that the dependencies is to be determined by an asymptotic-consistency principle discussed in Diaconis and Friedman (1986) [9], i.e. "what if" guidelines. This method scrutinizes the choice of prior by asking whether, given a particular data set, the posterior distribution makes a meaningful update of the prior. (The name of the method alludes to the question, "What if the data came out that way?"). For present purposes, to check for "meaning" it is sufficient to consider the extreme limits of the posterior null probability: $\mathbf{P}_n\left[H_0|\widetilde{\mathbf{Y}}\right] \to 1$ is here to mean "overwhelming evidence for H_0" and $\mathbf{P}_n\left[H_0|\widetilde{\mathbf{Y}}\right] \to 0$ is to mean "overwhelming evidence against H_0". The limits here are "almost sure" with respect to the model; this mode of convergence is stronger than convergence "in probability" and is required by the "what if" method's focus on data rather than probabilities. By applying "what if" guidelines, the resulting tests are sensitive to the data in high dimensions and meaningful in the sense of being proper Bayes or limits of proper Bayes procedures. Moreover, the proposed priors are non-informative, but avoid the need to specify arbitrary normalizing constants.

The connection between the choice of the prior and the "what if" guidelines is made through the following parameterization where $\rho_{0,n}$ is represented in terms of other parameters:

$$\rho_{0,n} = \left\{1 + exp\left\{-\frac{1}{2}\sum_{j=1}^{p_n}\left[log\left(1-\upsilon_{n,j}\right) + \upsilon_{n,j}\right]\right\}\right\}^{-1}. \tag{3.3}$$

Note that this parameterization dose not constrain the prior in any way.

3.1.3 Specification of the Matrix of Weights $\mathbf{W}_n$

To incorporate the underlying smoothness assumption, suppose $\boldsymbol{\theta}$ is an element of a Sobolev ellipsoid in continuous space, the restriction can be expressed as

$$\boldsymbol{\theta} = (\theta_j, j = 1, 2, \cdots) \in \mathcal{B}_{s,M}$$
$$\mathcal{B}_{s,M} = \left\{ \left(\widetilde{\theta}_1, \widetilde{\theta}_2, \cdots\right) : \sqrt{\sum_{j=1}^{\infty} j^{2s} \widetilde{\theta}_j^2} \leq M \right\}$$

, where $\mathcal{B}_{s,M}$ is a Sobolev ellipsoid of radius M in infinite-dimensional discrete space, and $M > 0$, $s > 1/2$ are fixed constants, larger s makes the restriction stronger.

Spitzner (2008) [27] set up the Bayesian mathematical framework similar to the frequentist "rate of testing" theory as discussed in Section 1.5. Consider the sequence $\delta_n \to 0$ satisfying

$$\sup_{\boldsymbol{\theta} \in H_1(\delta_n/\delta_n^\star; s, M)} P_{\boldsymbol{\theta}} \left[H_0 | \widetilde{\mathbf{Y}} \right] \to 0, \forall \delta_n^\star \to 0$$
$$H_1(\delta; s, M) = \left\{ \boldsymbol{\theta} \in \mathcal{B}_{s,M} : \sqrt{\sum_{j=1}^{\infty} \theta_j^2} \geq \delta \right\} \tag{3.4}$$

The parameter δ_n defines the range of an indistinguishable region, which collects these θs which are consistent with H_1 and do not yield $P_{\boldsymbol{\theta}} \left[H_0 | \widetilde{\mathbf{Y}} \right] \to 0$ almost surely. The objective of "rate of testing" theory is to keep the indistinguishable region as small as possible. δ_n, which is referred to as "rate" in the "rate of testing" theory, gives the boundary of the indistinguishable region, and so faster rate of testing identifies a smaller indistinguishable region. This leads to the guideline that δ_n goes to 0 as fast as possible.

Thus the favorable setting for weights $w_{n,1}, \cdots, w_{n,p_n}$ are those for which the associated δ_n goes to 0 with the fastest rate among all tapering mechanism.

Note that in the formula for the Bayes factor as in Equation (3.1), the data appear in the quadratic form $\sum_{j=1}^{p_n} \upsilon_{n,j} \widetilde{Y}_j^2$, which is similar to our test statistic $\widetilde{Q}_n = \sum_{j=1}^{p_n} j^{-1/2} \widetilde{Y}_j^2$ as in Section 2.3 for Model 3. Spitzner (2008) [27] showed in Appendix 3 that the current problem of finding optimal setting for $w_{n,1}, \cdots, w_{n,p_n}$ based on "rate of testing" criteria in Equation (3.4) is mathematically parallel to the ideas in frequentist context where the only tests considered are those which use a quadratic form as a test statistic. Even the tapering test is originally proposed for testing in Functional linear model, not for the testing of equal spectral densities in time series, we can still

borrow the idea of tapering and try to identify suitable definitions of those quantities for time series problems under our settings in Chapter 2. Thus the same results can also be applied in our context, as our Bayesian setting is mathematically parallel to frequentist setting of the "rate of testing" theory. So we can draw results for our Bayesian context similar to Remark (2.1.3), Remark (2.2.2), and Remark (2.3.4) in Chapter 2.

Consequently, the optimal rate will be achieved by the settings of $\upsilon_{n,j}$ such that

$$\upsilon_{n,j} = j^{-1/2}, \quad j = 1, \cdots, p_n. \tag{3.5}$$

As discussed in Chapter 2, this rate is adaptively optimal among tests based on quadratic forms. Thus this weight setting achieves the prior that, from an adaptive "rate of testing" viewpoint, reduce the indistinguishable region to the greatest extent.

Additionally, as we defined before, $\upsilon_{n,j} = \left\{1 + \frac{1}{\tau_n^2 w_{n,j}}\right\}^{-1}$, so the definition of $\upsilon_{n,j}$ as in Equation (3.5) will also help us to skip the specification of overall scale parameter τ_n, which could be very difficult for Bayesian testing setup.

Finally, if we plug the settings of weights $\upsilon_{n,j}$ as in Equation (3.3) into the setup of prior parameter $\rho_{0,n}$ as in Equation (3.3), we will get the adaptive optimal settings of priors based on parameters $\rho_{0,n}$, τ_n and $\mathbf{W}_n$ under the Bayesian context of our Model 3 or Model 4. Under Model 3, Bayes factor and posterior null probability can also be derived based on Equation (3.1) and Equation (3.2), i.e.

$$BF_0(\widetilde{\mathbf{Y}}) = \exp\left\{-\frac{1}{2\sigma_n'^2}\sum_{j=1}^{p_n} j^{-1/2}\widetilde{Y}_j^2 - \frac{1}{2}\sum_{j=1}^{p_n} log\left(1 - j^{-1/2}\right)\right\}. \tag{3.6}$$

Under Model 4, Bayes factor and posterior null probability can also be derived based on Equation (3.1) and Equation (3.2), i.e.

$$BF_0(\widetilde{\mathbf{Y}}) = \exp\left\{-\frac{1}{2\sigma_n'^2/m}\sum_{j=1}^{p_n} j^{-1/2}\widetilde{Y}_j^2 - \frac{1}{2}\sum_{j=1}^{p_n} log\left(1 - j^{-1/2}\right)\right\}. \tag{3.7}$$

We can make decisions of whether to reject H_0 based on values of $2\log BF$ and/or $P\left(H_0|\widetilde{\mathbf{Y}}\right)$.

3.2 Bayesian Multiple Testing

In Session 3.1.3, the Bayes factor in Equation (3.6) are written as a closed form which is the monotone transformation of the test statistic in our newly proposed "optimal" tapering test. Another benefit of writing the testing procedure into Bayesian framework is that form of Bayes factor helps a lot in setting up Bayesian multiple testing framework.

Suppose we need to do simultaneous testing of M hypotheses and we want to identify the tests that we have enough evidence to argue that the null hypothesis is true. For each of these $k = 1, \cdots, M$ tests, hypothesis is

$$H_{0,k} : \boldsymbol{\theta}_k = 0 \quad \text{v.s.} \quad H_{1,k} : \text{ not } H_{0,k}.$$

Scott and Berger (2010) [24] studies the multiplicity-correction effect of standard Bayesian variable-selection priors in linear regression. They clarify how multiplicity correction enters Bayesian variable selection by allowing the choice of prior model probabilities to depend upon the data in an appropriate way, i.e., each variable is presumed to be in the model independently with an unkown common probability q.

Following Scott and Berger's (2010) [24] idea, let's define a binary vector

$$\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_M)$$

be a set of "model indicators", with each entry

$$\gamma_k = \begin{cases} 1 & \text{if } H_{0,k} \text{ is true} \\ 0 & \text{if } H_{1,k} \text{ is true} \end{cases}.$$

Define n_γ as the number of tests with $\gamma_k = 1$, which means the number of tests with null hypothesis is true.

Then the multiple testing problem can thus be formulated as a model selection problem: to choose among 2^M models indicated by all possible values of $\boldsymbol{\gamma}$.

As indicated by Scott and Berger (2010) [24], the standard practice in Bayesian model selection problems is to treat each inclusion as exchangeable Bernoulli trials with common success probability

p, which implies that the prior probability of a model indicated by $\boldsymbol{\gamma}$ given the success probability p is

$$p\left(\boldsymbol{\gamma}|p\right) = p^{n_\gamma}\left(1-p\right)^{M-n_\gamma}$$

with unknown parameter p has a Beta distribution,

$$p \sim Beta(a,b)$$

The default choice is $a = b = 1$, implying a Uniform(0,1) prior on p. Then the prior of the model indicated by $\boldsymbol{\gamma}$ reduces to

$$p\left(\boldsymbol{\gamma}\right) = \frac{\left(n_\gamma!\right)\left(M-n_\gamma\right)!}{\left(M+1\right)\left(M!\right)} = \frac{1}{M+1}\binom{M}{n_\gamma}^{-1}.$$

And the likelihood is

$$f\left(\mathbf{Y}|\boldsymbol{\gamma}\right) = \prod_{k=1}^{M} f\left(\mathbf{Y}_k|\gamma_k\right) = \prod_{\gamma_k=1} N(\mathbf{0}, \Sigma_n^{H_0}) \prod_{\gamma_k=0} N(\mathbf{0}, \Sigma_n^{H_1}).$$

Here $\Sigma_n^{H_0} = \sigma_n'^2 I$, and $\Sigma_n^{H_1} = \sigma_n'^2 \times \text{diag}\left(\left(1+j^{-1/2}\right)^{-1}\right)$, according to the kernel smoothing model in Equation (2.14); $\Sigma_n^{H_0} = \frac{\sigma_n'^2}{m} I$, and $\Sigma_n^{H_1} = \frac{\sigma_n'^2}{m} \times \text{diag}\left(\left(1+j^{-1/2}\right)^{-1}\right)$, according to the replicated kernel smoothing model in Equation (2.34).

Then the posterior for the model indicator $\boldsymbol{\gamma}$ will be

$$p\left(\boldsymbol{\gamma}|\mathbf{Y}\right) \propto p\left(\boldsymbol{\gamma}\right) f\left(\mathbf{Y}|\boldsymbol{\gamma}\right).$$

And posterior inclusion probability of each test can be derived from the posterior probability of the model indicator $\boldsymbol{\gamma}$. We are interested in the median-probability model, the model that includes those tests having posterior inclusion probability at least 0.5. According to Barbieri and Berger (2004) [1], under many circumstances, this model has greater predictive power than the most probable model.

Additionally, common Bayesian procedures assign vanishingly small posterior probabilities to all models in high-dimensional settings, even when the maximum probability model assigns relatively high probability to the true model. It is for this reason that articles describing Bayesian model selection algorithms usually do not report the posterior probability assigned to the most probable model, often opting instead to report the marginal probabilities that individual covariates

were included in models sampled from the posterior distribution, i.e. posterior inclusion probability which is defined as

$$p_k = Pr\left(\gamma_k = 1|\mathbf{Y}\right) = \sum_{\boldsymbol{\gamma}} 1_{\gamma_k=1} p\left(\boldsymbol{\gamma}|\mathbf{Y}\right).$$

To get the posterior inclusion probability, we need samples from posterior distribution $p\left(\boldsymbol{\gamma}|\mathbf{Y}\right)$. However, direct simulation from $p\left(\boldsymbol{\gamma}|\mathbf{Y}\right)$ is too difficult since we do not have the closed form for it, instead, we will run Markov Chain Monte Carlo (MCMC) to simulate samples from the posterior empirical distribution, which eventually converge to the stationary distribution $p\left(\boldsymbol{\gamma}|\mathbf{Y}\right)$.

Steps to run MCMC are described as following:

1. Define the initial state $\boldsymbol{\gamma}^{(0)} = \left(\gamma_1^{(0)}, \cdots, \gamma_M^{(0)}\right)$.
2. For the current state $\boldsymbol{\gamma}^{(l)}$, update each $\gamma_k^{(l)}$ separately by generating a candidate $t_k^{(l)}$ from the symmetric proposal distribution, Bernoulli(0.5).
3. Define the current state $\mathbf{S}_k^{(l)} = \left(\gamma_1^{(l+1)}, \cdots, \gamma_{k-1}^{(l+1)}, \gamma_k^{(l)}, \gamma_{k+1}^{(l)}, \cdots, \gamma_M^{(l)}\right)$ and the candidate state $\mathbf{T}_k^{(l)} = \left(\gamma_1^{(l+1)}, \cdots, \gamma_{k-1}^{(l+1)}, t_k^{(l)}, \gamma_{k+1}^{(l)}, \cdots, \gamma_M^{(l)}\right)$ with $\mathbf{T}_k^{(l)}$ only changes one element in $\mathbf{S}_k^{(l)}$. Then calculate the acceptance ratio $\alpha = \min\left\{1, \frac{p\left(\mathbf{T}_k^{(l)}|\mathbf{Y}\right)}{p\left(\mathbf{S}_k^{(l)}|\mathbf{Y}\right)}\right\}$.
4. Accept $t_k^{(l+1)}$ and set $\gamma_k^{(l+1)} = t_k^{(l+1)}$ with probability α, otherwise, $\gamma_k^{(l+1)} = \gamma_k^{(l)}$.
5. After updating all M elements of $\boldsymbol{\gamma}^{(l)}$, we will get the new state as $\boldsymbol{\gamma}^{(l+1)}$.
6. Repeat Step 2-Step 5 for L times, to get $\left\{\boldsymbol{\gamma}^{(1)}, \cdots, \boldsymbol{\gamma}^{(L)}\right\}$ as L posterior samples of $\boldsymbol{\gamma}$ and calculate the posterior inclusion probability for each test.

The critical aspect of the MCMC steps is to calculate the acceptance Ratio,

$$\alpha = \min\left\{1, \frac{p\left(\mathbf{T}_k^{(l)}|\mathbf{Y}\right)}{p\left(\mathbf{S}_k^{(l)}|\mathbf{Y}\right)}\right\}.$$

The Posterior Ratio can be derived from our settings of priors,

$$\begin{aligned}\frac{p\left(\mathbf{T}_k^{(l)}|\mathbf{Y}\right)}{p\left(\mathbf{S}_k^{(l)}|\mathbf{Y}\right)} &= BF_{\mathbf{TS}_k} \times \frac{p\left(\mathbf{T}_k^{(l)}\right)}{p\left(\mathbf{S}_k^{(l)}\right)} \\ &= BF_{\mathbf{TS}_k} \times \frac{n_{\mathbf{T}^{(l)}}!\left(M - n_{\mathbf{T}^{(l)}}\right)!}{n_{\mathbf{S}^{(l)}}!\left(M - n_{\mathbf{S}^{(l)}}\right)!},\end{aligned}$$

where $BF_{\mathbf{TS}_k}$ is the Bayes factor of comparing model indicated by $\mathbf{T}_k^{(l)}$ with the model indicated by $\mathbf{S}_k^{(l)}$. It is equivalent to the Bayes Factor of comparing hypothesis indicated by $\gamma_k^{(l)}$ with the hypothesis indicated by $t_k^{(l)}$. Additionally,

$$\begin{aligned}
&\text{if } t_k^{(l)} = 1, \gamma_k^{(l)} = 0, \quad BF_{\mathbf{TS}_k} = BF_{0k} \\
&\text{if } t_k^{(l)} = 0, \gamma_k^{(l)} = 1, \quad BF_{\mathbf{TS}_k} = 1/BF_{0k} \\
&\text{if } t_k^{(l)} = \gamma_k^{(l)}, \quad BF_{\mathbf{TS}_k} = 1.
\end{aligned}$$

BF_{0k} is the Bayes factor we derived in Equation (3.6) or Equation (3.7). *Therefore, the closed form for Bayes factor for the tapering test make it more convenient to do Bayesian multiple testing process.*

Chapter 4

Power Study

In Chapter 2, three new testing procedures are proposed based on three different methods of estimating the spectral density. Each test is derived from the optimizing criteria "rate of testing" theory, a framework to find the rate at which the power is retained under geometric smoothness constraints. Since our tests are derived to optimize the asymptotic power, we also want to assess the empirical power of our newly proposed tests and compare the performance with existing tests. The empirical powers of these tests are investigated here in this chapter by a comprehensive simulation study.

4.1 Outline of Study

In this chapter, empirical power of the tests described in Chapter 2 based on raw periodogram, Fourier transformations, and discrete Uniform Kernel smoothing as well as some currently popular tests for testing equal spectral density in literatures are compared. Details about those tests to be compared are listed as below:

- Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios, with the test statistic
$$R := \max\left\{\log\frac{I_1(\omega_j)}{I_2(\omega_j)}\right\} - \min\left\{\log\frac{I_1(\omega_j)}{I_2(\omega_j)}\right\}.$$
- Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized

cumulative periodograms, with the test statistic

$$D_m = \sup \mid F_1(\omega) - F_2(\omega) \mid$$
$$F_1(\omega_j) = \sum_{i=1}^{j} I_1(\omega_i) \,/\, \sum_{i=1}^{p_n} I_1(\omega_i)$$
$$F_2(\omega_j) = \sum_{i=1}^{j} I_2(\omega_i) \,/\, \sum_{i=1}^{p_n} I_2(\omega_i).$$

- Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms, with the test statistic

$$\overline{D} := \frac{1}{(n/2-1)} \sum_{j=1}^{n/2-1} \left| \log \frac{I_1(\omega_j)}{I_2(\omega_j)} \right|.$$

- Lu and Li (2013) [17]'s frequency domain test applying Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test, with the test statistic

$$\overline{LL}^{\star} := \max_{1 \le k \le n_m} \frac{1}{\sqrt{k\hat{\sigma}_2^2}} \sum\nolimits_{i=1}^{k} \left(\left(\bar{D}_{m,i}^{\star} \right)^2 - \hat{\sigma}_1^2 \right)$$
$$\overline{D}_{m,i} = ln \frac{m^{-1} \sum\limits_{k=(i-1)m+1}^{im} I_1(\omega_k)}{m^{-1} \sum\limits_{k=(i-1)m+1}^{im} I_2(\omega_k)} \quad i = 1, \cdots, N_m$$
$$\overline{D}_{m,i}^{\star} = \frac{1}{N_m} \sum_{i=1}^{N_m} \overline{D}_{m,i} \psi_{j,i}$$
$$\widehat{\sigma}_1^2 = \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}_{m,i}^{\star} \right)^2 - \left\{ \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \overline{D}_{m,i}^{\star} \right\}^2$$
$$\widehat{\sigma}_2^2 = \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}_{m,i}^{\star} \right)^4 - \left\{ \frac{1}{N_m - I_{N_m}} \sum_{i=I_{N_m}+1}^{N_m} \left(\overline{D}_{m,i}^{\star} \right)^2 \right\}^2,$$

  where $\psi_{j,i}$ is the Fourier basis functions as defined before, $N_m = \left[\frac{n}{m}\right]$, $I_{N_m} = \left[\frac{N_m}{4}\right]$.

- Our tapering test based on Model 1 (Raw periodogram) as proposed in Section 2.1, with test statistic

$$\widetilde{Q}_n := \sum\nolimits_{j=1}^{p_n} j^{-1/2} Y_j^2.$$

- Our tapering test based on Model 2 (Fourier transform) as proposed in Section 2.2, with test statistic

$$Q_n^{\star} := \sum\nolimits_{j=1}^{p_n} j^{-1/2} Y_j^{\star 2}$$

- Our tapering test based on Model 3 (Kernel smoothing) as proposed in Section 2.3, with test statistic

$$\widetilde{Q}_n := \sum_{j=1}^{p_n} j^{-1/2}\widetilde{Y}_j^2$$

Note that the kernel used here is the Uniform kernel as defined before, and the rate of the bandwidth parameter to do Uniform Kernel smoothing is chosen according to the optimal rate derived in Equation (2.25) with leading constant B is set to be 1.

The simulation design is such that the models examined each have dimensionality p_n. The length of the time series are specified into two scenarios, short time series with $n = 256$, and long time series with $n = 1024$. We compared the powers of the tests by applying them to simulations of pairs of moving average of order q, MA(q), autoregressive of order p, AR(p), or long-order Gaussian seasonal $\text{AR}(1)_{12}$ processes:

- A MA(q) process $\{X_t\}$ is defined by

$$X_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$$

where $\theta_1, \cdots, \theta_1$ are parameters of the model, $\epsilon_t, \epsilon_{t-1}$ are the white nose error terms.

- An AR(p) process $\{X_t\}$ is defined by

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

where $\phi_1, \cdots, \phi_p$ are parameters of the model, ϵ_t is the white nose error term.

- An $\text{AR}(1)_{12}$ process $\{X_t\}$, which is a generalized exponential smoothing models that can incorporate long-term trends and seasonality, is defined by

$$X_t = \phi_{12} X_{t-12} + \epsilon_t$$

where ϕ_{12} is the parameter of the model, ϵ_t is the white noise error term.

Simulated powers are calculated at each of the following settings:

Setting 1: MA(1) with $\theta_1 = 0.1$ versus MA(1) with various θ_1' values.

Setting 2: AR(1) with $\phi_1 = 0.1$ versus AR(1) with various ϕ_1' values.

Setting 3: AR(1) with $\phi_1 = 0.5$ versus long-order Gaussian seasonal AR$(1)_{12}$ with various ϕ_1' values.

Setting 4: AR(1) with $\phi_1 = \{-0.75, -0.5, -0.25, 0.25, 0.5, 0.75\}$ versus MA(1) with $\theta_1 = \pm\sqrt{\phi_1^2/\left(1-\phi_1^2\right)}$, where the plus sign is chosen if $\phi_1 > 0$ and the minus sign is chosen when $\phi_1 < 0$. This choice of θ_1 makes the variances of the two series identical; moreover, the lag-one auto-covariances have the same sign.

Setting 5: MA(1) with $\theta_1 = 1$ versus MA(2) with $\theta_1' = 1, \theta_2' = \{0.1, 0.25, 0.5\}$. Hence, the two time series differ only in their second-order moving-average coefficient.

Note that Setting 1-3 are designed by ourself, and Setting 4-5 cover those simulation scenarios in Lund et al. (2009) [18] and Lu and Li (2013) [17]. Other different settings were also explored, but similar results can be derived with current settings. Therefore, Setting 1-5 are presented as examples.

4.2 Empirical Comparisons

The first phase of the study consisted of calculating the empirical critical values for each test. For this phase, we performed 10,000 replicate simulations of each pair of processes to give reasonably precise estimates of actual significance levels. Firstly, independent pairs of series of length $n =$ 256 or 1024 are simulated. The series length is taken as the multiple of two so that the fast Fourier transform algorithm can be employed to rapidly compute the $I_1(\omega_j)$ and $I_2(\omega_j)$. *It is very important to explore multiple settings of p_n (at 256 and 1024), since it helps to check our choice of the leading constant B.* For calculating the empirical critical values, both series have the same parameter values $\theta_1, \cdots, \theta_q$ or $\phi_1, \cdots, \phi_p$, and this parameter is varied within the causal model range of $|\phi| < 1$. The innovations (errors) are normally distributed with a unit variance. Hence, the two series have equivalent spectral densities. Table 4.1 and Table 4.2 reports the empirical critical values for the Coates and Diggle (1986) [8]'s test (i.e. R), Diggle and Fisher (1991) [10]'s test (i.e. D_m), Lund et al. (2009) [18]'s test (i.e. $\overline{D}$), Lu and Li (2013) [17]'s test (i.e. $\overline{LL}^{\star}$), Tapering test based on raw periodograms (i.e. Q_n), Tapering test based on Fourier transform (i.e. $Q_n^{\star}$), Tapering test based on kernel smoothing (i.e. $\widetilde{Q}_n$), with the significance level 5%.

Remark 4.2.1. *Note that for our tapering test based on Model 3 (Kernel smoothing), i.e. $\widetilde{Q}_n$, the selection of the "optimal" bandwidth as in Equation (2.25) is also dependent on the leading*

constant B. So the selection of the value for B becomes critical for the calculation of $\widetilde{Q}_n$. Here for simplicity, in our empirical work, we will set $B = 1$ as it is often used by convention, see Ma and He (2016) [19] for similar specifications. And the simulation results listed as following help us to check our choice of the leading constant B.

Table 4.1: *Table of Empirical Critical Values for short series ($n = 256$) under different simulation scenarios described in Section 4.1.*

	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
MA(1) with $\theta_1 = 0.1$	14.06	0.20	1.56	11.71	91.94	0.40	2.11
AR(1) with $\phi_1 = 0.1$	14.31	0.28	1.55	11.95	92.07	0.39	2.10
AR(1) with $\phi_1 = 0.5$	14.55	0.44	1.56	15.58	93.45	0.41	2.41
AR(1) with $\phi_1 = -0.75$	14.21	0.30	1.56	10.76	95.30	0.41	2.20
AR(1) with $\phi_1 = -0.5$	14.03	0.22	1.56	11.92	93.48	0.39	2.18
AR(1) with $\phi_1 = -0.25$	14.12	0.18	1.56	10.26	94.31	0.40	2.12
AR(1) with $\phi_1 = 0.25$	14.09	0.17	1.57	9.99	93.72	0.40	2.15
AR(1) with $\phi_1 = 0.5$	14.15	0.21	1.56	12.07	93.13	0.39	2.13
AR(1) with $\phi_1 = 0.75$	14.04	0.31	1.56	12.43	96.21	0.39	2.20
MA(1) with $\theta_1 = 1$	14.23	0.19	1.58	12.14	94.36	0.43	2.20

Table 4.2: *Table of Empirical Critical Values for for long series ($n = 1024$) under different simulation scenarios described in Section 4.1.*

	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
MA(1) with $\theta_1 = 0.1$	17.02	0.10	1.47	8.60	171.83	0.17	3.50
AR(1) with $\phi_1 = 0.1$	17.06	0.15	1.47	7.94	170.54	0.17	3.54
AR(1) with $\phi_1 = 0.5$	17.05	0.11	1.48	7.56	170.70	0.17	3.61
AR(1) with $\phi_1 = -0.75$	16.90	0.17	1.47	7.39	169.77	0.17	3.55
AR(1) with $\phi_1 = -0.5$	17.04	0.11	1.47	7.24	172.28	0.17	3.71
AR(1) with $\phi_1 = -0.25$	16.58	0.09	1.47	9.06	172.13	0.17	3.62
AR(1) with $\phi_1 = 0.25$	16.92	0.09	1.47	8.14	169.90	0.17	3.54
AR(1) with $\phi_1 = 0.5$	16.80	0.11	1.47	7.84	169.13	0.17	3.60
AR(1) with $\phi_1 = 0.75$	16.82	0.16	1.48	7.71	171.66	0.17	3.53
MA(1) with $\theta_1 = 1$	16.93	0.10	1.48	9.69	170.60	0.18	3.61

The second phase consisted of obtaining estimates of the power when the two underlying processes are different. For this phase, we performed 1,000 replicate simulations of each pair of processes since good coverage of a range of cases was more important than precise estimation of power in any particular case. We try to look at different scenarios as introduced in Section 4.1. These scenarios includes pairs of processes with values of parameters adjusted so that $\{X_{1,t}\}$ and $\{X_{2,t}\}$ have the same variance. The power of the test therefore derives from shape differences between the two underlying spectra. Scenarios that pairs of processes differed only in their variances, in order to assess the robustness of the test to purely scale differences in the underlying spectra, are also included. The empirical powers are summarized in the following tables and figures.

4.2.1 Comparison of Empirical Powers under Setting 1

Figure 4.1: *Figure of the spectral density for MA(1) with different parameter values.*

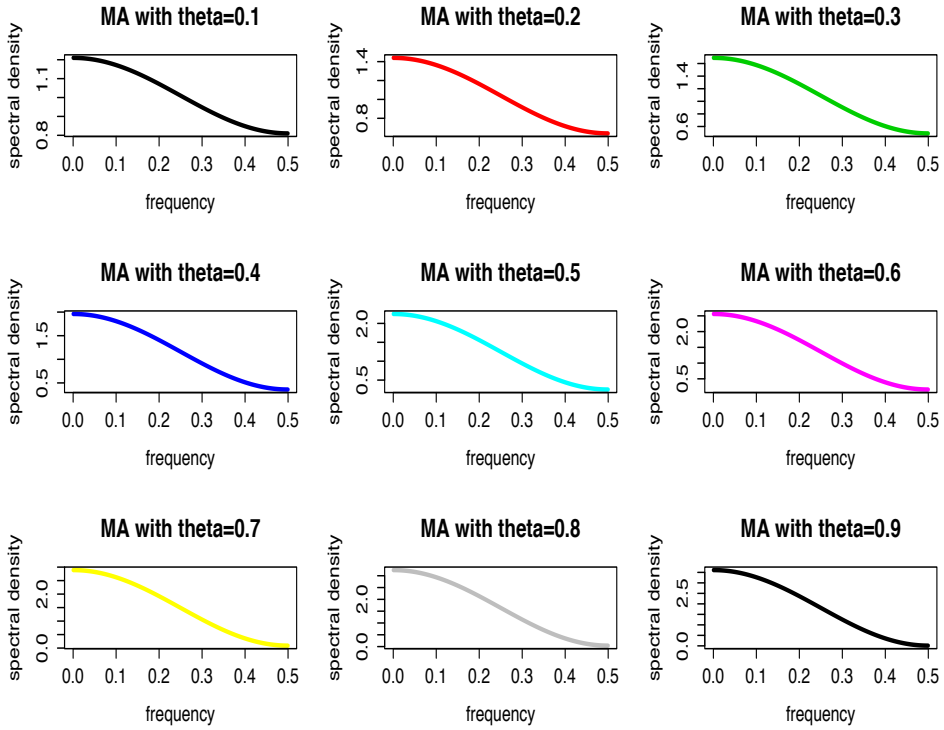


Table 4.3: *Empirical powers of 5% tests for short series ($n = 256$) and long series ($n = 1024$) under Setting 1 for MA(1) with $\theta_1 = 0.1$ vs MA(1) with different values of θ_1'. R is for Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios; D_m is for Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized cumulative periodograms; $\overline{D}$ is for Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms; $\overline{LL}^\star$ is for Lu and Li (2013) [17]'s frequency domain test which applies Jianqing Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test; Q_n, $Q_n^\star$, and $\widetilde{Q}_n$ are tapering tests proposed in this paper.*

(a) short series: n=256

θ_1'	R	D_m	$\overline{D}$	$\overline{LL}^\star$	Q_n	$Q_n^\star$	$\widetilde{Q}_n$
0.2	0.04	0.05	0.06	0.04	0.06	0.07	0.12
0.3	0.05	0.23	0.08	0.08	0.07	0.10	0.24
0.4	0.06	0.51	0.14	0.17	0.10	0.21	0.50
0.5	0.05	0.76	0.21	0.32	0.13	0.38	0.77
0.6	0.09	0.89	0.36	0.50	0.18	0.64	0.93
0.7	0.10	0.94	0.56	0.66	0.25	0.87	0.99
0.8	0.10	0.97	0.80	0.75	0.42	0.98	1.00
0.9	0.17	0.97	0.93	0.85	0.59	1.00	1.00

(b) long series: n=1024

θ_1'	R	D_m	$\overline{D}$	$\overline{LL}^\star$	Q_n	$Q_n^\star$	$\widetilde{Q}_n$
0.15	0.04	0.06	0.05	0.06	0.06	0.06	0.38
0.2	0.06	0.29	0.06	0.10	0.07	0.08	0.57
0.25	0.04	0.61	0.07	0.25	0.06	0.13	0.78
0.3	0.05	0.87	0.11	0.55	0.08	0.29	0.93
0.35	0.06	0.98	0.16	0.82	0.10	0.54	0.98
0.4	0.06	1.00	0.20	0.95	0.13	0.75	1.00
0.45	0.04	1.00	0.32	0.99	0.16	0.93	1.00
0.5	0.07	1.00	0.50	1.00	0.21	0.98	1.00

Figure 4.2: *Figure of simulated powers under Setting 1 for comparing MA(1) with* $\theta_1 = 0.1$ *vs MA(1) with* θ_1' *for short series* ($n = 256$) *and long series* ($n = 1024$).

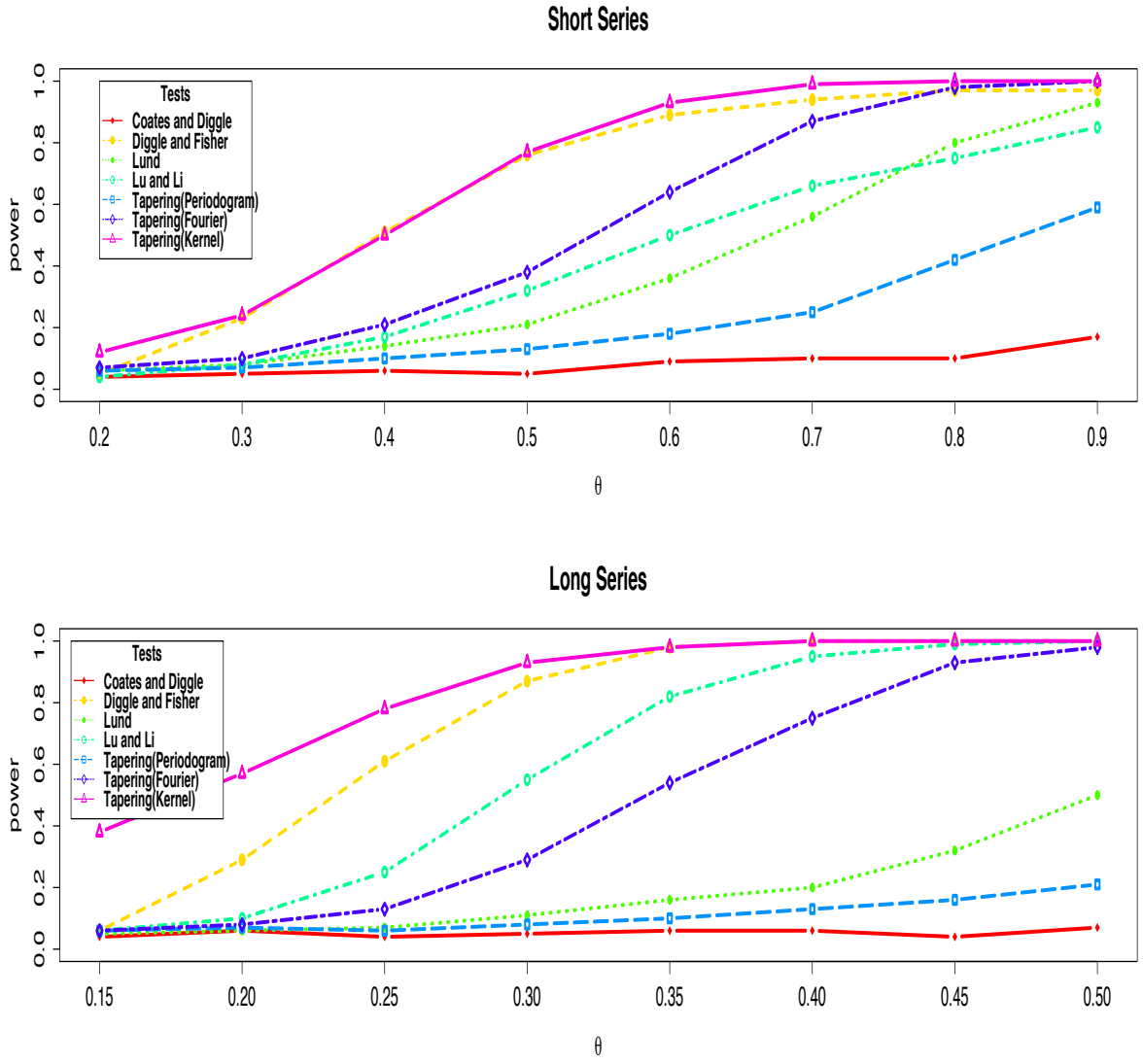


Table 4.3 summarized empirical powers of these seven tests we want to compare as introduced before, when both time series are moving average process of order 1 but with different parameter values as listed in the table. Results are summarized in this table for both short series and long series. Also note that for here, we fix $\theta_1 = 0.1$ and try different θ_1' values. We also tried other values of θ_1, but still get similar results. Thus we will only show results of $\theta_1 = 0.1$ as an example. Note that parameter settings for long series are designed to have narrower ranges of parameter values to reflect the generally higher power of tests. The corresponding spectral densities under

each parameter setting are plotted in Figure 4.1. As in the Figure 4.1, the spectral densities of MA(1) process are quite smooth and very similar under different parameter values. Additionally, to better visualize the power comparison, we also plot these empirical powers in Figure 4.2. From Figure 4.1, Table 4.3, and Figure 4.2, we find that long series generally will have higher empirical power for all tests compared with results of short series. Additionally, even the spectral densities of these two time series are very similar, for both short series and long series, the newly proposed tapering test based on Model 3 (Kernel Smoothing) still performs very good with the highest power; Diggle and Fisher (1991) [10]'s test also performs very good; the tapering test based on Model 2 (Fourier Transformation) as well as Lu and Li (2013) [17]'s test based on adaptive Neyman test performs quite similar with second highest empirical power; Lund et al. (2009) [18]'s test performs worse than those tests but better than the tapering test based on Model 1 (Raw Periodogram); and Coates and Diggle (1986) [8]'s test performs the worst with the lowest empirical power.

4.2.2 Comparison of Empirical Powers under Setting 2

Figure 4.3: *Figure of the spectral density for AR(1) with different parameter values.*

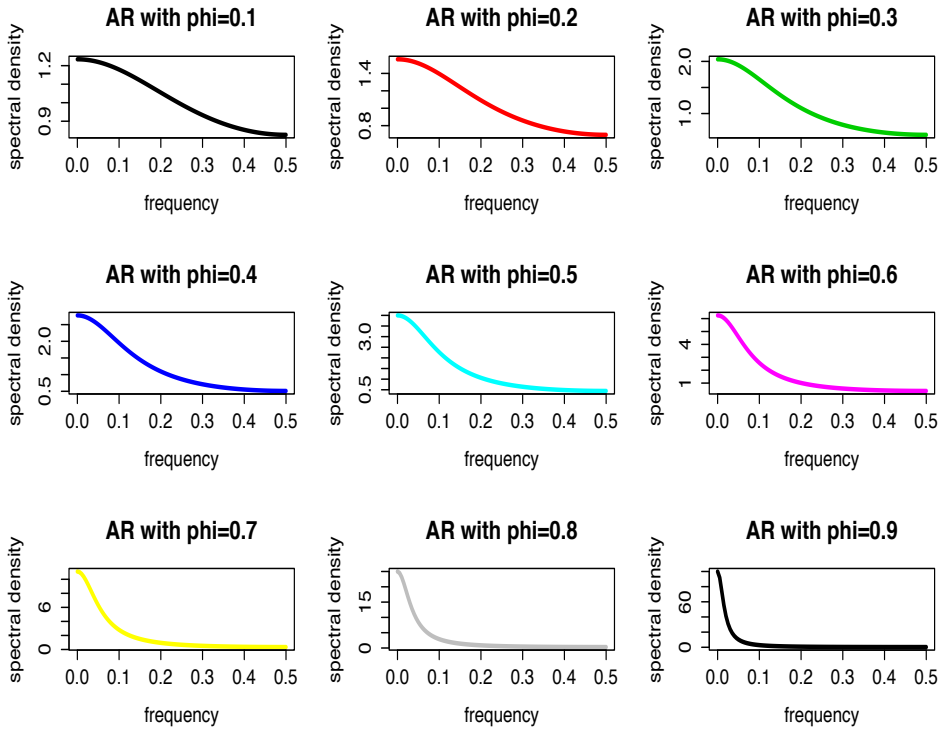


Table 4.4: *Empirical powers of 5% tests for short series ($n = 256$) and long series ($n = 1024$) under Setting 2 for AR(1) with $\phi_1 = 0.1$ vs AR(1) with different values of ϕ_1'. R is for Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios; D_m is for Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized cumulative periodograms; $\overline{D}$ is for Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms; $\overline{LL}^{\star}$ is for Lu and Li (2013) [17]'s frequency domain test which applies Jianqing Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test; Q_n, $Q_n^{\star}$, and $\widetilde{Q}_n$ are tapering tests proposed in this paper.*

(a) short series: n=256

ϕ_1'	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
0.2	0.04	0.05	0.06	0.04	0.06	0.07	0.12
0.3	0.05	0.23	0.08	0.08	0.07	0.10	0.24
0.4	0.06	0.51	0.14	0.17	0.10	0.21	0.50
0.5	0.05	0.76	0.21	0.32	0.13	0.38	0.77
0.6	0.09	0.89	0.36	0.50	0.18	0.64	0.93
0.7	0.10	0.94	0.56	0.66	0.25	0.87	0.99
0.8	0.10	0.97	0.80	0.75	0.42	0.98	1.00
0.9	0.17	0.97	0.93	0.85	0.59	1.00	1.00

(b) long series: n=1024

ϕ_1'	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
0.15	0.04	0.00	0.06	0.05	0.05	0.07	0.07
0.2	0.05	0.02	0.07	0.11	0.06	0.08	0.20
0.25	0.05	0.13	0.07	0.29	0.06	0.16	0.44
0.3	0.05	0.52	0.13	0.59	0.09	0.32	0.75
0.35	0.07	0.89	0.15	0.84	0.12	0.57	0.95
0.4	0.06	0.99	0.24	0.96	0.19	0.80	0.99
0.45	0.06	1.00	0.37	1.00	0.29	0.95	1.00
0.5	0.07	1.00	0.51	1.00	0.43	1.00	1.00

Figure 4.4: *Figure of simulated powers under Scenario 2 for comparing AR(1) with* $\phi_1 = 0.1$ *vs AR(1) with* ϕ_1'.

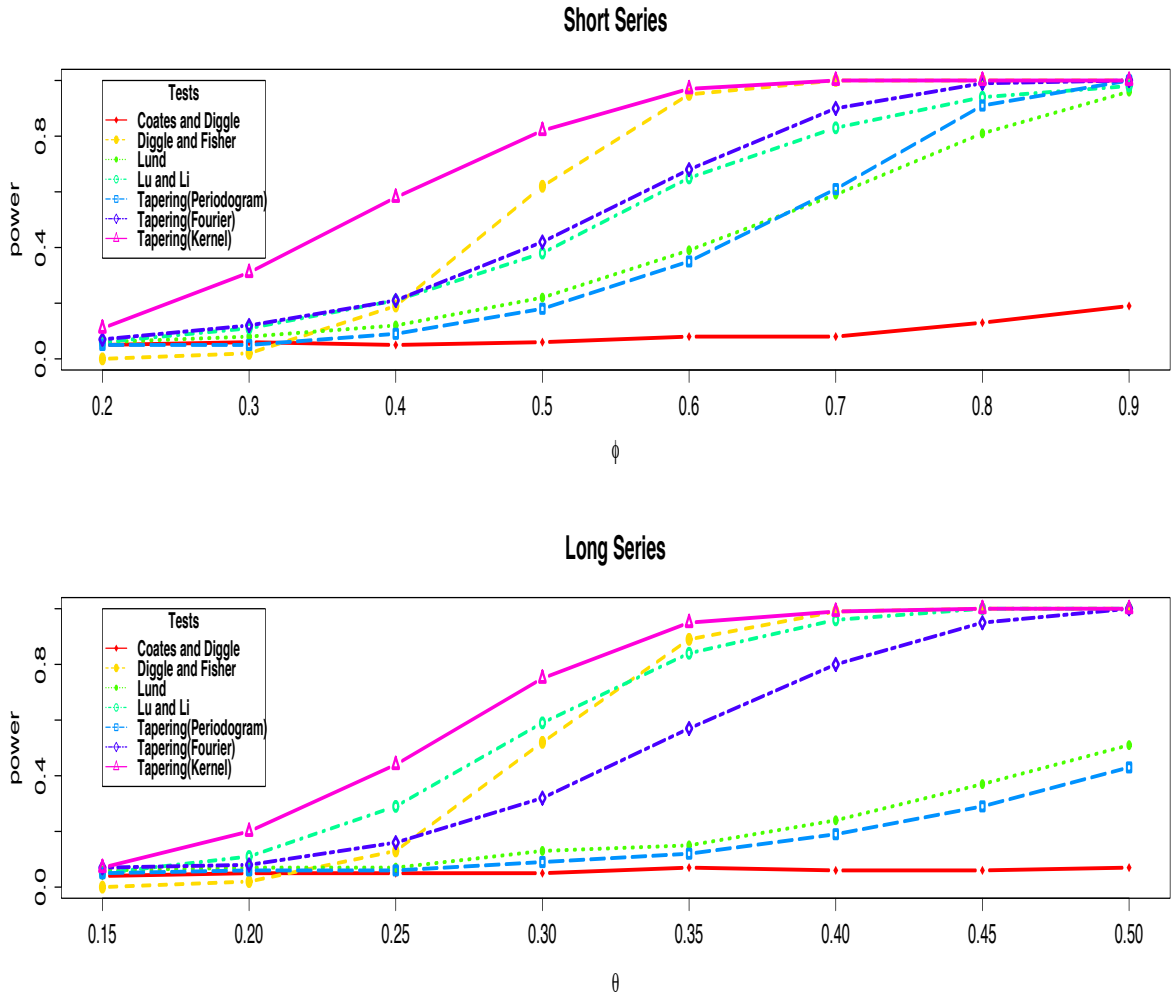


Table 4.4 summarized empirical powers of these seven tests we want to compare as introduced before, when both time series are auto regressive process of order 1 but with different parameter values as listed in the table. Results are summarized in this table for both short series and long series. Also note that for here, we fix $\phi_1 = 0.1$ and try different ϕ_1' values. We also tried other values of ϕ_1, but still get similar results. Thus we will only show results of $\phi_1 = 0.1$ as an example. Note that parameter settings for long series are designed to have narrower ranges of parameter values to reflect the generally higher power of tests. The corresponding spectral densities under each parameter setting are plotted in Figure 4.3. As in the Figure 4.3, the spectral densities of AR(1) process are quite smooth and have different shapes under different parameter values. Additionally,

to better visualize the power comparison, we also plot these empirical powers in Figure 4.4. From Figure 4.3, Table 4.4, and Figure 4.4, we find that long series generally will have higher empirical power for all tests compared with results of short series. And we can can declare that, for both short series and long series, the newly proposed tapering test based on Model 3 (Kernel Smoothing) still performs very good with the highest power; the performance of Diggle and Fisher (1991) [10]'s test, the tapering test based on Model 2 (Fourier Transformation) and Lu and Li (2013) [17]'s test based on adaptive Neyman test are very similar, with Lu and Li (2013) [17]'s test will have higher power for short series; Lund et al. (2009) [18]'s test as well as the tapering test based on Model 1 (Raw Periodogram) perform worse than those tests; and Coates and Diggle (1986) [8]'s test performs the worst with the lowest empirical power.

4.2.3 Comparison of Empirical Powers under Setting 3

Figure 4.5: *Figure of the estimated spectral density for long-order Gaussian seasonal $AR(1)_{12}$ with different parameter values.*

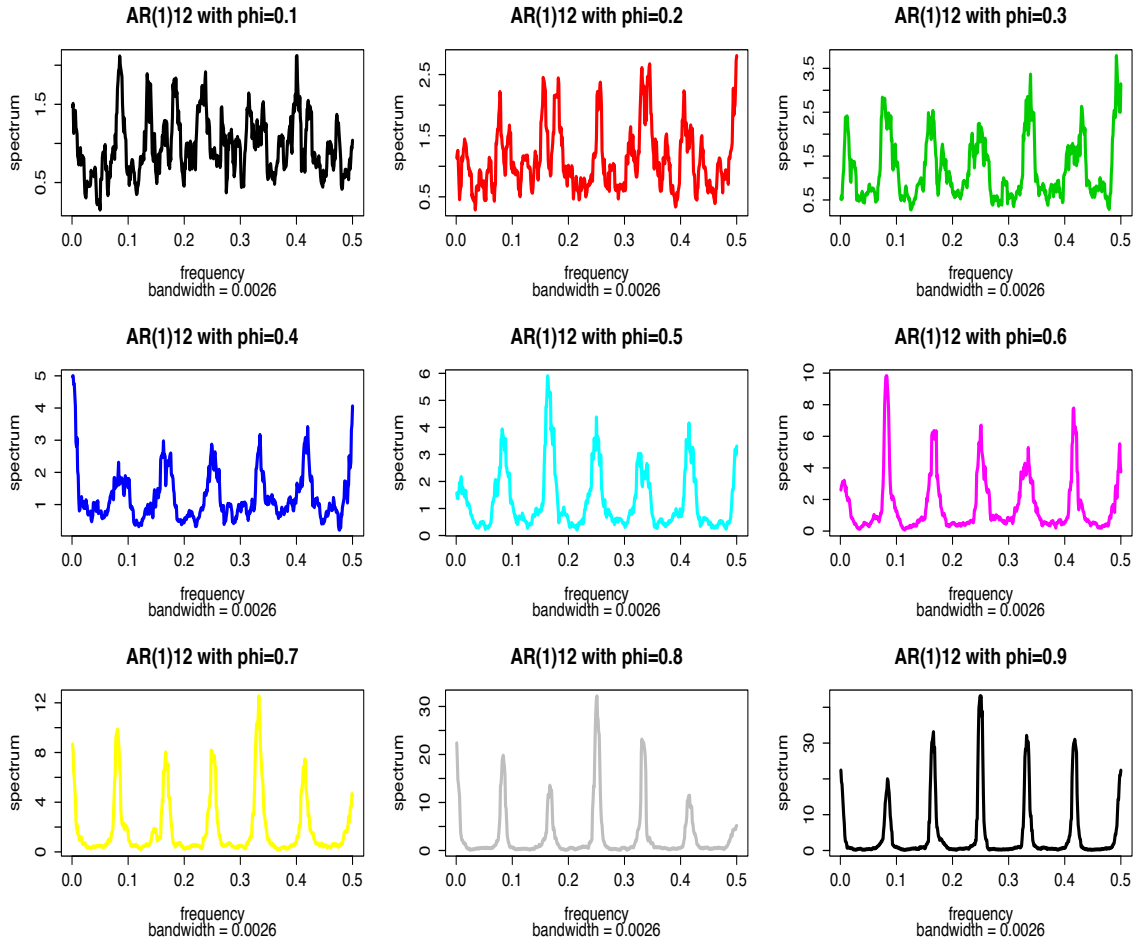


Table 4.5: *Empirical powers of 5% tests for short series ($n = 256$) and long series ($n = 1024$) under Setting 3 for AR(1) with $\phi_1 = 0.5$ vs long order Gaussian seasonal AR(1)$_{12}$ with different values of ϕ_1'. R is for Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios; D_m is for Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized cumulative periodograms; $\overline{D}$ is for Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms; $\overline{LL}^{\star}$ is for Lu and Li (2013) [17]'s frequency domain test which applies Jianqing Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test; Q_n, $Q_n^{\star}$, and $\widetilde{Q}_n$ are tapering tests proposed in this paper.*

(a) short series: n=256

ϕ_1'	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
0.1	0.07	0.05	0.31	0.55	0.34	0.53	0.94
0.2	0.08	0.05	0.34	0.44	0.33	0.56	0.93
0.3	0.08	0.07	0.42	0.34	0.38	0.64	0.94
0.4	0.10	0.09	0.50	0.17	0.42	0.69	0.94
0.6	0.13	0.16	0.80	0.02	0.62	0.90	0.94
0.7	0.14	0.22	0.90	0.00	0.74	0.96	0.95
0.8	0.19	0.33	0.97	0.00	0.86	1.00	0.94
0.9	0.25	0.44	0.99	0.00	0.95	1.00	0.96

(b) long series: n=1024

ϕ_1'	R	D_m	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$
0.3	0.10	1.00	0.93	1.00	0.71	1.00	1.00
0.35	0.10	1.00	0.97	1.00	0.76	1.00	1.00
0.4	0.09	1.00	0.98	1.00	0.80	1.00	1.00
0.45	0.09	1.00	0.99	1.00	0.88	1.00	1.00
0.55	0.14	1.00	1.00	1.00	0.95	1.00	1.00
0.6	0.14	1.00	1.00	1.00	0.98	1.00	1.00
0.65	0.12	1.00	1.00	1.00	0.99	1.00	1.00
0.7	0.18	1.00	1.00	1.00	1.00	1.00	1.00

Figure 4.6: *Figure of simulated powers under Scenario 3 for comparing AR(1) with* $\phi_1 = 0.5$ *vs long-order Gaussian seasonal AR(1)*$_{12}$ *with* ϕ_1'.

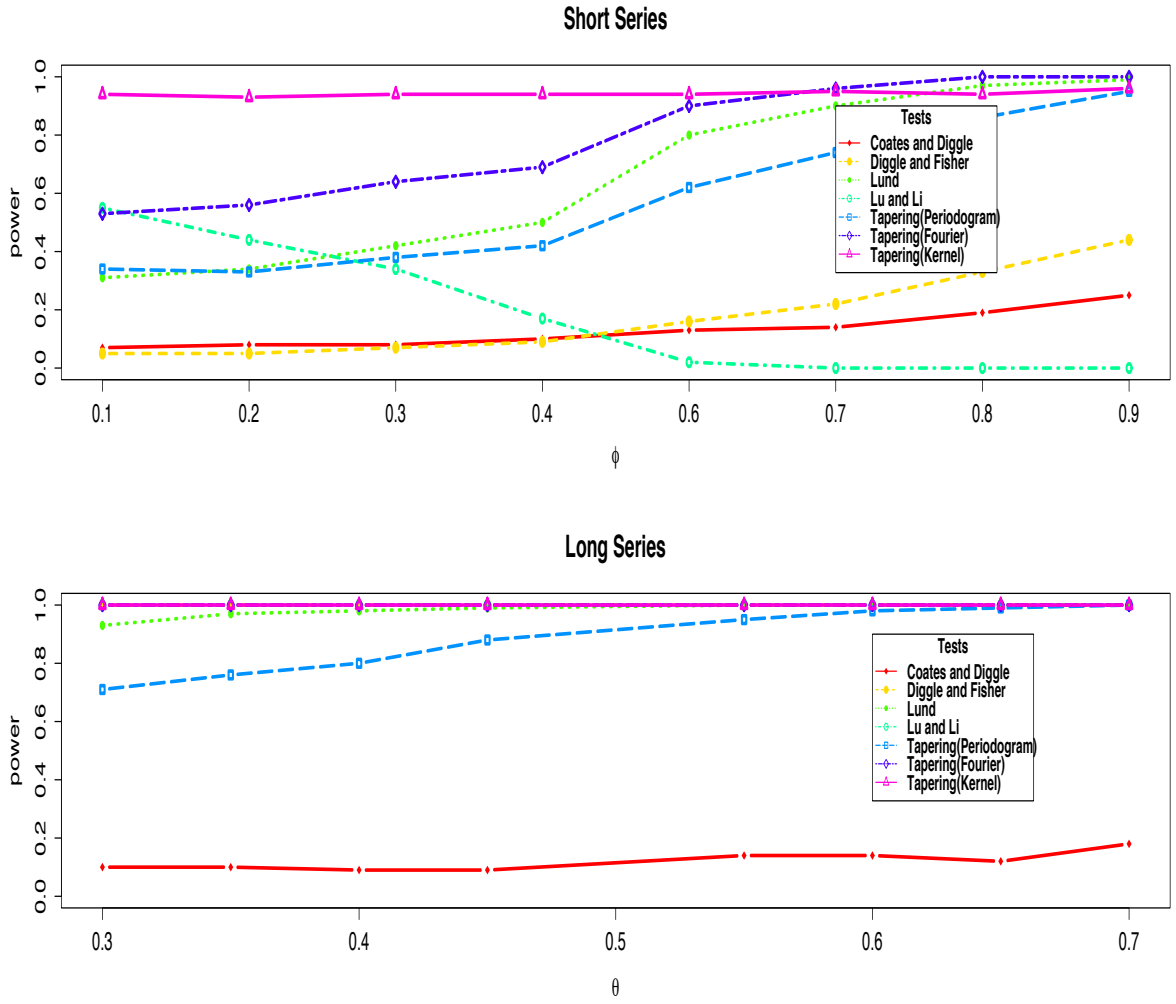


Table 4.5 summarized empirical powers of these seven tests we want to compare as introduced before, when one time series is auto regressive process of order 1 with parameter value 0.5 and the other time series is long-order Gaussian seasonal auto regressive process with different parameter values as listed in the table. Results are summarized in this table for both short series and long series. Also note that for here, we fix $\phi_1 = 0.5$ and try different ϕ_1'. We also tried other values of ϕ_1, but still get similar results. Thus we will only show results of $\phi_1 = 0.5$ as an example. Note that parameter settings for long series are designed to have narrower ranges of parameter values to reflect the generally higher power of tests. The corresponding spectral densities under each parameter setting are plotted in Figure 4.5. the spectral densities of AR(1)$_{12}$ process are

not smooth and have a lot of spikes. Compared with the smooth spectral density curve of AR(1) with $\phi_1 = 0.5$ as plotted in Figure 4.3, these tests should detect the difference with higher power. Additionally, to better visualize the power comparison, we also plot these empirical powers in Figure 4.6. From Table 4.5 and Figure 4.6, it is clear that when comparing the time series with spikes in its spectral density curve with the time series with smooth spectral density curve, these tests generally have higher empirical powers. Additionally, we can find that long series generally will have higher empirical power for all tests compared with results of short series. Comparing among those tests, the newly proposed tapering test based on Model 3 (Kernel Smoothing) also have the highest power; the tapering test based on Model 2 (Fourier Transformation) as well as Diggle and Fisher (1991) [10]'s test also performs very good for long series, however, for short series, the tapering test based on Model 2 (Fourier Transformation) has the secondly highest powers while Diggle and Fisher (1991) [10]'s test performs extremely bad; Lund et al. (2009) [18]'s test performs a little bit better than the tapering test based on Model 1 (Raw Periodogram); and Coates and Diggle (1986) [8]'s test performs the worst with lowest empirical power. Note that the performance of Lu and Li (2013) [17]'s test is not good for short series. It is mainly because that even the adaptive Neyman test should always outperform tapering test asymptotically, but the way they apply the adaptive Neyman test as in the paper is not proper. They waste a lot of information from the data. That's also the reason we want to propose the tapering test based on Model 2 (Fourier Transformation) which is an improvement of their test.

4.2.4 Comparison of Empirical Powers under Setting 4

Figure 4.7: *Figure of the estimated spectral density for AR(1) and MA(1) with different parameter values.*

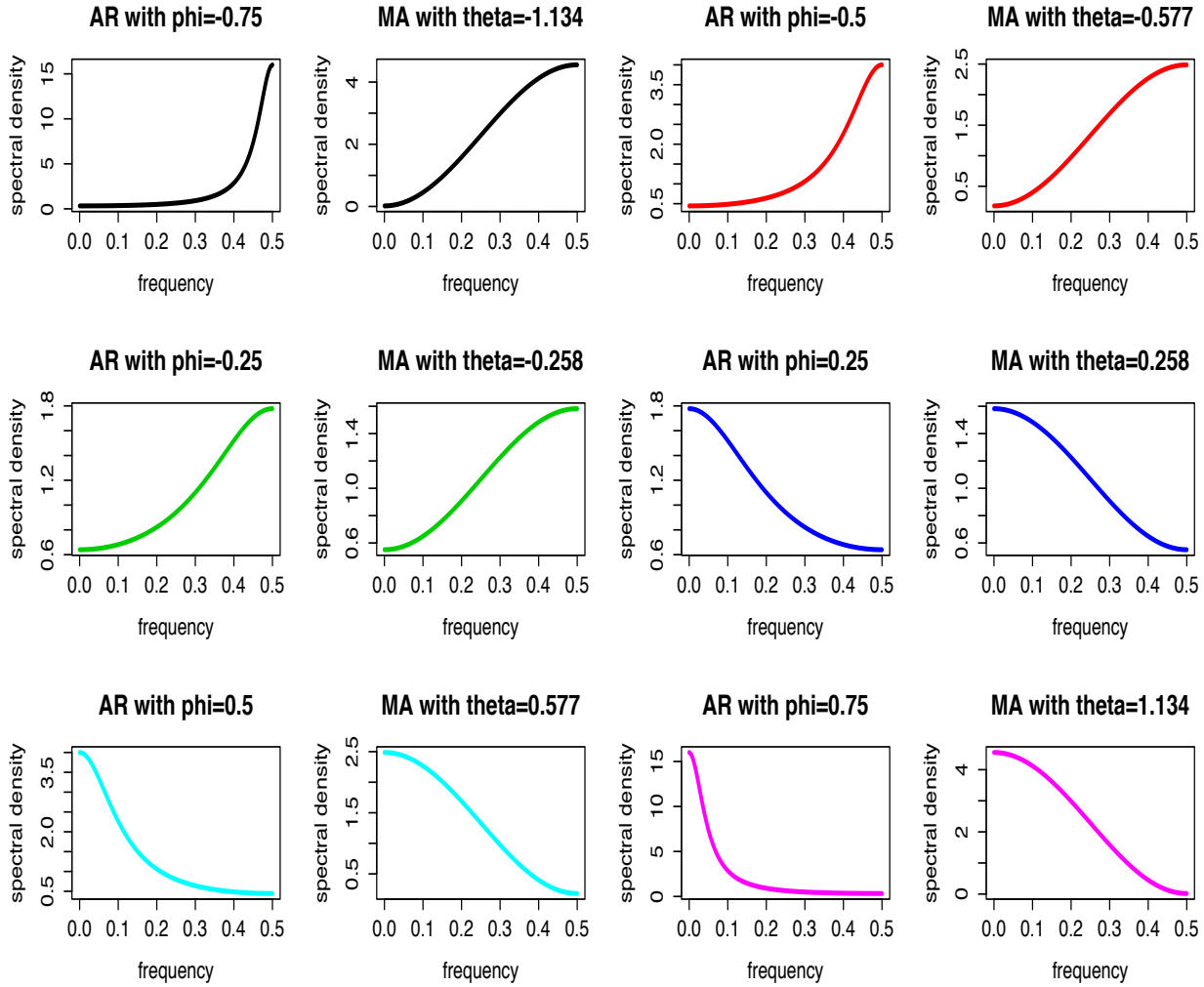


Table 4.6: *Empirical powers of 5% tests for short series ($n = 256$) and long series ($n = 1024$) under Setting 4 for AR(1) with various values of ϕ_1 versus MA(1) with $\theta_1 = \pm\sqrt{\phi_1^2/(1-\phi_1^2)}$. R is for Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios; D_m is for Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized cumulative periodograms; $\overline{D}$ is for Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms; $\overline{LL}^{\star}$ is for Lu and Li (2013) [17]'s frequency domain test which applies Jianqing Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test; Q_n, $Q_n^{\star}$, and $\widetilde{Q}_n$ are tapering tests proposed in this paper.*

(a) short series: n=256

ϕ_1	θ_1	R	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$	
-0.75	-1.13	0.12	0.88	0.71	1.00	0.63	0.99	0.99
-0.5	-0.58	0.05	0.26	0.08	0.40	0.11	0.21	0.27
-0.25	-0.26	0.06	0.06	0.06	0.07	0.08	0.08	0.08
0.25	0.26	0.06	0.07	0.06	0.05	0.05	0.06	0.07
0.5	0.58	0.05	0.28	0.12	0.27	0.06	0.25	0.19
0.75	1.13	0.11	0.86	0.73	0.99	0.39	0.99	0.98

(b) long series: n=1024

ϕ_1	θ_1	R	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$	
-0.75	-1.13	0.08	1.00	1.00	1.00	1.00	1.00	1.00
-0.5	-0.58	0.07	0.87	0.21	1.00	0.20	0.90	0.99
-0.25	-0.26	0.06	0.09	0.06	0.09	0.05	0.07	0.11
0.25	0.26	0.05	0.12	0.06	0.11	0.04	0.08	0.11
0.5	0.58	0.05	0.87	0.21	1.00	0.10	0.92	0.97
0.75	1.13	0.13	1.00	1.00	1.00	0.88	1.00	1.00

Figure 4.8: *Figure of simulated powers under Scenario 4 for comparing AR(1) with ϕ_1 versus MA(1) with $\theta_1 = \pm\sqrt{\phi_1^2/(1-\phi_1^2)}$.*

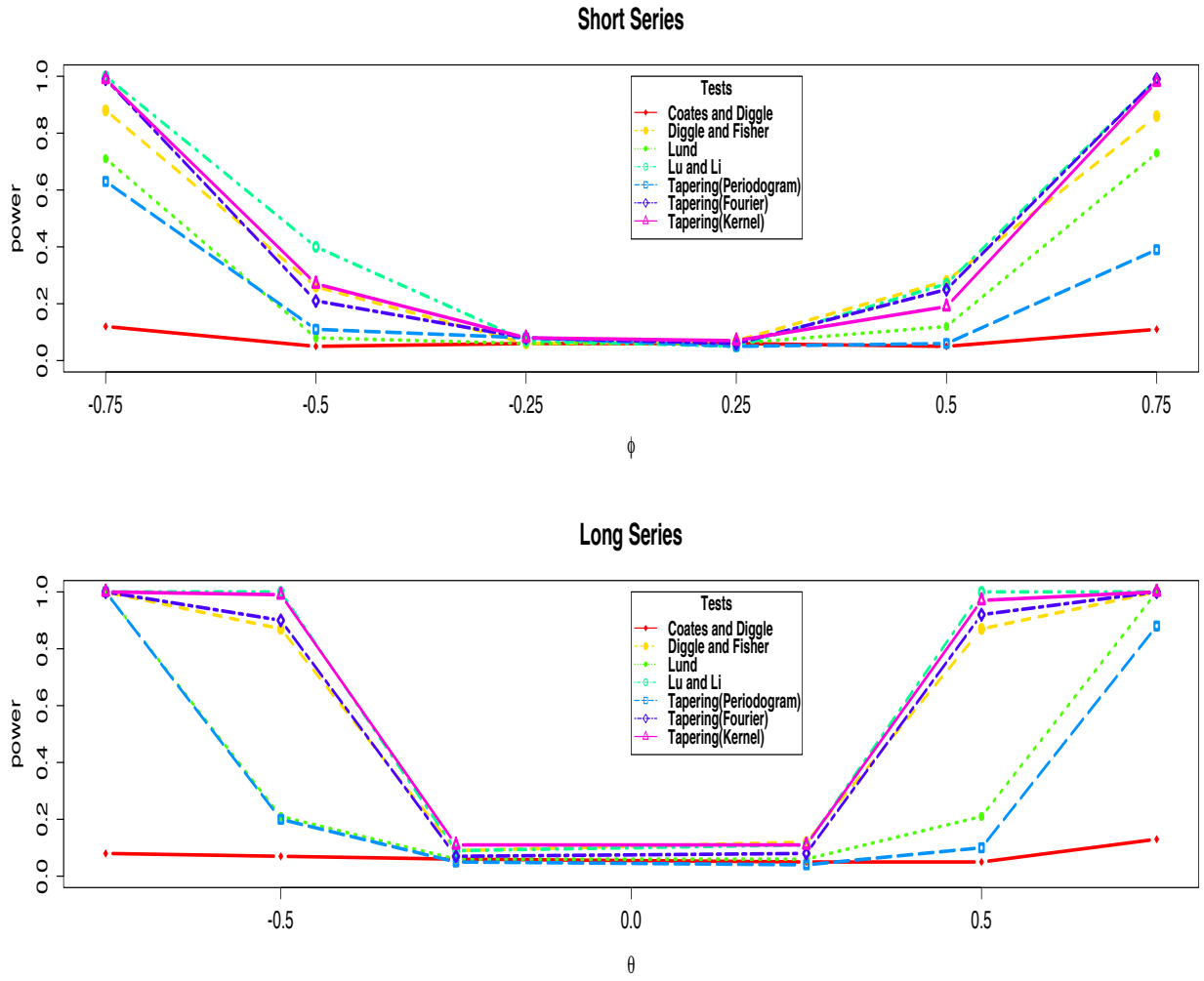


Table 4.6 summarized empirical powers of these seven tests we want to compare as introduced before, when one time series is auto regressive process of order 1 and the other time series is moving average process of order one with different parameter values as listed in the table. Results are summarized in this table for both short series and long series. Except those values listed in the table, we also tried other values of ϕ_1, but still get similar results. Thus we only show results of ϕ_1 values defined in the table as an example. Also note that this choice of parameter values for these two time series makes the variances of the two series identical; moreover, the lag-one auto-covariances have the same sign. The corresponding spectral densities are plotted in Figure 4.7. As

in the Figure 4.7, the spectral densities are quite smooth and have different shapes under different parameter values. But when $\phi_1 = -0.25$ and 0.25, the spectral density curve for these two time series are quite similar. Then these tests have low powers under these conditions as shown in Table 4.6. Additionally, to better visualize the power comparison, we also plot these empirical powers in Figure 4.8. From Figure 4.7, Table 4.6 and Figure 4.8, we can declare that long series generally will have higher empirical power for all tests compared with results of short series. Comparing among those tests, the newly proposed tapering test based on Model 3 (Kernel Smoothing) still performs very good; Lu and Li (2013) [17]'s test, the tapering test based on Model 2 (Fourier Transformation) and Diggle and Fisher (1991) [10]'s test also perform very good with similar empirical powers with tapering test based on Model 3 (Kernel Smoothing); Lund et al. (2009) [18]'s test and the tapering test based on Model 1 (Raw Periodogram) have similar performance which is not so good; and Coates and Diggle (1986) [8]'s test performs the worst with lowest empirical power.

4.2.5 Comparison of Empirical Powers under Setting 5

Figure 4.9: *Figure of the estimated spectral density for MA(1) and MA(2) with different parameter values.*

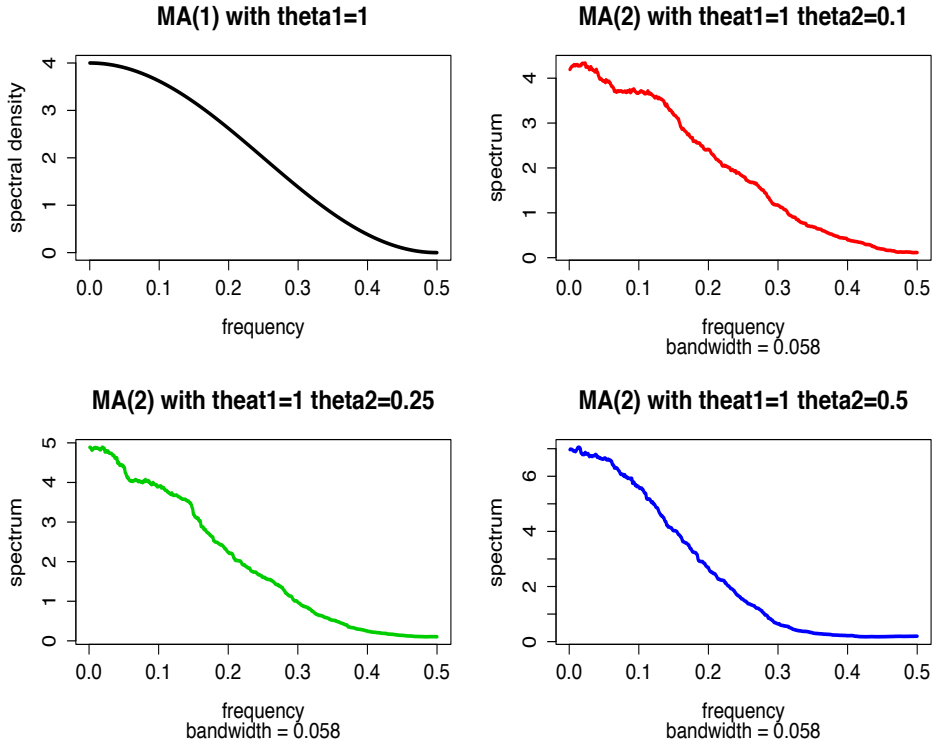


Table 4.7: *Empirical powers of 5% tests for short series ($n = 256$) and long series ($n = 1024$) under Setting 5 for MA(1) with $\theta_1 = 1$ versus MA(2) with $\theta_1' = 1$, and various values of θ_2'. R is for Coates and Diggle (1986) [8]'s test based on the range of periodogram ratios; D_m is for Diggle and Fisher (1991) [10]'s test based on Kolmogorov-Smirnov statistics of the normalized cumulative periodograms; $\overline{D}$ is for Lund et al. (2009) [18]'s frequency domain test based on the average of log ratio of periodograms; $\overline{LL}^{\star}$ is for Lu and Li (2013) [17]'s frequency domain test which applies Jianqing Fan (1996) [11]'s adaptive Neyman tests idea to the Lund et al. (2009) [18]'s test; Q_n, $Q_n^{\star}$, and $\widetilde{Q}_n$ are tapering tests proposed in this paper.*

(a) short series: n=256

θ_2'	R	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$	
0.1	0.07	0.07	0.06	0.07	0.06	0.05	0.06
0.25	0.07	0.22	0.15	0.24	0.10	0.20	0.16
0.5	0.15	0.54	0.59	0.73	0.23	0.88	0.62

(b) long series: n=1024

θ_2'	R	$\overline{D}$	$\overline{LL}^{\star}$	Q_n	$Q_n^{\star}$	$\widetilde{Q}_n$	
0.1	0.05	0.18	0.08	0.13	0.05	0.12	0.14
0.25	0.07	0.73	0.47	0.97	0.14	0.93	0.94
0.5	0.17	1.00	1.00	1.00	0.67	1.00	1.00

Figure 4.10: *Figure of simulated powers under Scenario 5 for comparing MA(1) with* $\theta_1 = 1$ *versus MA(2) with* $\theta_1' = 1, \theta_2'$.

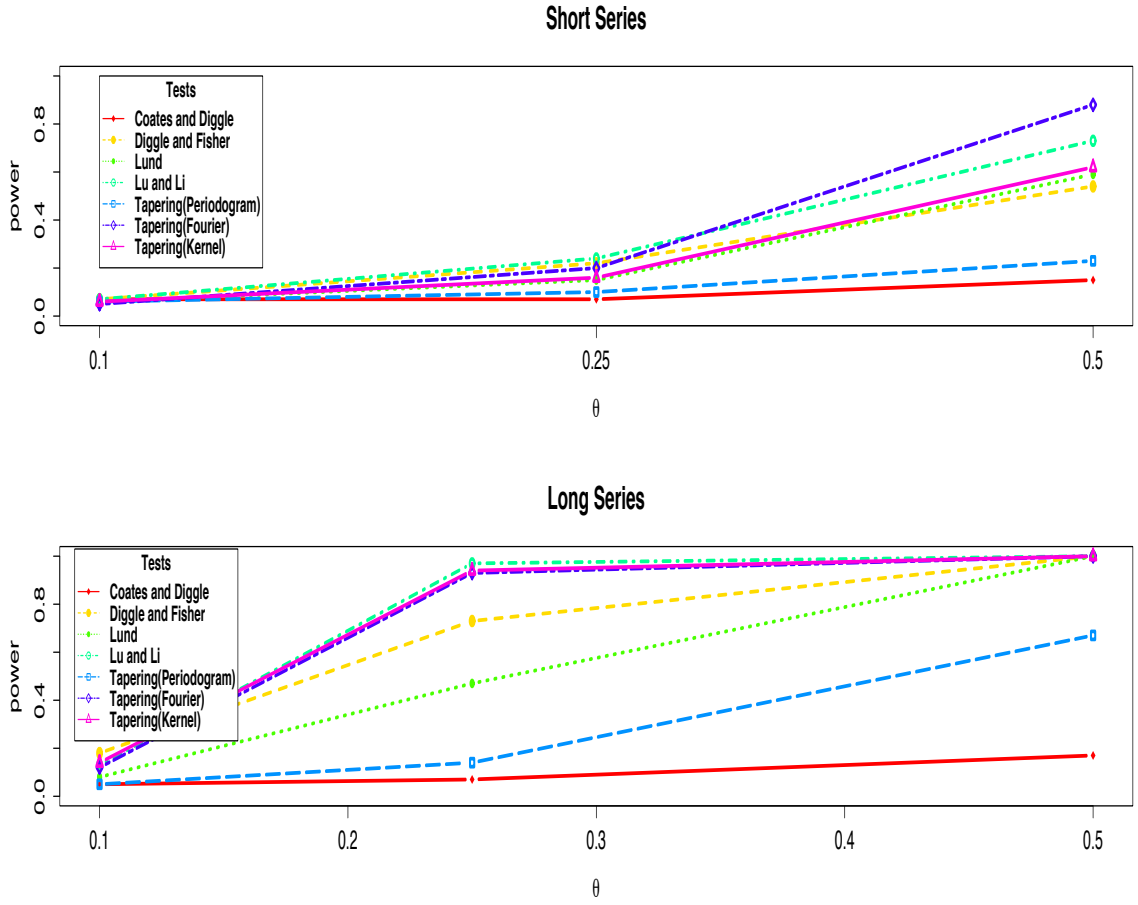


Table 4.7 summarized empirical powers of these seven tests we want to compare as introduced before, when one time series is moving average process of order 1 and the other time series is moving average process of order two with different parameter values as listed in the table. Results are summarized in this table for both short series and long series. This choice of parameter values for these two time series leads to the difference of them only in their second-order moving-average coefficient. Also note that for here, we fix $\theta_1 = \theta_1' = 1$ and try different θ_2' values. We also tried other values of these parameters, but still get similar results. Thus we will only show these results as the example. The corresponding spectral densities are plotted in Figure 4.9. As in the Figure 4.9, the spectral densities have similar shapes under different parameter values. Then these tests have lower powers as shown in Table 4.7. Additionally, to better visualize the power comparison, we

also plot these empirical powers in Figure 4.10. From Figure 4.9, Table 4.7 and Figure 4.10, we can declare that long series generally will have higher empirical power for all tests compared with results of short series. Comparing among those tests, the newly proposed tapering test based on Model 3 (Kernel Smoothing), the newly proposed tapering test based on Model 2 (Fourier Transformation) and the Lu and Li (2013) [17]'s test perform the best with highest empirical powers; Diggle and Fisher (1991) [10]'s test as well as Lund et al. (2009) [18]'s test have second highest empirical power; the tapering test based on Model 1 (Raw Periodogram) performs a little bit worse than Lund et al. (2009) [18]'s test; and Coates and Diggle (1986) [8]'s test performs the worst with lowest empirical power.

4.2.6 Summary

Based on the Table 4.3 to Table 4.7 as well as Figure 4.2 to Figure 4.10, we can make our final conclusion that our results from the comprehensive simulation study show that the newly proposed tapering test based on discrete compact Uniform Kernel smoothing with optimal tuning parameter, $\widetilde{Q}_n$, generally performs the best with the highest empirical power. Even it is observed that sometimes Diggle and Fisher (1991) [10]'s test or the recently proposed Lu and Li (2013) [17]'s test $\overline{LL}^{\star}$ or the tapering test based on the Fourier transform $Q_n^{\star}$ performs a little bit better, however, the difference is very small and the performance of the tapering test based on "optimal" kernel smoothing is the most stable test with very high power under each of these five settings, which covers a wide variety of testing scenarios. Consequently, the tapering test based on "optimal" kernel smoothing $\widetilde{Q}_n$ is recommended for testing equal spectral density for two independent stationary time series, which is very easy to implement in practice while has very good asymptotic and empirical performance. Additionally, the strong empirical powers shown here also provide supports for our argument of setting leading constant B equals to one.

Chapter 5

EEG Data Study

In Chapter 4, we conducted simulations to assess the empirical power of these newly proposed tapering test, comparing with these popular tests in literature. From simulation results, we learned that the newly proposed tapering test based on kernel smoothing always performs the best, also this test is also regarded as the most reasonable one from the point of view of time series. Based on the Bayesian tapering test framework derived in Chapter 3.1 from the Model 3 (Kernel Smoothing), we can now answer the scientific questions we proposed regarding to the motivating example EEG dataset.

5.1 Motivating Example: EEG Data Study

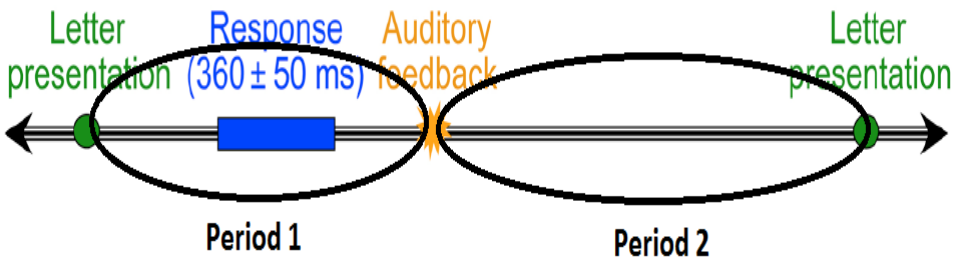
5.1.1 Outline of Study

From the simulation results in Section 4.2, we find that the Tapering test based on discrete Uniform Kernel smoothing with optimal rate of tuning parameter, i.e. $\widetilde{Q}_n$ performs best in general. Thus, for this section, we will further implement the Bayesian version of the tapering test $\widetilde{Q}_n$ as discussed in Chapter 3.1 to solve our real data problem in the Reward Two-back Study as discussed in Section 1.6. We will focus on EEG observations from one trial as shown in Figure 1.1. The duration of the trial is around 1.5 seconds, from letter shown up on the screen until the next letter shown up. Since the sampling rate is 256Hz, the length of the EEG observations will be 340 for the trial. In

addition to looking at the whole trial, we also split the trial into two periods, as in Figure 5.1. Period 1 is from letter shown up on the screen till the auditory feedback, and Period 2 is from the auditory feedback till the end of the trial. We will do separate testing for equal spectral densities under the following three scenarios:

- The whole trial: 340 observations
- Only Period 1: 116 observations
- Only Period 2: 224 observations

Figure 5.1: *Visualization of the two periods of one whole trial. The duration of one trial is around 1.5 second, from the letter shown up on the screen until the next letter shown up. The duration of Period 1 is around 0.7 seconds, from the letter shown up on the screen until the auditory feedback. The duration of Period 2 is around 0.8 seconds, from auditory feedback until the next letter shown up on the screen.*



As discussed in Section 1.6, among the four lobes on the brain, we will only focus on Frontal Lobe, Temporal Left Lobe and Occipital Lobe, which are highly active during the Reward Two-back study. From the data, we picked up series for each lobe:

- Temporal Left Lobe (TL): 10 series
- Occipital Lobe (OL): 15 series
- Frontal Lobe (FL): 15 series

We will do Bayesian tapering test on each pair of these three lobes, and do the same process separately for the whole trail, Period 1 and Period 2. After setting up the structure, two main questions we are trying to answer are

1. Similar patterns between lobes while performing cognitive tasks as in Reward Two-back study?
    - TL versus OL
    - TL versus FL
    - OL versus FL

2. Will this pattern changes while we are looking at these three different time periods?
    - Whole trial
    - Period 1
    - Period 2

As for each lobe, we have multiple time series, however, our Bayesian tapering test as in Chapter 3.1 is for testing the equal spectral density for two time series. Additionally, dealing directly with a high-dimensional spectral matrix itself is somewhat cumbersome because it is a function into the set of complex, nonnegative-definite, Hermitian matrices. Our problem can be restate as suppose have a stationary, p-dimensional, vector-valued process X_t and we are only able to keep a univariate process y_t such that, when needed, we may reconstruct the vector-valued process, X_t, according to an optimality criterion. So we further do spectral domain Principle Components Analysis (PCA) on the data matrix of each lobe. The topics of principal components in the frequency domain are rigorously presented in Brillinger (2001) [5] (Chapters 9 and 10) and many of the details concerning these concepts can be found in Shumway and Stoffer (2013) [26].

For the case of time series, suppose we have a zero mean, $p \times 1$, stationary vector process X_t that has a $p \times p$ spectral density matrix given by $f_{xx}(\omega)$. Recall $f_{xx}(\omega)$ is a complex-valued, nonnegative-definite, Hermitian matrix. Using the analogy of classical principal components, suppose, for a fixed value of ω, we want to find a complex-valued univariate process $y_t(\omega) = \mathbf{c}(\omega) * X_t$, where $\mathbf{c}(\omega)$ is a complex, such that the spectral density of $y_t(\omega)$ is maximized at frequency ω, and $\mathbf{c}(\omega)$ have unit length. Because at frequency ω, the spectral density of $y_t(\omega)$ is $f_y(\omega) = \mathbf{c}(\omega) f_{xx}(\omega) \mathbf{c}(\omega)$, it is equivalent to find the complex vector $\mathbf{c}(\omega)$ such that

$$\underset{\mathbf{c}(\omega) \neq 0}{max} \frac{\mathbf{c}(\omega) f_{xx}(\omega) \mathbf{c}(\omega)}{\mathbf{c}(\omega) \mathbf{c}(\omega)} \tag{5.1}$$

Suppose $\{(\lambda_1(\omega), \mathbf{e}_1(\omega)), \cdots, (\lambda_p(\omega), \mathbf{e}_p(\omega))\}$ are pairs of eigenvalues and eigenvectors of spectral density matrix $f_{xx}(\omega)$, where $\lambda_1(\omega) > \cdots > \lambda_p(\omega) > 0$ and eigenvectors $\mathbf{e}_1(\omega), \cdots, \mathbf{e}_p(\omega)$ are of unit length. Note that the eigenvalues of a Hermitian matrix are real, thus, $\lambda_1(\omega), \cdots, \lambda_p(\omega)$ are real values. The solution of Equation (5.1) would be $\mathbf{c}(\omega) = \mathbf{e}_1(\omega)$. Then the linear combination would be $y_t(\omega) = \mathbf{e}_1(\omega) * X_t$ and

$$\max_{\mathbf{c}(\omega)\neq 0} \frac{\mathbf{c}(\omega) f_{xx}(\omega) \mathbf{c}(\omega)}{\mathbf{c}(\omega)\mathbf{c}(\omega)} = \lambda_1(\omega)$$

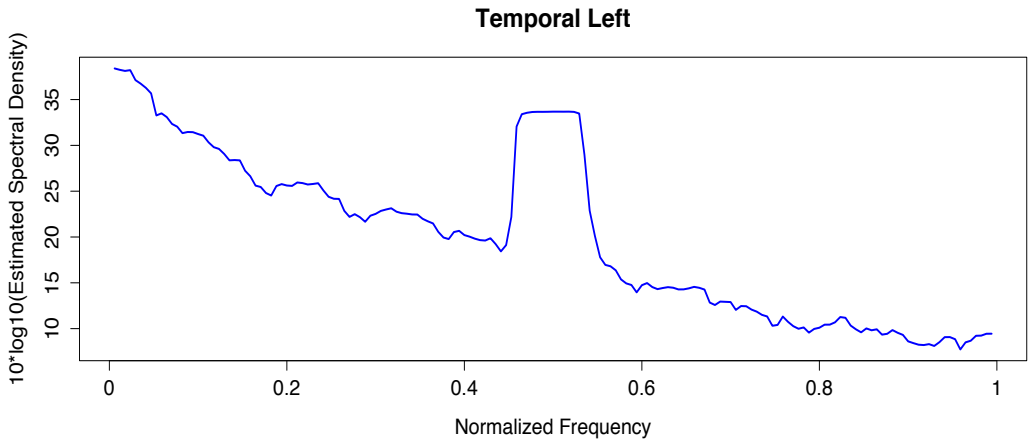
This process may be repeated for any frequency ω, and the complex-valued process, $y_{t1}(\omega)$, is called the first principal component series at frequency ω. The k-th principal component series at frequency ω, for $k = 1, \cdots, p$, is the complex-valued time series $y_{tk}(\omega) = \mathbf{c}_k(\omega) * X_t$, in analogy to the classical case. In this case, the spectral density of $y_{tk}(\omega)$ at frequency ω is $\lambda_k(\omega)$.

5.1.2 Results

For each of these three lobes, we firstly estimate the spectral density matrix for each frequency. For example, for the whole trial, Temporal Left Lobe, we will get a 10×10 estimated spectral density matrix for each of these 169 frequencies $\omega_1, \cdots, \omega_{169}$.

Then the second step is for each frequency, for the 10×10 estimated spectral density matrix, calculate the eigenvalues. In this step, for each lobe under whole trial, Period1 and Period 2, for each frequency ω, we checked the cumulative contribution rates of the principal components of the estimated spectral density matrix. Each of the first principal components have cumulative contribution rate higher than 90%. Thus, it is reasonable to consider the estimated spectral density of the first principle component series of each lobe to represent the spectral characteristics of the lobe, which is actually the first eigenvalue of each estimated spectral density matrix. In Figure 5.2, we plot one example of the estimated spectral density of the first principle component series of Temporal Left Lobe. For each of the other two lobes, we also get one series of the spectral density.

Figure 5.2: *One example of the estimated spectral density of the first principle component series of Temporal Left Lobe. Spectral density matrix is estimated based on the discrete Uniform Kernel smoothing with the optimal rate of tuning parameter.*



After the second step, for each lobe, we get one estimated spectral density series. For example for the whole trial, we will get

$$\left\{\widehat{f}_{TL}(\omega_1), \cdots, \widehat{f}_{TL}(\omega_{169})\right\}$$
$$\left\{\widehat{f}_{OL}(\omega_1), \cdots, \widehat{f}_{OL}(\omega_{169})\right\}$$
$$\left\{\widehat{f}_{FL}(\omega_1), \cdots, \widehat{f}_{FL}(\omega_{169})\right\}.$$

We will get similar series for Period 1 with length 57, and Period 2 with length 111. Then the third step is for each pair of the comparison in Question 1, calculate the log-ratio of the estimated spectral density, and then do the discrete Uniform Kernel smoothing with optimal rate of tuning parameter we derived in Section **??** on these log-ratios. Then we will get those $\widetilde{Y}_j$s as in Chapter 3.1.

The last step will be calculating the Bayes factor as discussed in Chapter 3.1. Following the setup and the prior specifications of Chapter 3.1, we will get Bayes factor values for each of the comparisons in Question 1 and for each scenario in Question 2.

$$BF_0(\widetilde{\mathbf{Y}}) = \frac{m\left(\widetilde{\mathbf{Y}}|H_0\right)}{m\left(\widetilde{\mathbf{Y}}|H_1\right)}.$$

Decisions will be made based on the values of $2\log BF$, and decision rules are discussed in Chapter

3.1. Results are listed in Table 5.1. From Table 5.1, we will find that, only Temporal Left Lobe

Table 5.1: *Results of doing Bayesian tapering test based on the discrete Uniform Kernel smoothing with the optimal rate of tuning parameter. This table is to answer Question 1 and Question 2 with each row is one comparison under different scenarios and each column is one scenario for different comparisons. These values listed in table are values of* $2 \log BF$ *and posterior null probabilities (inside the parentheses).*

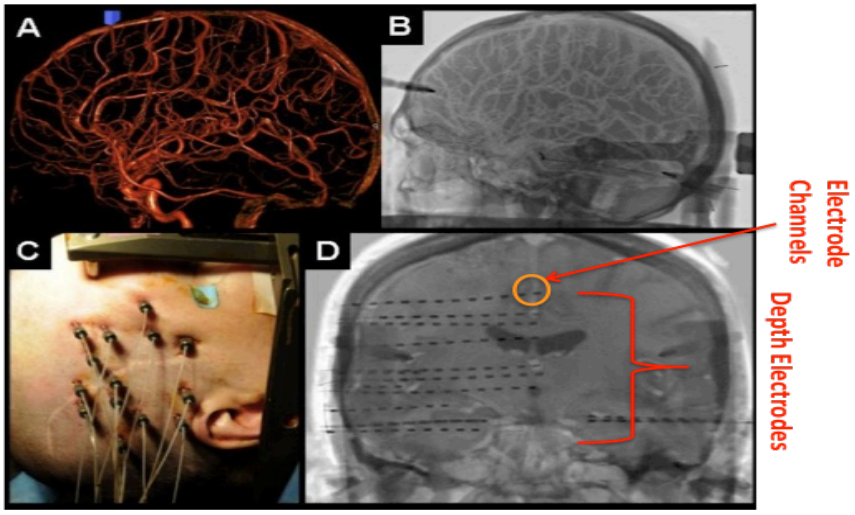
Comparison	Whole trial	Period 1	Period 2
TL vs OL	-144 (≈0)	-84 (≈0)	-112 (≈0)
TL vs FL	-1 (0.1)	10 (0.98)	2 (0.38)
OL vs FL	-254 (≈0)	-94 (≈0)	-184 (≈0)

and Frontal Lobe at Period 1 will have strongly similar pattern, and this pattern dose not appear in Period 2 and the whole trial. So an very interesting phenomenon of the brain lobes can be identified from our results: when the Temporal Left Lobe, Occipital Lobe and Frontal Lobe are all very active, the Temporal Left Lobe tends to have similar patterns of activities with Frontal Lobe, however, this phenomenon cannot be observed if all these lobes are not very active, and we also cannot observe phenomenon if we look at the combined period of highly active and not active of these three lobes. And this provide us a very interesting perspective to explore further in the future.

5.2 Gambling Task: SEEG Data Study

Stereotactic electroencephalography (SEEG) is often used to localize a seizure focus in patients who are candidates for surgery. It uses needlelike electrodes implanted in the depths of specific brain areas redundant to record from structures within that region. Visualization of how these depth electrodes are implemented can be found in Figure 5.3.

Figure 5.3: *Depth electrodes to record SEEG.*



Statistical spectral analysis treats SEEGs as time series data generated by stationary stochastic (random) processes. Recently more and more advanced mathematical tools may be brought to bear on the problem of understanding the link between brain activity, as seen in the SEEG, and function as determined via the implementation of experimental protocols. The task oriented brain activity analysis and classification is a prime issue in SEEG signal processing these days. The similar attempt has been done here to estimate the brain activity on the basis of power spectrum analysis. The data are from the Neuromedical Control Systems Laboratory (NCSL) at John Hopkins University. They were able to examine high temporal resolution (2kHZ) electrophysiological data in the subject who are performing the task involves a gambling decision, by taking advantage of SEEG recordings in epileptic patients at the Cleveland Clinic. The subject played a game of high card with a computer where they would win virtual money if their card beat the computer's card. The cards ranged from 2 to 10, even numbers only. After seeing their card, the subject had to decide whether to bet high($20) or low ($5). Finally the computer's card was revealed and the results of bet were shown to the subject. The timeline of one trial in gambling task is shown in Figure 5.4. We have SEEG recordings for around 30 minutes from a 41 year old female seizure patient. 104 electrodes were implanted on 13 depth electrodes, which were inserted in twenty-seven different brain areas through skull. In total, 162 different trials were recorded.

Figure 5.4: *The timeline of one trial for the gambling task. After fixation, subjects are shown their card. Once the bets are shown, subjects select one of the choices and then are shown the computer's card following a delay. Feedback is provided afterwards.*

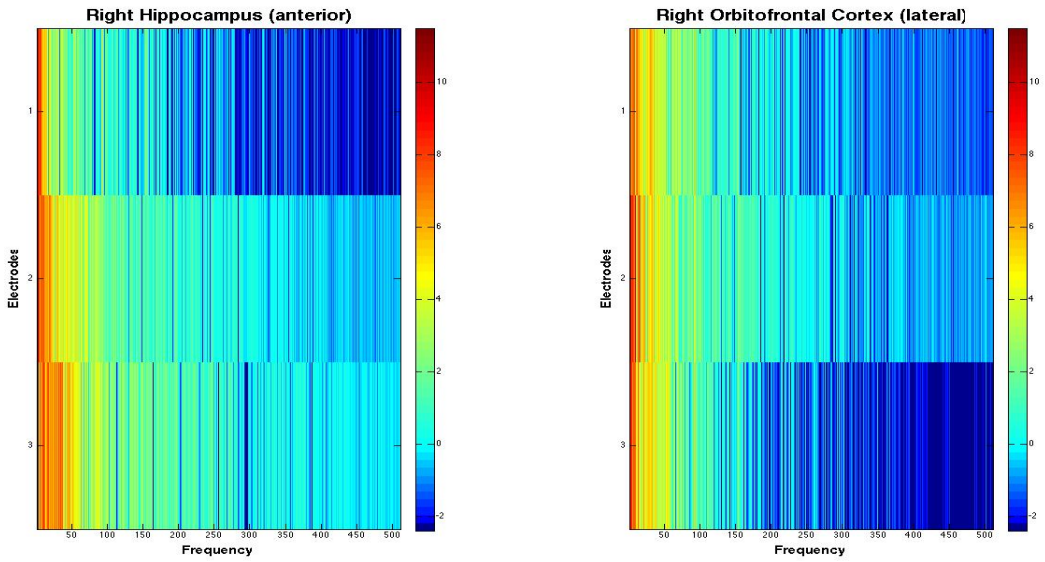
SEEG spectral analysis (decomposing a signal into its constituent frequency components) is an important method to investigate brain activity. Spectral analysis is a mathematical approach to quantify the EEG. It have been used for a long time in the analysis of EEG signals, which referred to as frequency domain analysis or spectral density estimation. Its purpose is the decomposition of signals such as the SEEG, into its constituting frequency components. The fast Fourier transform is a widely applied method for obtaining the SEEG spectrum. The power density spectrum or power spectrum displays the distribution of power or variance over the frequency components of a signal. It is defined as the Fourier transform of the autocorrelation function. Spectral content reveals neural oscillation. And characterizing such neural oscillatory behavior gives insight into the underlying dynamics of brain systems.

The task related SEEG recordings are assumed to be realizations of a specific type of non-stationary random process, namely, a locally stationary process. Thus, within short segments of time, the process exhibits stationary, even though the global characteristics vary throughout the trial. Our focus is on the SEEG recordings of around 0.5 seconds (i.e. length of each recording is 1024) after showing card, which are regarded as the period of processing card information and making decision on whether to bet high or low in gambling task. The first question we are trying to address is how these 27 regions communicate or connect during the decision making process. Two regions have the same spectral characteristics over this period probably have certain connections. Therefore, to address this question, our newly proposed "optimal" tapering test for replicated model introduced in Section 2.5 can be applied for each pair of these 27 regions and the brain network can thus be constructed based on the connections identified by the procedure. To implement

the testing procedure, in the replicated model (2.5), recordings over 162 trials are regarded as replicated samples and the Uniform kernel smooth function (simple moving average) is defined based on the optimal bandwidth $\widehat{b}_n$, which has been proved in this paper to be powerful in both asymptotical and empirical perspectives. Furthermore, to test between any pair of these 27 brain regions to see whether they have same power spectrum over this period means to perform 351 tests simultaneously. The multiplicity adjustment is considered and thus the Bayesian multiple testing procedure introduced in Section 3.2 was implemented. Finally, the brain network was constructed based on the connections identified by this procedure. The median probability model is built based on MCMC posterior samples and the brain network graph is plotted in Figure 5.11. This figure provides insights into how these 27 regions are connected in the brain while the subject is making the bet decisions.

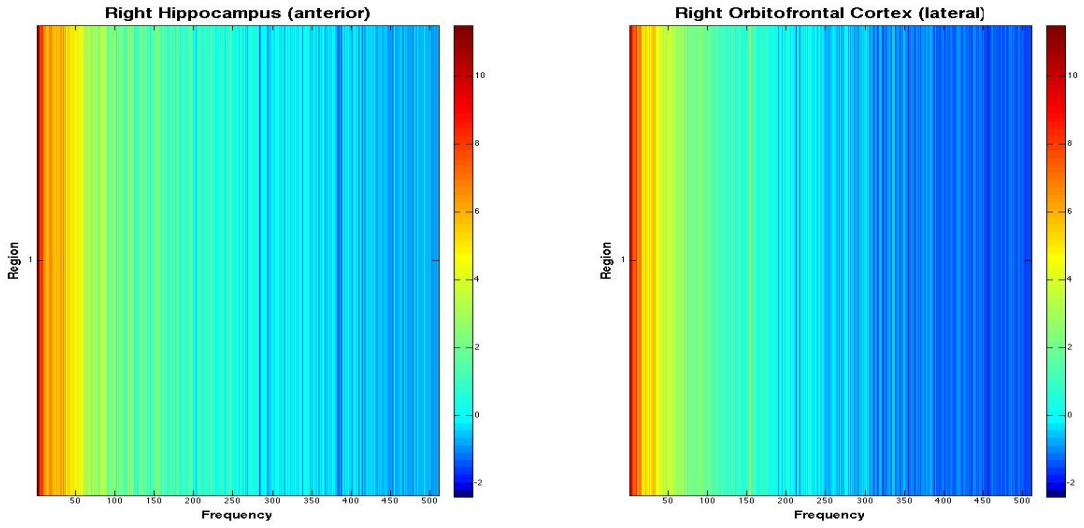
We will then explain step by step how our testing procedure helps to address this question and why our procedure is attractive by plots. Firstly, two regions are randomly picked from those 27 regions in SEEG data. Each of these two regions have three electrodes, from each electrode in each region, remember that we have the data for 162 trials, for simplicity, we take the data for the first trial for visualization and extract recordings of 0.5 seconds after showing card. Then the log-peirodogram is calculated and plotted as following,

Figure 5.5: *Heatmap for the raw log-periodograms for those two randomly picked regions. Each row represents each electrode in those regions. These two heatmaps are in the same scale.*



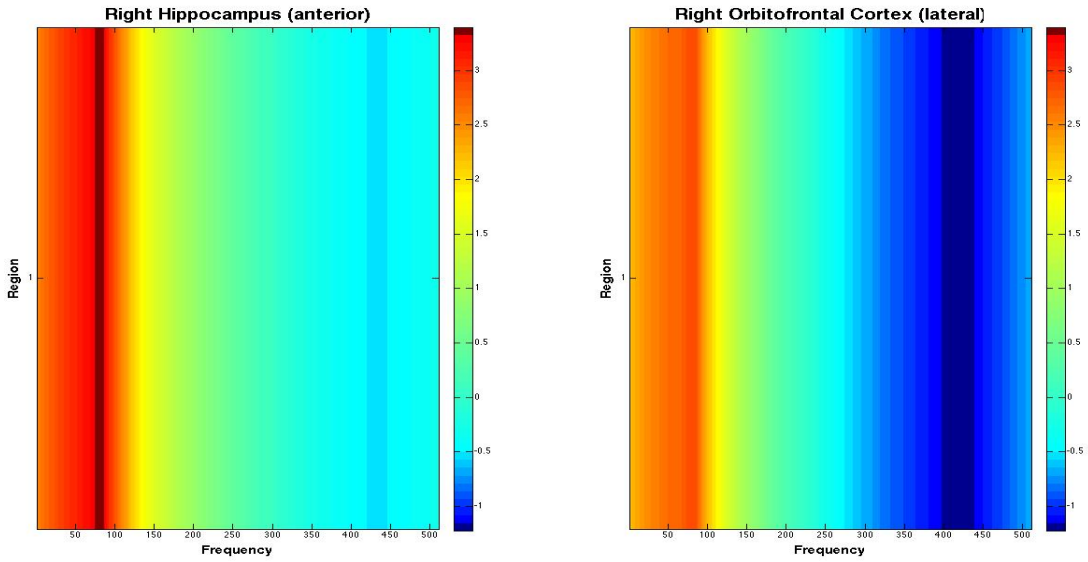
From the above plot, even for one single trial, we cannot tell whether these two regions perform similarly or not regarding to the spectral densities, since there are multiple electrodes in each region. Thus, we average over the electrodes in the same region and plot the averaged raw log-periodograms as following,

Figure 5.6: *Heatmap for the averaged raw log-periodograms for those two randomly picked regions, after averaging over the electrodes in the same region. These two heatmaps are in the same scale.*



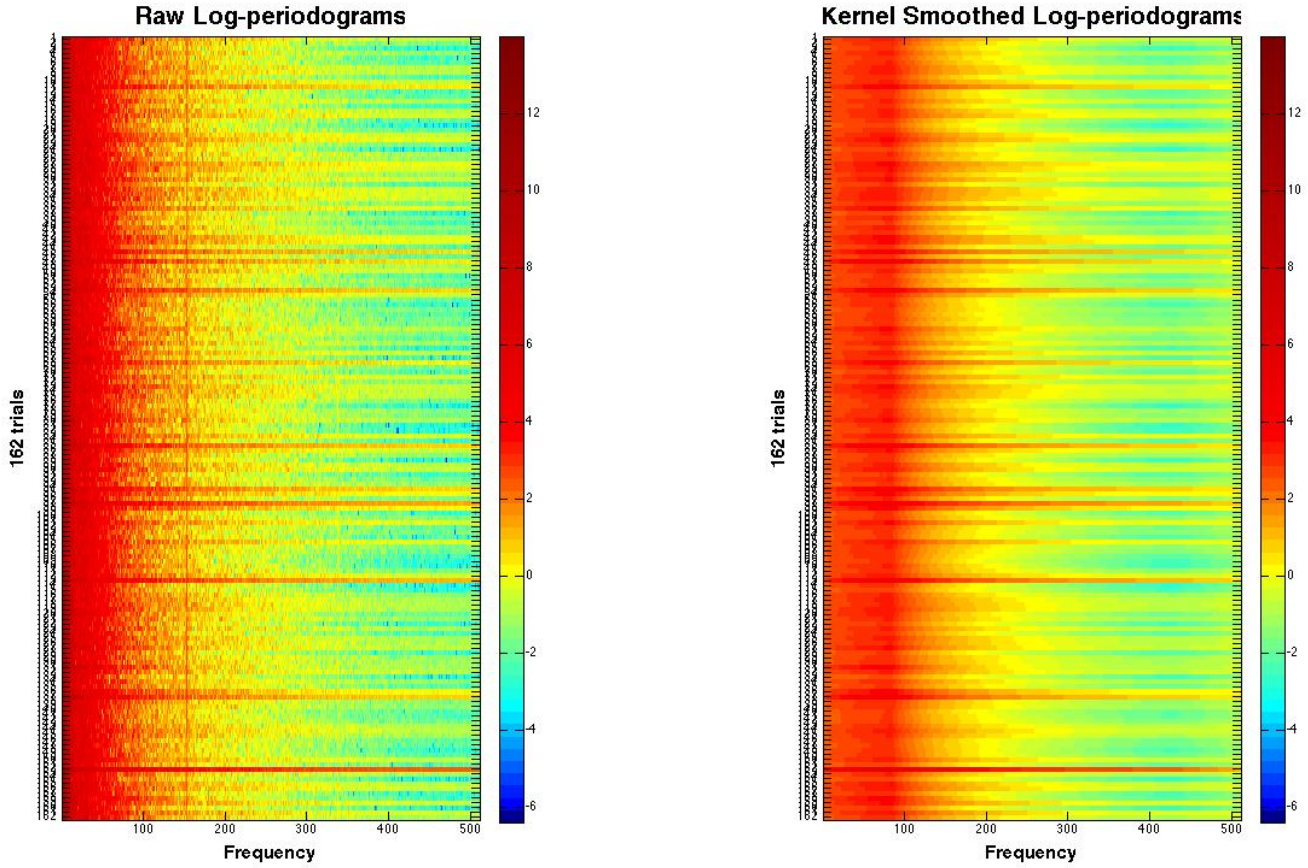
From the above plot, we can compare the left and right side heatmap. However, we can find that the variability in these two plots are very large. That's the reason that we need to do kernel smoothing. The way we make it optimal is to use "optimal" bandwidth in the kernel function, which is derived from "rate of testing" theory. Then kernel smoothing is performed on the averaged log-periodograms for each region and the heatmap for optimal kernel smoothed log-periodogram is plotted as following,

Figure 5.7: *Heatmap for the optimal kernel smoothed log-periodograms for those two randomly picked regions, for the first trial. These two heatmaps are in the same scale.*



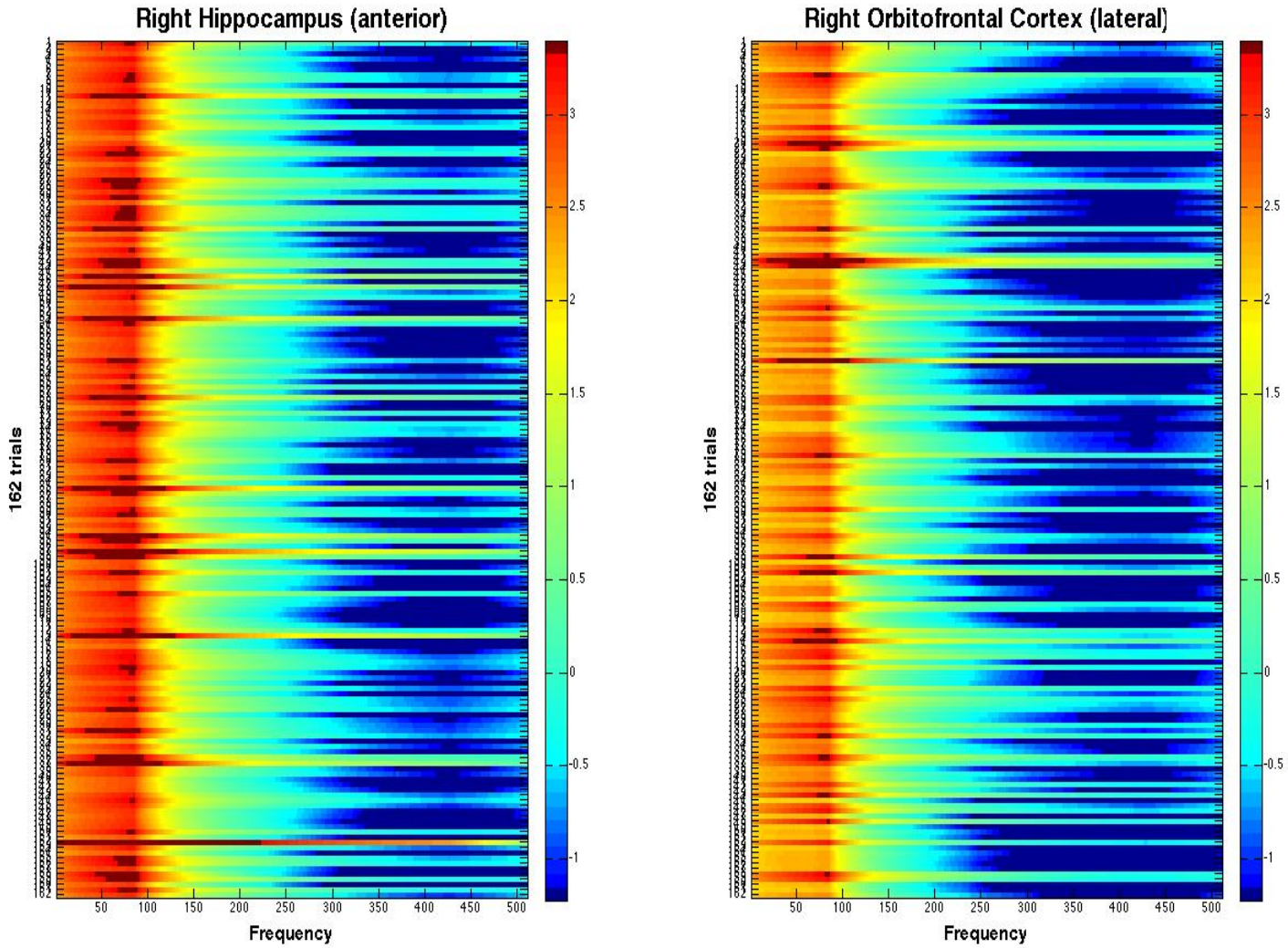
However, this is only for one trial. There are 162 trials in total. We further plot the comparison of the raw log-periodograms and optimal kernel smoothed log-periodograms for all these 162 trials for Right Hippocampus Anterior region to illustrate the effect of kernel smoothing in reducing the variability.

Figure 5.8: *Heatmap for the comparison of raw log-periodograms and kernel smoothed log-periodograms with the "optimal bandwidth". Region "HA", i.e. right hippocampus (anterior) is randomly picked to showcase the effect of "optimal kernel smoothing". Each row represents each of those 162 trials, and each column represents each frequency.*



The comparison of the optimal kernel smoothed log-periodograms for these two regions is also plotted as follows,

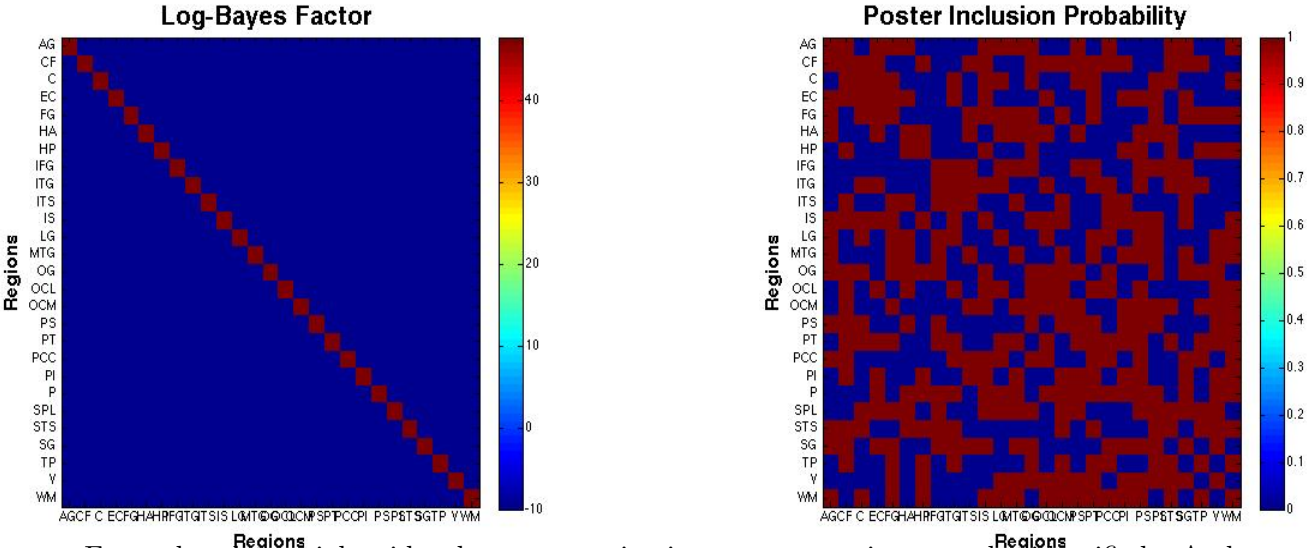
Figure 5.9: *Heatmap for the comparison of kernel smoothed log-periodograms with the "optimal bandwidth" for these two randomly picked regions. Each row represents each of those 162 trials, and each column represents each frequency.*



From the above plot, we cannot tell whether these two regions performs similarly or not, i.e., they have communications or connections or not during the decision making process, since there are replicated trials. Thus, the models for replications and the corresponding testing procedures are needed to perform test between these two regions. The Bayesian testing procedures for replications as introduced in Chapter 2 are applied for each of 351 pairwise comparisons of those 27 regions. The corresponding heatmap for the Log-Bayes Factor are plotted in the following left side figure. In

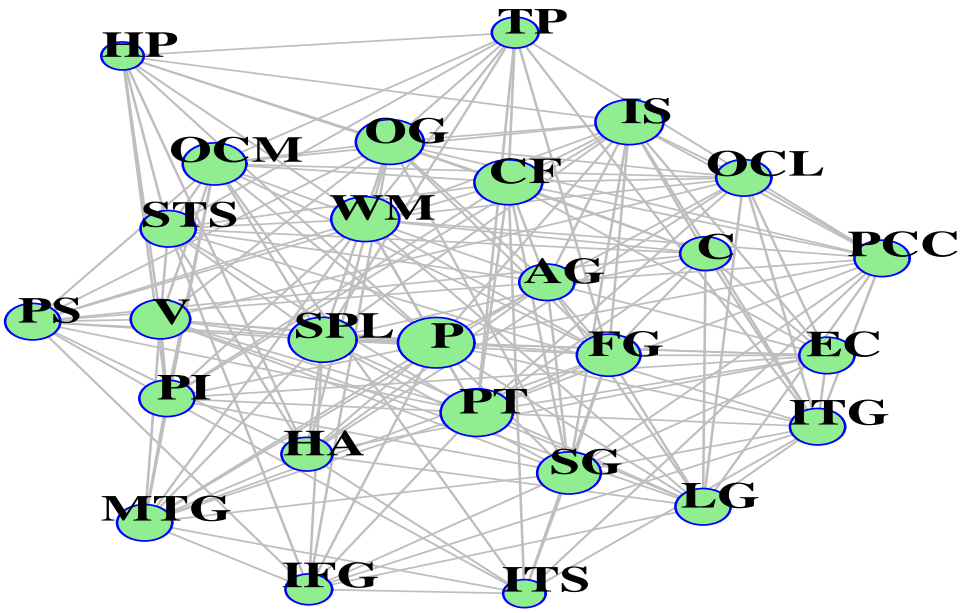
this figure, the red color indicates communication or connection exists between the corresponding regions in the row and column. So the conclusion from this figure is that no communications or connections can be identified. That's not the result we want. The main reason is that we are making 351 comparisons simultaneously without multiplicity adjustment. Therefore, Bayesian multiple testing procedure based on the kernel smooth model with replications are further applied and the corresponding results are shown in the right side of the following plot.

Figure 5.10: *Heatmap for the log Bayes factor and posterior inclusion probability of comparing each pair of 27 brain regions to see whether they have connection or communication during the decision making process. The left side plot is for the log Bayes factor without multiplicity adjustment, and the red color indicates connection or communication between the corresponding two regions in the raw and column. The right side plot is for the posterior inclusion probability for each pair, and the red color indicates connection or communication between the corresponding two regions in the raw and column. Each region is named by its acronyms: AG means "right angular gyrus'"; CF means "right calcarine fissure"; C means "right cuneus"; EC means "right entorhinal complex"; FG means "right fusiform gyrus"; HA means "right hippocampus (anterior)' HA; HP means "right hippocampus (posterior)"; IFG means "right inferior frontal gyrus"; ITG means "right inferior temporal gyrus"; ITS means "right inferior temporal sulcus; IS means "right intraparietal sulcus; LG means "right lingulal gyrus"; MTG means "right middle temporal gyrus"; OG means "right occipital gyrus"; OCL means "right orbitofrontal cortex (lateral)"; OCM means "right orbitofrontal cortex (mesial)"; PS means "right parieto-occipital sulcus'" PT means "right planum temporale"; PCC means "right posterior cingulate cortex"; PI means "right posterior insula" P means "right precuneus"; SPL means "right superior parietal lobule"; STS means "right superior temporal sulcus; SG means "right supramarginal gyrus"; TP means "right temporal pole"; V means "right ventricle"; WM means "white matter".*



From the above right side plot, communications or connections can be identified. And to summarize the results, the brain network graph is plotted as following,

Figure 5.11: *The brain network graph for 0.5 seconds after showing card. Each dot in the graph represents a region, each line connects the pair of regions that have connections during this period and the larger dot means the corresponding region has more connections with other regions.*



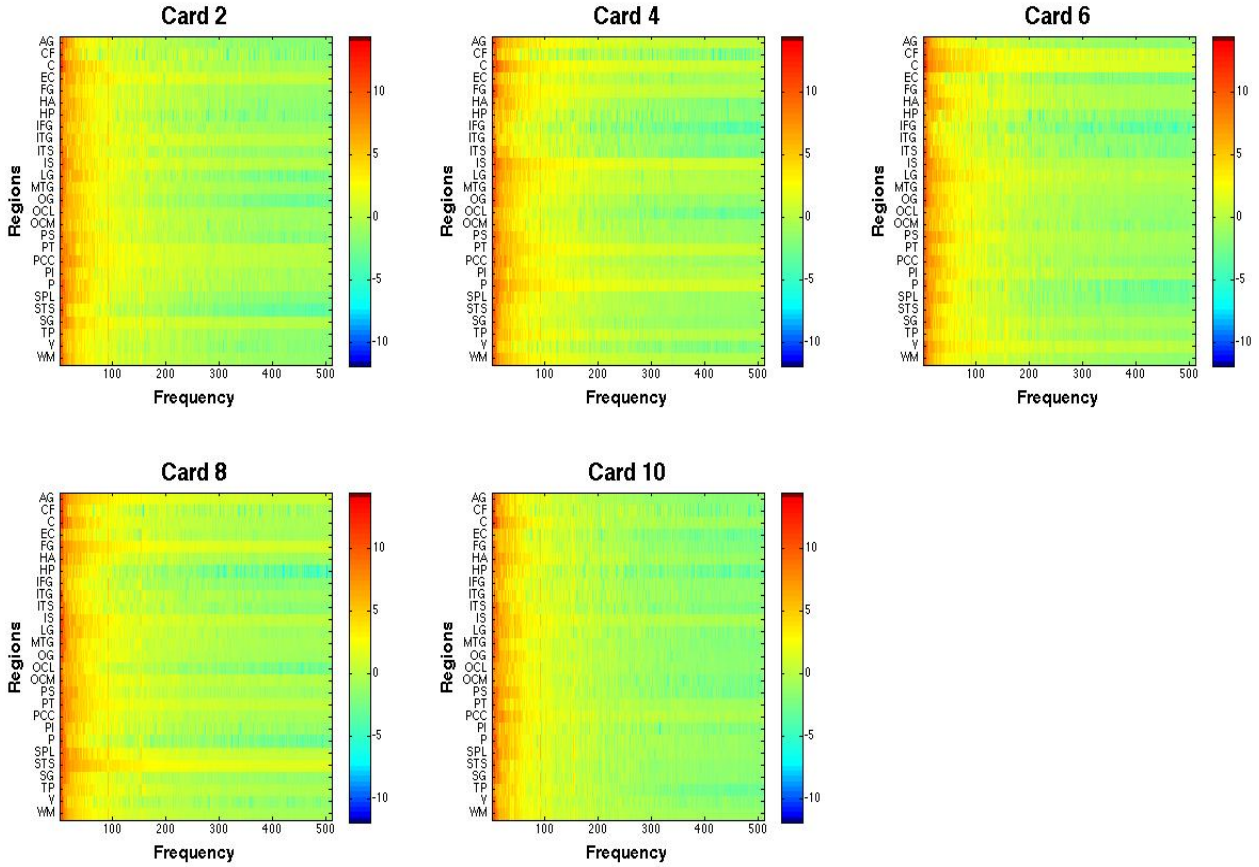
The second question of interest is to identify the regions that are involved in encoding card numbers. For a certain region, it is probably related with encoding the card number if it performs differently under different card numbers. For each region, SEEG recordings of around 0.5 seconds after showing card are extracted separately for different card number trials. Gambling trials with card number 2 are compared with card 4, card 6, card 8 and card 10 for each region and regions with the trend of having different spectral characteristics under these comparisons are regarded as involved in encoding card numbers. Therefore, Bayesian tapering test introduced in Section 3.1 was applied based on the replicated model as discussed in Section 2.5 for each of those 27 regions. The heatmap of corresponding $2 \log BF$ values for each region and for each comparison are plotted in Figure 5.14. This heatmap suggests that Right Cuneus (C), Right Entorhinal Complex (EC), Right Hippocampus (anterior and posterior) (HA and HP), Right Inferior Temporal Sulcus (ITS), Right Superior Temporal Sulcus (STS), Right Temporal Pole (TP) tend to perform differently

regarding to the power spectrum under those comparisons of different card number and thus they are probably actively involved in encoding card numbers.

To explain step by step how our proposed Bayesian tapering testing procedure based on kernel smoothing model with or without replications helps to address this question, we plot those figures in the following.

Firstly, for each of these card numbers, we plot the raw log-periodograms of the recordings for the first trial of these 162 trials.

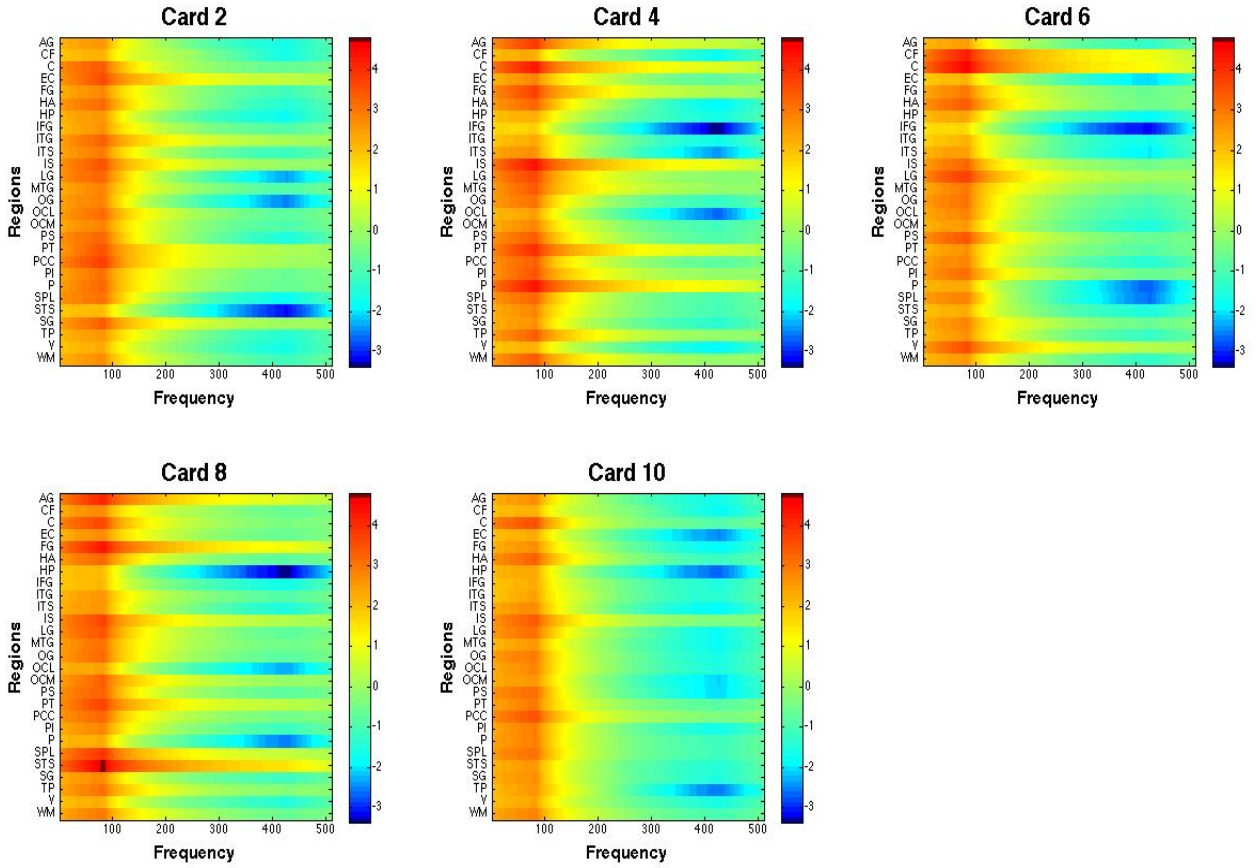
Figure 5.12: *Heatmap of raw log-periodograms for each of these 27 regions for the recordings of the first trial. Each row represents each region, and each column means each frequency. Those heatmaps are in the same scale.*



From the above plot, we can find that the varibility is vary large. Then the "optimal" kernel

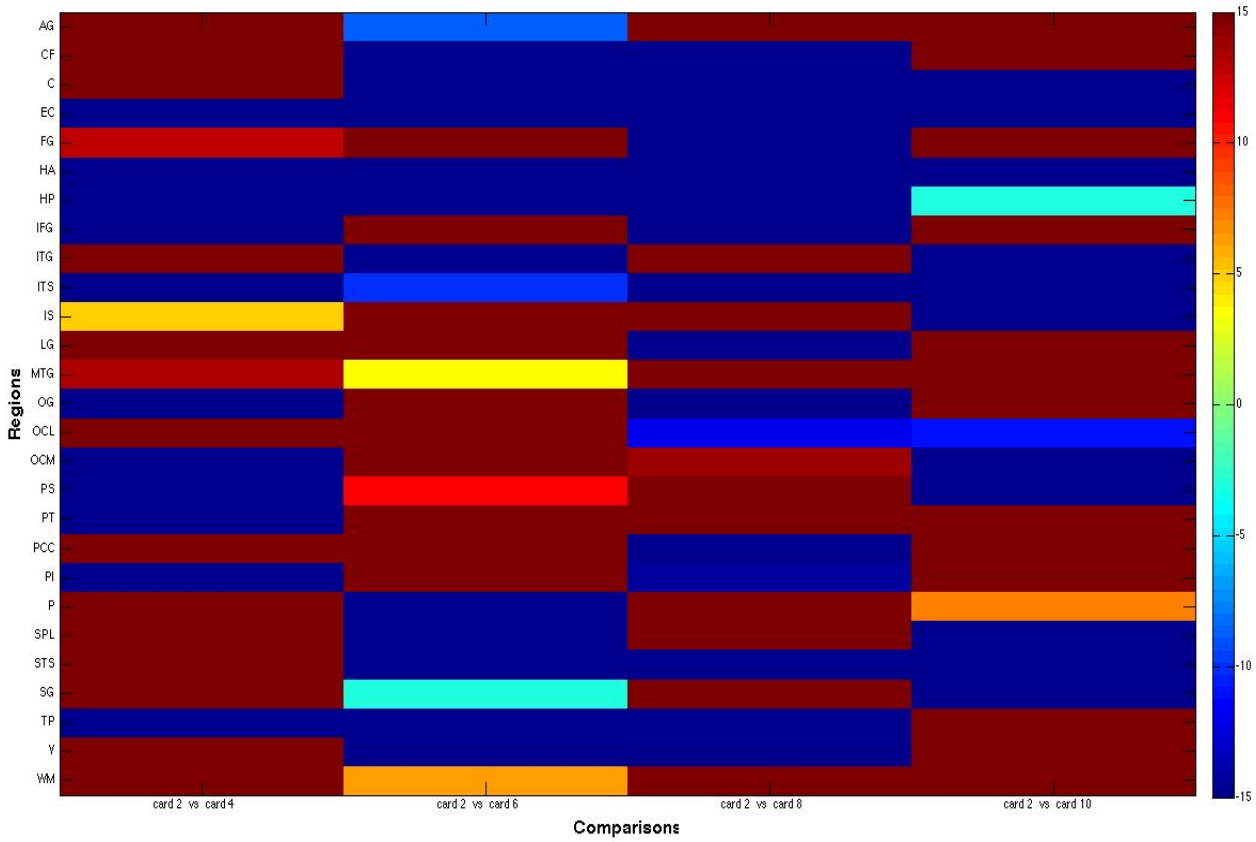
smoothing is performed and the smoothed log-periodograms for each card number for the first trial are plotted as following,

Figure 5.13: *Heatmap of "optimal" kernel smoothed log-periodograms for each of these 27 regions for the recordings of the first trial. Each row represents each region, and each column means each frequency. Those heatmaps are in the same scale.*



From the above plot, we can find that the variability is improved a lot. Now we cam make comparisons between any two card numbers. However, this is only for the first trial, and we have 162 trials in total. Considering 162 plots, we cannot tell whether any of the region performs differently under different card numbers. Thus, the tapering testing procedures based on the kernel smoothing model with replications is needed to address this question. The results of comparisons between each card number is plotted as the heatmap in the following,

Figure 5.14: *Heatmap of* $2\log BF$ *values for each of these 27 regions and for each comparison. Each row represents each region, and each column means each comparison. Values smaller than -6 means strong evidence for this region being involved in encoding card numbers.*



From the above plot, regions that performs differently under different card number trials can be identified. $2\log(BF)$ values provides insights of testing results: smaller than -6 indicate "strong" evidence for this region being involved in encoding card number; and smaller than -10 means "very strong" evidence. While the positive values indicate the strength of evidence for this region not being involved in encoding card number. The trend of how differently each regions will perform under different card numbers over the period of around 500 milliseconds can be identified from these heatmaps. Right Cuneus (C), Right Entorhinal Complex (EC), Right Hippocampus (anterior and posterior) (HA and HP), Right Inferior Temporal Sulcus (ITS), Right Superior Temporal Sulcus (STS), Right Temporal Pole (TP) are identified to be actively involved in encoding card numbers.

In summary, the illustration of how our tapering testing procedures can be used to extract useful information regarding to how brain performs when the subject is doing different tasks is shown by two EEG datasets. The functions of some of these brain regions are still not well understood,

and our method provides a new powerful way to gain insights into the brain activities.

Chapter 6

Summary and Future Work

Motivated by an application example of EEG data, a new methodology of Bayesian tapering test and corresponding Bayesian multiple testing procedures for assessing whether two independent stationary time series have same spectral density has been developed. Non-parametric time series methods are relevant to a variety of applied contexts. Related literatures describe applications to economic time series of wheat prices, medical monitoring data of hormone levels in blood, and meteorology data of temperature and precipitation levels.

In Chapter 2, tapering test based on the raw log-periodograms, Fourier transformation of the log-periodograms, and "optimal" kernel smoothing of log-periodograms are explored. The problem of whether kernel smoothing is still needed when the replication exists is further investigated. The limiting behavior is defined with respect to $n \to \infty$. The theoretical powers of these tests are optimized based on "rate-of-testing" theory, which minimizes the "indistinguishable region", i.e. the region that the our test cannot detect with high asymptotic power, by letting the rate of the parameter δ_n that controls the size of this region goes to zero as fast as possible. The optimal weights and optimal rate of bandwidth parameter is derived from the fastest rate $\delta_n \to 0$. Theoretical theorems and proofs under each model are explained in details in Chapter 2 and as a result, the "optimal" kernel smoothing model and corresponding tapering test procedures are recommended wether the replication exists or not.

In Chapter 3, driven by the needs of the application SEEG dataset, the corresponding Bayesian tapering test framework is constructed and the Bayesian multiple test procedure is further investi-

gated based on Beta-Binomial prior. The benefits of Bayesian tapering test procedures are mainly in its convenience of construct the acceptance ratio in constructing MCMC posterior estimates.

In Chapter 4, the empirical performance of these tests are further investigated by a comprehensive simulation study. The newly proposed tests are compared with existing tests in literature under different truth scenarios. All these three tests performs pretty good, and especially the test based on the Kernel smoothing model generally performs the best, which is consistent with the theoretical results.

In Chapter 5, how the newly proposed tapering test procedures based on "optimal" kernel smoothing model can be applied to answer the questions proposed regarding to how different regions of human brain performs during the subject is doing different tasks is illustrated by the motivating Scalp EEG dataset and the Stereotactic EEG dataset. Some interesting and clinical meaningful phenomenon are identified, which provides insights into human brain.

To sum up, this newly proposed tapering test based on "optimal" kernel smoothing model with or without replications achieves the target of our research, which is to propose a new testing procedure which is powerful and empirically as well as easy to implement in practice even for clinicians.

There are still remaining issues that needs discussion and further exploration. The multitaper method is recently proposed by David Thomson in [31] to get consistent estimator of spectral density, as discussed in Section 1.3. This estimation method addresses simultaneously the issues of bias and variance in an optimal fashion. We have some preliminary theoretical results that starts from the "rate-of-testing" theory to derive the "optimal" tapering test based on the multitaper spectral density estimates. The optimality is constructed by deriving the "optimal" number of tapers in multitaper method. But we didn't get too much progress. Considering the fact that the underlying theory of choosing optimal type of tapering sequence in multitaper method is very complicated and it is not quite easy to implement in practice, we decided that that's not the correct direction we want to go in this research. Probably in the future, when the multitaper method gets more and more popular and easier to understand, the tapeirng test framework based on the multitaper spectral density estimation method can be further explored.

Additionally, the cumulative periodogram, which is closely related to the cumulative spectrum,

is a useful tool for describing the overall behavior of the periodogram. It is defined as

$$F_x(\omega_j) = \sum_{i=1}^{j} I_x(\omega_i) \Big/ \sum_{i=1}^{p_n} I_x(\omega_i).$$

It is a direct application of the periodogram for testing the hypothesis that a particular time series is a white noise sequence or it is a purely random series. Bartlett (1954) [2] proposed a plot of $F_x(\omega_j)$ against ω_j to assess departure from white noise. See also Jenkins and Watts (1968) [14]. Diggle and Fisher (1991) [10] extended this idea to the comparison of two sample periodograms via a plot of $F_y(\omega_j)$ against $F_x(\omega_j)$. Probably the asymptotic results for the cumulative periodogram is more sound. The question comes as whether we can propose our tapering tests based on cumulative periodogram. However, in literatures, only Diggle and Fisher (1991) [10]'s graphic method is proposed based on the cumulative periodogram in the current context. No formal testing procedures can be found based on the cumulative periodograms. Probably it is because that the underlying asymptotic properties of the cumulative periodogram is much more complicated than the periodogram and it is not easy to implement in practice. We explored the tapering test based on cumulative periodogram based on "rate of testing" theory. But we didn't get too much progress. Maybe in the future, when the asymptotic properties of the cumulative periodogram is more deeply investigated and thus attracts more attentions, we work more on developing the tapering test procedures based on the cumulative periodogram.

Bibliography

[1] Maria Maddalena Barbieri and James O Berger. Optimal predictive model selection. *Annals of Statistics*, pages 870–897, 2004.

[2] Maurice Stevenson Bartlett. Processus stochastiques ponctuels. In *Annales de l'institut Henri Poincaré*, volume 14, pages 35–60, 1954.

[3] Thomas Bengtsson and Joseph E Cavanaugh. State-space discrimination and clustering of atmospheric time series data based on kullback information measures. *Environmetrics*, 19(2):103–121, 2008.

[4] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[5] David R Brillinger. *Time series: data analysis and theory*, volume 36. Siam, 2001.

[6] Peter J Brockwell and Richard A Davis. Time series: theory and methods. *New York: Springer*, 66:119–128, 1991.

[7] Jorge Caiado, Nuno Crato, and Daniel Peña. A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50(10):2668–2684, 2006.

[8] DS Coates and PJ Diggle. Tests for comparing two estimated spectral densities. *Journal of Time Series Analysis*, 7(1):7–20, 1986.

[9] Persi Diaconis and David Freedman. On the consistency of bayes estimates. *The Annals of Statistics*, pages 1–26, 1986.

[10] Peter J. Diggle and Nicholas I. Fisher. Nonparametric comparison of cumulative periodograms. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(3):pp. 423–434, 1991.

[11] Jianqing Fan. Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association*, 91(434):pp. 674–688, 1996.

[12] Hsiao-Yun Huang, Hernando Ombao, and David S Stoffer. Discrimination and classification of nonstationary time series using the slex model. *Journal of the American Statistical Association*, 99(467):763–774, 2004.

[13] Y. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives i-iii. *Mathematical Methods of Statistics*, 1993.

[14] Gwilym M Jenkins and Donald G Watts. Spectral analysis. 1968.

[15] Yoshihide Kakizawa, Robert H Shumway, and Masanobu Taniguchi. Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340, 1998.

[16] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[17] Kewei Lu and Linyuan Li. On fan's adaptive neyman tests for comparing two spectral densities. *Journal of Statistical Computation and Simulation*, 83(9):1585–1601, 2013.

[18] Robert Lund, Hany Bassily, and Brani Vidakovic. Testing equality of stationary autocovariances. *Journal of Time Series Analysis*, 30(3):332–348, 2009.

[19] Shujie Ma and Xuming He. Inference for single-index quantile regression models with profile optimization. *Annals of Statistics*, 2016.

[20] Maurice Bertram Priestley. Spectral analysis and time series. 1981.

[21] Barry G Quinn. Statistical methods of spectrum change detection. *Digital Signal Processing*, 16(5):588–596, 2006.

[22] Christian P Robert. A note on jeffreys-lindley paradox. *Statistica Sinica*, 3(2):601–608, 1993.

[23] Christian P Robert. The bayesian choice: From decision-theoretic foundations to computational implementation (springer texts in statistics) by. 2001.

[24] James G Scott and James O Berger. Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619, 2010.

[25] Robert H Shumway. Time-frequency clustering and discriminant analysis. *Statistics & probability letters*, 63(3):307–314, 2003.

[26] Robert H Shumway and David S Stoffer. *Time series analysis and its applications*. Springer Science & Business Media, 2013.

[27] Dan J Spitzner. An asymptotic viewpoint on high-dimensional bayesian testing. *Bayesian Analysis*, 3(1):121–160, 2008.

[28] Dan J. Spitzner. A powerful test based on tapering for use in functional data analysis. *Electronic Journal of Statistics*, 2:939–962, 2008.

[29] Dan J Spitzner. Testing in functional data analysis using quadratic forms. *Under review*, 2008.

[30] Vladimir G Spokoiny et al. Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498, 1996.

[31] David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.