THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

HYPOTHESIS TESTING FOR TWO-SAMPLE FUNCTIONAL/LONGITUDINAL DATA

JIN YANG

Ph.D

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of Applied Mathematics

Hypothesis Testing for Two-Sample Functional/Longitudinal Data

Jin Yang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Sep. 2017

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

__________________________(Signature)

Jin Yang (Name of student)

Abstract

During recent two decades, functional data commonly arise from many scientific fields such as transportation flow, climatology, neurological science and human mortality among others. The corresponding data recorded may be in the form of curves, shapes, images and functions that may be correlated, multivariate, or both. The intrinsic infinite dimensionality of functional data poses challenges in the development of theory, methodology and computation for functional data analysis. Tests of significance are essential statistical problems and are challenging for functional data due to the demands coming from real world applications. Motivated by requirements in real-world data analysis, we have focused on two topics of study. 1) Multivariate functional data have received considerable attention. It is natural to validate whether two mean surfaces are homogeneous but existing work is few. 2) In existing literature, most testing methods were designed for validity of dense and regular functional data samples, whereas in practice, functional samples may be sparse and irregular or even partly dense. In such functional data setting, there is rare work for testing equality of covariance functions or mean curves. To address these problems, we aim to two targets: 1) We propose novel sequential and parallel projection testing procedures that can detect the difference in mean surfaces powerfully. Furthermore, we apply the idea to present testing statistics for test of equality of mean curves for two functional data samples irrespective of the data type. Furthermore, the other related work takes auxiliary information into consideration. We propose a new func-

tional regression model to characterize the conditional mean of functional response given covariates. 2) We derive a novel test procedure for test of equality of covariance functions that can deal with any functional data type, even irregular or sparse data. In addition, by using the stringing technique, once a high-dimensional data can map into functional data, we excogitate a testing procedure for comparison of covariance matrices under the high-dimensional data setting. Our method outperforms the existing testing methods in high-dimensional data testing procedures. Almost all work mentioned above include asymptotic theory and rigorous theorem proof, intensive numerical experiments and real-world data analysis.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Catherine Chunling Liu, for her motherly patient care and generous support throughout my Ph.D. period. She continued to push me to go ahead, helped me to solve the problem both in study and life and many more things than I could have expected from a supervisor. I sincerely appreciate all the time and effort she spent on training my research abilities. The independent, creative and critical thinking she taught me will be the treasure for me, not only in my Ph.D. study, but also in my future life.

My deepest thanks also goes to my academic uncle, Dr. Tao Zhang, an intelligent, enthusiastic and energetic researcher. In past three years, he showed his power in exploring the novel ideas, and broadened my horizon in statistics. Especially in real data study, he not only presented his talent in analyzing, but also let me realize our research has closely connection with our real life which enlighten my strong interest in research. Furthermore, he showed his solid mathematical foundation in proof. I really learned and benefited a lot from him.

I am thankful to Prof. Liugen Xue, Dr. Junshan Shen, Dr. Chuanhua Wei, Dr. Jicai Liu, Dr. Xiaolin Chen for their helpful discussions and valuable suggestions on my research when they visited PolyU.

My sincere acknowledgements goes to my academic brother: Mr. Sheng Xu, who is a clever and diligent boy. We often discussed together about our research and he gave me a lot of help in computing.

My special thanks goes to my best friend, Dr. Lei Yang, who always gave remarkably kind assistance during my Ph.D. study as well as life. As a genius, he usually presented his enthusiasm and energy in solving my the mathematical problem.

I would also like to thank all the friends I have met in Hong Kong for their friendships. In particular, I would like to especially acknowledge Dr. Yang Zhou and Dr. Hong Wang, for their sincere friendship constant support over the years.

Last but far from least, my deepest gratitude and love belong to my mother Hongjing Tan, my father Xinmin Yang and all my family for their unconditional love and consistent support through my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

During recent two decades, functional data commonly arise from many scientific fields such as transportation flow, climatology, neurological science and human mortality, among others. The corresponding data recorded may be in the form of curve, shape, image and functions that may be correlated, multivariate, or both. Wang et al. (2015a) presented the so-called first and next generation functional data by the criterion that the random sample of real-valued functions are random trajectory/curve data (Gasser et al., 1984; Rice and Silverman, 1991; Gasser and Kneip, 1995) or part of complex data objects like hypersurface data. For instance, the neuroimage data stated on page 23 of the report for the London workshop on the topic the Future of Statistical Sciences held in November 2013, refer to `http://www.worldofstatistics.org/wos/pdfs/Statistics&Science-TheLondonWorkshopReport.pdf`. In a word, functional data has already established itself as an important and dynamic area of statistics.

The term "functional data analysis" was coined by Ramsay (1982) and Ramsay and Dalzell (1991), and the related history of this field can be tracked back to Grenander (1950) and Rao (1958). It offers new effective tools and has stimulated new theoretical and methodological development. The book by Ramsay and Silverman (2005) gives a clear account of the basic considerations of FDA. The first advances in nonparametric FDA are described by Ferraty and Vieu (2006). Other monographs include Bosq (2000), Ramsay and Silverman (2002), Ferraty and Vieu (2006), Ramsay et al. (2009), Ferraty and Romain (2011), Shi and Choi (2011), Horváth and Kokoszka (2012), and Kokoszka and Reimherr (2017). Comprehensive reviewing papers include Hall and Hosseini-Nasab (2006), Müller (2014) and Müller (2016), among others. For functional data analysis, when it comes to the comparison between two samples, two natural yet essential questions are to check the heterogeneity and heteroscedasticity. In the past decade, much effort has been made in developing powerful testing procedures to detect the difference in mean functions or deviations in covariance structure for two- or multiple- sample functional data. The related reference regarding tests of significance may go to the introduction parts in Chapters 2, 3 and 5 under functional data setting. In this thesis, we concentrate on 1) testing the equality of mean surface functions for hypersurface functional samples; and 2) testing covariance functions for two functional data samples that are no necessary dense and regular. Next we introduce the original practical motivation for the theme in this thesis. It has come from exploration of two data sets.

Testing equality of two mean surfaces is motivated from detecting change in precipitation affected by both the spatial and temporal effects in the Midwest of the United States. The data is collected in cohort for climate monitoring and stored in Global historical climatological network database, refer to the National Oceanic and Atmospheric Administration: “ `(https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/)`”, which collects main climate parameters such as daily maximum and minimum temperature, amount of precipitation (liquid equivalent), amount of snow fall and snow depth, and so on. We investigate into daily precipitation amounts recorded at 59 terrestrial observatories spread over 12 states during the period 1941-2000, refer to Fig. 2.1 in Chapter 2.

Gromenko et al. (2017) detected the annual patten change along the temporal domain by a spatially indexed way. Their method could detect annual change point when precipitation amount is looked as if observed on the curve with observations on all 59 spatial locations. Extremely Berkes et al. (2009) fixed their test on an individual station and thus detected no significance for any pair year segments during the time course 60 years. Unfortunately, due to purpose of detection, such existing methods can neither test the significance affected by the joint affects from both spatial and temporal domains nor test treatment effect impacted by the marginal spacial domain. However, both questions are important for climate monitoring. This inspires our study in Chapter 2.

Testing the equality of covariance functions of two functional data samples is motivated by investigation of CD4 cell measurements in an AIDS clinical study by the AIDS clinical trials group (ACTG) monitored for 2.1 years, say the data from Fischl et al. (2003). There are CD4 cell responses collected after three different treatment arms. Such functional data is sparse and irregularly spaced due to various reasons. It is important to provide solid and reliable statistical support to justify whether the covariance functions are identical instead of a naive visualization of sample covariance function surfaces. This drives our study in Chapter 5.

After testing the mean surfaces for multivariate functional data, we have extension research in two aspects. On one hand, employing the idea of mean surface testing, we also present testing procedures to detect change in two mean curve functions in univariate functional data scenario. Our method performs well regardless dense, sparse or mixed functional samples. On the other hand, we notice that in analysis of the aforementioned daily precipitation in the Midwest of the United States, there are quite a lot of auxiliary information such as temperature, pressure normal, wind, cloudiness and other climate indexes. Naturally, we may use mean regression to investigate how the auxiliary information affects the mean surface or curve. This brings us considering new modeling of functional regression models where conditional mean response involves both main and interaction effects.

Furthermore, notice the stringing technique by Chen et al. (2011b) which may transform high-dimensional vectors into functional data under some assumptions. This motivates us to extend the method in testing equality of covariance functions for functional data samples into the method to test the covariance matrices for high-dimensional data samples provided that the high-dimensional data can map into functional data.

Literature review for each piece of work are depicted in each chapter. Chapters 2, 3 and 4 are based on the work Zhang et al. (2017b), Yang and Zhang (2017) and Zhang et al. (2017a). Chapters 5 and 6 are based on our work: Yang et al. (2017).

For the contribution of the thesis, we summarize into trifold. First, our testing procedure for mean surface detection may be the very early work on equality test in hypersurface functional data. Second, our proposed testing procedures for checking equality of mean curve or covariance functions perform well in both size and power not only for sparse and irregular random curves but also apply well for dense and regular functional trajectories. Third, borrowing the strength of stringing perception, we develop powerful testing statistic to detect the equality of covariance matrices in high-dimensional setting. This will enlighten more effective transplanting of methodology between functional data and high-dimensional data.

The rest of the thesis covers three parts. Chapters 2 and 3 attribute to part I. Chapter 2 discusses how we develop two sequential and parallel projection testing procedures to detect change of the joint effects impacted by both domains. Chapters 3 and 4 are extension work on test equality of mean curves and new modeling of functional mean regression with auxiliary information. Chapters 5 and 6 composes part II discussing testing equality of covariance functions for functional data and its extension in high-dimensional setting to test equality of covariance matrices. Part III includes Chapter 7 which has a brief discussion in ongoing work and future work.

Chapter 2

Testing Equality of Mean Surface for Two-sample Functional Data

2.1 Introduction

In the multivariate functional stochastic process $X(u)$, there has increasing research interest in data type that is both functional and multidimentional. That is, $u = (s, t)$ has two arguments where $s \in \mathcal{S} \subset \mathbb{R}^{d_1}$ and $t \in \mathcal{T} \subset \mathbb{R}^{d_2}$ with d_1 and d_2 being positive integers. Here s and t inherently belong to distinct domains $\mathcal{S}$ and $\mathcal{T}$ in terms of scientific meaning or research design. For example, $X(s, t)$ may be the mortality rate of age s during year t in a given country. A typical example of such data comes from neuroimaging studies using functional magnetic resonance imaging (fMRI), in which the so-called voxels data, i.e. brain activity like blood flow changes are discrepantly recorded at a large number of locations at irregular time units (Lindquist, 2008; Aston and Kirch, 2012). Spatiotemporal study is no doubt another important application of this kind of data where t is defined on a temporal domain and s is defined on a spatial domain. Although functional data of afore structure are encountered in many applications, there is rare progress in inferential aspect for such data (Gromenko et al., 2017; Aston et al., 2017). In the present work, we plan to investigate the profile and globe tests of mean surfaces for two bivariate functional samples.

A practical motivation for this research comes from precipitation data in Midwest of the United States, where the daily data of precipitation from 1941 to 2000 are collected at 59 spatial locations scattered over 12 states in the Midwest of USA. For ease of reference, we provide a map of Midwest states with the locations of the climate monitoring stations in Fig. 2.1. The Midwest is a breadbasket of the United States and its agriculture has continued to play a major role in the economy of the region (Pryor, 2013). The agriculture in the Midwest is vulnerably affected by the climate, of which precipitation is a vital component. To monitoring the future agricultural activities, it therefore has long been recognized as an important problem to reveal how the change of precipitation takes place for different locations, different regions, or different years in the same region.

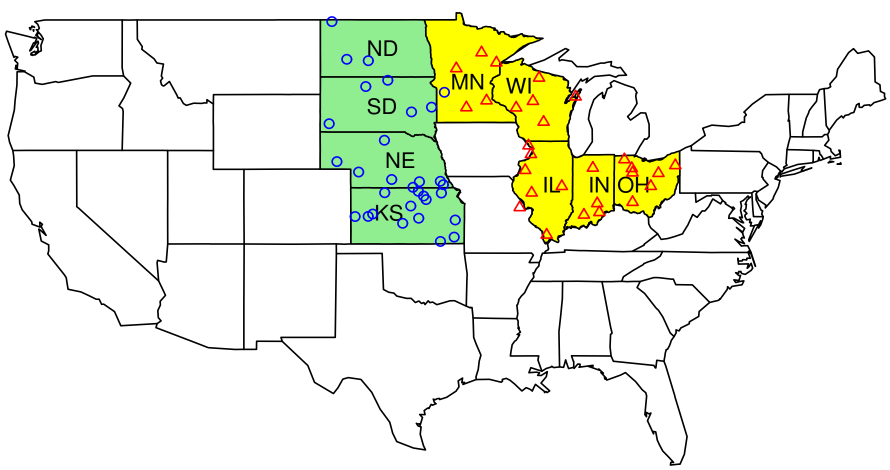

Figure 2.1: Light green region: 4 states from the Great Plains; Blue circle ○ indicates location of a station; Yellow region: 5 states from the Great Lakes; Red triangle △ indicates location of a station.

The study of the precipitation data has led to several interesting findings. For instance, Berkes et al. (2009) detected no changes during the period 1941-2000 for only individual station. However, it is difficult to implement if we sequently tested for every station when the number of stations were large. Gromenko et al. (2017) used cumulative sum paradigm to expose the fact that, the mean precipitation curves before and after 1966 were different over the whole region. Nevertheless their method was particularly designed to detect the temporal change but not applicable to detect the difference in spatio domain, not to mention the joint spatiotemporal effect on the precipitation. Looking into analysis of heatmaps of yearly sample mean surfaces where $X_i(s,t), i = 1941, \cdots, 1967$, corresponds to the precipitation of the tth day in the ith year at the sth station, intuitively we have observed that the yearly sample mean surface of precipitation in the Great Plains is different from that in the Great Lakes, refer to Fig. 2.2. Also, we can recognize from Fig. 2.2 that some profiles of mean surface are same but others are different. These motivate us to develop more powerful inferential procedures to detect if mean surfaces or its profiles have significant difference for either different regions or different individual stations.

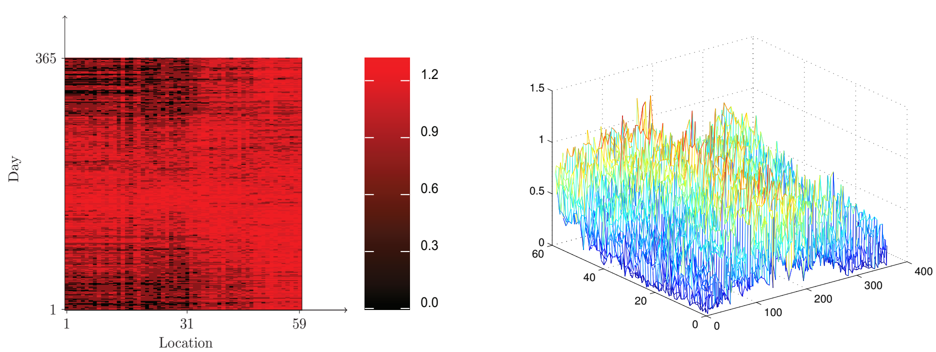

Figure 2.2: The heatmap of sample mean surface of precipitation during the time 1941-1967 in the Midwest, where the first 31 stations are located in the Great Plains and the latter 28 stations are located in the Great Lakes.

Tracking back testing procedures for the equality of mean functions in the functional data setting, existing works mainly focus on detecting the curve equality for univariate functional data. In the two-sample testing scenario, Benko et al. (2009) presented bootstrap procedures for testing the equality of mean curves through the eigenelements for two independent functional samples. Under the Gaussian assumption, Zhang et al. (2010) considered the two-sample test based on L^2-norm. Fremdt et al. (2014) derived mean functions comparison through a normal approximation method but only applicable to dense functional data samples. Pomann et al. (2016) still solved testing the curve equality problem though in bivariate (two-dimensional by their words) functional data setting and for distribution function testing. Regarding the k-sample testing or the one-way ANOVA for functional data, works include HANOVA (Fan and Lin, 1998), Cramér-von Mises type test (Cuevas et al., 2004; Estévez-Pérez and Vilar, 2013), F-type test (Ramsay and Silverman, 2005; Zhang, 2013; Zhang and Liang, 2014), B-spline test (Górecki and Smaga, 2015), and Mahalanobis distance (Ghiglietti et al., 2017), among others. In the case of within-curve dependence in each sample, Aston and Kirch (2012) detected the mean curve variation using L^2-norm criterion. Staicu et al. (2014) and its multiple group extension Staicu et al. (2015) worked on parametric testing relying on quite strong assumptions. Notice that, throughout our literature review, since our awareness concentrates on testing the equality of mean functions, we leave out other inferential topics such as testing the equality of coefficient operators or testing independency within a sample, and etc.

It has series of work in functional time series literature on testing the equality of mean functions, where weak dependence between or within two samples are accommodated in reality. Testing mean function difference in such functional time series study had still been on comparison of mean curve functions (Zhang et al., 2011; Horváth et al., 2013, 2014; Horváth and Rice, 2015a,b; Torgovitski, 2015, among others).

Aforementioned literature in both functional curve samples and functional time series have all inclined to testing the equality of mean curve functions, i.e. the inferential target is on univariate functional data. However, for comparison between samples of multivariate functional data, there have been few works by far. Only Gromenko et al. (2017) raised testing the equality of the mean surfaces of bivariate functional data, but eventually the equality of mean curves indexed at all locations were tested. Also to the best of our knowledge, the profile test of mean surfaces has not been considered for two bivariate functional data samples. Although the profile test of mean surfaces may belong to the curve test scope, it attributes to two different topics due to the different subjects. Above dire need in real-world data analysis and literature review motivates us to develop valid tests for equality of means surfaces and the corresponding profile test for bivariate functional data samples.

To address the problem in demand, firstly, we obtain the marginal eigen-function of the pooled sample by marginal functional principal component analysis (FPCA) and project the profiles of mean surfaces on marginal eigenfunctions. The profile testing statistic measures the distance of the profile of mean surfaces for two bivariate functional samples. Once the marginal eigenfunctions are obtained, the eigensurfaces of the pooled sample can be constructed by further FPCA. The distance between mean surfaces for two samples can be measured by the globe test statistic using the analogous projection ideas. Consequently, our proposed profile testing procedures can be implemented for every profile of the mean surface, which corresponds to simultaneously test whether mean precipitation curves have significant difference for every station. The globe test performs well in terms of both the size and the power in that it includes the information of two domains effectively.

The major contribution of this paper is threefold. Firstly, the presented methodology may be the first one to detect difference of mean surfaces and its profile for two-sample bivariate functional data. In contrast to the literature that we can search out by far, of which the focus has almost all been on testing the equality of mean curves as a matter of fact. When one argument is fixed, our profile test methodology can also simultaneously detect the mean difference in the other domain. Secondly, our testing procedures are interpretable and easily implemented. This will help fill out some theoretical gaps in functional inference and facilitate the real application and interpretation in statistical perspective. Finally, asymptotic distributions of the test statistics under null hypotheses has been derived. The consistency of test procedure has been proved. In addition, simulation studies show that the proposed tests have a good control of the type I error by the size and can detect difference in mean surfaces and its profile effectively in terms of power in finite samples.

The rest of the paper is organized as follows. In Section 2.2, we describe the model and data structure. The profile test procedure of mean surfaces for two bivariate functional data samples is presented in Section 2.3, while globe test procedure is proposed in Section 2.4. The finite sample performance for several representative scenarios is investigated in Section 2.5. In Section 2.6, we demonstrate two applications associated with the precipitation changes affected jointly by time and locations in the Midwest of USA, and the trends in human mortality from European period life tables. Theory proofs are included in Section 2.7.

2.2 Model and data structure

Let $L^2(\mathcal{S} \times \mathcal{T})$ be the separable Hilbert space. $\{X^{(m)}(s,t) : (s,t) \in \mathcal{S} \times \mathcal{T}\}$ is a square integrable stochastic process on $L^2(\mathcal{S} \times \mathcal{T})$ with mean function $\mu_m(s,t) = E\{X^{(m)}(s,t)\}$ and covariance function

$$C^{(m)}\{(s,v),(u,t)\} = \mathrm{E}\{X^{(m)c}(s,v)X^{(m)c}(u,t)\},$$

where $X^{(m)c}(s,t) = X^{(m)}(s,t) - \mu_m(s,t)$, for $m = 1,2$, respectively. With this notation, we can decompose $X^{(m)}(s,t)$ into

$$X^{(m)}(s,t) = \mu_m(s,t) + \varepsilon^{(m)}(s,t),\ m = 1,2,$$

where $\varepsilon^{(m)}(s,t)$ is the stochastic part of $X^{(m)}(s,t)$ with $\mathrm{E}\{\varepsilon^{(m)}(s,t)\} = 0$ and covariance function $C^{(m)}\{(s,v),(u,t)\}$.

Functional samples $\{X_i^{(m)}(s,t), m = 1,2;\ i = 1,\cdots,n_m\}$ may usually be modeled as independent realizations of the underlying stochastic process $X^{(m)}(s,t)$. In practice, $\{X_i^{(m)}(s,t), m = 1,2; i = 1,\cdots,n_m\}$ can not be observed, but rather, measurements are taken at discrete time points. In this paper, we assume $\{X_i^{(m)}(s,t), m = 1,2; i = 1,\cdots,n_m\}$ are recorded on a regular and dense grid of time points as follows,

$$\begin{aligned}
&X_i^{(m)}(s_{il_1},t_{il_2}) = \mu_m(s_{il_1},t_{il_2}) + \varepsilon_i^{(m)}(s_{il_1},t_{il_2});\\
&m = 1,2;\ i = 1,\cdots,n_m;\ l_1 = 1,\cdots,N;\ l_2 = 1,\cdots,M.
\end{aligned}$$

In this paper, we are firstly interested in profile test of bivariate functional data samples, i.e. for every fixed $t^* \in \mathcal{T}$,

$$H_0^{\mathcal{S}} : \mu_1(s,t^*) = \mu_2(s,t^*) \text{ vs. } H_1^{\mathcal{S}} : \mu_1(s,t^*) \neq \mu_2(s,t^*),\ s \in \mathcal{S}, \tag{2.1}$$

or for every fixed $s^* \in \mathcal{S}$,

$$H_0^{\mathcal{T}} : \mu_1(s^*,t) = \mu_2(s^*,t) \text{ vs. } H_1^{\mathcal{T}} : \mu_1(s^*,t) \neq \mu_2(s^*,t),\ t \in \mathcal{T}. \tag{2.2}$$

Then we go to the second target to present a globe test procedure for bivariate functional data samples with hypothesis below,

$$H_0 : \mu_1(s,t) = \mu_2(s,t) \text{ vs. } H_1 : \mu_1(s,t) \neq \mu_2(s,t),\ s \in \mathcal{S}, t \in \mathcal{T}. \tag{2.3}$$

The equality in hypothesis (2.1) means that $\int_{\mathcal{S}}\{\mu_1(s,t^*) - \mu_2(s,t^*)\}^2 ds = 0$ for every fixed $t^* \in \mathcal{T}$, and the alternative means that $\int_{\mathcal{S}}\{\mu_1(s,t^*) - \mu_2(s,t^*)\}^2 ds > 0$. Analogously meaning can be interpreted for (2.2). However, null hypothesis of (2.3) implies $\int_{\mathcal{S}}\int_{\mathcal{T}}\{\mu_1(s,t) - \mu_2(s,t)\}^2 dtds = 0$ while the alternative means that $\int_{\mathcal{S}}\int_{\mathcal{T}}\{\mu_1(s,t) - \mu_2(s,t)\}^2 dtds > 0$. For statistical inference of bivariate functional data, marginal FPCA is a widely used tool, which often assumes that bivariate functional data can project onto finite-dimensional eigensurfaces (Li and Guan, 2014; Park and Staicu, 2015; Aston et al., 2017). It is our start point for the proposed profile and globe test procedures.

2.3 Profile test of bivariate functional data

Profile test of bivariate functional data is an important problem, as it allows to provide multiple insight from multiple angles, and also is of interest in many applications. For example, in analysis of precipitation, the testing problem (2.2) corresponds to test whether mean precipitation curves have significant difference before and after 1966 for every station, while the testing problem (2.1) means to test whether different stations have significant difference for every day. Berkes et al. (2009) considered detection the difference only on an individual station. However, it is difficult to implement when the number of stations is large if we sequentially test for every station by their method. So, we propose the profile test of mean functions which is easy to implement and can simultaneously detect difference of all stations. In this section, we address the test problem (2.1) only as (2.2) can be analogously implemented.

As a first step, the marginal covariance function is denoted to be $G_{\mathcal{S}}^{(m)}(s,u) = \int_{\mathcal{T}} C^{(m)}\{(s,t),(u,t)\}dt$, as the form of (5) in Chen et al. (2017), and may be estimated by

$$\widehat{G}_{\mathcal{S}}^{(m)}(s_h, s_l) = \frac{1}{n_m M}\sum_{i=1}^{n_m}\sum_{k_2=1}^{M} \widehat{X}_i^{(m)c}(s_h, t_{ik_2})\widehat{X}_i^{(m)c}(s_l, t_{ik_2}), \tag{2.4}$$

where $\widehat{X}_i^{(m)c}(s,t) = X_i^{(m)}(s,t) - \overline{X}^{(m)}(s,t)$ with $\overline{X}^{(m)}(s,t) = \dfrac{1}{n_m}\sum_{i=1}^{n_m} X_i^{(m)}(s,t)$.

Denote

$$\widehat{G}_{\mathcal{S}}(s,u) = \frac{n_2}{n_1+n_2}\widehat{G}_{\mathcal{S}}^{(1)}(s,u) + \frac{n_1}{n_1+n_2}\widehat{G}_{\mathcal{S}}^{(2)}(s,u),\ \ s,u \in \mathcal{S}.$$

It is easy to see $\widehat{G}_{\mathcal{S}}(s,u) \xrightarrow{p} (1-\theta)G_{\mathcal{S}}^{(1)}(s,u) + \theta G_{\mathcal{S}}^{(2)}(s,u) \equiv G_{\mathcal{S}}(s,u)$, where θ is defined in Assumption 6 stated in next section and $G_{\mathcal{S}}(s,u)$ is the pooled covariance function. Consequently, it has orthogonal eigenfunctions $\{\psi_j\}_{j\geqslant 1}$ and non-negative eigenvalues $\{\nu_j\}_{j\geqslant 1}$ satisfying

$$\int_{\mathcal{S}} G_{\mathcal{S}}(s,u)\psi_j(u)du = \nu_j\psi_j(s),\ \ s,u \in \mathcal{S},\ \ j = 1,2,\ldots.$$

Such eigencomponents can be numerically estimated by suitably discretized eigenequations,

$$\int_{\mathcal{S}} \widehat{G}_{\mathcal{S}}(s,u)\widehat{\psi}_j(u)du = \widehat{\nu}_j\widehat{\psi}_j(s),\ \ j = 1,2,\ldots, \tag{2.5}$$

with orthogonal constraints on $\{\widehat{\psi}_j\}_{j\geqslant 1}$.

Once the estimators of marginal eigen-functions $\widehat{\psi}_j(s),\ j = 1,2,\ldots$, are obtained, we project the observations onto the marginal eigenfunctions and obtain the profile estimators of mean functions as follows: for every fixed $t^* \in \mathcal{T}$,

$$\widehat{\mu}_m(s,t^*) = \sum_{j=1}^{J} \widehat{\eta}_j^{(m)}(t^*)\widehat{\psi}_j(s),\ \ m = 1,2, \tag{2.6}$$

with

$$\widehat{\eta}_j^{(m)}(t^*) = \frac{1}{n_m}\sum_{i=1}^{n_m}\widehat{\eta}_{ij}^{(m)}(t^*),\ \ \widehat{\eta}_{ij}^{(m)}(t^*) = \frac{1}{N}\sum_{l_1=1}^{N} X_i^{(m)}(s_{il_1},t^*)\widehat{\psi}_j(s_{il_1}).$$

For practical implementation, one has to decide the magnitude of J. A practical strategy is $J = \min\{j : \frac{\widehat{\nu}_1+\widehat{\nu}_2+\cdots+\widehat{\nu}_k}{\widehat{\nu}_l+\widehat{\nu}_2+\cdots} > q\}$, where $\widehat{\nu}_l,\ l = 1,2,\cdots$ are defined in (2.5). We find that $q = 90\%$ threshold works well for our numerical examples.

Based on above discussion, we propose the following profile test statistic

$$\widehat{\mathrm{TP}}(t^*) = \frac{n_1 n_2}{n_1+n_2}\sum_{j=1}^{J}\frac{\left(\widehat{\eta}_j^{(1)}(t^*) - \widehat{\eta}_j^{(2)}(t^*)\right)^2}{\widehat{\lambda}_j(t^*)},$$

where $\widehat{\lambda}_j(t^*) = \frac{n_2}{n_1+n_2}\widehat{\lambda}_j^{(1)}(t^*)+\frac{n_1}{n_1+n_2}\widehat{\lambda}_j^{(2)}(t^*)$ with $\widehat{\lambda}_j^{(m)}(t^*) = n_m^{-1}\sum_{i=1}^{n_m}\left\{\widehat{\eta}_{ij}^{(m)}(t^*) - \widehat{\eta}_j^{(m)}(t^*)\right\}^2$, $m = 1, 2$.

Remark 2.1. *It is easy to see that* $\frac{n_1 n_2}{n_1+n_2}\int[\widehat{\mu}_1(s,t^*) - \widehat{\mu}_2(s,t^*)]^2 dt \xrightarrow{p} U_{n_1,n_2} = \frac{n_1 n_2}{n_1+n_2}\sum_{l=1}^{K}(\widehat{\eta}_l(t^*) - \widehat{\eta}_2(t^*))^2$. *However, the variance of* U_{n_1,n_2} *may be unnecessarily inflated by the presence of, possibly many, very small estimates* $\widehat{\mu}_1(s,t^*) - \widehat{\mu}_2(s,t^*)$. *This drawback can be remedied by giving a divisor to* $\widehat{\lambda}_j(t^*)$.

We then establish asymptotic behaviors of the test statistic $\widehat{\mathrm{TP}}(t^*)$ under the null hypothesis $H_0^{\mathcal{S}}$ and the alternative one $H_1^{\mathcal{S}}$. To derive the asymptotic properties of profile test statistic, we make the following assumptions.

Assumption 2.1. $\nu_1 > \nu_2 > \cdots$ *where* $\{\nu_j\}_{j=1,2,\ldots}$ *are the eigenvalues of covariance operates* $G_S(s,u)$.

Assumption 2.2. *For every fixed* t^*, $\mu_m(s,t^*), m = 1, 2$ *may be written as* $\mu_m(s,t^*) = \sum_{j=1}^{\infty}\eta_j^{(m)}(t^*)\psi_j(s)$, *where* $\eta_j^{(m)}(t^*) = \int_0^1 \mu_m(s,t^*)\psi_j(s)ds$.

Assumption 2.3. *Assume* $\sup_{(s,t)\in\mathcal{S}\times\mathcal{T}}\mu_m^2(s,t), m = 1, 2$ *are bounded and* $E(\sup|\varepsilon^m(s,t)|^4)$, $m = 1, 2$ *are bounded.*

Assumption 2.4. *The grid point* $\{t_{il_1} : l_1 = 1, \ldots, N\}$ *and* $\{s_{il_2} : l_2 = 1, \ldots, M\}$ *are equidistant. We assume* $n_1/N^2 = o(1)$, $n_1/M^2 = o(1)$, $n_2/N^2 = o(1)$ *and* $n_2/M^2 = o(1)$.

Assumption 2.5. $\min\{n_1, n_2\} \to \infty$, $n_1/(n_1+n_2) \to \theta$ *for a fixed constant* $\theta \in (0, 1)$.

Assumptions 2.1 and 2.3 are regular conditions. One needs these conditions to uniquely (up to signs) choose $\psi_j(s)$ and obtain the bound of $\widehat{\psi}_j(s) - \psi_j(s)$. Assumption 2.2 means that the profiles of mean surface are projected onto a space that is generated by a large set of basis functions. Assumption 2.4 requires that functional data are recorded on dense grid. Assumption 2.5 is of standard for two-sample asymptotic inference.

Theorem 2.1. *Under Assumptions 2.1-2.5 and $H_0^{\mathcal{S}}$, we have $\widehat{\mathrm{TP}}(t^*) \xrightarrow{d} \chi_J^2$, where χ_J^2 stands for a χ^2 -distributed random variable with J degrees of freedom. Under $H_1^{\mathcal{S}}$ and $0 < \theta < 1$, we have $\widehat{\mathrm{TP}}(t^*) \xrightarrow{p} \infty$.*

From the expression of $\widehat{\mathrm{TP}}(t^*)$ and remark 2.1, we can see that $\widehat{\mathrm{TP}}(t^*)$ depends on sample sizes n_1, n_2, and $\widehat{\eta}_j^{(1)}(t^*) - \widehat{\eta}_j^{(2)}(t^*), j = 1, \cdots, J$, which reflects the difference of profile mean functions $\mu_1(s, t^*)$ and $\mu_2(s, t^*)$. Intuitively, $\sqrt{\frac{n_1 n_2}{n_1+n_2}}(\widehat{\eta}_j^{(1)}(t^*) - \widehat{\eta}_j^{(2)}(t^*))\widehat{\lambda}_j^{-1/2}$ has a limiting standard normal distribution under $H_0^{\mathcal{S}}$. Theorem 2.1 shows that $\widehat{\mathrm{TP}}(t^*)$ asymptotically follows the chi-square distribution with J degrees of freedom if $H_0^{\mathcal{S}}$ holds. Furthermore, $\widehat{\mathrm{TP}}(t^*)$ is consistent under $H_1^{\mathcal{S}}$. The proof of this theorem is provided in Section 2.7.

2.4 Globe test of bivariate functional data

Compared with the profile test, the globe test of bivariate functional data attempts to detect the joint effects impacted by both domains. In this section, we develop a globe test method for bivariate functional data which aims to detect whether mean surfaces of precipitation have significant difference over a specific time window and/or a specific area, or whether two regions exist significant difference during different time windows.

Based on the estimated marginal eigenfunctions $\widehat{\psi}_j(s)$ in Section 2.3, we next estimate the marginal functional principal component scores $\widetilde{\xi}_{j,i}^{(m)}(t)$. The traditional integral estimates of $\widetilde{\xi}_{j,i}^{(m)}(t)$ based on the definition

$$\widetilde{\xi}_{j,i}^{(m)}(t) = \int_{\mathcal{S}} X_i^{(m)c}(s,t)\widehat{\psi}_j(s)ds,\ i = 1,\dots,n_m;\ j = 1,2,\dots.$$

are

$$\begin{aligned}\widehat{\xi}_{j,i}^{(m)}(t) &= \sum_{l=2}^{N} X_i^{(m)c}(s_l,t)\widehat{\psi}_j(s_l)(s_l - s_{l-1}),\\ i &= 1,\dots,n_m;\ j = 1,2,\dots.\end{aligned} \tag{2.7}$$

where N is the number of measurements for $X_i^{(m)c}(s,t)$ in the direction $\mathcal{S}$.

Notice that each score function $\widehat{\xi}_{j,i}^{(m)}(t)$ is a centered new random curve. Denote the covariance function of $\xi_{j,i}^{(m)}(t)$ by $G_{\mathcal{T},j}^{(m)}(v,t) = E\{\xi_{j,i}^{(m)}(v)\xi_{j,i}^{(m)}(t)\}$. Then, the estimator of $G_{\mathcal{T},j}^{(m)}$ is denoted as,

$$\widehat{G}_{\mathcal{T},j}^{(m)}(t_h,t_l) = \frac{1}{n_m}\sum_{i=1}^{n_m} \widehat{\xi}_{j,i}^{(m)}(t_h)\widehat{\xi}_{j,i}^{(m)}(t_l),\ t_h,t_l \in \mathcal{T};\ j = 1,2,\dots.$$

Let

$$\widehat{G}_{\mathcal{T},j}(v,t) = \frac{n_2}{n_1+n_2}\widehat{G}_{\mathcal{T},j}^{(1)}(v,t) + \frac{n_1}{n_1+n_2}\widehat{G}_{\mathcal{T},j}^{(2)}(v,t),\ v,t \in \mathcal{T}; j = 1,2,\dots.$$

It is easy to see $\widehat{G}_{\mathcal{T},j}(v,t) \xrightarrow{p} (1-\theta)G_{\mathcal{T},j}^{(1)}(v,t) + \theta G_{\mathcal{T},j}^{(2)}(v,t) \equiv G_{\mathcal{T},j}(v,t)$ where $G_{\mathcal{T},j}(v,t)$ is the covariance function and has orthogonal eigenfunctions $\{\phi_{jk}\}_{k\geqslant 1}$ and non-negative eigenvalues $\{\nu_{jk}\}_{k\geqslant 1}$ satisfying

$$\int_{\mathcal{T}} G_{\mathcal{T},j}(v,t)\phi_{jk}(v)dv = \nu_{jk}\phi_{jk}(t),\ v,t \in \mathcal{T};\ k,j = 1,2,\dots.$$

Then estimators of eigenvalues and eigenfunctions $\{(\nu_{jk},\phi_{jk}(t)) : j,k \geqslant 1\}$ are obtained by the following equations,

$$\int_{\mathcal{T}} \widehat{G}_{\mathcal{T},j}(v,t)\widehat{\phi}_{jk}(v)dv = \widehat{\nu}_{jk}\widehat{\phi}_{jk}(t),\ k,j = 1,2,\dots, \tag{2.8}$$

with orthogonal constraints on $\{\widehat{\phi}_{jk}\}_{k\geqslant 1}$.

Denote $\varphi_{jk}(s,t) \equiv \phi_{jk}(t)\psi_j(s)$ and its consistent estimator by $\widehat{\varphi}_{jk}(s,t) = \widehat{\phi}_{jk}(t)\widehat{\psi}_j(s)$. We propose estimators of the mean surfaces which are projection of observations onto a hyperspace spanned from the pooled eigensurfaces $\{\widehat{\varphi}_{jk}(s,t) : j,k \geqslant 1\}$, written as

$$\widehat{\mu}_m(s,t) = \sum_{j=1}^{J}\sum_{k=1}^{K_j} \widehat{\eta}_{jk}^{(m)}\widehat{\varphi}_{jk}(s,t),\ m = 1,2, \tag{2.9}$$

with

$$\widehat{\eta}_{jk}^{(m)} = \tfrac{1}{n_m}\sum_{i=1}^{n_m}\widehat{\eta}_{ijk}^{(m)},\ \widehat{\eta}_{ijk}^{(m)} = \tfrac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N} X_i^{(m)}(s_{il_1}, t_{il_2})\widehat{\varphi}_{jk}(s_{il_1}, t_{il_2}),$$

where selection of J is the same to in Section 2.3 and K_j can be decided by analogous procedure. In details, we select $K_j = \min\{k : \frac{\widehat{\nu}_{j1}+\widehat{\nu}_{j2}+\cdots+\widehat{\nu}_{jk}}{\widehat{\nu}_{j1}+\widehat{\nu}_{j2}+\cdots} > 0.9\}$, where $\widehat{\nu}_{jl}$, $l = 1, 2, \cdots$ are defined in (2.8).

It is natural to take into consideration the term $\widetilde{\mathrm{TC}} \equiv \int_{\mathcal{S}}\int_{\mathcal{T}}\{\mu_1(s,t)-\mu_2(s,t)\}^2 dtds$ to measure the distance between two estimated mean surfaces.

It is readily seen that $\widetilde{\mathrm{TC}} \xrightarrow{p} \sum_{j=1}^{J}\sum_{k=1}^{K_j}\left(\widehat{\eta}_{jk}^{(1)} - \widehat{\eta}_{jk}^{(2)}\right)^2$. Therefore, H_0 will be rejected if $\widetilde{\mathrm{TC}}$ is large. Similarly, the variance of $\widetilde{\mathrm{TC}}$ may be unnecessarily inflated by the presence of, possibly many, very small estimates $\widehat{\eta}_{jk}^{(1)} - \widehat{\eta}_{jk}^{(2)}$. This drawback can also be remedied by giving a divisor to their variance.

Based on the above steps, we propose the following test statistic

$$\widehat{\mathrm{TM}} = \frac{n_1 n_2}{n_1+n_2}\sum_{j=1}^{J}\sum_{k=1}^{K_j}\frac{\left(\widehat{\eta}_{jk}^{(1)} - \widehat{\eta}_{jk}^{(2)}\right)^2}{\widehat{\lambda}_{jk}},$$

where $\widehat{\lambda}_{jk} = n_2(n_1+n_2)^{-1}\widehat{\lambda}_{jk}^{(1)} + n_1(n_1+n_2)^{-1}\widehat{\lambda}_{jk}^{(2)}$ with $\widehat{\lambda}_{jk}^{(m)} = (n_m-1)^{-1}\sum_{i=1}^{n_m}\left(\widehat{\eta}_{ijk}^{(m)} - \widehat{\eta}_{jk}^{(m)}\right)^2$, $m = 1, 2$.

From (2.9), we can see that $X_i^{(1)}(\cdot,\cdot)$ and $X_i^{(2)}(\cdot,\cdot)$ are directly projected on the common basis surface and obtain $\widehat{\eta}_{ijk}^{(1)}$ and $\widehat{\eta}_{ijk}^{(2)}$. $\widehat{\eta}_{jk}^{(1)}$ and $\widehat{\eta}_{jk}^{(2)}$, which are the average of such projection, and hence can be viewed as the scores of projection that two mean surfaces $\mu_1(s,t)$ and $\mu_2(s,t)$ project on the same basis function space, respectively. The representation of $\widehat{\mathrm{TM}}$ measures the total such deviation between two samples. Therefore, the proposed method has a nice explanation and easy to implement.

Next we establish asymptotic behavior of the test statistic $\widehat{\mathrm{TM}}$ under hypotheses (2.3). Additionally, we need the following assumptions.

Assumption 2.6. *$\nu_{j1} > \nu_{j2} > \cdots$ where $\{\nu_{jk}\}_{k=1,2,\ldots;j=1,2,\ldots}$ are the eigenvalues of the covariance function $G_T(v,t)$.*

Assumption 2.7. *Assume $\mu_m(s,t), m = 1,2$ may be written as $\mu_m(s,t) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\eta_{jk}^{(m)}\varphi_{jk}(s,t)$, where $\eta_{jk}^{(m)} = \int_0^1\int_0^1 \mu_m(s,t)\varphi_{jk}(s,t)dsdt$.*

Assumption 2.6 along with Assumption 2.3 in Section 2.3 ensures the bound of $\widehat{\phi}_{jk}(t) - \widehat{\phi}_{jk}(t)$. The interpretation of Assumption 2.7 is similar to Assumption 2.2 in Section 2.3.

Theorem 2.2. *Under Assumptions 2.1-2.7 and H_0, we have*

$$\widehat{\mathrm{TM}} \xrightarrow{d} \chi^2_{\sum\limits_{j=1}^{J} K_j},$$

where $\chi^2_{\sum_{j=1}^{J} K_j}$ stands for a χ^2-distributed random variable with $\sum_{j=1}^{J} K_j$ degrees of freedom. Under H_1 and $0 < \theta < 1$, we have $\widehat{\mathrm{TM}} \xrightarrow{p} \infty$.

Intuitively $\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left(\widehat{\eta}_{jk}^{(1)} - \widehat{\eta}_{jk}^{(2)}\right)\widehat{\lambda}_{jk}^{-1/2}$ has a limiting standard normal distribution under H_0. Theorem 2.2 shows that $\widehat{\mathrm{TM}}$ asymptotically follows the chi-square distribution with $\sum\limits_{j=1}^{J} K_j$ degrees of freedom under H_0. The consistency of $\widehat{\mathrm{TM}}$ is also illustrated under H_1, which together provides clear theoretical justification of the empirical properties of the proposed test. The proof of this theorem is provided in 2.7.

2.5 Simulation studies

We conduct extensive simulation studies and report two representative examples here. Examples 2.1 and 2.2 evaluate two proposed testing procedures in terms of empirical size and power when covariance functions of two samples are identical or distinct, separately. The data grid for argument s consists of 100 equispaced points on $[0, 1]$, and the grids for argument s consists of 50 equispaced points on $[0, 1]$. Each pair of data-generated processes was replicated 1000 times.

Example 2.1. *Identical covariance functions.*

In this example, we consider the following model

$$\begin{aligned} X_i^{(1)}(s,t) &= \varepsilon_i^{(1)}(s,t),\ i = 1, \ldots, n_1, \\ X_i^{(2)}(s,t) &= \delta(s+t) + \varepsilon_i^{(2)}(s,t),\ i = 1, \ldots, n_2, \end{aligned} \tag{2.10}$$

where $\varepsilon_i^{(1)}(s,t)$ and $\varepsilon_i^{(2)}(s,t)$ are independently generated from

$$\varepsilon(s,t) = \sum_{j=1}^{2} \xi_j(t)\psi_j(s),\ s \in [0,1], t \in [0,1],$$

with $\psi_1(s) = s^2$ and $\psi_2(s) = s^3$, $s \in [0, 1]$. $\xi_j(t)$ is generated from

$$\xi_j(t) = \sum_{k=1}^{2} \chi_{jk}\phi_{jk}(t),\ j = 1, 2,$$

with $\phi_{11}(t) = \phi_{21}(t) = -\sqrt{2}\cos(2\pi t)$, $\phi_{12}(t) = \phi_{22}(t) = \sqrt{2}\sin(2\pi t)$, $t \in [0, 1]$; $\chi_{11} \sim N(0, 3)$, $\chi_{12} \sim N(0, 1.5)$, $\chi_{21} \sim N(0, 2)$, and $\chi_{22} \sim N(0, 1)$.

Example 2.2. *Distinct covariance functions.*

To compare with Example 2.1, we consider the following model

$$X_i^{(1)}(s,t) = \varepsilon_i^{(1)}(s,t),\ i = 1,\ldots,n_1,$$
$$X_i^{(2)}(s,t) = \delta(s+t) + \varepsilon_i^{(2)}(s,t),\ i = 1,\ldots,n_2, \tag{2.11}$$

where $\varepsilon_i^{(1)}(s,t)$ is generated from

$$\varepsilon^{(1)}(s,t) = \textstyle\sum_{j=1}^{2} \xi_j(t)\psi_j(s),\ s \in [0,1], t \in [0,1],$$

and $\varepsilon_i^{(2)}(s,t)$ from

$$\varepsilon^{(2)}(s,t) = \xi_1(t)\psi_1(s),\ s \in [0,1], t \in [0,1],$$

with $\psi_1(s) = s^2$ and $\psi_2(s) = s^3$, $s \in [0,1]$. $\xi_j(t)$ is generated from

$$\xi_j(t) = \sum_{k=1}^{2} \chi_{jk}\phi_{jk}(t),\ j = 1, 2,$$

with $\phi_{11}(t) = \sqrt{2}\cos(2\pi t)$, $\phi_{21}(t) = \sqrt{2}\sin(2\pi t)$, $\phi_{12}(t) = 2\cos(4\pi t)$, $\phi_{22}(t) = 2\sin(4\pi t)$, $t \in [0,1]$; $\chi_{11} \sim N(0,3)$, $\chi_{12} \sim N(0,1.5)$, $\chi_{21} \sim N(0,2)$, and $\chi_{22} \sim N(0,1)$.

Example 2.1 can be seen as two-sample tests where covariance functions are identical, while covariance functions of Example 2.2 are distinct. The sample size pair is taken to be $(n_1,n_2) = (25,75)$, $(50,150)$, $(100,300)$, $(50,50)$, $(100,100)$, and $(200,200)$, respectively. The empirical sizes of profile test are computed for different s and t. To save space, we here only present the results of different s for $(n_1,n_2) = (100,100)$ in Fig. 2.3. Next, we can also compute the empirical sizes of the globe test. The results are reported in Table 2.1. The empirical power can be evaluated when $\delta \neq 0$. The empirical power at $\delta = 0.4, 0.6, 0.8$ of profile tests are displayed in Fig. 2.4 while the results of globe tests at $\delta = 0.2, 0.4, 0.6, 0.8, 1.0, 1.2$ are scatter plotted in Fig. 2.5.

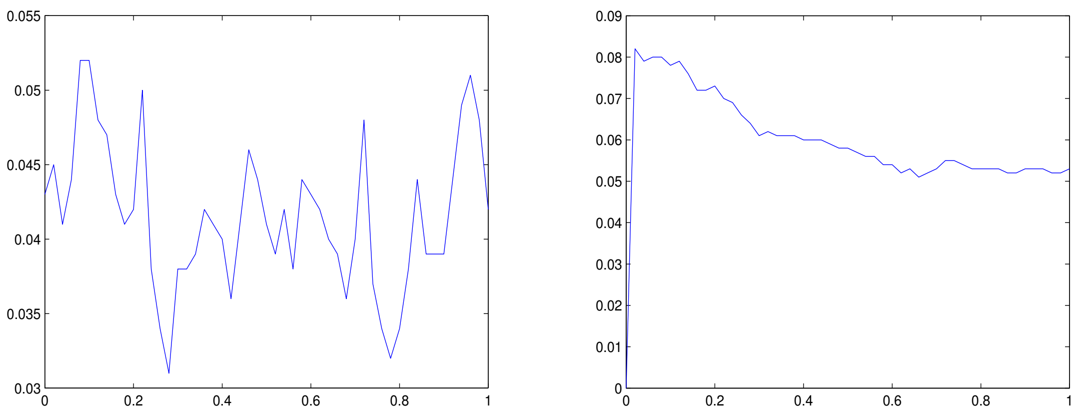

Figure 2.3: The results of empirical size when covariance functions of two samples are identical (left column) and distinct (right column).

Table 2.1: Empirical sizes of two proposed test procedures in Examples 2.1 and 2.2.

(n_1, n_2)	(50,50)	(100,100)	(200,200)	(25,75)	(50,150)	(100,300)
Example 2.1	0.079	0.060	0.050	0.111	0.085	0.061
Example 2.2	0.074	0.064	0.048	0.080	0.062	0.048

Several observations can be concluded from Fig. 2.3 and Fig. 2.4. Firstly, the profile tests have a good control of the type I error. The empirical sizes of identical covariance scenarios are better than that of distinct covariance cases. Secondly, the empirical power of the test becomes larger when δ increases from 0.4 to 0.8, which is expected. Lastly, the empirical power for the same covariance case is slightly larger than that of the different covariance function cases.

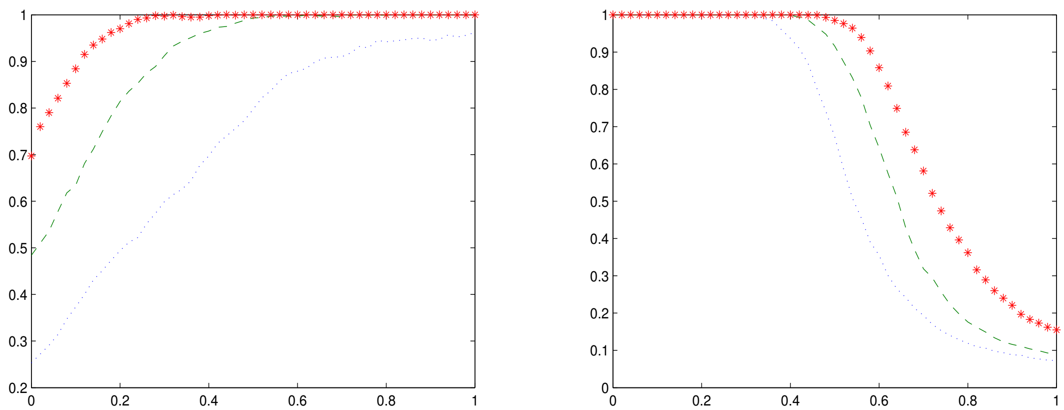

Figure 2.4: The results of empirical power when covariance functions of two samples are identical (left column) and distinct (right column).

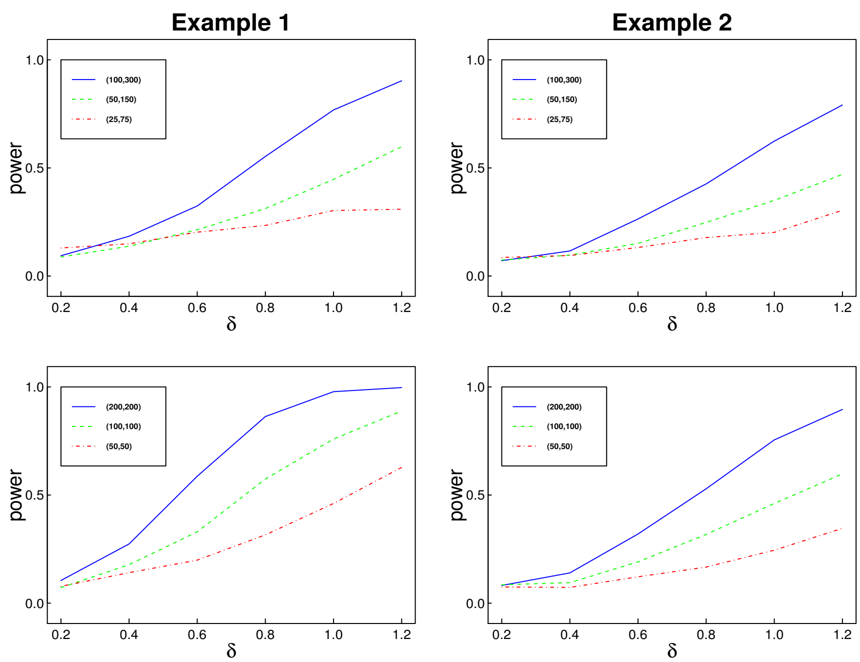

Figure 2.5: The results of empirical power when covariance functions of two samples are identical (left column) and distinct (right column). Top: The results of empirical power of $(n_1, n_2) = (25, 75)$ (red), $(50, 150)$ (green) and $(100, 300)$ (blue). Bottom: The results of empirical power of $(n_1, n_2) = (50, 50)$ (red), $(100, 100)$ (green) and $(200, 200)$ (blue).

We may observe from Table 2.1 and Fig. 2.5 that the globe test approach can keep steady empirical size even at pairs of small sample sizes $(n_1, n_2) = (25, 75)$

or $(50, 50)$. The empirical power of two test methods increases as the sample size increases. When δ increases from 0.2 to 1.2, the empirical power of the test becomes more and more large, which is evidence of the consistency of the testing procedures. Also the empirical power of equal sample size scenario is slightly better than that of unequal sample size one.

2.6 Real data examples

To illustrate profile and globe tests methods, we analyse the historical precipitation data in the Midwest of USA and the period lifetables in Europe for human mortality trend analysis.

2.6.1 Precipitation data

The first example is used to analyze the changes of precipitation during 1941-2000 or in different regions in the Midwest of USA. Berkes et al. (2009) detected no changes during the period 1941-2000 for only one station while Gromenko et al. (2017) detected the change of precipitation during 1941-2000 over the whole region.

The precipitation data is available from the global historical climatological network database. The comprehensive U.S. Climate Normals dataset includes various derived products including daily air temperature normals, precipitation normals and hourly normals. The dataset that we analyzed in this paper can be downloaded directly from GHCN (Global Historical Climatology Network)-Daily, an integrated public database of NOAA (https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/) by an R interface. Our interest is daily precipitation records from Midwestern states including Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. In Fig. 2.1, totally 59 locations of the climate monitoring stations are indicated with blue circles ○ in 4 states from the Great Plains (light green region), and with red triangles △ in 5 states from the Great Lakes (yellow region). Notice that there is no climate monitoring stations in Iowa, Michigan, and Missouri. We target to detect whether the changes of average precipitation took place for different time phases or regions.

Let $Y_i(s,t)$ be the precipitation of the tth day in the ith year of the sth station. Before we apply the proposed method, we need to do registration with the data. To remove the effects due to the heavy tail distribution, we apply the transformation

$$Z_i(s,t) = \log_{10}\{Y_i(s,t) + 1\},$$

where $\{Y_i(s,t)\}$ are original records. After the transformation, we pre-smooth data by using the cubic splines function. It is noted that the data of every climate monitoring stations from 1941 to 2000 can be constituted into a time series with length 21900(365 day by 60 year). Then, the data of the 59 climate monitoring stations can be seen as a sample with sample size being 21900 and variables being 59. According to the empirical Pearson correlation of 59 variables, the 59 climate monitoring stations is stringed into a function by the stringing method in Chen et al. (2011b). Consequently the spatiotemporal data $\{Y_i(s,t)\}$ are converted into the bivariate functional data $\{X_i(s,t)\}$. Notice that the difference between the spatiotemporal data $Y_i(s,t)$ and the bivariate functional data $X_i(s,t)$ is that the argument s in the former expression has no order but it is ranked in the latter.

Gromenko et al. (2017) studied the data $Y_i(s,t)$ and detected out the change of the average precipitation at about 1967. In this subsection, we firstly apply the profile test to check if the profile of mean surfaces are equal during the periods 1941-1967 and 1968-2000. It corresponds to test whether the average precipitation of every station has changes during these periods. The p-values of the profile tests are computed and results are displayed in Fig. 2.6. As can be seen from Fig. 2.6, most of the p-values are less than 0.05 or significant except 11 stations. For ease of reference, we list the latitude and longitude in Table 2.2 for 11 stations. This displays that the

Table 2.2: The latitude and longitude of stations where the p-value of profile test are more than 0.1.

Code	latitude	longitude
USC00148235	38.4661	-101.7758
USC00250050	42.5522	-99.8556
USC00252145	41.4086	-102.9661
USC00255090	40.8508	-101.5428
USC00325479	46.8128	-100.9097
USC00394007	43.4378	-103.4739
USC00398307	45.4283	-101.0764
USC00394007	43.4378	-103.4739
USC00392797	45.7644	-99.6353
USC00321871	48.9075	-103.2944
USC00327530	46.8886	-102.3192

average precipitation of most locations had changed during the periods 1941-1967 and 1968-2000.

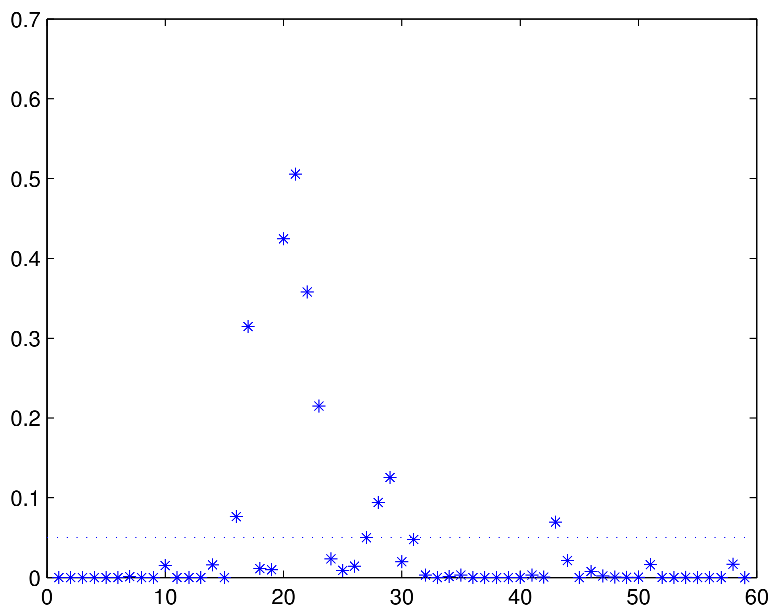

Figure 2.6: The p-value of the profile tests for every station.

Next, we implement the following globe test

$$H_0^{\text{Midwest}} : \mu_1(s,t) = \mu_2(s,t) \text{ vs. } H_1^{\text{Midwest}} : \mu_1(s,t) \neq \mu_2(s,t), s \in \mathbb{R}^{59}, t \in \mathbb{R}^{365}.$$

From the globe test procedure presented in Section 2.4 together with the asymptotic distribution of the test statistic $\widehat{\text{TM}}$, we calculate the corresponding p-value to be 0.001. This result is consistent with the conclusion of Gromenko, Kokoszka and Reimherr. That is, the patterns of mean surfaces are different over the whole Midwest region between before 1967 and after 1967. Intuitively, according to the results of the profile test, the precipitation had changed in most of locations which lead to the variations of whole region.

The heatmaps in Fig. 2.2 leak the information that sample mean values of annual precipitation in the Great Lakes (GL) based on 28 stations are more than that in the Great Plains (GP). This motivates us to further explore how the mean functions of bivariate functional data $\{X_i(s,t)\}$ was affected by temporal and spatial effects from both domains. It is natural to test the equality of two mean surfaces of the precipitation for the 31 stations located in the GP and the 28 stations located in the GL during the periods 1941-1967 and 1968-2000, respectively by

$$H_0^{1967-} : \mu^{\text{GP}} = \mu^{\text{GL}} \text{ vs. } H_1^{1967-} : \mu^{\text{GP}} \neq \mu^{\text{GL}},$$

and

$$H_0^{1967+} : \mu^{\text{GP}} = \mu^{\text{GL}} \text{ vs. } H_1^{1967+} : \mu^{\text{GP}} \neq \mu^{\text{GL}}.$$

All the p-values by globe test procedures for above two hypotheses are tiny approaching to zero indicating rejecting the null hypotheses but in favor of the alternative one. It is consistent with the intuition that the mean patterns of precipitation at Great Plains and at Great Lakes are different.

Furthermore, for the 28 stations located in the GL, we test the mean surfaces of

Table 2.3: Results of the tests based on statistics $\widehat{\text{TM}}$.

statistic	the observed value of a statistic	p-value
	$H_0^{\text{Midwest}} : \mu^{1967-}(s,t) = \mu^{1967+}(s,t)$	
$\widehat{\text{TM}}$	123.6	**0.001**
	$H_0^{\text{GP}} : \mu^{1967-}(s,t) = \mu^{1967+}(s,t)$	
$\widehat{\text{TM}}$	59.7175	0.5677
	$H_0^{\text{GL}} : \mu^{1967-}(s,t) = \mu^{1967+}(s,t)$	
$\widehat{\text{TM}}$	108.20	**0.0163**
	$H_0^{1967-} : \mu^{\text{GP}}(s,t) = \mu^{\text{GL}}(s,t)$	
$\widehat{\text{TM}}$	973.11	**0.0000**
	$H_0^{1967+} : \mu^{\text{GP}}(s,t) = \mu^{\text{GL}}(s,t)$	
$\widehat{\text{TM}}$	1116.4	**0.0000**

precipitation before and after 1967, denoted by

$$H_0^{\text{GL}} : \mu^{1967-} = \mu^{1967+} \text{ vs. } H_1^{\text{GL}} : \mu^{1967-} \neq \mu^{1967+}.$$

The p-value is 0.0163. The null hypothesis would be rejected at 0.05 significance level. Testing equality of the mean surfaces of precipitation before and after 1967 is also implemented for the 31 stations located in the GP, denoted by

$$H_0^{\text{GP}} : \mu^{1967-} = \mu^{1967+} \text{ vs. } H_1^{\text{GP}} : \mu^{1967-} \neq \mu^{1967+}.$$

The p-values by globe testing method are 0.5677. The null hypothesis would not be rejected at 0.05 significance level. That is, averagely speaking, the precipitation in the Great Lakes changed before 1967 and after 1967, whereas the mean pattern of precipitation in the Great Plains had no change before 1967 and after 1967. *Therefore, our analysis provides evidence that change in the mean function of precipitation was mainly due to the Great Lakes but the Great Plains may be affected little. By looking up the map, we find that all the stations in Table 2.2 are located in the Great Plains. It further verify the reliability of the proposed methods.* All testing results are presented in Table 2.3.

2.6.2 European human mortality rate data

In the second example, we will analyse the trends in human mortality based on the records in the period life tables during the calendar years 1960-2006 for Europe countries. A period life table represents the mortality conditions at a specific moment in time. It is approachable from the Human Mortality Database via the website linkage `www.mortality.org` (Wilmoth et al., 2007). The analysis of trends in human mortality is important to recover the demographic impacts. Results of such research will benefit the prediction and forecasting of future cohort mortality (Vaupel et al., 1998; Oeppen and Vaupel, 2002). We focus on comparison of different countries or genders, specifically on the older ages over 50 years old.

There are 32 countries included in the European period life tables. It contains five Eastern European countries, Belarus, Bulgaria, Russia, Ukraine and Lithuania, and the remaining 27 Western European countries. Following the notation introduced in Section 2.3, $X_i^{(1)}(s,t), i = 1, \ldots, 5$, denotes the mortality rate of the five Eastern European countries for subjects at age s and calendar year t, where $50 \leqslant s \leqslant 90$, focusing on the death rates of older individuals, and on a recent block of 47 years, $1960 \leqslant t \leqslant 2006$. Similarly, $X_i^{(2)}(s,t), i = 1, \ldots, 27$, denotes the mortality rate for other countries. The sample mean function $\widehat{\mu}_1(s,t) = \sum_{i=1}^{5} X_i^{(1)}(s,t)$ and $\widehat{\mu}_2(s,t) = \sum_{i=1}^{27} X_i^{(2)}(s,t)$ for two clusters of countries are visualized in Fig. 2.7. The heatmaps and sample mean surfaces show obvious opposite trend of mortality rates particularly for very aged people in Eastern and Western European countries as the calendar year passed 1980 or so. We apply the profile and globe test procedures to test if the two underlying mean surfaces and its profile are different.

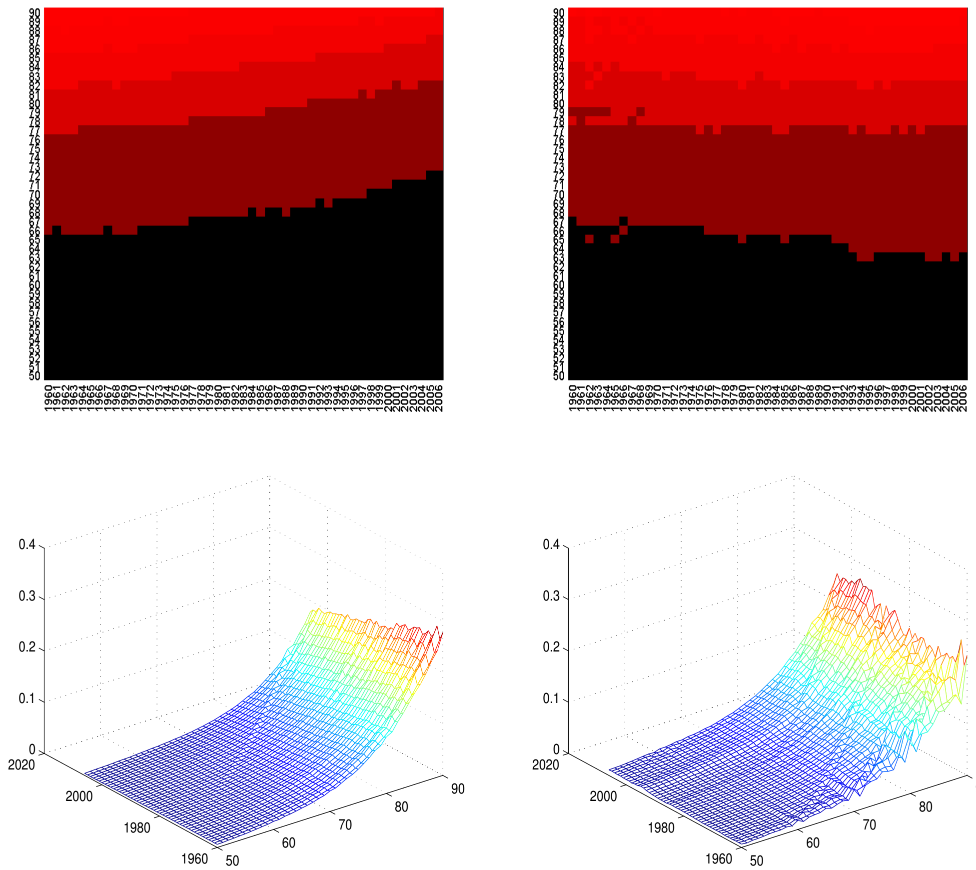

Figure 2.7: The heatmap of sample mean surface of precipitation during the time 1941-1967 in the Midwest, where the first 31 stations are located in the Great Plains and the latter 28 stations are located in the Great Lakes.

According to profile test method introduced in Section 2.3, we implement the tests (2.1) and (2.2). The p-values for fixed s^* or t^* are calculated, respectively. The results are presented in Fig. 2.8. For every fixed age s^*, we find that all of p-values are approaching to zero. This indicates that the mean mortality rates of the Eastern and Western European is different for every age $s^* = 50, \cdots, 90$. For every fixed year t^*, almost all p-values are less than 0.05 except for years $t^* = 1978$ and 1986. Sequentially, we implement the globe test for the mean mortality rates of the Eastern and Western European. The numbers of included components is

$J = 2, K_1 = 2, K_2 = 2$ are chosen by the fraction of variance explained (FVE) criterion with the threshold 0.90. Based on the asymptotic distribution of the test statistic $\widehat{\mathrm{TM}}$, the p-value is calculated to be 0. It coincides with the intuition on images in Fig. 2.7 and is evidence that the mean surfaces of the mortality rates are different between the Eastern and Western European countries. Also, it is consistent with the conclusion of the profile test because almost of $H_0^{\mathcal{S}}$ and $H_0^{\mathcal{T}}$ are rejected for fixed s^* and t^*.

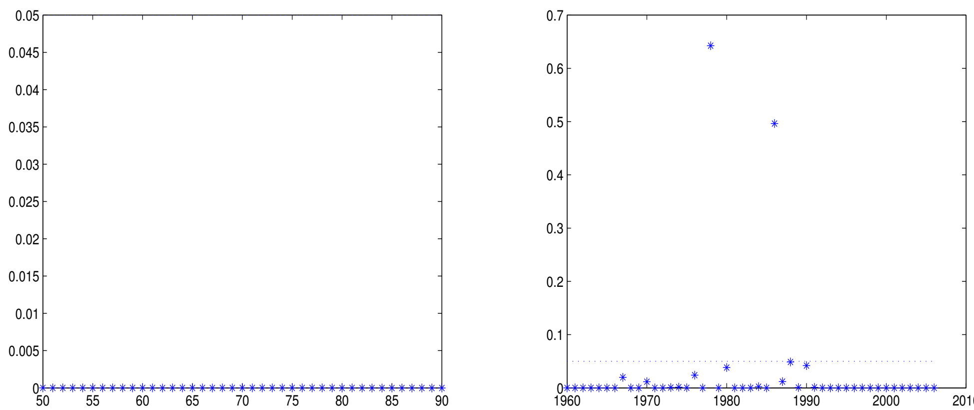

Figure 2.8: The p-value of the profile tests for every age (left) and year (right).

Next we examine the equality of mean surfaces and its profile between female and male clusters in West Europe. The heatmaps and sample mean surfaces for male and female clusters are displayed in Fig. 2.9. Intuitively it does not show obvious difference. However, all the p-values of profile tests are zero for fixed s^* and t^*. Furthermore, we also implement globe test and obtain the p-value that is 0. Therefore, the mean surface and its profile are different in Western Europe for aged people in different gender type.

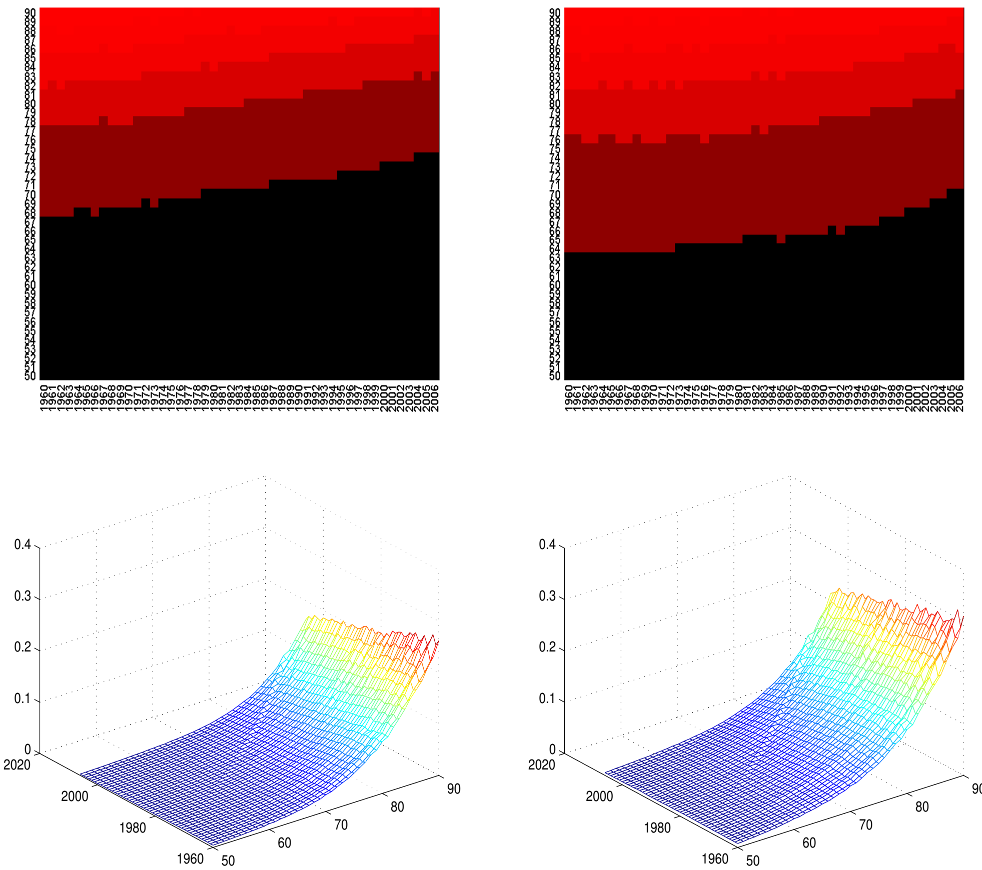

Figure 2.9: Top: Sample means of the mortality rate of male. Bottom: Sample means of the mortality rate of female.

2.7 Proof of the theorems

In order to prove the Theorems 2.1 and 2.2, we first introduce several lemmas.

Lemma 2.1. *Under Assumptions 2.1 and 2.3, we have*

$$\max_{1\leqslant j\leqslant J} \|\widehat{\psi}_j(s) - \hat{c}_j\psi_j(s)\| = O_p\{(n_1+n_2)^{-1/2}\},$$

where $\widehat{c}_j = sign(\widehat{\psi}_j, \psi_j)$.

Lemma 2.2. *Under Assumptions 2.3 and 2.6, we have*

$$\max_{1\leqslant j\leqslant J;1\leqslant k\leqslant K_j} \|\widehat{\phi}_{jk}(t)-\hat{d}_{jk}\phi_{jk}(t)\| = O_p\{(n_1+n_2)^{-1/2}\},$$

where $\widehat{d}_{jk} = sign(\widehat{\phi}_{jk}, \phi_{jk})$.

The proof of Lemmas 2.1 and 2.2 can easily be obtained by the Lemma 4.3 of Bosq (2000).

Lemma 2.3. *Under Assumptions 2.1, 2.3 and 2.6, we have*

$$\max_{1\leqslant j\leqslant J;1\leqslant k\leqslant K_j} \|\widehat{\varphi}_{jk}(s,t)-\varphi_{jk}(s,t)\| = O_p\{(n_1+n_2)^{-1/2}\}$$

The proof of Lemma 2.3 can easily be obtained by Lemmas 2.1 and 2.2.

Proof of Theorem 2.1

Firstly, we prove

$$\begin{aligned} {n_1}^{1/2}\left(\widehat{\eta}_j^{(1)}(t^*)-\eta_j^{(1)}(t^*)\right) &\xrightarrow{d} N\left(0,\lambda_j^{(1)}(t^*)\right), \\ {n_2}^{1/2}\left(\widehat{\eta}_j^{(2)}(t^*)-\eta_j^{(2)}(t^*)\right) &\xrightarrow{d} N\left(0,\lambda_j^{(2)}(t^*)\right), \end{aligned} \tag{2.12}$$

where

$$\lambda_j^{(1)}(t^*) = \int_0^1\int_0^1 \psi_j(s)C^{(1)}\{(s,t^*),(u,t^*)\}\psi_j(u)dsdu,$$

$$\lambda_j^{(2)}(t^*) = \int_0^1\int_0^1 \psi_j(s)C^{(2)}\{(s,t^*),(u,t^*)\}\psi_j(u)dsdu,$$

For the term ${n_1}^{1/2}\left(\widehat{\eta}_j^{(1)}(t^*)-\eta_j^{(1)}(t^*)\right)$, it can be observed that

$$
\begin{aligned}
\widehat{\eta}_j^{(1)}(t^*) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} X_i^{(1)}(s_{il_1},t^*)\psi_j(s_{il_1}) \\
&\quad + \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} X_i^{(1)}(s_{il_1},t^*)\widehat{\psi}_j(s_{il_1})\right. \\
&\quad \left. - \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} X_i^{(1)}(s_{il_1},t^*)\psi_j(s_{il_1})\right\} \\
&\equiv A_1 + A_2.
\end{aligned} \tag{2.13}
$$

For A_1, we have

$$
\begin{aligned}
A_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} \varepsilon_i^{(1)}(s_{il_1},t^*)\psi_j(s_{il_1}) \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} \mu_1(s_{il_1},t^*)\psi_j(s_{il_1}) \\
&\equiv A_{11} + A_{12}.
\end{aligned} \tag{2.14}
$$

It is easy to see that A_{11} is the average of independent and identically distributed random variables with mean $E(A_{11}) = 0$ and variance $\mathrm{var}(A_{11}) = \lambda_j^{(1)}(t^*)$. By the central limit theorem, we obtain

$$
{n_1}^{1/2}A_{11} \xrightarrow{d} N\left(0, \lambda_j^{(1)}(t^*)\right). \tag{2.15}
$$

For A_{12}, according to Assumption 2.2, we have

$$
\begin{aligned}
A_{12} - \eta_j^{(1)}(t^*) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N} \mu_1(s_{il_1},t^*)\psi_j(s_{il_1}) - \int_0^1 \mu_1(s,t^*)\psi_j(s)ds \\
&\equiv O\left(\frac{1}{N}\right).
\end{aligned} \tag{2.16}
$$

Combing (2.13), (2.14), (2.15), (2.16) and Assumption 2.4, we obtain

$$ {n_1}^{1/2}\left(A_1-\eta_{jk}^{(1)}\right)\xrightarrow{d}N\left(0,\lambda_j^{(1)}(t^*)\right). \tag{2.17} $$

For A_2, we have

$$ \begin{aligned} A_2 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N}\mu_1(s_{il_1},t^*)\{\widehat{\psi}_j(s_{il_1})-\psi_j(s_{il_1})\} \\ &+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N}\varepsilon_i^{(1)}(s_{il_1},t^*)\{\widehat{\psi}_j(s_{il_1})-\psi_j(s_{il_1})\} \\ &\equiv A_{21}+A_{22}. \end{aligned} \tag{2.18} $$

It is easy to see

$$ \begin{aligned} E(A_{21}^2) &\leqslant \frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{N^2}\sum_{l_1=1}^{N}\{\mu_1^2(s_{il_1},t^*)\mathrm{E}[\widehat{\psi}_j(s_{il_1})-\psi_j(s_{il_1})]^2\} \\ &+\frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{N(N-1)}\sum_{l_1\neq l_1'}\{\mu_1(s_{il_1},t^*)\mu_1(s_{il_1'},t^*) \\ &\times E[\widehat{\psi}_j(s_{il_1})-\psi_j(s_{il_1})][\widehat{\psi}_j(s_{il_1'})-\psi_j(s_{il_1'})]\} \\ &\equiv A_{211}+A_{212} \end{aligned} $$

For A_{211}, by Assumption 2.3 and Lemma 2.1, we have

$$ \begin{aligned} A_{211} &\leqslant \sup_{(s,t)\in\mathcal{S}\times\mathcal{T}}\mu_1^2(s,t)\frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{N^2}\sum_{l_1=1}^{N}\mathrm{E}[\widehat{\psi}_j(s_{il_1})-\psi_j(s_{il_1})]^2\} \\ &=O\left(\frac{1}{n_1N}\right) \end{aligned} \tag{2.19} $$

According to Cauchy-Schwarz inequality, we obtain

$$ A_{212}=O\left(\frac{1}{n_1N}\right) \tag{2.20} $$

By (2.18), (2.19), (2.20) and Assumption 2.4, we have $A_{21} = o_p(n_1^{-1/2})$. Similarity, we can obtain $A_{22} = o_p(n_1^{-1/2})$.

Combing above discussion, we have

$$A_2 = o_p\left(n_1^{-1/2}\right). \tag{2.21}$$

By (2.13), (2.17) and (2.21), we obtain $n_1^{1/2}\left(\widehat{\eta}_j^{(1)}(t^*) - \eta_j^{(1)}(t^*)\right) \xrightarrow{d} N\left(0, \lambda_j^{(1)}(t^*)\right)$. Similarly, we can prove $n_2^{1/2}\left(\widehat{\eta}_j^{(2)}(t^*) - \eta_{jk}^{(2)}(t^*)\right) \xrightarrow{d} N\left(0, \lambda_j^{(2)}(t^*)\right)$. The proof of (2.12) is completed.

Secondly, we prove

$$\widehat{\lambda}_j^{(1)}(t^*) \xrightarrow{p} \lambda_j^{(1)}(t^*) \quad \widehat{\lambda}_j^{(2)}(t^*) \xrightarrow{p} \lambda_j^{(2)}(t^*). \tag{2.22}$$

It can be observed that

$$\begin{aligned}
\widehat{\lambda}_j^{(1)}(t^*) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N^2}\sum_{l_1=1}^{N}[X_i(s_{l_1}, t^*)\widehat{\psi}_j(s_{l_1}) - \widehat{\eta}_j^{(1)}(t^*)]\sum_{l_1=1}^{N}[X_i(s_{l_1}, t^*)\widehat{\psi}_j(s_{l_1}) - \widehat{\eta}_j^{(1)}(t^*)] \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\sum_{l_1=1}^{N}\sum_{l_1'=1}^{N}\varepsilon_i(s_{l_1}, t^*)\varepsilon_i(s_{l_1'}, t^*)\widehat{\psi}_j(s_{l_1})\widehat{\psi}_j(s_{l_1'})}{N^2} \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{1}{N}\sum_{l_1=1}^{N}\varepsilon_i(s_{l_1}, t^*)\widehat{\phi}_j(s_{l_1})\left[\frac{1}{N}\sum_{l_1'=1}^{N}\mu(s_{l_1'}, t^*)\widehat{\psi}_j(s_{l_1'}) - \widehat{\eta}_j^{(1)}(t^*)\right]\right\} \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n}\left\{\frac{1}{N}\sum_{l_1'=1}^{N}\varepsilon_i(s_{l_1'}, t^*)\widehat{\psi}_j(s_{l_1'})[\frac{1}{N}\sum_{l_1=1}^{N}\mu(s_{l_1}, t^*)\widehat{\psi}_j(s_{l_1}) - \widehat{\eta}_j^{(1)}(t^*)]\right\} \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{1}{N^2}\left[\sum_{l_1=1}^{N} l\mu(s_{l_1}, t^*)\widehat{\psi}_j(s_{l_1}) - \widehat{\eta}_j^{(1)}(t^*)][\sum_{l_1'=1}^{N}\mu(s_{l_1'}, t^*)\widehat{\psi}_{l'}(s_{l_1'}) - \widehat{\eta}_j^{(1)}(t^*)\right]\right\} \\
&\equiv B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

It is easy to see that $B_1 \xrightarrow{p} \lambda_j^{(1)}(t^*)$. Next, we analyze the term B_2. In fact, by (2.12) and Lemma 2.1, we have

$$\frac{1}{N}\sum_{l_1'=1}^{N}\mu(s_{l_1'},t^*)\widehat{\psi}_j(s_{l_1'})-\widehat{\eta}_j^{(1)}(t^*)=o_p(1)$$

According to (2.15) and Lemma 2.1, we have

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N}\sum_{l_1=1}^{N}\varepsilon_i(s_{l_1},t^*)\widehat{\phi}_j(s_{l_1})=O_p(1).$$

Then, we have $B_2 = o_p(1)$. Using the arguments similar to that of B_2, we have

$$B_3 = o_p(1).$$

Similarity, we can prove

$$B_4 = o_p(1).$$

So, the proof of (2.22) is completed. By (2.12) and (2.22), together with Slutsky's lemma, the firstly part of Theorem 2.1 have been proved.

Next, we prove $\widehat{TP}(t^*) \xrightarrow{p} \infty$ under $H_A^{\mathcal{S}}$. According to the results in the above proof, we have for $j = 1,\dots,J$,

$$\widehat{\eta}_j^{(1)}(t^*) \xrightarrow{p} \eta_j^{(1)}(t^*), \widehat{\eta}_j^{(2)}(t^*) \xrightarrow{p} \eta_j^{(2)}(t^*).$$

Then it yields that

$$\sum_{j=1}^{J}\frac{[\widehat{\eta}_j^{(1)}(t^*)-\widehat{\eta}_j^{(2)}(t^*)}{\widehat{\lambda}_j(t^*)}-\sum_{j=1}^{J}\frac{[\eta_j^{(1)}(t^*)-\eta_j^{(1)}(t^*)]^2}{\lambda_j(t^*)} \xrightarrow{p} 0.$$

Therefore, under $H_A^{\mathcal{S}}$ and Assumption 2.5, we have

$$\widehat{TP}(t^*)=\frac{n_1n_2}{n_1+n_2}\sum_{j=1}^{J}\frac{(\widehat{\eta}_j^{(1)}(t^*)-\widehat{\eta}_j^{(2)}(t^*))^2}{\widehat{\lambda}_j(t^*)} \xrightarrow{p} \frac{n_1n_2}{n_1+n_2}\sum_{j=1}^{J}\frac{(\eta_j^{(1)}(t^*)-\eta_j^{(2)}(t^*))^2}{\lambda_j(t^*)} \to \infty.$$

Then, Theorem 2.1 is proved.

Proof of Theorem 2

The proof of Theorem 2.2 is similar to that of Theorem 2.1, and so we only outline the main section. If we can prove

$$n_1{}^{1/2}\left(\widehat{\eta}_{jk}^{(1)}-\eta_{jk}^{(1)}\right)\xrightarrow{d}N\left(0,\lambda_{jk}^{(1)}\right),\ n_2{}^{1/2}\left(\widehat{\eta}_{jk}^{(2)}-\eta_{jk}^{(2)}\right)\xrightarrow{d}N\left(0,\lambda_{jk}^{(2)}\right),\tag{2.23}$$

where

$$\lambda_{jk}^{(1)}=\int_0^1\int_0^1\int_0^1\int_0^1\varphi_{jk}(s,v)C^{(1)}\{(s,u),(v,t)\}\varphi_{jk}(u,t)dsdudvdt,$$

$$\lambda_{jk}^{(2)}=\int_0^1\int_0^1\int_0^1\int_0^1\varphi_{jk}(s,v)C^{(2)}\{(s,u),(v,t)\}\varphi_{jk}(u,t)dsdudvdt,$$

then together with Slutsky's lemma, Theorem 2.2 can be easily proved.

For the term $n_1{}^{1/2}\left(\widehat{\eta}_{jk}^{(1)}-\eta_{jk}^{(1)}\right)$, it can be observed that

$$\begin{aligned}\widehat{\eta}_{jk}^{(1)}&=\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}X_i^{(1)}(s_{il_1},t_{il_2})\varphi_{jk}(s_{il_1},t_{il_2})\\&\quad+\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}X_i^{(1)}(s_{il_1},t_{il_2})\widehat{\varphi}_{jk}(s_{il_1},t_{il_2})\right.\\&\quad\left.-\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}X_i^{(1)}(s_{il_1},t_{il_2})\varphi_{jk}(s_{il_1},t_{il_2})\right\}\\&\equiv D_1+D_2.\end{aligned}\tag{2.24}$$

For D_1, we have

$$\begin{aligned}D_1&=\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\varepsilon_i^{(1)}(s_{il_1},t_{il_2})\varphi_{jk}(s_{il_1},t_{il_2})\\&\quad+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\mu_1(s_{il_1},t_{il_2})\varphi_{jk}(s_{il_1},t_{il_2})\\&\equiv D_{11}+D_{12}.\end{aligned}\tag{2.25}$$

It is easy to see that D_{11} is the average of independent and identically distributed random variables with mean $E(D_{11}) = 0$ and variance $\mathrm{var}(D_{11}) = \lambda_{jk}^{(1)}$. By the central limit theorem, we obtain

$$n_1{}^{1/2} D_{11} \xrightarrow{d} N\left(0, \lambda_{jk}^{(1)}\right). \tag{2.26}$$

For D_{12}, according to Assumption 2.7, we have

$$\begin{aligned} D_{12} - \eta_{jk}^{(1)} &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\mu_1(s_{il_1}, t_{il_2})\varphi_{jk}(s_{il_1}, t_{il_2}) - \int_0^1\int_0^1 \mu_1(s,t)\varphi_{jk}(s,t)dsdt \\ &\equiv O\left(\frac{1}{MN}\right). \end{aligned} \tag{2.27}$$

Combing (2.25), (2.26), (2.27) and Assumption 2.4, we obtain

$$n_1{}^{1/2}\left(D_1 - \eta_{jk}^{(1)}\right) \xrightarrow{d} N\left(0, \lambda_{jk}^{(1)}\right). \tag{2.28}$$

For D_2, we have

$$\begin{aligned} A_2 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\mu_1(s_{il_1}, t_{il_2})\{\hat{\varphi}_{jk}(s_{il_1}, t_{il_2}) - \varphi_{jk}(s_{il_1}, t_{il_2})\} \\ &+ \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{MN}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\varepsilon_i^{(1)}(s_{il_1}, t_{il_2})\{\hat{\varphi}_{jk}(s_{il_1}, t_{il_2}) - \varphi_{jk}(s_{il_1}, t_{il_2})\} \\ &\equiv A_{21} + A_{22}. \end{aligned} \tag{2.29}$$

It is easy to see

$$\begin{aligned} E(A_{21}^2) &\leqslant \frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{M^2N^2}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}\{\mu_1^2(s_{il_1}, t_{il_2})\mathrm{E}[\hat{\varphi}_{jk}(s_{il_1}, t_{il_2}) - \varphi_{jk}(s_{il_1}, t_{il_2})]^2\} \\ &+ \frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{M(M-1)N(N-1)}\sum_{l_2\neq l_2'}\sum_{l_1\neq l_1'}\{\mu_1(s_{il_1}, t_{il_2})\mu_1(s_{il_1'}, t_{il_2'}) \\ &\times E[\hat{\varphi}_{jk}(s_{il_1}, t_{il_2}) - \varphi_{jk}(s_{il_1}, t_{il_2})][\hat{\varphi}_{jk}(s_{il_1'}, t_{il_2'}) - \varphi_{jk}(s_{il_1'}, t_{il_2'})]\} \\ &\equiv A_{211} + A_{212} \end{aligned} \tag{2.30}$$

For A_{211}, by Assumption 2.4 and Lemma 2.3, we have

$$\begin{aligned} A_{211} &\leqslant \sup_{(s,t)\in\mathcal{S}\times\mathcal{T}} \mu_1^2(s,t)\frac{c}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{M^2N^2}\sum_{l_2=1}^{M}\sum_{l_1=1}^{N}E[\hat{\varphi}_{jk}(s_{il_1},t_{il_2})-\varphi_{jk}(s_{il_1},t_{il_2})]^2\} \\ &= O\left(\frac{1}{n_1MN}\right) \end{aligned} \tag{2.31}$$

According to Cauchy-Schwarz inequality, we obtain

$$A_{212} = O\left(\frac{1}{n_1MN}\right) \tag{2.32}$$

By (2.30), (2.31), (2.32) and Assumption 2.4, we have $A_{21} = o_p(n_1^{-1/2})$. Similarly, we can obtain $A_{22} = o_p(n_1^{-1/2})$.

Combing the above discussions, we have

$$A_2 = o_p\left(n_1^{-1/2}\right). \tag{2.33}$$

By (2.24), (2.28) and (2.33), we obtain $n_1^{1/2}\left(\widehat{\eta}_{jk}^{(1)} - \eta_{jk}^{(1)}\right) \xrightarrow{d} N\left(0, \lambda_{jk}^{(1)}\right)$. Similarly, we can prove $n_2^{1/2}\left(\widehat{\eta}_{jk}^{(2)} - \eta_{jk}^{(2)}\right) \xrightarrow{d} N\left(0, \lambda_{jk}^{(2)}\right)$. The proof of (2.23) is then completed.

According to (2.23), we have

$$\hat{\eta}_{jk}^{(1)} \xrightarrow{p} \eta_{jk}^{(1)},\ \hat{\eta}_{jk}^{(2)} \xrightarrow{p} \eta_{jk}^{(2)}.$$

Under H_1, we obtain

$$\widehat{TM} \xrightarrow{p} \frac{n_1n_2}{n_1+n_2}\sum_{j=1}^{J}\sum_{k=1}^{K_j}\frac{\left(\eta_{jk}^{(1)}-\eta_{jk}^{(2)}\right)^2}{\lambda_{jk}} \xrightarrow{p} \infty.$$

Then, Theorem 2.2 is proved.

Chapter 3

Testing Equality of Mean Curve for Two-sample Functional Data

3.1 Introduction

Over the last two decades, functional data analysis has established itself as an important and dynamic area of statistics. It offers effective new tools and has stimulated new methodological and theoretical developments. The field has become very broad and specialized directions of research. Many areas of functional data analysis have been developing rapidly over the last decade. For a summary of some of these developments, we refer to Ramsay and Silverman (2005) and Ferraty and Vieu (2006). More recently, see Ferraty and Romain (2011) and Horváth and Kokoszka (2012).

Functional data which are referred to as curve data in the early days was pioneered by Castro et al. (1986) and was further developed in Rice and Silverman (1991). Functional Data Analysis (FDA) dealing with curve data is concerned for the data that are repeated measurements of the same subject. The repeated measurements are often recorded over a period of time, say on an interval $\mathcal{T}$. Generally, there exist two different approaches to treating them, depending on whether the measurements are available on a dense grid of time points, or whether they are recorded relatively sparsely. Dense functional data allow the number of observations for every subject tending to infinity and a conventional estimation approach is to smooth each individual curve and then infer. For a summary of some of these development, we refer to Hall et al. (2006) and Zhang and Chen (2007). In the case of sparse functional data, every subject is often observed at a small number of time points, and often irregularly, spaced measurements on human or other biological subjects, they are typically termed longitudinal data. For an introduction to this area, see Yao et al. (2005), Hall et al. (2006), Yao (2007), and Ma et al. (2012).

For dense or sparse functional data, a lot of regression models have been extensively studied. For example, functional linear model (Yuan and Cai, 2010; He et al., 2010; Hall and Horowitz, 2007; Lee and Park, 2012), functional nonparametric model (Ferraty and Vieu, 2002, 2006; Ferraty et al., 2012), Semiparametric functional model (Chen et al., 2011a; Chiou et al., 2003; Jiang and Wang, 2011). However, most of inferential procedures based on it assume that mean function is the same for all subjects. If, in fact, mean functions are different, the results of inference may be confounded. So, it is important to consider the two-sample or multi-sample problems of functional data. Despite the above problem are important for functional data, they have received little attention.

In the setting of dense functional data, Horváth et al. (2009) compared linear operators in two functional regression models. Horváth et al. (2013) developed and asymptotically justified testing procedures for the equality of means in two functional samples exhibiting temporal dependence. Fremdt et al. (2014) considered a normal approximation method to derive statistics that used segments of observations and segments of the FPC's and then applied results to derive inferential procedures for the mean function. However, all these research are based on the assumption which the repeated measurements take place on the dense and regular time points for each subject. In the setting of sparse functional data, less attention has been paid to this area.

In practice, it is hard to decide when the observations are dense or sparse. In some functional studies it is possible that we have dense observations on some subjects and sparse observations on the others. It is thus useful to develop a unified methodology which can test if two sample or multi-sample have the same mean function for functional data no matter they are dense or sparse. A direct motivation for the research of this chapter comes from a two-sample problem in which we wish to test whether the mean functions of two functional observation sample are equal without the information that the data are dense or sparse.

We propose a significance test for testing the null hypothesis of having the same mean function against the alternative of different mean functions. One particular advantage of the proposed method is that we do not have to discern data type: dense or sparse.

This chapter is organized as follows. In Section 3.2, we present the proposed testing method. Asymptotic theory of the proposed procedures are also developed in this section. While Section 3.3 is devoted to a report on simulation results. In Section 3.4, we analyze two real data sets to illustrate the proposed procedures. All proofs are displayed to Section 3.5.

3.2 Methodology and main results

Consider two independent samples:

$$\begin{aligned} &Y_i^{(m)}(t_{il_m}) = \mu_m(t_{il_m}) + v_i^{(m)}(t_{il_m}) + \varepsilon_{il_m}^{(m)}, \\ &m = 1,2;\ i = 1,\cdots,n_m;\ l_m = 1,\cdots,N_i;\ t \in \mathcal{T}, \end{aligned} \tag{3.1}$$

where $\mu_m(t),\ m = 1,2,$ are the fixed population means of $Y_i^{(m)}(t),\ m = 1,2.$ $v_i^{(m)}(t),\ m = 1,2,$ are the subject-specific random trajectories of $Y_i^{(m)}(t)$ with $E\{v_i^{(m)}(t)\} = 0$ and covariance function $\gamma_m(t,s) = \mathrm{cov}\{v_i^{(m)}(t), v_i^{(m)}(s)\}$. N_{n_1} and N_{n_2} are the number of measurements collected from two subjects, refer to Horváth et al. (2013). $\varepsilon_i(t)$s are i.i.d. random error process independent of $v_i(t)$, refer to Shi et al. (1996), Zhang and Chen (2007), and Horváth and Kokoszka (2012), among others.

In this chapter, we want to test if two samples have the same mean function. Thus, we are interested in testing

$$H_0 : \mu_1(t) = \mu_2(t) \text{ vs. } H_1 : \ \mu_1(t) \neq \mu_2(t), \ t \in \mathcal{T}.$$

To perform the test, we first estimate the mean functions. The following is the procedure to obtain the estimators.

Step 3.1. *Obtain initial estimators of mean functions $\mu_1(t)$ and $\mu_2(t)$.*

To estimate the mean function $\mu_1(t)$ by local linear scatterplot smoothers, one minimizes

$$\sum_{i=1}^{n_1} \sum_{l_1=1}^{N_i} K^{(1)} \left(\frac{t_{il_1} - t}{b_0^*} \right) \left\{ Y_i^{(1)}(t_{il_1}) - d_0 - d_1(t_{il_1} - t) \right\}^2,$$

with respect to d_0 and d_1 to obtain $\hat{\mu}_1^0(t) = \hat{d}_0(t)$, where the kernel $K^{(1)}(\cdot)$ is assumed to be a smooth symmetric density function and b_0^* is a bandwidth. Analogously, one may define the estimator of the mean function $\mu_2(t)$, say $\hat{\mu}_2^0(t)$.

Step 3.2. *Obtain the estimation of covariances functions $\gamma_1(t,s)$ and $\gamma_2(t,s)$.*

Let $G_{1,i}(t_{il_{11}}, t_{il_{12}}) = \{Y_i^{(1)}(t_{il_{11}}) - \hat{\mu}_1^0(t_{il_{11}})\}\{Y_i^{(1)}(t_{il_{12}}) - \hat{\mu}_1^0(t_{il_{12}})\}$. Define the local linear surface smoother for $\gamma_1(t,s)$ by minimizing

$$\sum_{i=1}^{n_1} \sum_{1 \leqslant l_{11} \neq l_{12} \leqslant N_i} K^{(2)} \left(\frac{t_{il_{11}} - t}{h_Y^*}, \frac{t_{il_{12}} - s}{h_Y^*} \right) \{ G_{1,i}(t_{il_{11}}, t_{il_{12}}) - f(\alpha, (t,s), (t_{il_{11}}, t_{il_{12}})) \}^2,$$

with respect to $\alpha = (\alpha_0, \alpha_{11}, \alpha_{12})$ where $f(\alpha, (t,s), (t_{il_{11}}, t_{il_{12}})) = \alpha_0 + \alpha_{11}(t - t_{il_{11}}) + \alpha_{12}(s - t_{il_{12}})$, yielding $\hat{\gamma}_1^0(t,s) = \hat{\alpha}_0(t,s)$. Here, the kernel $K^{(2)}$ is a two-dimensional smooth density with zero mean and finite covariance and h_Y^* is a bandwidth. An essential feature is the omission of the diagonal elements $l_{11} = l_{12}$ which are contaminated with the measurement errors. Analogously, we can obtain the estimator of the covariance function $\gamma_2(t, s)$, say $\hat{\gamma}_2^0(t, s)$. Let $\hat{\gamma}^0(t, s) = \frac{n_2}{n_1+n_2}\hat{\gamma}_1^0(t, s) + \frac{n_1}{n_1+n_2}\hat{\gamma}_2^0(t, s)$. It is easy to see $\hat{\gamma}^0(t, s) \xrightarrow{p} \gamma(t, s) \equiv (1-\theta)\gamma_1(t, s) + \theta\gamma_2(t, s)$ and $\gamma(t, s)$ is a covariance function where θ is defined in the following assumption 3.5. Consequently, it has orthonormal eigenfunctions $\{\phi_j\}_{j\geqslant 1}$ and non-negative eigenvalues $\{\nu_j\}_{j\geqslant 1}$ satisfying:

$$\int_{\mathcal{T}} \gamma(t, s)\phi_j(s)ds = \nu_j\phi_j(t).$$

Step 3.3. *Estimates the eigenvalues and eigenfunctions $\{\nu_j, \phi_j\}_{j\geqslant 1}$ of $\gamma(t, s)$.*

Estimation of eigenvalues and eigenfunctions $\{\nu_j, \phi_j\}_{j\geqslant 1}$ are obtained by numerical solutions $\{\hat{\nu}_j, \hat{\phi}_j\}_{j\geqslant 1}$ of the following suitably discretized eigenequations,

$$\int_{\mathcal{T}} \hat{\gamma}^0(t, s)\hat{\phi}_j(s)ds = \hat{\nu}_j\hat{\phi}_j(t),$$

with orthonormal constraints on $\{\hat{\phi}_j\}_{j\geqslant 1}$.

In order to obtain final estimators of $\mu_m(t),\ m = 1, 2$, we make the following assumption.

Assumption 3.1. *Assume $\int_0^1 \mu_m^2(t)dt < \infty,\ m = 1, 2$ and $\mu_m(t)$ may be written as $\mu_m(t) = \sum_{j=1}^\infty \eta_j^{(m)}\phi_j(t)$ where $\eta_j^{(m)} = \int_0^1 \mu_m(t)\phi_j(t)dt$.*

Step 3.4. *Obtain the projection estimator onto eigenfunctions.*

Estimator of the mean function is a projection estimator onto a space that is generated by a set of eigenfunctions,

$$\hat{\mu}_1(t) = \sum_{j=1}^J \hat{\eta}_j^{(1)}\hat{\phi}_j(t),$$

where

$$\hat{\eta}_j^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\eta}_{ij}^{(1)},\ \hat{\eta}_{ij}^{(1)} = \frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_i^{(1)}(t_{il_1})\hat{\phi}_j(t_{il_1}),$$

with $\hat{\phi}_j(t),\ j = 1, \cdots, J$ are the eigenfunctions. The number J is a tuning parameter. A practical strategy to select J will be discussed in Section 3.3.

Similarly, we can define

$$\hat{\mu}_2(t) = \sum_{j=1}^{J}\hat{\eta}_j^{(2)}\hat{\phi}_j(t),$$

where

$$\hat{\eta}_j^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\eta}_{ij}^{(2)},\ \hat{\eta}_{ij}^{(2)} = \frac{1}{N_i}\sum_{l_2=1}^{N_i} Y_i^{(2)}(t_{il_2})\hat{\phi}_j(t_{il_2}).$$

The difference between our estimating approach and that of Fremdt et al. (2014) is typically that every curve of the proposed method is directly projected on the common basis function space and obtain $\hat{\eta}_j^{(1)}$ and $\hat{\eta}_j^{(2)}$ which are the means of projection. $\hat{\eta}_j^{(1)}$ and $\hat{\eta}_j^{(2)}$ can be viewed as the scores of projection that two mean functions $\mu_1(t)$ and $\mu_2(t)$ project on the basis function space, respectively. Advantages of this method lies in two aspects. On the one hand, global statistics can be provided from the expression of $\mu_1(t)$ and $\mu_2(t)$. On the other hand, whatever functional data is sparse or dense, regular or irregular, the means of projection can be always obtained from the expressions of $\hat{\eta}_j^{(1)}$ and $\hat{\eta}_j^{(2)}$ which lead to the wide applicability of our tests.

It is natural to take an empirical version of the integrated square deviation between two mean curves $\int_{\mathcal{T}}\{\mu_1(t) - \mu_2(t)\}^2 dt$ to measure the distance between two estimated mean curves. Consider the statistic

$$\mathrm{TM} = \frac{n_1 n_2}{n_1 + n_2}\int_{\mathcal{T}}\{\hat{\mu}_1(t) - \hat{\mu}_2(t)\}^2\, dt, \tag{3.2}$$

H_0 will be rejected if TM is large.

From lemma 3.1 in Section 3.5, it is easy to see that

$$
\begin{aligned}
\mathrm{TM} &= \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{T}}\left\{\sum_{j=1}^{J}\hat{\eta}_j^{(1)}\hat{\phi}_j(t)-\sum_{j=1}^{J}\hat{\eta}_j^{(2)}\hat{\phi}_j(t)\right\}^2 dt \\
&= \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{T}}\left[\sum_{j=1}^{J}\left\{\hat{\eta}_j^{(1)}-\hat{\eta}_j^{(2)}\right\}\right]^2 \phi_j^2(t)dt \\
&\quad + \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{T}}\left[\sum_{j=1}^{J}\left\{\hat{\eta}_j^{(1)}-\hat{\eta}_j^{(2)}\right\}\right]^2 \left\{\hat{\phi}_j^2(t)-\phi_j^2(t)\right\} dt \\
&\equiv \frac{n_1 n_2}{n_1+n_2}\sum_{j=1}^{J}\left\{\hat{\eta}_j^{(1)}-\hat{\eta}_j^{(2)}\right\}^2 + o_p(1).
\end{aligned}
$$

We consider the following test statistic for testing the hypothesis $H_0 : \mu_1(t) = \mu_2(t),\ t \in \mathcal{T}$,

$$
\widehat{\mathrm{TM}} = \left[\frac{n_1 n_2}{n_1+n_2}\sum_{j=1}^{J}\frac{\left\{\hat{\eta}_j^{(1)}-\hat{\eta}_j^{(2)}\right\}^2}{\hat{\lambda}_j} - J\right] \Big/ \sqrt{2J}
$$

where $\hat{\lambda}_j = \frac{n_2}{n_1+n_2}\hat{\lambda}_j^{(1)} + \frac{n_1}{n_1+n_2}\hat{\lambda}_j^{(2)}$ with

$$
\hat{\lambda}_j^{(1)} = \frac{1}{n_1-1}\sum_{i=1}^{n_1}\left\{\hat{\eta}_{ij}^{(1)}-\hat{\eta}_j^{(1)}\right\}^2
$$

$$
\hat{\lambda}_j^{(2)} = \frac{1}{n_2-1}\sum_{i=1}^{n_2}\left\{\hat{\eta}_{ij}^{(2)}-\hat{\eta}_j^{(2)}\right\}^2.
$$

We demand the following assumptions before showing the theorems 3.1 and 3.2.

Assumption 3.2. *There exists positive constant C and $\alpha > 1$ such that*

$$
\nu_j - \nu_{j+1} \geqslant C j^{-\alpha-1},
$$

where $\{\nu_j\}_{j\leqslant 1}$ are the eigenvalues of covariance function $\gamma(t,s)$.

Assumption 3.3. *$n_m^{-1}J^{4\alpha+4} \to 0$, where $m = 1, 2$.*

Assumption 3.4. *$\{v_i^{(1)}(\cdot)\}_i$, $\{t_{il_1}\}_{i,l_1}$ and $\{\varepsilon_{il_1}\}_{i,l_1}$ are independent and identically distributed and mutually independent. So are $\{v_i^{(2)}(\cdot)\}_i$, $\{t_{il_2}\}_{i,l_2}$ and $\{\varepsilon_{il_2}\}_{i,l_2}$.*

Assumption 3.5. *We assume $\tau = E(1/n_1) = E(1/n_2)$ and $\frac{n_1}{n_1+n_2} \to \theta$ for some $0 < \theta < 1$ as $\min(n_1, n_2) \to \infty$.*

Assumption 3.2 requires that the spacings between the eigenvalues are not too small. It implies that each ν_j is greater than a constant multiple of $j^{-\alpha}$. One needs this condition to get the bound of $\hat{\phi}_j(t) - \phi_j(t)$. Assumption 3.3 requires that the number of principle component for two samples are not too large since $J = \max(J_1, J_2)$, where J_1 and J_2 will be described in Section 3.3. Assumption 3.4 is a regular condition and assumption 3.5 requires that the type of the observations for two samples are same.

Theorems 3.1 and 3.2 shown below establish the asymptotic behaviors of the statistic $\widehat{\mathrm{TM}}$ under hypotheses H_0 and H_1, respectively. The proofs of these theorems are provided in Section 3.5.

Theorem 3.1. *Under assumptions 3.1-3.5 and H_0, we have*

$$\widehat{\mathrm{TM}} \xrightarrow{d} N(0, 1)$$

The null hypothesis $H_0 : \mu_1(t) = \mu_2(t),\ t \in \mathcal{T}$ is rejected if $|\widehat{\mathrm{TM}}| > q_\alpha$, where q_α is the upper-α quantile of $N(0, 1)$.

Theorem 3.2. *Under assumptions 3.1-3.5 and H_1, we have*

$$\widehat{\mathrm{TM}} \xrightarrow{p} \infty.$$

Both Theorems 3.1 and 3.2 provide clear theoretical justification of the empirical properties of the proposed test.

3.3 Simulation studies

To evaluate the finite sample performances of the proposed test method, we conducted some simulation studies.

3.3.1 Tuning parameter selection

For practical implementation, one has to decide the values of the tuning parameters J. We can select J using the following method, Firstly, we select J_1 as the minimum number of FPCs that explain 99% of the total variation for the sample one by the PACE (principal analysis by conditional estimation) package and J_2 for sample two. In detail, we refer to Yao et al. (2005) for a complete description. Then, we select $J = \max(J_1, J_2)$. Using the asymptotic developed in this chapter, selecting J is not essential. We can observe that different J gives the similarly conclusion on the empirical size and power in the following simulation.

The choice of bandwidth is a very important topic in nonparametric regression estimation for our Steps 3.1 and 3.2. The popular method such as cross-validation, generalized cross-validation (GCV) and the rule of thumb can be used to select the optimal bandwidth for the estimators of $\mu_m(t)$ and $\gamma_m(t, s)$, $m = 1, 2$. Here, we recommend using GCV to determine the optimal bandwidth.

3.3.2 Test of mean function

We consider combinations of sample sizes $(n_1, n_2) = (100, 100)$ and $(200, 200)$, each pair of data-generated processes was replicated 1000 times. In this section, $v(t)$ was generated from

$$v(t) = \sum_{j=1}^{2} \xi_j \phi_j(t),$$

with $\phi_1(t) = -\sqrt{2}\cos(2\pi t)$, $\phi_2(t) = \sqrt{2}\sin(2\pi t)$, $t \in [0,1]$. ξ_j, $j = 1,2$ were generated from $N(0,\nu_k)$ with $\nu_1 = 1$ and $\nu_2 = 0.5$.

Example 3.1. *Sparse design with the same covariance function.*

To illustrate the adaptivity of our test method to sparse design, we firstly consider the following model

$$\begin{aligned} Y_i^{(1)}(t_{il_1}) &= v_i^{(1)}(t_{il_1}) + \varepsilon_{il_1}^{(1)}, \ i = 1,\cdots,n_1; \ l_1 = 1,\cdots,N_i, \\ Y_i^{(2)}(t_{il_2}) &= at_{il_2} + v_i^{(2)}(t_{il_2}) + \varepsilon_{il_2}^{(2)}, \ i = 1,\cdots,n_2; \ l_2 = 1,\cdots,N_i, \end{aligned} \tag{3.3}$$

where $v_i^{(m)}(t)$, $m = 1,2$ were generated from $v(t)$. The parameter a regulates the difference between the means of two samples. The number N_{n_m}, $m = 1,2$ of measurements for each curve were selected from $\{5,\cdots,9\}$ with equal probability in $[0,1]$. The measurement error is $\varepsilon_{il_1}^{(1)} \sim N(0,1)$, so is $\varepsilon_{il_2}^{(2)}s$. Model (3.3) can be seen to be sparse design with the same covariance function $\gamma_1(t,s) = \mathrm{cov}(v_i^{(1)}(t), v_i^{(1)}(s)) = \mathrm{cov}(v_i^{(2)}(t), v_i^{(2)}(s)) = \gamma_2(t,s)$ in this example. The empirical sizes can be calculated when $a = 0$ and the empirical power can be calculated when $a \neq 0$. The empirical size and power of the test are reported in Tables 3.1-3.3.

Example 3.2. *Sparse design with different covariance functions.*

In this example, we consider the sparse design with different covariance functions for comparing with Example 3.1. We consider Model (3.3) except $v_i^{(2)}(t)$ was generated from

$$v^{(2)}(t) = \sum_{j=1}^{2} \xi_j^{(2)} \phi_j(s),$$

with $\xi_j^{(2)}$ generated from $N(0,\nu_j)$ for $j = 1$ and 2 with $\nu_1 = 2$ and $\nu_2 = 1$. The number of measurements for each subject and the measurement error are the same

Table 3.1: The empirical sizes of the test in Examples 3.1 and 3.2.

J	$(\boldsymbol{n_1, n_2}) = (\mathbf{100, 100})$			$(\boldsymbol{n_1, n_2}) = (\mathbf{200, 200})$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			Example 3.1			
2	0.031	0.049	0.067	0.025	0.039	0.062
3	0.017	0.039	0.060	0.026	0.049	0.068
4	0.023	0.043	0.063	0.020	0.043	0.061
5	0.02	0.049	0.066	0.015	0.041	0.056
6	0.018	0.038	0.053	0.021	0.045	0.061
7	0.022	0.044	0.068	0.018	0.044	0.071
8	0.019	0.052	0.084	0.023	0.039	0.064
			Example 3.2			
2	0.027	0.049	0.066	0.031	0.063	0.082
3	0.019	0.035	0.056	0.019	0.040	0.064
4	0.015	0.039	0.057	0.019	0.039	0.063
5	0.016	0.040	0.056	0.013	0.034	0.059
6	0.023	0.042	0.065	0.024	0.048	0.073
7	0.018	0.036	0.076	0.020	0.045	0.066
8	0.014	0.038	0.055	0.015	0.037	0.065

as in Example 3.1. This example can be seen to be the sparse design with the different covariance functions $r_1(t,s) = \mathrm{cov}(v_i^{(1)}(t), v_i^{(1)}(s))$ and $r_2(t,s) = \mathrm{cov}(v_i^{(2)}(t), v_i^{(2)}(s))$. The empirical size and power of the test are also reported in Tables 3.1-3.3.

Several observations can be made from Tables 3.1-3.3. Firstly, the empirical size does not depend on J at all level from Table 3.1. The test based on proposed method has asymptotically correct empirical size at the 5% level, overrejects by about 3% at the 10% level and slightly higher than nominal (about 1% at 1% level). Secondly, when a increases from 1 to 1.5 and $J > 3$, the empirical power of the test does not depend on J at all level and become large from Tables 3.2 and 3.3, which is expected. Thirdly, from the simulations of Examples 3.1 and 3.2, we find that the empirical power of the test increases as the sample size increases. Lastly, the empirical power for the same covariance case is slightly larger than that for the case of different covariance functions. All in all, the proposed method work well for the sparse design even in the case of different covariance functions.

Example 3.3. *Dense design with same/different covariance functions.*

Table 3.2: The empirical power of the test for $a = 1$.

J	$(n_1, n_2) = (100, 100)$			$(n_1, n_2) = (200, 200)$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			Example 3.1			
2	0.311	0.417	0.477	0.619	0.717	0.756
3	0.682	0.762	0.794	0.887	0.914	0.929
4	0.797	0.845	0.879	0.946	0.962	0.971
5	0.780	0.841	0.858	0.930	0.954	0.965
6	0.770	0.822	0.846	0.936	0.960	0.972
7	0.760	0.812	0.839	0.926	0.952	0.964
8	0.738	0.802	0.830	0.936	0.960	0.970
			Example 3.2			
2	0.229	0.297	0.340	0.456	0.550	0.595
3	0.557	0.638	0.679	0.807	0.857	0.875
4	0.691	0.765	0.809	0.876	0.908	0.924
5	0.649	0.730	0.772	0.892	0.929	0.946
6	0.639	0.729	0.763	0.887	0.921	0.940
7	0.634	0.710	0.756	0.865	0.894	0.911
8	0.586	0.673	0.721	0.851	0.908	0.923

In order to compare the proposed method (denoted by $\widehat{\mathrm{TM}}$) with the testing method $\widehat{\mathrm{TF}}$ due to Fremdt et al. (2014) and evaluate the influence of the number of measurements for testing, we consider the following model

$$\begin{aligned} Y_i^{(1)}(t_{il_1}) &= v_i^{(1)}(t_{il_1}) + \varepsilon_{il_1}^{(1)},\ i = 1, \cdots, n_1;\ l_1 = 1, \cdots, N_i, \\ Y_i^{(2)}(t_{il_2}) &= at_{il_2}(1 - t_{il_2}) + v_i^{(2)}(t_{il_2}) + \varepsilon_{il_2}^{(2)},\ i = 1, \cdots, n_2;\ l_2 = 1, \cdots, N_i, \end{aligned} \tag{3.4}$$

where $v_i^{(m)}(t)$, $m = 1, 2$ were generated from standard Brownian motions. The locations of measurements for each curve were selected at 100 equidistant time points in $[0, 1]$. We compute the empirical size ($a = 0$) and power ($a = 1$ and $a = 1.5$) of the test for sample sizes $(n_1, n_2) = (100, 100)$ and $(200, 200)$ in Tables 3.4-3.6.

Table 3.3: The empirical power of the test for $a = 1.5$.

J	$(\boldsymbol{n_1, n_2}) = (\mathbf{100, 100})$			$(\boldsymbol{n_1, n_2}) = (\mathbf{200, 200})$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			Example 3.1			
2	0.642	0.718	0.752	0.916	0.949	0.964
3	0.889	0.920	0.932	0.986	0.994	0.994
4	0.928	0.950	0.959	0.991	0.995	0.997
5	0.940	0.959	0.965	0.994	0.999	0.999
6	0.945	0.967	0.972	0.991	0.994	0.996
7	0.934	0.959	0.971	0.992	0.995	0.996
8	0.922	0.952	0.960	0.990	0.996	0.996
			Example 3.2			
2	0.492	0.597	0.649	0.802	0.848	0.875
3	0.810	0.848	0.875	0.958	0.977	0.984
4	0.879	0.902	0.915	0.976	0.984	0.986
5	0.896	0.919	0.938	0.978	0.984	0.988
6	0.870	0.909	0.931	0.969	0.982	0.987
7	0.861	0.907	0.930	0.977	0.985	0.988
8	0.859	0.903	0.922	0.964	0.977	0.983

Table 3.4: The empirical sizes of the test in Example 3.3.

J		$\widehat{\mathrm{TM}}$			$\widehat{\mathrm{TF}}$	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			$(\boldsymbol{n_1, n_2}) = (\mathbf{100, 100})$			
2	0.030	0.054	0.074	0.032	0.058	0.075
3	0.033	0.063	0.083	0.035	0.063	0.086
4	0.018	0.042	0.070	0.020	0.042	0.072
5	0.023	0.052	0.071	0.025	0.056	0.074
6	0.020	0.043	0.071	0.020	0.044	0.073
7	0.024	0.046	0.073	0.025	0.050	0.076
8	0.019	0.038	0.075	0.019	0.043	0.074
			$(\boldsymbol{n_1, n_2}) = (\mathbf{200, 200})$			
2	0.036	0.057	0.069	0.038	0.058	0.073
3	0.017	0.047	0.066	0.018	0.047	0.067
4	0.022	0.047	0.075	0.022	0.049	0.078
5	0.018	0.041	0.066	0.018	0.043	0.068
6	0.021	0.050	0.071	0.021	0.051	0.074
7	0.020	0.051	0.076	0.024	0.054	0.079
8	0.019	0.056	0.081	0.021	0.060	0.083

From Tables 3.4-3.6, we can see that both the tests based on $\widehat{\mathrm{TM}}$ and $\widehat{\mathrm{TF}}$ can control the type I error and do not depend on J. We can also see that the empirical powers based on $\widehat{\mathrm{TM}}$ and $\widehat{\mathrm{TF}}$ are comparable in the setting of dense design. In a word, the proposed method works well for the case of the dense design.

Table 3.5: The empirical power of the test for $a = 1$ in Example 3.3.

J		$\widehat{\mathrm{TM}}$			$\widehat{\mathrm{TF}}$	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			$(\boldsymbol{n_1, n_2}) = (\mathbf{100, 100})$			
2	0.841	0.884	0.911	0.843	0.886	0.917
3	0.826	0.883	0.899	0.826	0.884	0.899
4	0.807	0.872	0.900	0.813	0.875	0.902
5	0.763	0.841	0.881	0.766	0.845	0.885
6	0.762	0.848	0.883	0.771	0.852	0.886
7	0.735	0.814	0.852	0.745	0.819	0.853
8	0.749	0.836	0.877	0.756	0.840	0.881
			$(\boldsymbol{n_1, n_2}) = (\mathbf{200, 200})$			
2	0.991	0.996	0.997	0.991	0.996	0.999
3	0.992	0.997	0.999	0.992	0.998	0.999
4	0.990	0.994	0.996	0.990	0.994	0.996
5	0.984	0.992	0.996	0.984	0.992	0.997
6	0.992	0.998	0.999	0.992	0.998	0.999
7	0.982	0.989	0.993	0.984	0.989	0.993
8	0.972	0.990	0.996	0.973	0.991	0.998

Table 3.6: The empirical power of the test for $a = 1.5$ in Example 3.3.

J		$\widehat{\mathrm{TM}}$			$\widehat{\mathrm{TF}}$	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
			$(\boldsymbol{n_1, n_2}) = (\mathbf{100, 100})$			
2	0.996	0.998	0.998	0.997	0.998	0.998
3	0.996	1.000	1.000	0.996	1.000	1.000
4	0.996	0.998	0.999	0.996	0.998	0.999
5	0.996	0.999	1.000	0.996	0.999	1.000
6	0.993	0.996	0.999	0.993	0.997	0.999
7	0.993	0.997	0.998	0.993	0.997	0.999
8	0.993	0.996	0.997	0.995	0.996	0.998
			$(\boldsymbol{n_1, n_2}) = (\mathbf{200, 200})$			
2	0.996	0.998	0.998	0.997	0.998	0.998
3	0.996	1.000	1.000	0.996	1.000	1.000
4	0.996	0.998	0.999	0.996	0.998	0.999
5	0.996	0.999	1.000	0.996	0.999	1.000
6	0.993	0.996	0.999	0.993	0.997	0.999
7	0.993	0.997	0.998	0.993	0.997	0.999
8	0.993	0.996	0.997	0.995	0.996	0.998

3.4 Real data examples

In this section, we analyzed two real data sets to illustrate the proposed method. The first example is the longitudinal CD4 data set for samples of AIDs patients, and we analyzed it to illustrate the application of the proposed method to the sparse and irregular function data. The second example is the Nitrogen Oxide Emission Level data set, which is analyzed to illustrate the proposed method for the case of the dense and regular functional data.

3.4.1 CD4 data

We applied the proposed test to an AIDS clinical study developed by the AIDS clinical trials group (ACTG) that can be found at `http://www.urmc.rochester.edu/biostat/people/\faculty/wusite/datasets/ACTG388.cfm`. The study enrolled 517 HIV-1-infected patients in three antiviral treatments, denote as A, B and C, respectively. Every group has 166, 171 and 176 patients, respectively. Patients were treated with an highly active antiretroviral therapy (HAART) for 128 weeks during which CD4 cell counts were monitored at weeks 4, 8 and every 8 weeks thereafter. However, each individual patient might not exactly follow the designed schedule and missing clinical visits for CD4 cell measurements frequently occurred which made the data set to be a typical longitudinal data set. The CD4 cell count data during 128 weeks of treatment are plotted for three groups in Fig. 3.1.

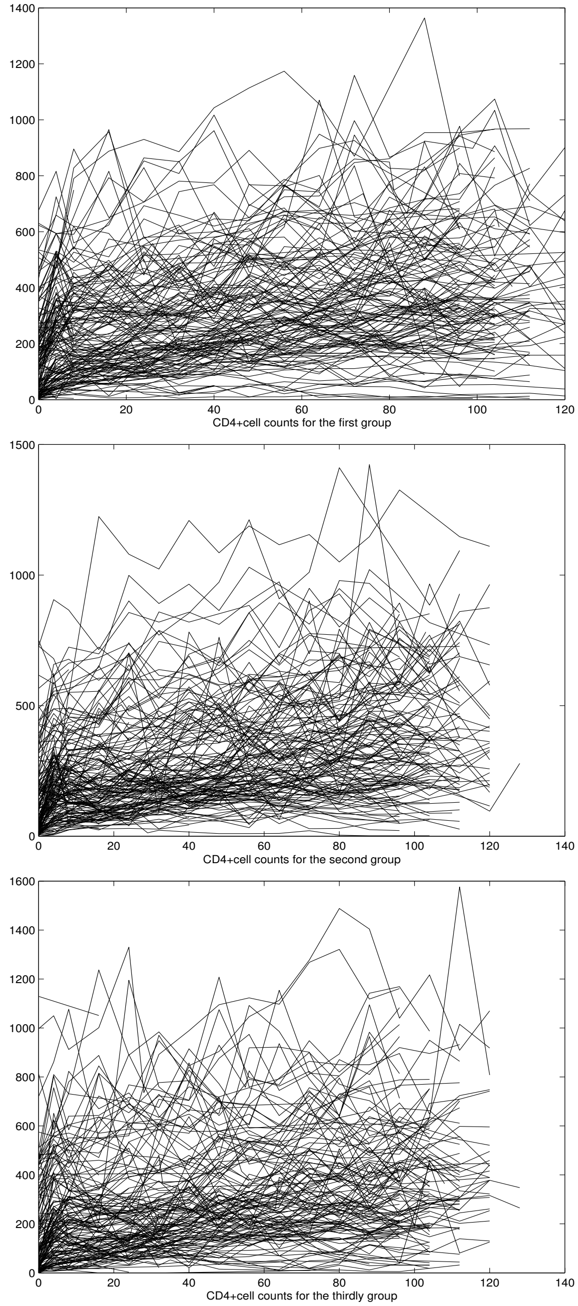

Figure 3.1: The CD4 cell count data during 128 weeks for group A, B and C, respectively.

We wanted to test if the two underlying mean functions of (A, B), (A, C) and (B, C) are different, which motivated a two-sample mean function testing problem.

Three combinations of the estimated mean functions were displayed in Fig. 3.2 for group A, B and C.

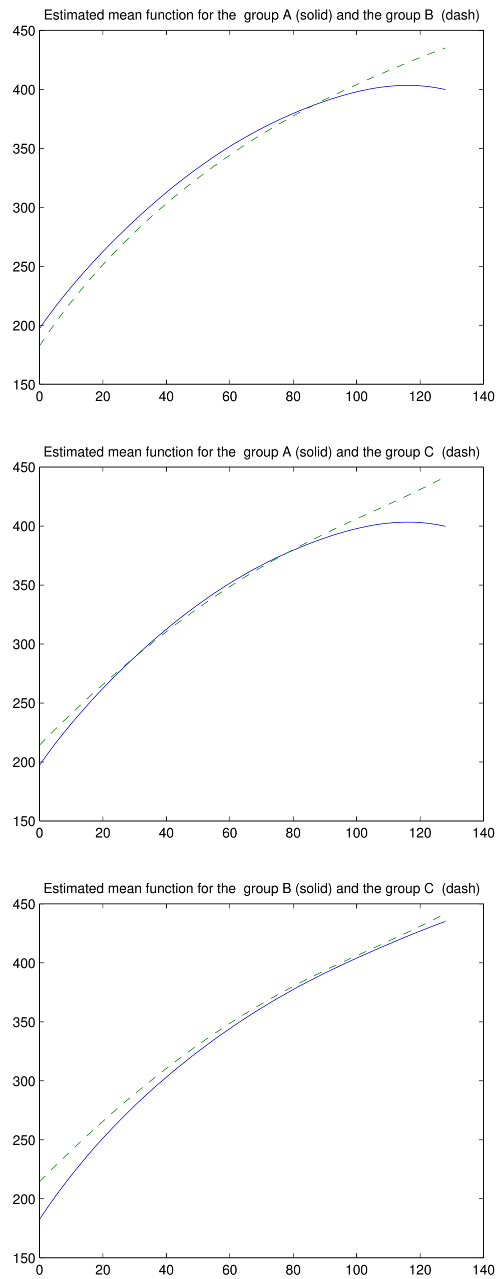

Figure 3.2: Three combination of the estimated mean functions for group A, B and C.

We firstly considered the combination (A, B) and computed the p-value of test from $J = 2$ to $J = 6$ and obtained which all of value are 0. This indicates that the mean function of group A and B is different. For the combinations (A, C) and (B, C), we obtained the same conclusion as the combination (A, B).

3.4.2 Nitrogen oxide emission level data

Nitrogen Oxides (NOx) are known to be among the most major pollutants, precursors of ozone formation, and contributors to global warming (Febrero et al., 2008). NOx is primarily caused by combustion processes in sources that burn fuels such as motor vehicles, electric utilities, and industries, among others. Fig. 3.3 shows NOx emission levels for seventy-six working days and thirty-nine non-working days, respectively, which were measured by an environmental control station close to an industrial area in Poblenou, Barcelona, Spain. The control station measured NOx emission levels in g/m3 every hour per day from February 23 to June 26 in 2005. The hourly measurements in one day (24 hours) formed a natural NOx emission level curve of the day. It is seen that within one day, the NOx levels increased in morning, attained

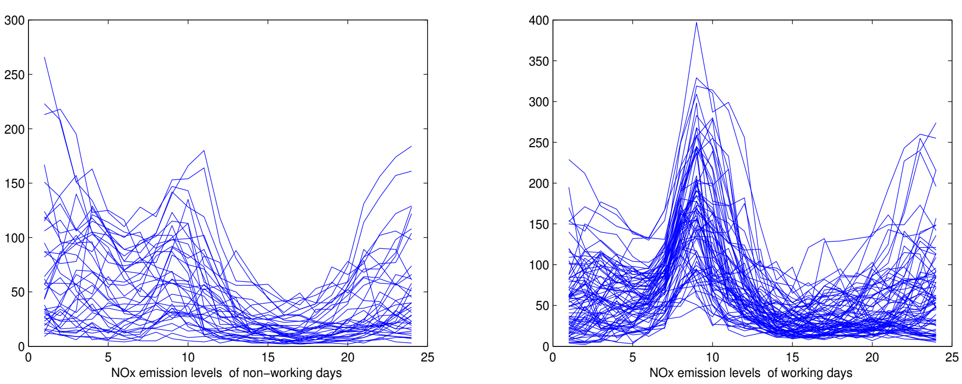

Figure 3.3: NOx emission levels for seventy-six working days and thirty-nine non-working days.

their extreme values around 8 a.m., then decreased until 2 p.m. and increased again in the evening. The influence of traffic on the NOx emission levels is not ignorable as the control station is located at the city center. It is not difficult to notice that the NOx emission levels of working days are generally higher than those of non-working days. This is why these NOx emission level curves were divided into two groups as pointed out by Febrero et al. (2008). Of interest is to test if the mean NOx emission level curves of working and non-working days are significantly different.

The fitted sample mean functions of two groups were displayed in Fig. 3.4. Using

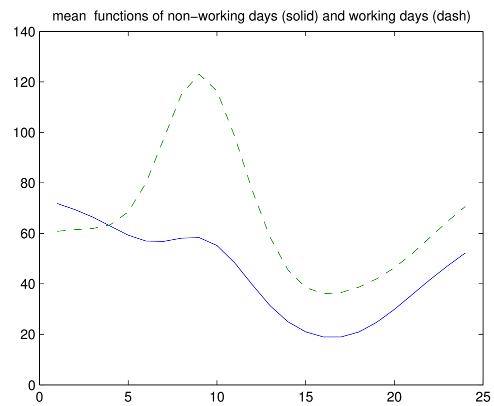

Figure 3.4: The estimated mean functions for seventy-six working days and thirty-nine non-working days.

the method similar in Subsection 3.4.1, the p-value is 0. This implies that the mean functions of the NOx levels are significantly different between working and non-working days.

3.5 Proofs of main results

Lemma 3.1. *Under assumptions 3.2-3.5, we have*

$$\|\hat{\phi}_j(s) - \phi_j(s)\| = O_p\left(\frac{J^{2\alpha+2}}{n}\right).$$

Using assumption 3.2 and equation (5.2) in Hall and Horowitz (2007), we can obtain proof of Lemma 3.1.

Lemma 3.2. *Under assumptions 3.1-3.5, We have*

$$\sqrt{n_1}\left(\hat{\eta}_j^{(1)} - \eta_j^{(1)}\right) \xrightarrow{d} N\left(0, \lambda_j^{(1)}\right),$$

$$\sqrt{n_2}\left(\hat{\eta}_j^{(2)} - \eta_j^{(2)}\right) \xrightarrow{d} N\left(0, \lambda_j^{(2)}\right).$$

Proof of Lemma 3.2

It can be observed

$$\begin{aligned}
&\hat{\eta}_j^{(1)} - \eta_j^{(1)} \\
&= \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_i^{(1)}(t_{il_1})\phi_j(t_{il_1}) - \eta_j^{(1)}\right\} \\
&\quad + \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_i^{(1)}(t_{il_1})\hat{\phi}_j(t_{il_1}) - \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_i^{(1)}(t_{il_1})\phi_j(t_{il_1})\right\} \\
&\equiv A_1 + A_2.
\end{aligned} \tag{3.5}$$

For A_1, we have

$$\begin{aligned}
A_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\left\{v_i^{(1)}(t_{il_1}) + \varepsilon_{il_1}^{(1)}\right\}\phi_j(t_{il_1}) \\
&\quad + \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\mu_1(t_{il_1})\phi_j(t_{il_1}) - \eta_j^{(1)}\right\} \\
&\equiv A_{11} + A_{12}.
\end{aligned} \tag{3.6}$$

It is easy to see that A_{11} is the average of independent and identically distributed random variables with mean $E(A_{11}) = 0$ and variance $\mathrm{var}(A_{11}) = \lambda_j^{(1)}/n_1$ where

$$\lambda_j^{(1)} = (1-\tau)\Gamma_j^{(1)} + \tau\left\{\int_0^1 \gamma_1(t,t)\phi_j^2(t)dt + \sigma^2\right\}$$

with

$$\Gamma_j^{(1)} = \int_0^1 \int_0^1 \phi_j(t)\gamma_1(t,s)\phi_j(s)dtds.$$

By the central limit theorem, we obtain

$$A_{11} \xrightarrow{d} N\left(0, \frac{1}{n_1}\lambda_j^{(1)}\right). \tag{3.7}$$

For A_{12}, according to assumption 3.5, we have

$$A_{12} = o(n_1^{-1/2}). \tag{3.8}$$

By (3.6), (3.7) and (3.8), we obtain

$$A_1 \xrightarrow{d} N\left(0, \frac{1}{n}\lambda_j^{(1)}\right). \tag{3.9}$$

For A_2, we have

$$\begin{aligned} A_2 = & \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\mu_1(t_{il_1})\left\{\hat{\phi}_j(t_{il_1}) - \phi_j(t_{il_1})\right\} \\ & + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}v_i^{(1)}(t_{il_1})\left\{\hat{\phi}_j(t_{il_1}) - \phi_j(t_{il_1})\right\} \\ & + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\varepsilon_{il_1}^{(1)}\left\{\hat{\phi}_j(t_{il_1}) - \phi_j(t_{il_1})\right\} \\ \equiv & A_{21} + A_{22} + A_{23}. \end{aligned} \tag{3.10}$$

According to Cauchy-Schwarz inequality, Assumption 3.1 and lemma 3.1, we have

$$\begin{aligned} A_{21} \leqslant & \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\left[\sum_{l_1=1}^{N_i}\mu_1^2(t_{il_1})\sum_{l_1=1}^{N_i}\left\{\hat{\phi}_j(t_{il_1}) - \phi_j(t_{il_1})\right\}^2\right]^{1/2} \\ = & o_p\left(n_1^{-1/2}\right). \end{aligned} \tag{3.11}$$

For A_{22}, we have

$$
\begin{aligned}
&E(A_{22}^2)\\
&=\frac{1}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left[\left\{\hat{\phi}_j(t_{il_1})-\phi_j(t_{il_1})\right\}\right]^2 Ev_i^{(1)2}(t_{il_1})\\
&\quad+\frac{1}{n_1^2}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{1\leqslant l_{11}\neq l_{12}\leqslant N_i}\left\{\hat{\phi}_j(t_{il_{11}})-\phi_j(t_{il_{11}})\right\}\left\{\hat{\phi}_j(t_{il_{12}})-\phi_j(t_{il_{12}})\right\}E\left\{v_i^{(1)}(t_{il_{11}})v_i^{(1)}(t_{il_{12}})\right\}\\
&=o(n_1^{-1}).
\end{aligned}
\tag{3.12}
$$

Using the arguments similar to that of (3.12), it can be shown

$$E(A_{23}^2)=o\left(n_1^{-1}\right). \tag{3.13}$$

By (3.10), (3.11), (3.12), and (3.13), we have

$$A_2=o_p\left(n_1^{-1/2}\right). \tag{3.14}$$

By (3.5), (3.9) and (3.14), we obtain $\sqrt{n_1}(\hat{\eta}_j^{(1)}-\eta_j^{(1)})\xrightarrow{d}N(0,\lambda_j^{(1)})$. Similarly, we can prove $\sqrt{n_2}(\hat{\eta}_j^{(2)}-\eta_j^{(2)})\xrightarrow{d}N(0,\lambda_j^{(2)})$. The proof is then completed.

Proof of theorem 3.1 According to lemma 3.1 and Slutsky theory, we can easy obtain the the conclusion of theorem 3.1. The proof is then completed.

Proof of theorem 3.2 According to lemma 3.1, we have

$$\hat{\eta}_j^{(1)}\xrightarrow{p}\eta_j^{(1)},\ \hat{\eta}_j^{(2)}\xrightarrow{p}\eta_j^{(2)}.$$

Under H_1, we obtain

$$\widehat{\mathrm{TM}}\xrightarrow{p}\frac{n_1n_2}{n_1+n_2}\sum_{j=1}^{J}\frac{\left\{\eta_j^{(1)}-\eta_j^{(2)}\right\}^2}{\lambda_j}\rightarrow\infty.$$

Chapter 4

Interaction Models with Nonlinear Link for Functional Regression

4.1 Introduction

Functional regression is modeled by allowing functional random trajectories of either response or covariate or both. It has attracted more and more research interest raised from real data analysis, among which one class of modeling extensively investigated is the scenario with scalar response and functional covariates, refer to Ramsay and Silverman (2005), Li and Hsing (2007), Ma (2016), Usset et al. (2016), among others. For this kind of functional regression model, in practice there used to involve multiple functional covariates. A motivating example is the daily precipitation which is affected by temperature curve, pressure normals curve, wind curve, cloudiness curve, and other climate indices, etc.

Existing literature tends to model the association between scalar response and functional covariate in two ways. One just considers the main effects of functional covariates additively but rarely take into consideration of the interaction effects, refer to Cardot et al. (2003) and Ramsay and Silverman (2005). This intuitively will lead to inappropriate conclusions due to biased or even inaccurate estimation of the model parameters. The other is to directly assume a nonparametric link on the mean regression, refer to Yao et al. (2005), Cai and Hall (2006), Hall and Horowitz (2007), and Li and Hsing (2007). This is robust but bit of losing information. Usset et al. (2016) might be the first one to incorporate the interaction effect besides the individual main effect of functional covariates. Their estimation procedure conducted the inference for main effects using penalized regression splines and for the interaction effect by a tensor product basis. However, it is idealistic that the interaction part in the model of Usset et al. (2016) is assumed to be linear. This motivates us to consider a more general model with non-linear interaction part instead. This is implemented mathematically by adding an unknown link function structure on the interaction part. Consequently it makes the corresponding statistical inference much more complicated. We address the statistical problem via the widely used functional principal components (FPC) and the minimum average variance estimation (MAVE) methods by Xia and Härdle (2006).

The remaining of this chapter is as follows. In Section 4.2, we introduce the conditional mean regression models with nonparametric single-index interaction, all of the estimation procedures are discussed in Section 4.3. We describe the asymptotic theory of the procedure in Section 4.4, while Section 4.5 is devoted to a report on simulation results, followed by a description of one application to regression for climate data in Section 4.6. Some details of estimation, assumptions and all proofs are included in Supplementary material.

4.2 Model alternative based on K-L representation

The data observed is $\{(Y_i, X_i, Z_i), i = 1, \cdots, n\}$, where Y_i is a scalar response, $X_i(\cdot)$ and $Z_i(\cdot)$ are independent non-stationary smooth random functions in $L^2[0,1]$. The conditional mean regression given the covariates is described as follows when taking interaction effects into account,

$$\begin{aligned} E[Y|X,Z] = \alpha + \int_{\mathcal{T}} \beta_X(s) X^c(s) ds + \int_{\mathcal{T}} \beta_Z(t) Z^c(t) dt \\ + g\left(\int_{\mathcal{T}} \int_{\mathcal{T}} \gamma(s,t) X^c(s) Z^c(t) dt ds \right), \end{aligned} \tag{4.1}$$

where α is an intercept and $g(\cdot)$ is an unknown link function, $X^c(s) = X(s) - \mu_X(s)$ and $Z^c(t) = Z(t) - \mu_Z(t)$ denote the centered predictor processes. The regression parameter function $\beta_X(s)$ and $\beta_Z(t)$ are assumed to be smooth and square integrable. $\gamma(s,t)$ is a real valued bi-variate function defined on $\mathcal{T}^2$. It is quite general to include many other important models as special examples. For instance, if $\gamma(s,t) \equiv 0$ in (4.1), it reduces to the conditional mean regression models, refer to Ait-Saïdi et al. (2008). If $X(s) \equiv Z(s)$ and $g(\cdot)$ is an identity function, (4.1) becomes the functional quadratic regression model which has been studied by Yao and Müller (2010). When $\beta_X(s) \equiv 0, \beta_Z(t) \equiv 0$ and $Z^c(t) \equiv constant$, (4.1) becomes the functional single-index regression model which has been investigated by Chen et al. (2011a). In addition, if $\beta_X(s) \equiv 0, \beta_Z(t) \equiv 0$ and $\gamma(s,t)Z^c(t)$ is a semiparametric function with a single-index structure, (4.1) reduces to the generalized functional linear model with semiparametric single-index interaction considered by Li et al. (2010) as a special case. It is noted that the proposed model (4.1) included also the model of Usset et al. (2016) as the special situation when $g(\cdot)$ is an identity function. However, our model distinguishes from model of Ma (2016) where interaction of two functional data are not considered.

Denote mean and auto-covariance functions of two predictor processes are smooth.

$$\begin{aligned} &E\{X(s)\} = \mu_X(s), E\{Z(t)\} = \mu_Z(t); \\ &\text{cov}\{X(s_1), X(s_2)\} = G_X(s_1, s_2), \text{cov}\{Z(t_1), Z(t_2)\} = G_Z(t_1, t_2). \end{aligned}$$

For prediction processes $X(\cdot)$ and $Z(\cdot)$, their Karhunen-Loève expansions (Ash and

Gardner, 1975) are

$$X(s) = \mu_X(s) + \sum_{j=1}^{\infty} \xi_j \phi_j(s),\ Z(t) = \mu_Z(t) + \sum_{k=1}^{\infty} \zeta_k \psi_k(t),$$

where ϕ_j and ψ_k are sequences of orthonormal eigenfunctions of associated auto-covariance operators that form a basis of the function space and are associated with sequences of non-increasing eigenvalues λ_j and τ_k, satisfying $\sum_j \lambda_j < \infty$, $\sum_k \tau_k < \infty$, $G_X(s_1, s_2) = \sum_j \lambda_j \phi_j(s_1)\phi_j(s_2),\ s_1, s_2 \in \mathcal{T}$ and $G_Z(t_1, t_2) = \sum_k \tau_k \psi_k(t_1)\psi_k(t_2),\ t_1, t_2 \in \mathcal{T}$, and the coefficients ξ_j and ζ_k are referred to as functional principal component (FPC) score. They are sequences of uncorrelated random variables, respectively, with means $E(\xi_j) = 0$, $E(\zeta_k) = 0$ and variances $\mathrm{Var}(\xi_j) = \lambda_j$, $\mathrm{Var}(\zeta_k) = \tau_k$.

Since the eigenfunctions $\phi_j, j = 1, 2, \ldots$ and $\psi_k, k = 1, 2, \ldots$ of the processes X and Z form a complete basis, the regression parameter functions in (4.1) can be represented in this basis,

$$\beta_X(s) = \sum_{j=1}^{\infty} \beta_X^j \phi_j(s),\ \beta_Z(t) = \sum_{k=1}^{\infty} \beta_Z^k \psi_k(t),\ \gamma(s,t) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \gamma_{jk}\phi_j(s)\psi_k(t), \tag{4.2}$$

for suitable sequences $\{\beta_X^j\}_{j=1,2,\cdots}$, $\{\beta_Z^k\}_{k=1,2,\cdots}$ and $\{\gamma_{jk}\}_{j,k=1,2,\cdots}$ with $\sum_j \beta_X^j < \infty$, $\sum_k \beta_Z^k < \infty$ and $\sum_{j,k} \gamma_{jk} < \infty$.

Substituting (4.2) into (4.1) and applying the orthonormality property of the eigenfunctions, one finds that model (4.1) can be alternatively expressed as a function of the scores ξ_j and ζ_k of predictor processes X and Z,

$$E[Y|X, Z] = \alpha + \sum_{j=1}^{\infty} \beta_X^j \xi_j + \sum_{k=1}^{\infty} \beta_Z^k \zeta_k + g\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \gamma_{jk}\xi_j\zeta_k\right). \tag{4.3}$$

Model (4.3) can be alternatively expressed as a function of a finite number of the scores ξ_j and ζ_k of predictor processes X and Z, if we skip the intercept or let $\alpha = 0$

for identifiability purposes,

$$E[Y|X, Z] = \sum_{j=1}^{K_1} \beta_X^j \xi_j + \sum_{k=1}^{K_2} \beta_Z^k \zeta_k + g\left(\sum_{k=1}^{K_2}\sum_{j=1}^{K_1} \gamma_{jk}\xi_j\zeta_k\right) \equiv \beta^\top U + g\left(\theta^\top W\right), \quad (4.4)$$

where $\beta^\top = (\beta_X^1, \cdots, \beta_X^{K_1}, \beta_Z^1, \cdots, \beta_Z^{K_2})$, $\theta^\top = (\gamma_{1,1}, \cdots, \gamma_{1,K_2}, \cdots, \gamma_{K_1,K_2})$, $U^\top = (\xi_1, \cdots, \xi_{K_1}, \zeta_1, \cdots, \zeta_{K_2})$ and $W^\top = (\xi_1\zeta_1, \cdots, \xi_1\zeta_{K_2}, \cdots, \xi_{K_1}\zeta_{K_2})$. The first equivalence indicates that Y mainly depends on the leading K_1 principal components in $X(\cdot)$ and the leading K_2 principal components in $Z(\cdot)$. The justification lies on in the fact that estimation of high order principal components is highly unstable and difficult to interpret in functional data analysis, refer to the comments in Rice and Silverman (1991) and Hall and Hosseini-Nasab (2006). The second equivalence is in the form of classical partially linear single-index model. For model identification, let θ satisfy $\|\theta\| = 1$ and $\gamma_{1,1} > 0$.

4.3 Estimation of coefficient functions of all functional covariates

Since the scores ξ_j and ζ_k are unknown, we cannot estimate β, θ and γ based on the model (4.4) directly. Thus, to estimate functional principal component scores is necessary. Our starting point for modeling are the actual observations, which consist either of densely spaced and non-random measurement (dense design) or alternatively of sparse and randomly (irregularly) spaced repeated measurements (sparse design) of the predictor trajectories X_i and Z_i. Let R_{ij} and V_{ik} denote the observations of the random trajectories X_i and Z_i at fixed or random time points S_{ij} and T_{ik} contaminated with measurement errors ε_{ij} and ϵ_{ik}, respectively. The errors ε_{ij} and ϵ_{ik} are assumed to be independent and identically distributed with zero means and variances σ_X^2 and σ_Z^2, respectively. Meanwhile, the errors are independent of $X_i(s)$

and $Z_i(t)$.

$$\begin{aligned} R_{ij} &= X_i(S_{ij}) + \varepsilon_{ij} = \mu_X(S_{ij}) + \sum_{j=1}^{\infty} \xi_{ij}\phi_j(S_{ij}) + \varepsilon_{ij}, \\ V_{ik} &= Z_i(T_{ik}) + \epsilon_{ik} = \mu_Z(T_{ik}) + \sum_{k=1}^{\infty} \zeta_{ik}\psi_k(T_{ik}) + \epsilon_{ik}, \end{aligned} \tag{4.5}$$

where $E\{\xi_{ij}\} = E\{\zeta_{ik}\} = 0$, $E\{\xi_{ij}\xi_{ij'}\} = E\{\zeta_{ik}\zeta_{ik'}\} = 0$ $(j \neq j',\ k \neq k')$, $E\{\xi_{ij}^2\} = \lambda_j$ and $E\{\zeta_{ik}^2\} = \tau_k$. Estimates $\hat{\mu}_X$, $\hat{G}_X$, $\hat{\lambda}_j$, $\hat{\phi}_j$ and $\hat{\sigma}_X^2$ (respectively, $\hat{\mu}_Z$, $\hat{G}_Z$, $\hat{\tau}_k$, $\hat{\psi}_k$ and $\hat{\sigma}_Z^2$) of the underlying population mean function μ_X, covariance function G_X, eigenvalues λ_j, eigenfunctions ϕ_j and error variance σ_X^2 (respectively, mean function μ_Z, covariance function G_Z, eigenvalues τ_k, eigenfunctions ψ_k and error variance σ_Z^2) are easily obtained by applying a nonparametric functional approach. Estimates $\hat{\xi}_j$ and $\hat{\zeta}_k$ of the FPC scores ξ_j and ζ_k can be obtained by the traditional integral estimates for the dense design case or the conditional expectation approach of Yao et al. (2005) for the sparse design case. Some additional details are given in Part a of Section 4.7.

Once these preliminary estimates are in hand, we adopt the minimum average variance estimation (MAVE) method by Xia and Härdle (2006) to estimate β and θ. We briefly describe the method as following. Let

$$\hat{U}_i = \left(\hat{\xi}_{i1}, \cdots, \hat{\xi}_{iK_1},\ \hat{\zeta}_{i1}, \cdots, \hat{\zeta}_{iK_2}\right)^\top,$$

and

$$\hat{W}_i = \left(\hat{\xi}_{i1}\hat{\zeta}_{i1}, \cdots, \hat{\xi}_{i1}\hat{\zeta}_{iK_2}, \cdots, \hat{\xi}_{iK_1}\hat{\zeta}_{iK_2}\right)^\top,\ i = 1, \cdots, n.$$

For $\hat{W}_i$ which are close to w, we have the following local linear approximation

$$Y_i - \beta^\top \hat{U}_i - g\left(\theta^\top \hat{W}_i\right) \approx Y_i - \beta^\top \hat{U}_i - g\left(\theta^\top w\right) - g'\left(\theta^\top w\right) \hat{W}_{i0}^\top \theta,$$

where $\hat{W}_{i0} = \hat{W}_i - w$. Following the idea of local linear smoothing, we may estimate $g(\theta^\top w)$ and $g^{'}(\theta^\top w)$ by the argument $(a, d)^\top$ that minimizes

$$\sum_{i=1}^{n} \left\{ Y_i - \left(\beta^\top \hat{U}_i + a + d\hat{W}_{i0}^\top \theta \right) \right\}^2 \hat{H}_{i0}, \tag{4.6}$$

where $\hat{H}_{i0} \geqslant 0, i = 1, \cdots, n$, are some weights with $\sum_{i=1}^{n} \hat{H}_{i0} = 1$, typically centering at w. Let a_j and d_j be the estimate of $g(\theta^\top \hat{W}_j)$ and $g^{'}(\theta^\top \hat{W}_j)$, respectively. Our estimating procedure is to minimize

$$\sum_{j=1}^{n} G\left(\theta^\top \hat{W}_j\right) I_n\left(\hat{W}_j\right) \sum_{i=1}^{n} \left\{ Y_i - \left(\beta^\top \hat{U}_i + a_j + d_j \hat{W}_{ij}^\top \theta \right) \right\}^2 \hat{H}_{ij}, \tag{4.7}$$

with respect to (a_j, d_j) and (β, θ), where $G(\cdot)$ is another weight function that controls the contribution of $(\hat{U}_i, \hat{W}_i, Y_i),\ i = 1, \cdots, n$ to the estimation of (β, θ), $I_n(\cdot)$ is employed here for technical purpose to handle the boundary points, and $\hat{H}_{ij}$ is local weight function and $\hat{W}_{ij} = \hat{W}_i - \hat{W}_j$.

We use two sets of weights. In the initial stage, let $\hat{H}_{ij} = \hat{H}_{b,i}(\hat{W}_j) / \sum_{l=1}^{n} \hat{H}_{b,l}(\hat{W}_j)$, where $\hat{H}_{b,i}(\hat{W}_j) = b^{-K_1K_2} H(\hat{W}_{i,j}/b)$, with $H(\cdot)$ is a K_1K_2-dimensional kernel function and b is a bandwidth. This will enable us to find a consistent estimator $(\check{\beta}, \check{\theta})$ based on (4.7). We then switch to a set of refined weights to gain more efficiency. In the second stage, we carry out the same iteration steps but let $\hat{H}_{ij}^{\check{\theta}} = \hat{K}_{b_1,i}^{\check{\theta}}(\check{\theta}^\top \hat{W}_j) / \sum_{l=1}^{n} \hat{K}_{b_1,l}^{\check{\theta}}(\check{\theta}^\top \hat{W}_j)$, where $\hat{K}_{b_1,i}^{\check{\theta}}(w) = b_1^{-1} K\{(\check{\theta}^\top \hat{W}_i - w)/b_1\}$, with $K(\cdot)$ is an univariate kernel function, b_1 is the bandwidth and $\check{\theta}$ is the estimated value of θ from the previous iteration. Denote the final value by $(\hat{\check{\beta}}, \hat{\check{\theta}})$, where

$$\hat{\check{\beta}} = \left(\hat{\check{\beta}}_X^1, \cdots, \hat{\check{\beta}}_X^{K_1}, \hat{\check{\beta}}_Z^1, \cdots, \hat{\check{\beta}}_Z^{K_2} \right)^\top,\ \hat{\check{\theta}} = \left(\hat{\check{\gamma}}_{1,1}, \cdots, \hat{\check{\gamma}}_{1,K_2}, \cdots, \hat{\check{\gamma}}_{K_1,K_2} \right)^\top. \tag{4.8}$$

The estimator of $\beta_X(\cdot)$, $\beta_Z(\cdot)$ and $\gamma(\cdot,\cdot)$ are then given by

$$\hat{\beta}_X(s) = \sum_{j=1}^{K_1} \hat{\tilde{\beta}}_X^{(j)} \hat{\phi}_j(s),\ \hat{\beta}_Z(t) = \sum_{k=1}^{K_2} \hat{\tilde{\beta}}_Z^{(k)} \hat{\psi}_j(s),\ \hat{\gamma}(s,t) = \sum_{j=1}^{K_1}\sum_{k=1}^{K_2} \hat{\tilde{\gamma}}_{jk} \hat{\phi}_j(s)\hat{\psi}_k(t). \quad (4.9)$$

Meanwhile, we can also estimate $g(v)$ by the solution of a_j in (4.7) with $\hat{\tilde{\theta}}^\top \hat{W}_j$ replaced by v and denote its estimate by $\hat{g}(v)$.

4.4 Asymptotic theory

To establish the relevant asymptotic results, we require studying the relationship between the true FPC scores ξ_{ij} and η_{ik} with their estimates $\hat{\xi}_{ij}$ and $\hat{\eta}_{ik}$ since the estimates of the conditional mean regression models with single-index interaction need to be based on the estimated scores. A key step in the mathematical analysis is to establish exact upper bounds of $|\hat{\xi}_{ij} - \xi_{ij}|$ and $|\hat{\eta}_{ik} - \eta_{ik}|$. The convergence properties of the estimated conditional mean regression models follow from those upper bounds since these estimates are obtained by applying MAVE method to $\{\hat{\xi}_{ij}, \hat{\eta}_{ik}, Y_i\}$ for $i = 1, \cdots, n$, $j = 1, \ldots, K_1$ and $k = 1, \ldots, K_2$.

We consider the consistency rate of the estimated regression functions in a functional setting where the number of FPCs depends on the sample size n, i.e. $K_1 = K_1(n)$ and $K_2 = K_2(n)$, and tends to infinity as $n \to \infty$. In practice the choice of $K_1 = K_1(n)$ and $K_2 = K_2(n)$ depends on the intrinsic structural complexity and estimating accuracy of the covariance structure.

We are ready to present the asymptotic results of the proposed estimators. Theorems 4.1 and 4.2 below establish the consistency of the estimators of the parameter function and the nonparametric function, respectively. The proofs of these theorems are provided in Part c of Section 4.7.

Theorem 4.1. *Under all the assumptions listed in Part b of Section 4.7, we have*

$$\|\hat{\beta}_X - \beta_X\| \xrightarrow{p} 0, \ \|\hat{\beta}_Z - \beta_Z\| \xrightarrow{p} 0, \ \|\hat{\gamma} - \gamma\| \xrightarrow{p} 0$$

where $\hat{\beta}_X$, $\hat{\beta}_Z$ and $\hat{\gamma}$ are defined in (4.9).

The rates of convergence of $\|\hat{\beta}_X - \beta_X\|$, $\|\hat{\beta}_Z - \beta_Z\|$, and $\|\hat{\gamma} - \gamma\|$ can be found in the (4.29) of Part b of Section 4.7.

Theorem 4.2. *Under all the assumptions listed in Part b of Section 4.7, if the density function $f_\theta(v)$ of $\theta^\top W$ is positive, the derivative of $E[\varepsilon^2|\theta^\top W = v]$ exists, $b_1 \sim n^\delta$ with $1/6 < \delta < 1/4$, and $E[\varepsilon_i|U_j, W_j, Y_j, j < i] = 0$ almost surely, we then have*

$$|\hat{g}(v) - g(v)| \xrightarrow{p} 0.$$

Additional results on the rates of convergence of $|\hat{g}(v) - g(v)|$ can be found in the (4.30) of Part b of Section 4.7. The proof of Theorems 4.1 and 4.2 is in Section 4.7

4.5 Simulation studies

We conducted some Monte Carlo simulation studies to evaluate the performance of our proposed estimators for finite samples.

Simulation 1. Additive model for functional data has been studied by Ferraty and Vieu (2009). We used the same model as that of Ferraty and Vieu (2009) in the simulation. The predictor functions were generated by

$$X_i^1(t) = \exp\left\{\sin^2(\omega_i t)\right\} + (a_i + 2\pi)\, t^3 + b_i,$$
$$X_i^2(t) = \cos(\omega_i t) + (c_i + 2\pi)\, t^2 + d_i,$$

where $i = 1, \cdots, 100$, $t \in [-1, 1]$. ω_i are random real numbers generated from uniform distributions on $[0, 2\pi]$ and a_i, b_i, c_i and d_i are generated from $U[0, 1]$.

The responses have been generated according to different regression models of the form

$$Y_i = \alpha_1(a_i + 2\pi) + \frac{(1 - \cos\omega_i + \alpha_2|\sin\omega_i|)}{\omega_i} + \alpha_2 d_i + \epsilon_i,$$

where ϵ_i simulated from a $N(0,1)$ distribution.

These models have been chosen to contain three terms, one depending only on the first covariate (this is controlled by the parameter α_1), one depending only on the second covariate (this is controlled by the parameter α_2) and the third one depends on both covariates. Indeed, just by changing the values of the parameters α_1 and α_2 one can change significantly the structure of the model.

For each model we simulated two samples: a learning sample of size 100 from which the estimates are computed and a testing sample of size 50 on which the prediction errors are calculated. For each predictor function, we sampled through 50 equidistantly spaced measurements in $[-1, 1]$. To show the usefulness of the conditional mean regression models with single-index interaction, we have also computed the predicted values on the testing sample by means of procedures in Section 4.3 and compared the proposed model with various nonparametric models:

- Method M^1_{Np}: kernel estimate based on the single covariate X_1;
- Method M^2_{Np}: kernel estimate based on the single covariate X_2;
- Method $M^{(1,2)}_{Np}$: kernel estimate based on the pair of (X_1, X_2);
- Method $M^{(1,2)}_{Add}$: two-step additive estimate starting with variable X_1;
- Method $M^{(2,1)}_{Add}$: two-step additive estimate starting with variable X_2;
- Method M^{auto}_{Add}: two-step additive estimate with automatic order choice;

- Method M_{inter}: the estimate based on the conditional mean regression models with single-index interaction.

The results are summarized in Table 4.1 which gives, for various values of α_1 and α_2, the comparative Mean of Square Residuals (MSR)

$$MSR = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2,$$

where $\hat{Y}_i$ is the prediction described just before.

Table 4.1: MSR for different method.

α_1	α_2	M^1_{Np}	M^2_{Np}	$M^{(1,2)}_{Np}$	$M^{(1,2)}_{Add}$	$M^{(2,1)}_{Add}$	M^{auto}_{Add}	M_{inter}
$\alpha_1 = 1$	$\alpha_2 = 1$	1.76	1.57	1.69	1.64	1.64	1.62	1.38
	$\alpha_2 = 3$	2.98	1.56	1.95	1.63	1.63	1.58	1.34
	$\alpha_2 = 5$	4.26	1.60	2.22	1.67	1.67	1.88	1.57
	$\alpha_2 = 7$	7.09	1.73	2.89	1.81	1.81	2.39	1.72
	$\alpha_2 = 9$	10.92	1.98	3.85	2.07	2.07	3.17	1.97
$\alpha_1 = 3$	$\alpha_2 = 1$	2.76	2.58	2.67	2.62	2.62	2.63	2.00
	$\alpha_2 = 3$	3.08	2.76	2.62	2.74	2.74	2.50	2.43
	$\alpha_2 = 5$	4.88	2.64	2.86	2.66	2.66	2.73	2.52
	$\alpha_2 = 7$	7.67	2.77	3.52	2.81	2.81	3.24	2.57
	$\alpha_2 = 9$	11.47	3.01	4.44	3.05	3.05	4.02	2.63
$\alpha_1 = 5$	$\alpha_2 = 1$	3.42	4.29	3.45	3.32	3.45	4.29	3.04
	$\alpha_2 = 3$	4.18	4.72	3.81	3.43	3.81	4.09	3.42
	$\alpha_2 = 5$	5.95	4.90	4.43	3.55	3.55	4.22	3.53
	$\alpha_2 = 7$	8.71	5.15	4.71	3.77	3.77	4.71	3.59
	$\alpha_2 = 9$	12.48	4.90	5.66	3.70	3.70	5.48	3.66
$\alpha_1 = 7$	$\alpha_2 = 1$	5.00	6.70	5.08	4.90	5.08	6.62	4.88
	$\alpha_2 = 3$	5.73	7.27	5.47	5.04	5.04	6.29	4.95
	$\alpha_2 = 5$	7.76	7.76	6.27	5.33	5.33	6.30	5.27
	$\alpha_2 = 7$	10.20	8.06	7.10	5.57	5.57	6.78	5.29
	$\alpha_2 = 9$	13.94	8.36	7.33	5.90	5.90	7.55	5.49
$\alpha_1 = 9$	$\alpha_2 = 1$	7.03	9.81	7.18	6.95	7.18	9.65	6.75
	$\alpha_2 = 3$	7.28	10.46	7.57	7.09	7.57	10.23	6.93
	$\alpha_2 = 5$	9.43	11.27	8.40	7.50	8.40	9.12	7.30
	$\alpha_2 = 7$	12.14	11.67	9.28	7.80	9.28	8.99	7.52
	$\alpha_2 = 9$	15.85	12.22	9.98	8.29	9.98	9.51	8.36

From Table 4.1, it is observed that the conditional mean regression n models with single-index interaction give very nice results compared with nonparametric methods based on one or two variables. Indeed, the conditional mean regression models with single-index interaction lead to the smallest prediction error in all situations.

Simulation 2. In this study, we consider functional data generated from the process

$$X_i(s) = \sum_{j=1}^{2} \xi_{ij}\phi_j(s),\ \ Z_i(t) = \sum_{k=1}^{3} \zeta_{ik}\phi_k(t),\ \ i = 1, \cdots, 100,$$

where $\phi_1(s) = -\sqrt{2}\cos(2\pi s)$, $\phi_2(s) = \sqrt{2}\sin(2\pi s)$, $\phi_3(s) = -\sqrt{2}\cos(4\pi s)$, $s \in [0,1]$, and $\xi_{ij} \sim N(0, \lambda_j)$ with $\lambda_1 = 4$, $\lambda_2 = 1/2$ and $\zeta_{ik} \sim N(0, \tau_k)$ with $\tau_1 = 1$, $\tau_2 = 1/2$ and $\tau_3 = 1/4$. Responses Y_i were obtained as:

$$Model(A) : Y_i = \int_{\mathcal{T}} \beta_X(s)X_i(s)ds + \int_{\mathcal{T}} \beta_Z(t)Z_i(t)dt + \exp\left\{\int_{\mathcal{T}}\int_{\mathcal{T}} \gamma(s,t)X_i(s)Z_i(t)dtds\right\} + e_i,$$

$$Model(B) : Y_i = \int_{\mathcal{T}} \beta_X(s)X_i(s)ds + e_i,$$

$$Model(C) : Y_i = \int_{\mathcal{T}} \beta_Z(t)Z_i(t)dt + e_i,$$

$$Model(D) : Y_i = \int_{\mathcal{T}} \beta_X(s)X_i(s)ds + \int_{\mathcal{T}} \beta_Z(t)Z_i(t)dt + e_i,$$

where

$$\beta_X(s) = -2\phi_1(s) + \phi_2(s),\ \ \beta_Z(t) = 2\phi_1(t) - \phi_2(t) + 3\phi_3(t),$$

$$\gamma(s,t) = \frac{1}{\sqrt{3}}\phi_1(s)\phi_1(t) + \frac{1}{\sqrt{6}}\phi_1(s)\phi_2(t) + \frac{1}{3}\phi_1(s)\phi_3(t) + \frac{1}{3}\phi_2(s)\phi_2(t) + \frac{1}{3}\phi_2(s)\phi_3(t)$$

and e_i were simulated as $N(0, 0.01)$. The measurement error in (4.5) is $\varepsilon_{ij} \sim N(0, 0.25)$ and $\epsilon_{ij} \sim N(0, 0.25)$, respectively.

The number of measurements where each trajectory was sampled was selected 100 equi-distant point in $[0, 1]$ for dense cases. We compared the performance with various models. The performance measure tabulated in Table 4.2 is Mean of Square Residuals $MSR = \sum_{i=1}^{100}(Y_i - \hat{Y}_i)^2/100$.

Table 4.2: MSR for different model.

True Model	Fitted Model	$25th$	$50th$	$75th$
A	A	1.2424	2.8452	29.5848
	B	19.6825	41.3847	126.0650
	C	19.7162	41.6527	125.8314
	D	19.2137	40.8378	125.5848
B	A	0.3868	0.3952	0.4042
	B	0.4030	0.4153	0.4268
C	A	5.00	6.70	5.08
	C	5.73	7.27	5.47
D	A	0.4106	0.4190	0.4268
	B	2.5623	2.6141	2.6665
	C	3.9933	4.0696	4.1604
	D	0.4508	0.4508	0.4791

Our conclusion is that the conditional mean regression with single-index interaction leads to similar prediction errors with functional linear regression when the underlying regression model is linear, while the conditional mean regression with single-index interaction performs better than the linear model in situations when the underlying regression relationship is nonlinear.

4.6 Real data example: Climate data

We now focus on the analysis of climate data from NOAA (`www.ncdc.noaa.gov`) to illustrate functional data regression procedures. The source-datasets directory contain all normals derived from hourly data, including temperature, dew point temperature, heat index, wind chill, wind, cloudiness, heating and cooling degree hours, and pressure normals, also contains all precipitation, snowfall, and snow depth normals files including percentiles, frequencies, and averages.

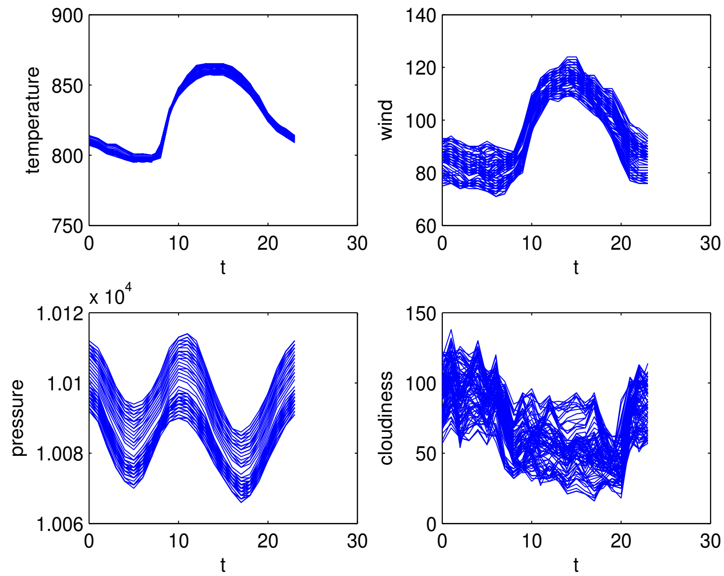

Figure 4.1: The four functional covariates for learning samples.

The scalar response Y is the precipitation every day. In order to study the possible influence, four functional covariates (temperature normals $X_1(t)$, wind chill $X_2(t)$, pressure $X_3(t)$, and cloudiness $X_4(t)$) which are daily curves each hour are studied. We select the climate data of one station over a period of 180 days. The 180 days have been split (randomly) into two subsamples: a learning one (of size $n = 90$) from which the various predictors are computed and a testing one (of size $n = 90$) on which the prediction errors are computed. The four functional covariates for learning samples are plotted in Fig. 4.1. According to method which is introduced in Part a, the numbers of included components $K_1 = 2$, $K_2 = 2$, $K_3 = 1$ and $K_4 = 2$ were chosen by the fraction of variance explained (FVE) with threshold 0.85, respectively.

Table 4.3: MSR of uni-functional linear model for different functional covariates.

Covariates	X_1	X_2	X_3	X_4
MSR	4.7510	12.5386	205.5929	344.8723

We firstly consider the functional linear model with only one functional covariate and compute the MSR in Table 4.3. Table 4.3 shows that the most influent functional variable is the temperature curve and wind chill while the others covariates have quite bad predictive power. However, even if these variables contain just a little information on the target Y it could be interesting to use them in order to improve the results given by the main explanatory variable. From Table 4.3, we can also seen, even the most influent variable (the temperature curve) has also a large MSR using simply linear model, this shows that uni-functional linear model can not reveal such a regression information.

To illustrate the performance of the proposed model, we compute the MSR in Table 4.4 based on model (4.1) for different functional covariates with or without interaction. In the case of with interaction, the best regression modeling is composed by the two covariates: wind chill and pressure curve, which is also consistent with the meteorological knowledge, i.e. the main factors affecting precipitation are wind band and pressure zone. For this regression model, the estimated univariate linear function $\beta_1(t)$ and $\beta_2(t)$ are plotted in Fig. 4.2 and Fig. 4.3 displays the estimated bivariate surface $\gamma(t, s)$. Meanwhile, we compare the regression models with and without the interaction effects by according to the MSR. In result, a large improvement has been obtained when regression models with interaction have been used.

Table 4.4: MSR of the proposed model and the functional linear model with two functional covariates for different functional covariates.

Covariates	With interaction	Without interaction
X_1 and X_2	17.6832	70.5021
X_1 and X_3	41.3530	126.0087
X_1 and X_4	40.6600	97.3201
X_2 and X_3	3.7300	387.3311
X_2 and X_4	66.7202	306.0660
X_3 and X_4	292.7686	744.1033

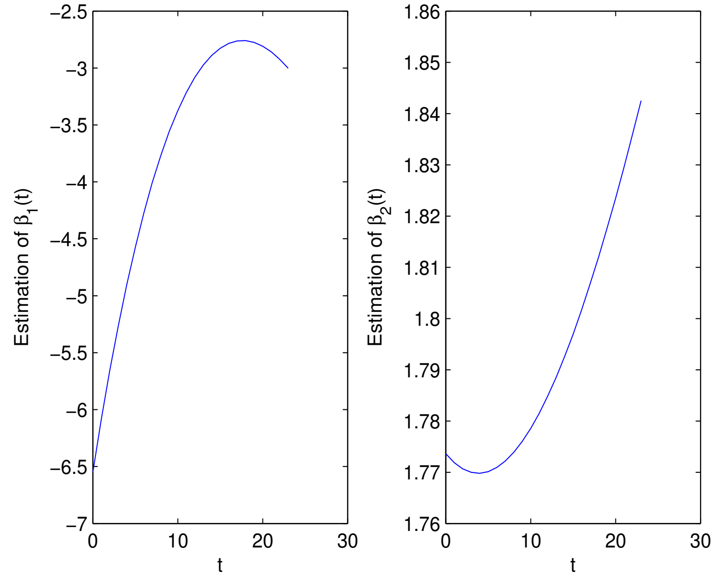

Figure 4.2: The estimated univariate linear function $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$.

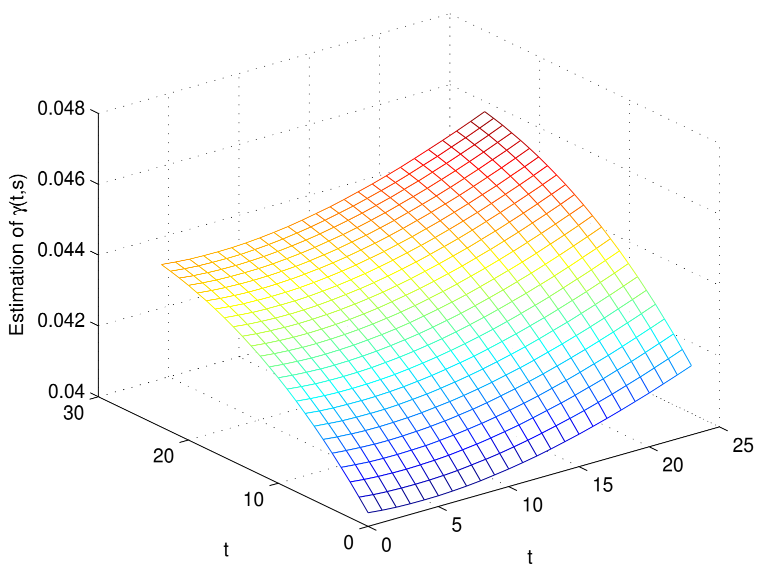

Figure 4.3: The estimated bivariate surface $\hat{\gamma}(t, s)$.

To conclude, the results of Tables 4.3 and 4.4 confirm that the wind chill is the most influent variable, while temperature $(X_1(t))$, pressure normals $(X_3(t))$ and cloudiness $(X_4(t))$ can lead to big and interesting additional information.

4.7 Some additional details and proofs of main results

4.7.1 Part a

To estimate the predictor mean function $\mu_X(s)$ by local linear scatterplot smoothers, one minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} K^{(1)}\left(\frac{s_{ij}-s}{b_X^*}\right)\{R_{ij}-d_0-d_1(s_{ij}-s)\}^2 \tag{4.10}$$

with respect to d_0 and d_1 to obtain $\hat{\mu}_X(s)=d_0(s)$, where the kernel $K^{(1)}$ is assumed to be a smooth symmetric density function and b_X^* is a bandwidth. Analogously, one may define the estimator of the mean function $\mu_Z(t)$.

Let $G_{X,i}(S_{ij},S_{il}) = \{R_{ij}-\hat{\mu}_X(S_{ij})\}\{R_{il}-\hat{\mu}_X(S_{il})\}$, and define the local linear surface smoother for $G_X(s,t)$ by minimizing

$$\sum_{i=1}^{n}\sum_{1\leqslant j\neq l\leqslant N_i} K^{(2)}\left(\frac{S_{ij}-s}{h_X^*},\frac{S_{il}-u}{h_X^*}\right)\left[G_{X,i}\left(S_{ij},S_{il}\right)-f\left\{\alpha,(s,t),(S_{ij},S_{il})\right\}\right]^2 \tag{4.11}$$

where $f(\alpha,(s,u),(S_{ij},S_{il})) = \alpha_0+\alpha_{11}(s-S_{ij})+\alpha_{12}(u-S_{il})$, with respect to $\alpha=(\alpha_0,\alpha_{11},\alpha_{12})$, yielding $\hat{G}_X(s,t)=\hat{\alpha}_0(s,t)$. Here, the kernel $K^{(2)}$ is a two-dimensional smooth density with zero mean and finite covariances and h_X^* is a bandwidth. An essential feature is the omission of the diagonal elements $j=l$ which are contaminated with the measurement errors. Analogously, we can obtain the estimator of the covariance function $G_Z(s,t)$.

Estimates of eigenvalues and eigenfunctions $\{\lambda_k,\phi_k\}_{k\geqslant 1}$ are obtained by numerical solutions $\{\hat{\lambda}_k,\hat{\phi}_k\}_{k\geqslant 1}$ of suitably discretized eigenequations,

$$\int_{\mathcal{T}} \hat{G}_X(s_1,s_2)\hat{\phi}_k(s_2)ds_2 = \hat{\lambda}_k\hat{\phi}_k(s_1) \tag{4.12}$$

with orthonormal constraints on $\{\hat{\phi}_k\}_{k\geqslant 1}$. Analogously the estimates of $\{\tau_k, \psi_k\}_{k\geqslant 1}$ can be obtained.

For the dense design case, the traditional integral estimates of the functional principal components ξ_j and ζ_k defined by

$$\xi_{ij} = \int \{X_i(s) - \mu_X(s)\}\phi_j(s)ds,\ \zeta_{ik} = \int \{Z_i(t) - \mu_Z(t)\}\psi_k(t)dt$$

are

$$\begin{aligned}
\hat{\xi}_{ij} &= \sum_{m=2}^{M_i} \{R_{im} - \hat{\mu}_X(s_{im})\}\hat{\phi}_j(s_{im})(s_{im} - s_{i,m-1}), \quad j = 1, 2, \cdots, \\
\hat{\zeta}_{ik} &= \sum_{q=2}^{N} \{V_{iq} - \hat{\mu}_Z(t_{iq})\}\hat{\psi}_k(t_{iq})(t_{iq} - t_{i,q-1}), \qquad k = 1, 2, \cdots,
\end{aligned} \tag{4.13}$$

For the sparse design case, to estimate ξ_{ij} and ζ_{ik}, we must first estimate σ_X^2 and σ_Z^2. To estimate σ_X^2, we first estimate $V_X(s) = G_X(s,s) + \{\mu_X(s)\}^2 + \sigma_X^2$, by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_1\left(\frac{s_{ij} - s}{\tilde{h}_X}\right)\{R_{ij}^2 - g_0 - g_1(s_{ij} - s)\}^2 \tag{4.14}$$

with respect to $g = (g_0, g_1)$, yielding $\hat{V}_X(s) = \hat{g}_0(s)$. Analogously we can obtain the estimator of $V_Z(t)$.

We then estimate σ_X^2 by

$$\hat{\sigma}_X^2 = \frac{1}{|\mathcal{T}|}\int_{\mathcal{T}} \left[\hat{V}_X(s) - \hat{G}_X(s,s) - \{\hat{\mu}_X(s)\}^2\right] ds \tag{4.15}$$

where $|\mathcal{T}|$ denote the length of interval $\mathcal{T}$. Analogously we can obtain the estimator of σ_Z^2.

Under the sparse design case, the best predictions of functional principal components ξ_{ij} and ζ_{ik} given observations $R_i = (r_{i1}, \cdots, r_{iN_i})^\top$ and $V_i = (v_{i1}, \cdots, v_{iL_i})^\top$

are the conditional expectation $E(\xi_{ij}|R_i)$ and $E(\zeta_{ik}|V_i)$, respectively, which under Gaussian assumptions are found to be

$$\hat{\xi}_{ij} = \hat{\lambda}_j \hat{\phi}_{ij}^T \hat{\Sigma}_{R_i}^{-1}(R_i - \hat{\mu}_{X_i}) \tag{4.16}$$

and

$$\hat{\zeta}_{ik} = \hat{\tau}_k \hat{\psi}_{ik}^T \hat{\Sigma}_{V_i}^{-1}(V_i - \hat{\mu}_{Z_i}), \tag{4.17}$$

where

$$\hat{\mu}_{X_i} = \left[\hat{\mu}_{X_i}(r_{i1}), \cdots, \hat{\mu}_{X_i}(r_{N_i})\right]^T, \ \hat{\mu}_{Z_i} = \left[\hat{\mu}_{Z_i}(v_{i1}), \cdots, \hat{\mu}_{Z_i}(v_{L_i})\right]^T,$$

$$\hat{\phi}_{ij} = \left[\hat{\phi}_j(r_{i1}), \cdots, \hat{\phi}_j(r_{iN_i})\right]^\top, \ \hat{\psi}_{ik} = \left[\hat{\psi}_k(v_{i1}), \cdots, \hat{\psi}_k(v_{iL_i})\right]^\top,$$

$$\left(\hat{\Sigma}_{R_i}\right)_{m,l} = \hat{G}_X(r_{im}, r_{il}) + \hat{\sigma}_X^2 \delta_{ml}, \ \left(\hat{\Sigma}_{V_i}\right)_{m,l} = \hat{G}_Z(v_{im}, v_{il}) + \hat{\sigma}_Z^2 \delta_{ml},$$

and $(\hat{\Sigma}_{R_i})_{m,l}$ and $(\hat{\Sigma}_{V_i})_{m,l}$ are the (m,l)th elements of $\hat{\Sigma}_{R_i}$ and $\hat{\Sigma}_{V_i}$, respectively.

For the choice of the number of included components K_1 and K_2, one may use cross-validation or model selection criteria such as pseudo-BIC (Bayesian information criterion); we adopt the latter, i.e. we may take K_1 and K_2 by minimizing

$$BIC(K_1) = \sum_{i=1}^{n}\sum_{j=1}^{N_i}\left[-\frac{1}{2\hat{\sigma}_X}^2\left\{R_{ij} - \hat{\mu}_X(s_{ij}) - \sum_{k_1=1}^{K_1}\hat{\xi}_{ik_1}\hat{\phi}_{k_1}(s_{ij})\right\}\right]^2 + K_1 log\left(\sum_{i=1}^{n} N_i\right) \tag{4.18}$$

and

$$BIC(K_2) = \sum_{i=1}^{n}\sum_{k=1}^{L_i}\left[-\frac{1}{2\hat{\sigma}_Z}^2\left\{V_{ij} - \hat{\mu}_Z(t_{ik}) - \sum_{k_2=1}^{K_2}\hat{\zeta}_{ik_2}\hat{\psi}_{k_2}(t_{ik})\right\}\right]^2 + K_2 log\left(\sum_{i=1}^{n} L_i\right) \tag{4.19}$$

respectively.

4.7.2 Part b

We require the following conditions for predictor processes.

Let $b_X^* = b_X^*(n)$ and $b_Z^* = b_Z^*(n)$ denote the bandwidths for estimators $\hat{\mu}_X$ and $\hat{\mu}_Z$ respectively. $h_X^* = h_X^*(n)$ and $h_Z^* = h_Z^*(n)$ denote the bandwidths for estimators $\hat{G}_X$ and $\hat{G}_Z$ respectively. $\tilde{h}_X = \tilde{h}_X(n)$ and $\tilde{h}_Z = \tilde{h}_Z(n)$ denote the bandwidths for estimators $\hat{\sigma}_X$ and $\hat{\sigma}_Z$ respectively. The following assumptions are needed.

(A1.1) $b_X^* \to 0$, $h_X^* \to 0$, $nb_X^{*4} \to \infty$, $nh_X^{*4} \to \infty$, $nb_X^{*6} < \infty$ and $nh_X^{*6} < \infty$,

(A1.2) $\tilde{h}_X \to 0$, $n\tilde{h}_X \to \infty$ and $nh_X^{*8} < \infty$.

For processes Z, analogous requirements are

(B1.1) $b_Z^* \to 0$, $h_Z^* \to 0$, $nb_Z^{*4} \to \infty$, $nh_Z^{*4} \to \infty$, $nb_Z^{*6} < \infty$ and $nh_Z^{*6} < \infty$,

(B1.2) $\tilde{h}_Z \to 0$, $n\tilde{h}_Z \to \infty$ and $nh_Z^{*8} < \infty$.

To obtain consistent functional principal component estimates for dense designs, we require both the pooled data across all subjects and the data from each subject to be dense in $\mathcal{T}$. For random process X, denote the sorted time points across all subjects by $a_0 \leqslant S_1 \leqslant S_2 \leqslant \cdots \leqslant S_{\tilde{N}} \leqslant b_0$, and $\Delta^X = max\{S_{(m)} - S_{(m-1)} : m = 1, \ldots, \tilde{N}+1\}$, where $\tilde{N} = \sum_{i=1}^n N_i$, $T = [a_0, b_o]$, $S_0 = a_0$ and $S_{(\tilde{N}+1)} = b_0$. For the ith subject, suppose that the time points S_{ij} have been ordered non-decreasingly. Let $\Delta_i^X = max\{S_{i,j} - S_{i,j-1} : j = 1, \ldots, N_i+1\}$, $\Delta^{X*} = max\{\Delta_i^X : i = 1, \ldots, n\}$ and $\bar{N} = \tilde{N}/n$. Put $N_{max}^X = max\{N_i : i = 1, \ldots, n\}$ and $N_{min}^X = min\{N_i : i = 1, \ldots, n\}$. Assume that

(A2.1) $\Delta^X = O(min\{n^{-1/2} b_X^{*-1}, n^{-1/2}\tilde{h}_X^{-1}, n^{-1/4} h_X^{*-1}\})$,

(A2.2) $\bar{N} \to \infty$, $N_{max}^X \leqslant c_2 \bar{N}$, $\Delta^{X*} = O(1/\tilde{N})$, for some $c_2 > 0$.

For processes Z, we analogously define the quantities $\Delta^Z, \tilde{L}, \Delta_i^Z, \Delta^{Z*}, \bar{L}_Z, L_{max}^Z, L_{min}^Z$, and assume that

(B2.1) $\Delta^Z = O(min\{n^{-1/2}b_Z^{*-1}, n^{-1/2}\tilde{h}_Z^{-1}, n^{-1/4}h_Z^{*-1}\})$,

(B2.2) $\bar{L} \to \infty$, $L_{max}^Z \leqslant c_1\bar{L}, \Delta^{Z*} = O(1/\tilde{L})$, for some $c_1 > 0$.

Denote the distribution that generates R_{ij} the ith subject at S_{ij} by $R_i(s) \sim R(s)$ with density $g_R(r;s)$. Let$g_R(r_1, r_2; s_1, s_2)$ be the density of $(R(s_1), R(s_2))$. Analogously for random process Z, denote the distribution that generates V_{ij}, the ith subject at T_{ij} by $V_i(t) \sim V(t)$ with density $f_V(v;t)$. Let$f_V(v_1, v_2; t_1, t_2)$ be the density of $(V(t_1), V(t_2))$. The following assumptions are for the case of dense designs. Let $f_W(w)$ and $f_U(u)$ be the densities of W and U, respectively.

(A3) $\sup_{s\in S} E[R^4(s)] < \infty$,

(B3) $\sup_{t\in T} E[V^4(t)] < \infty$,

and

(C2,1) $d^2/ds^2 g_R(r,s)$ is uniformly continuous on $\mathcal{R}^1 \times \mathcal{T}$; $d^2/ds_1^{l_1}ds_2^{l_2} g_R(r_1, r_2; s_1, s_2)$ is uniformly continuous on $\mathcal{R}^2 \times \mathcal{T}^2$ for $l_1 + l_2 = 2$, $0 \leqslant l_1, l_2 \leqslant 2$.

(C2,2) $d^2/dt^2 f_V(v,t)$ is uniformly continuous on $\mathcal{R}^1 \times \mathcal{T}$; $d^2/dt_1^{l_1}dt_2^{l_2} f_{V*}(v_1, v_2; t_1, t_2)$ is uniformly continuous on $\mathcal{R}^2 \times \mathcal{T}^2$ for $l_1 + l_2 = 2$, $0 \leqslant l_1, l_2 \leqslant 2$.

(C2,3) The second derivatives $f_W^{(2)}(w)$ exist and are continuous on R, and

$$f_W(w)b^{-K_1K_2} = o\{\min(n^{1/2}b_X^*, n^{1/2}b_Z^*, \bar{N}^{1/2})\}, f_W(w)b_1 = o\{\min(n^{1/2}b_X^*, n^{1/2}b_Z^*, \bar{N}^{1/2})\}.$$

The Fourier transforms of κ_1 and κ_2 are given by $\kappa_1^F(u) = \int exp(-iut)\kappa_1(t)dt$, $\kappa_1^F(u,v) = \int exp(-iut - ivs)\kappa_2(s,t)dsdt$ respectively.

(C2.4) κ_1^F and κ_2^F are absolutely integrable, $\int |\kappa_1^F(u)|du < \infty$, $\int |\kappa_2^F(u,v)|dudv < \infty$.

Let $\|f\|_\infty = sup_{t \in \mathcal{T}}|f(t)|$ for any function with support $\mathcal{T}$, and $\|g\| = \sqrt{\int_{\mathcal{A}} g^2(t)dt}$ for any $g \in L^2(\mathcal{A})$. The following assumptions are needed for Theorem 1.

(A4) $E(\|X'\|_\infty^2) < \infty$, $E(\|X'^2\|_\infty^2) = o(\bar{N})$, and $E(\xi_j^4) < \infty$ for any fixed j.

(B4) $E(\|Z'\|_\infty^2) < \infty$, $E(\|Z'^2\|_\infty^2) = o(\bar{L}_Z)$, and $E(\eta_k^4) < \infty$ for any fixed k.

(D1) $\{(U_i, W_i, Y_i)\}$ are a strongly mixing and stationary sequence with geometric decaying mixing rate $\alpha(k)$.

(D2) With Probability 1, W lies in a compact set Θ. Density functions f_θ of $\theta^\top W$ for any $\|\theta\| = 1$ have bounded derivatives. Regions $\{W : f(W) \geqslant c_0\}$ and $\{W : f_\theta(\theta^\top W) \geqslant c_0\}$ for all $\theta : \|\theta\| = 1$ are non-empty.

(D3) For any perpendicular unit norm vectors θ and ϑ, the joint density function $f(u_1; u_2)$ of $(\theta^\top W, \vartheta^\top W)$ satisfies $f(u_1; u_2)) < cf_{\theta^\top W}(u_1)f_{\vartheta^\top W}(u_2)$, where c is a constant.

(D4) Let $M = (U^\top, W^\top)^\top$, g has bounded, continuous third order derivative. The conditional expectations $E[U|W = w]$, $E[UU^\top|W = w]$, $E[M|\theta^\top W = v]$ and $E[MM^\top|\theta^\top W = v]$ have bounded derivatives. $E[Y^r|W = w]$, $E[|U|^r|W = w]$, $E[|U_l||U_1||W_1 = w_1, W_l = w_l]$ and $E[|U_l||U_1||\theta^\top W_1 = a, \theta^\top W_l = b]$ are bounded by a constant for all $l > 0$, w_1, w_l, w, a, b, where $r > 3$.

(D5) H is a density function with bounded derivative and compact support $\{|w| \leqslant d\}$ for some $d > 0$. K is a symmetric density function with bounded derivative and compact support $[-e_0, e_0]$ for some $e_0 > 0$.

(D6) Matrix $E[\{U - E(U|W)\}\{U - E(U|W)\}^\top]$ is a positive definite matrix.

4.7.3 Part c

Define

$$
\begin{aligned}
&D_X = \int_{\mathcal{T}^2} [\hat{G}_X(s,t) - G_X(s,t)]^2 dsdt,\ D_Z = \int_{\mathcal{T}^2} [\hat{G}_Z(s,t) - G_Z(s,t)]^2 dsdt, \\
&\delta_k^X = min_{1 \leqslant j \leqslant k}(\lambda_j - \lambda_{j+1}),\ \delta_k^Z = min_{1 \leqslant j \leqslant k}(\tau_j - \tau_{j+1}), \\
&\pi_j^X = 1/\lambda_j + 1/\delta_k^X,\ \pi_k^Z = 1/\tau_k + 1/\delta_k^Z.
\end{aligned}
\tag{4.20}
$$

The following lemmas give the weak uniform convergence rates for the estimators of the FPCs, setting the stage for the subsequent developments.

Lemma 4.1. *Suppose that Assumptions (A1.1)-(A3), (B1.1)-(B3) and (C2.1), (C2.2), and (C2.4) hold, we have*

$$
\sup_{t \in T} |\hat{\mu}_X(t) - \mu_X(t)| = O_p\left(\frac{1}{n^{1/2} b_X^*}\right),\ \sup_{t \in T} |\hat{\mu}_Z(t) - \mu_Z(t)| = O_p\left(\frac{1}{n^{1/2} b_Z^*}\right),
$$

$$
\sup_{s,t \in T} |\hat{G}_X(s,t) - G_X(s,t)| = O_p\left(\frac{1}{n^{1/2} h_X^{*2}}\right),\ \sup_{s,t \in T} |\hat{G}_Z(s,t) - G_Z(s,t)| = O_p\left(\frac{1}{n^{1/2} h_Z^{*2}}\right),
$$

and as a consequence, $\hat{\sigma}_X^2 - \sigma_X^2 = O_p(\frac{1}{n^{1/2} h_X^{*2}} + \frac{1}{n^{1/2} \tilde{h}_X})$ *and* $\hat{\sigma}_Z^2 - \sigma_Z^2 = O_p(\frac{1}{n^{1/2} h_Z^{*2}} + \frac{1}{n^{1/2} \tilde{h}_Z})$. *Considering eigenvalues* λ_j, τ_k *of multiplicity one,* $\hat{\phi}_j$ *and* $\hat{\psi}_k$ *can be chosen such that*

$$
P(\sup_{1 \leqslant j \leqslant \tilde{K}_1} |\hat{\lambda}_j - \lambda_j| \leqslant D_X) = 1,\ \sup_{t \in T} |\hat{\phi}_j(t) - \phi_j(t)| = O_p\left(\frac{\pi_j^X}{n^{1/2} h_X^*}\right),
$$

$$
P(\sup_{1 \leqslant k \leqslant \tilde{K}_2} |\hat{\tau}_k - \tau_k| \leqslant D_Z) = 1,\ \sup_{t \in T} |\hat{\psi}_k(t) - \psi_k(t)| = O_p\left(\frac{\pi_k^Z}{n^{1/2} h_Z^*}\right).
$$

The proof of Lemma 4.1 can be found in Müller and Yao (2008).

Recall that $\|f\|_\infty = sup_{t\in\tau}|f(t)|$ for any function with support T, and $\|g\| = \sqrt{\int_A g^2(t)dt}$ for any $g \in L^2(A)$ and define

$$\begin{array}{ll}
\rho_i^{(1)} = c_1\|X_i\| + c_2\|X_iX_i'\|_\infty + c_3, & F^{(1)} = \sup_{s\in S, 1\leqslant j\leqslant K_1}(|\hat{\phi}_j(s) - \phi_j(s)|), \\
\rho_i^{(2)} = 1 + \sup_{1\leqslant j\leqslant K_1} \|\phi_j\phi_j'\|_\infty \Delta^{X*}, & F^{(2)} = \sup_{s\in S} |\hat{\mu}_X(s) - \mu_X(s)|, \\
\rho_i^{(3)} = c_4\|X_i\|_\infty + c_5\|X_i'\|_\infty + c_6, & F^{(3)} = \sup_{1\leqslant j\leqslant K_1}(\|\phi_j'\|)_\infty \Delta^{X*}, \\
\rho_i^{(4)} = |\sum_{r=2}^{N_i} \varepsilon_{ir} \sup_{1\leqslant j\leqslant K_1} \|\phi_j\|_\infty (s_{i,r} - s_{i,r-1})|, & F_j^{(4)} \equiv 1, \\
\rho_i^{(5)} = |\sum_{r=2}^{N_i} |\varepsilon_{ir}|(s_{i,r} - s_{i,r-1}), & F^{(5)} \equiv F_j^{(1)},
\end{array} \tag{4.21}$$

for some positive constants $c_1, \cdots, c_6$ that do not depend on i or j. Similarly, define corresponding quantities for the process Z as follows

$$\begin{array}{ll}
\varrho_i^{(1)} = d_1\|Z_i\| + d_2\|Z_iZ_i'\|_\infty + d_3, & G^{(1)} = \sup_{t\in T, 1\leqslant k\leqslant K_2}(|\hat{\psi}_k(t) - \psi_k(t)|), \\
\varrho_i^{(2)} = 1 + \sup_{1\leqslant k\leqslant K_2}(\|\psi_k\psi_k'\|_\infty) \Delta^{Z*}, & G^{(2)} = \sup_{t\in T} |\hat{\mu}_Z(t) - \mu_Z(t)|, \\
\varrho_i^{(3)} = d_4\|Z_i\|_\infty + d_5\|Z_i'\|_\infty + d_6, & G^{(3)} = \sup_{1\leqslant k\leqslant K_2}(\|\psi_k'\|_\infty) \Delta^{Z*}, \\
\varrho_i^{(4)} = |\sum_{r=2}^{L_i} \varepsilon_{ir} \sup_{1\leqslant k\leqslant K_2}(\|\psi_k\|_\infty)(t_{i,r} - t_{i,r-1})|, & G^{(4)} \equiv 1, \\
\varrho_i^{(5)} = |\sum_{r=2}^{L_i} |\varepsilon_{ir}|(t_{i,r} - t_{i,r-1}), & G^{(5)} \equiv G^{(1)},
\end{array} \tag{4.22}$$

for some positive constants $d_1, \cdots, d_6$ that do not depend on i or k.

The next lemma is critical for the subsequent developments, providing exact upper bounds for the estimation errors $|\hat{\xi}_{ij} - \xi_{ij}|$ and $|\hat{\zeta}_{ik} - \zeta_{ik}|$ for the FPC estimates $\hat{\xi}_{ij}$ and $\hat{\zeta}_{ik}$ in 4.13.

Lemma 4.2. *For $\rho_{ij}^{(\ell)}$, $\varrho_{ik}^{(\ell)}$, $F_j^{(\ell)}$ and $G_k^{(\ell)}$ as defined in (12) and (13), suppose that Assumptions (A1.1)-(A3), (B1.1)-(B3) and (C2.1), (C2.2), and (C2.4) hold, then*

$$|\hat{\xi}_{ij} - \xi_{ij}| \leqslant \sum_{\ell=1}^{5} \rho_i^{(\ell)} F^{(\ell)}, \; |\hat{\zeta}_{ik} - \zeta_{ik}| \leqslant \sum_{\ell=1}^{5} \varrho_i^{(\ell)} G^{(\ell)}.$$

$$|\hat{\xi}_{ij}\hat{\zeta}_{ik} - \xi_{ij}\zeta_{ik}| \leqslant \sum_{\ell=1}^{5} \rho_i^{(\ell)} F^{(\ell)} + \sum_{\ell=1}^{5} \varrho_i^{(\ell)} G^{(\ell)}.$$

The proof of Lemma 4.2 can be found in Müller and Yao (2008).

Define

$$\begin{aligned}
&\rho = f_W(x)\left\{\tfrac{\max_{1\leqslant j\leqslant K_1}\pi_j^X}{\sqrt{n}h_X^*} + \tfrac{1}{\sqrt{n}b_X^*} + \sqrt{\Delta^{X*}}\right\}, && \varrho = f_W(w)\left\{\tfrac{\max_{1\leqslant j\leqslant K_1}\pi_k^Z}{\sqrt{n}h_Z^*} + \tfrac{1}{\sqrt{n}b_Z^*} + \sqrt{\Delta^{Z*}}\right\},\\
&W_i = (\xi_{i1}\zeta_{i1},\cdots,\xi_{i1}\zeta_{K_{i2}},\cdots,\xi_{K_{i1}}\zeta_{K_{i2}})^\top, && \hat{W}_i = \left(\hat{\xi}_{i1}\hat{\zeta}_{i1},\cdots,\hat{\xi}_{i1}\hat{\zeta}_{K_{i2}},\cdots,\hat{\xi}_{K_{i1}}\hat{\zeta}_{K_{i2}}\right)^\top,\\
&p = K_1K_2,
\end{aligned} \tag{4.23}$$

where Δ^{X*} is defined in Part b.

Recall that

$$\begin{aligned}
H_{b,i} &= H\{(w - W_i/b\}/(b^p), & \hat{H}_{b,i} &= H\{(w - \hat{W}_i/b\}/(b^p),\\
K_{b_1,i}^\theta &= K\{(v - \theta^\top W_i)/b_1\}/(b_1), & \hat{K}_{b_1,i}^\theta &= K\{(v - \theta^\top \hat{W}_i)/b_1\}/(b_1),
\end{aligned}$$

To evaluate $|\hat{\beta} - \beta|$ and $|\hat{\theta} - \theta|$, one has to quantify the order of the differences

$$\begin{aligned}
D_1 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i} - H_{b,i}), & D_2 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i} - H_{b,i})Y_i,\\
D_3 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i}\hat{\xi}_{ij} - H_{b,i}\xi_{ij}), & D_4 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i}\hat{\eta}_{ik} - H_{b,i}\eta_{ik})\\
D_5 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i}\hat{\xi}_{ij}\hat{\eta}_{ik} - H_{b,i}\xi_{ij}\eta_{ik}), & D_6 &= \textstyle\sum_{i=1}^n(\hat{H}_{b,i}\hat{\xi}_{ij}^2 - H_{b,i}\xi_{ij}^2).
\end{aligned}$$

Lemma 4.3. *Suppose that Assumptions (A1.1)-(A4), (B1.1)-(B4) and (C2.1)- (C2.4)hold, then*

$$\begin{aligned}
D_1 &= O_p(n(\rho+\varrho)b^{-p}), & D_2 &= O_p(n(\rho+\varrho)b^{-p}),\\
D_3 &= O_p(n(\rho+\varrho)b^{-p}), & D_4 &= O_p(n(\rho+\varrho)b^{-p})\\
D_5 &= O_p(n(\rho+\varrho)b^{-p}), & D_6 &= O_p(n(\rho+\varrho)b^{-p}).
\end{aligned}$$

Proof of Lemma 4.3 Considering D_1, without loss of generality, assume the compact support of H is $[-1,1]^p$. Since H is Lipschitz continuous on its support

$$D_1 \leqslant \frac{c}{b^{2p}}\sum_{i=1}^n\left|\hat{W}_i - W_i\right|\left\{I(|w - W_i| \leqslant b^p) + I(|w - \hat{W}_i| \leqslant b^p)\right\},$$

for some $c > 0$, where $I(\cdot)$ is an indicator function. We then have

$$\begin{aligned}
&\frac{c}{b^{2p}}\sum_{i=1}^n\left|\hat{W}_i - W_i\right| I\left(|w - W_i| \leqslant b^p\right)\\
&\leqslant \sum_{\ell=1}^5 G^{(\ell)}\frac{c}{b^{2p}}\sum_{i=1}^n\varrho_i^{(\ell)}I\left(|w - W_i| \leqslant b^p\right) + \sum_{\ell=1}^5 F^{(\ell)}\frac{c}{b^{2p}}\sum_{i=1}^n\rho_i^{(\ell)}I\left(|w - W_i| \leqslant b^p\right).
\end{aligned}$$

By the central limit theorem for a random number of summands (Billingsley (1995), page 380), observing $\sum_{i=1}^n I(|w-W_i|\leqslant b^p)/nb^p \xrightarrow{p} 2f_W(w)$, we have

$$\frac{c}{nb^p}\sum_{i=1}^n \rho_i^{(\ell)} I(|w-W_i|\leqslant b^p) \xrightarrow{p} 2f_W(w)E\left(\rho_i^{(\ell)}\right),$$

provided that $E(\rho_i^{(\ell)})<\infty$ for $\ell=1,\cdots,5$. Note that $E(\rho_i^{(1)})<\infty$, $E(\rho_i^{(3)})<\infty$ by (A4), $E(\rho_i^{(4)})\leqslant 2\sigma_X\sqrt{\Delta_X^*}$ and $E(\rho_i^{(5)})<|S|\sigma_X$ by the Cauchy-Schwarz inequality. Then

$$F^{(1)}\frac{1}{nb^{2p}}\sum_{i=1}^n \rho_i^{(1)} I(|w-W_i|\leqslant b^p) = O_p\left(\frac{\max_{1\leqslant j\leqslant K_1}\pi_j^X}{n^{1/2}h_X^{*2}b^p}\right)f_W(w),$$

$$F^{(2)}\frac{1}{nb^{2p}}\sum_{i=1}^n \rho_i^{(2)} I(|w-W_i|\leqslant b^p) = O_p\left(\frac{1}{n^{1/2}b_X b^p}\right)f_W(w),$$

$$F^{(3)}\frac{1}{nb^{2p}}\sum_{i=1}^n \rho_i^{(3)} I(|w-W_i|\leqslant b^p) = O_p\left(\frac{\sup_{1\leqslant j\leqslant K_1}\|\phi_j\|_\infty\Delta_X^*}{b^p}\right)f_W(w),$$

$$F^{(4)}\frac{1}{nb^{2p}}\sum_{i=1}^n \rho_i^{(4)} I(|w-W_i|\leqslant b^p) = O_p\left(\frac{\sqrt{\Delta_X^*}}{b^p}\right)f_W(w),$$

$$F^{(5)}\frac{1}{nb^{2p}}\sum_{i=1}^n \rho_i^{(5)} I(|w-W_i|\leqslant b^p) = O_p\left(\frac{\max_{1\leqslant j\leqslant K_1}\pi_k^X}{n^{1/2}h_X^{*2}b^p}\right)f_W(w).$$

We now obtain $\sum_{\ell=1}^5 F^{(\ell)}\frac{c}{nb^p}\sum_{i=1}^n \rho_i^{(\ell)} I(|w-W_i|\leqslant b^p) = O_p(\rho b^{-p})$, using similar arguments, we have $\sum_{\ell=1}^5 G^{(\ell)}\frac{c}{nb^p}\sum_{i=1}^n \varrho_i^{(\ell)} I(|w-W_i|\leqslant b^p) = O_p(\varrho b^{-p})$, then $\frac{1}{nb^{2p}}\sum_{i=1}^n |\hat{w}_i-W_i| I(|w-W_i|\leqslant b^p) = O_p((\rho+\varrho)b^{-p})$. The asymptotic rate of the second term can be derived analogously. Observing

$$\begin{aligned}
&\frac{1}{nb^p}\sum_{i=1}^n I\left(|w-\hat{W}_i|\leqslant b^p\right)\\
&\leqslant \frac{1}{nb^p}\sum_{i=1}^n\left[I\left(|w-W_i|\leqslant 2b^p\right)+I\left(\sum_{\ell=1}^5\left\{\rho_i^{(\ell)}F^{(\ell)}+\varrho_i^{(\ell)}G^{(\ell)}\right\}>b^p\right)\right] \xrightarrow{p} 4f_W(w).
\end{aligned}$$

This implies $\frac{1}{nb^{2p}}\sum_{i=1}^{n}|\hat{W}_i - w_i|I(|w - \hat{W}_i| \leqslant b^p) = O_p((\rho + \varrho)b^{-p})$. Then $D_1 = O_p((\rho + \varrho)b^{-p})$.

Analogously, one can show $D_2 = O_p((\rho+\varrho)b^{-p})$ by the Cauchy-Schwarz inequality for $\rho_i^{(\ell)}$ and $\varrho_i^{(\ell)}$, $\ell = 1, 3$ and observing the independence between Y_i and $\rho_i^{(\ell)}, \varrho_i^{(\ell)}$, $\ell = 1, 3$, given the moment condition (A4). For D_3, we have

$$
\begin{aligned}
D_3 &= \sum_{i=1}^{n}\left\{\left(\hat{H}_{b,i} - H_{b,i}\right)\xi_{ij} + \left(\hat{H}_{b,i} - H_{b,i}\right)\left(\hat{\xi}_{ij} - \xi_{ij}\right) + W_i\left(\hat{\xi}_{ij} - \xi_{ij}\right)\right\} \\
&= D_{31} + D_{32} + D_{33}.
\end{aligned}
$$

We have $D_{31} = O_p((\rho + \varrho)b^{-p})$, using the arguments similar to D_1. It is easy to see that $D_{32} = o_p(D_{31})$. Since $D_{33} \leqslant c\sum_{\ell=1}^{5} F_j^{(\ell)}\frac{1}{nb^p}\sum_{i=1}^{n}\rho_{ij}^{(\ell)}I(|w - W_i| \leqslant b^p) + \sum_{\ell=1}^{5} G_k^{(\ell)}\sum_{i=1}^{n}\varrho_{ik}^{(\ell)}I(|w - w_i| \leqslant b^p)$ for some $c > 0$, one also has $D_{33} = o_p(D_{31})$. This results in $D_3 = O_p((\rho + \varrho)b^{-p})$. Analogously, one shows $D_4 = O_p((\rho + \varrho_k)b^{-p})$, and $D_5 = O_p((\rho + \varrho)b^{-p})$. Observing $|\hat{\xi}_{ij}^2 - \xi_{ij}^2| \leqslant |\hat{\xi}_{ij} - \xi_{ij}||\xi_{ij}| + (\hat{\xi}_{ij} - \xi_{ij})^2$, one can show $D_6 = O_p((\rho + \varrho)b^{-p})$, using similar arguments to that of D_3.

Lemma 4.4. *Suppose that Assumptions (A1.1)-(A4), (B1.1)-(B4) and (C2.1)-(C2.4)* *hold, then*

$$E_1 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} - K_{b_1,i}^{\theta} \right) = O_p\left((\rho + \varrho) b_1 \right),$$

$$E_2 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} - K_{b_1,i}^{\theta} \right) Y_i = O_p\left((\rho + \varrho) b_1 \right),$$

$$E_3 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} \hat{\xi}_{ij} - K_{b_1,i}^{\theta} \xi_{ij} \right) = O_p\left((\rho + \varrho) b_1 \right),$$

$$E_4 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} \hat{\eta}_{ik} - K_{b_1,i}^{\theta} \eta_{ik} \right) = O_p\left((\rho + \varrho) b_1 \right),$$

$$E_5 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} \hat{\xi}_{ij} \hat{\eta}_{ik} - K_{b_1,i}^{\theta} \xi_{ij} \eta_{ik} \right) = O_p\left((\rho + \varrho) b_1 \right),$$

$$E_6 = \sum_{i=1}^{n} \left(\hat{K}_{b_1,i}^{\theta} \hat{\xi}_{ij}^2 - K_{b_1,i}^{\theta} \xi_{ij}^2 \right) = O_p\left((\rho + \varrho) b_1 \right).$$

The proof of Lemma 4.4 is similar to that of Lemma 4.3.

Let $\delta_\theta = |\theta - \theta_0|$, $\delta_\beta = |\beta - \beta_0|$, $\delta_\chi = \delta_\theta + \delta_\beta$, $p = K_1 K_2$, $\delta_{pn} = \{logn/(nb^p)\}^{1/2}$, $\tau_{pn} = b^2 + \delta_{pn}$, $\delta_n = \{logn/(nb_1)\}^{1/2}$, $\tau_n = b_1^2 + \delta_n$, $\Theta = \{\theta : |\theta| = 1\}$. Suppose A_n is a matrix. $A_n = O(a_n)$ (or $o(a_n)$) means every element in A_n is $O(a_n)$ (or $o(a_n)$) almost surely. We abbreviate$K_h(\theta^\top W_{i0})$ and $H_b(W_{i0})$ as $K_{h,i}^{\theta}(w)$ (or $K_{h,i}^{\theta}$) and $H_{b,i}(w)$ (or $H_{b,i}$) respectively in the following context. We take $G(\cdot) \equiv 1$ in the proofs for simplicity. We further assume that $\kappa_2 \equiv \int K(v)v^2 = 1$ and $\mathrm{H}_2 \equiv \int H(U)UU^\top dU = I_{p\times p}$.

In the following context, we abbreviate L for any function $L(x)$, and L_θ or $L_\theta(x)$ for any function $L_\theta(\theta^\top x)$. Let

$$\begin{array}{lll}
\hat{\varsigma}_0 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}, & \hat{S}_1 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{W}_{i,0}, & \hat{S}_2 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{W}_{i,0}\hat{W}_{i,0}^\top \\
\hat{T}_1 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{U}_i, & \hat{T}_2 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{U}_i\hat{U}_i^\top, & \hat{C}_2 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{W}_i\hat{U}_i^\top, \\
\hat{E}_1 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{U}_i Y_i, & \hat{D}_1 = \frac{1}{n}\sum_{i=1}^n \hat{H}_{b,i}\hat{W}_{i,0}Y_i, & \hat{L}_n = \hat{\varsigma}_0\hat{S}_2 - \hat{S}_1\hat{S}_1^\top.
\end{array}$$

and

$$\hat{\hat{\omega}}^{\theta}_{a,i}(w) = \left\{\theta^\top \hat{S}_2\theta\right\} \hat{H}_{b,i} - \theta^\top \hat{S}_1 \hat{H}_{b,i}\theta^\top \hat{W}_{i,0},\ \hat{\hat{\omega}}^{\theta}_{d,i}(w) = \hat{\varsigma}_0 \hat{H}_{b,i}\theta^\top \hat{W}_{i,0}.$$

We can obtain initial estimators of θ_0 and β_0 as follows. Choose a vector θ with norm 1 and any vector β. Let $\hat{\hat{\omega}}^{\theta}_j = \theta^\top \hat{L}_n(\hat{W}_j)\theta$ and calculate

$$\begin{aligned} \hat{\hat{a}}^{\theta}_j &= \{\hat{\hat{\omega}}^{\theta}_j\}^{-1} \sum_{i=1}^{n} \hat{\hat{\omega}}^{\theta}_{a,i}(\hat{W}_j)\{Y_i - \beta^\top \hat{U}_i\}, \\ \hat{\bar{d}}^{\theta}_j &= \{\hat{\hat{\omega}}^{\theta}_j\}^{-1} \sum_{i=1}^{n} \hat{\hat{\omega}}^{\theta}_{d,i}(\hat{W}_j)\{Y_i - \beta^\top \hat{U}_i\}, \end{aligned} \tag{4.24}$$

$$\begin{aligned} \begin{pmatrix} \hat{\hat{\beta}} \\ \hat{\hat{\theta}} \end{pmatrix} &= \{\hat{\hat{D}}^{\theta}_n\}^{-1} \sum_{j=1}^{n} I_n(\hat{W}_j) \begin{pmatrix} \hat{E}_1(\hat{W}_j) - \hat{\hat{a}}^{\theta}_j \hat{T}_1(\hat{W}_j) \\ \hat{\bar{d}}^{\theta}_j \hat{D}_1(\hat{W}_j) - \hat{\hat{a}}^{\theta}_j \hat{\bar{d}}^{\theta}_j \hat{S}_1(\hat{W}_j) \end{pmatrix} \Big/ \hat{\varsigma}_0(\hat{W}_j), \\ \check{\theta} &:= sgn_1 \check{\theta}/|\check{\theta}|. \end{aligned} \tag{4.25}$$

where sgn_1 is the sign of first entry of $\hat{\hat{\theta}}$ and

$$\hat{\hat{D}}^{\theta}_n = \textstyle\sum_{j=1}^{n} I_n(\hat{W}_j) \begin{pmatrix} \hat{T}_2(\hat{W}_j) & \bar{d}^{\theta}_j \hat{C}_2(\hat{W}_j) \\ \bar{d}^{\theta}_j \hat{C}_2^\top(\hat{W}_j) & (\bar{d}^{\theta}_j)^2 \hat{S}_2(\hat{W}_j) \end{pmatrix} \Big/ \hat{\varsigma}_0(\hat{W}_j),$$

and A^{-1} denotes the Moore-Penrose inverse of matrix A. Repeat the calculations in (4.24) and (4.25) with (θ, β) replaced by $(\hat{\hat{\theta}}, \hat{\hat{\beta}})$ until convergence. Denote the final value by $((\check{\theta}, \check{\beta})$.

Lemma 4.5. *Under all the Assumptions listed in Part b, as $b \to 0$ and $nb^{K_1K_2+2} \to \infty$, if we start the estimation procedure with θ such that $\theta^\top\theta_0 \neq 0$, then*

$$\check{\theta} - \theta_0 = O_p\left(b + b^{-1}\delta_{pn} + (\rho + \varrho)b^{-p}\right)$$

and

$$\check{\beta} - \beta_0 = O_p\left(b + b^{-1}\delta_{pn} + (\rho + \varrho)b^{-p}\right),$$

where ρ, ϱ are defined in (4.23), $p = K_1K_2$ and $\delta_{pn} = \{\log n/(nb^p)\}^{1/2}$.

Proof of Lemma 4.5 Let $\hat{\bar{\triangle}}_i(w) = Y_i - \check{a} - \beta_0^\top U_i - \check{d}W_{i0}^\top\theta_0$. We have

$$\begin{pmatrix} \check{\beta} \\ \check{\theta} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \theta_0 \end{pmatrix} + \hat{\bar{D}}_n^{-1}(\theta) \sum_{j=1}^n I_n\left(\hat{W}_j\right) \sum_{i=1}^n \hat{H}_{b,i}\left(\hat{W}_j\right) \begin{pmatrix} \hat{U}_i \\ \hat{W}_{ij}\hat{\bar{d}}_j \end{pmatrix} \hat{\bar{\triangle}}_i\left(\hat{W}_j\right) \Big/ \hat{\varsigma}_0\left(\hat{W}_j\right),$$

By Lemma 4.3, we can obtain

$$\hat{\bar{D}}_n^\theta - \bar{D}_n^\theta = O_p\left((\rho+\varrho)b^{-p}\right),$$

$$\hat{\varsigma}_0(\hat{W}_j) - \varsigma_0(W_j) = O_p\left((\rho+\varrho)b^{-p}\right),$$

$$\sum_{i=1}^n \hat{H}_{b,i} \begin{pmatrix} \hat{U}_i \\ \hat{W}_{ij}\hat{\bar{d}}_j \end{pmatrix} \hat{\bar{\triangle}}_i\left(\hat{W}_j\right) - \sum_{i=1}^n H_{b,i} \begin{pmatrix} U_i \\ W_{ij}\bar{d}_j \end{pmatrix} \bar{\triangle}_i\left(W_j\right) = O_p\left((\rho+\varrho)b^{-p}\right),$$

where $\bar{D}_n^\theta$, $\bar{d}_j$, $\bar{\triangle}_i$ and $\varsigma_0(W_j)$ are defined in Xia and Härdle (2006). It is easy to show

$$\begin{pmatrix} \check{\beta} \\ \check{\theta} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \theta_0 \end{pmatrix} + \bar{D}_n^{-1}(\theta) \sum_{j=1}^n I_n(W_j) \sum_{i=1}^n H_{b,i}(W_j) \begin{pmatrix} U_i \\ W_{ij}\bar{d}_j \end{pmatrix} \bar{\triangle}_i(W_j) \Big/ \varsigma_0(W_j)$$

$$+ O_p\left((\rho+\varrho)b\right).$$

Applying the method similar to Xia and Härdle (2006), we can obtain

$$\check{\theta} - \theta_0 = O_p(b + b^{-1} + (\rho+\varrho)b), \check{\beta} - \beta_0 = O_p(b + b^{-1} + (\rho+\varrho)b).$$

The proof is completed.

Next, we shall improve the efficiency of the estimators using a univariate kernel.

Let

$$\begin{aligned}
&\hat{\varsigma}_k^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\{\theta^\top\hat{W}_{i0}\}^k, k = 0,1,2,3, && \\
&\hat{\omega}_{a,i}^\theta = \hat{\varsigma}_2^\theta\hat{K}_{h,i}^\theta - \hat{\varsigma}_1^\theta\hat{K}_{h,i}^\theta\theta^\top\hat{W}_{i0}, && \hat{\omega}_{d,i}^\theta = \hat{\varsigma}_0^\theta K_{h,i}^\theta\theta^\top\hat{W}_{i0} - \hat{\varsigma}_1^\theta\hat{K}_{h,i}^\theta, \\
&\hat{\omega}^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{\omega}_{a,i}^\theta, && \hat{S}_1^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{W}_{i0}, \\
&\hat{S}_2^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta W_{i0}W_{i0}^\top, && \hat{T}_1^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{U}_i, \\
&\hat{E}_1^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{U}_iY_i, && \hat{D}_1^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{W}_{i0}Y_i, \\
&\hat{T}_2^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{U}_iU_i^\top, && \hat{C}_2^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\theta^\top W_{i0}\hat{U}_i^\top, \\
&\hat{S}_{1,1}^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\{\theta^\top\hat{W}_{i0}\}\hat{W}_{i0}, && \hat{S}_{2,1}^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\{\theta^\top\hat{W}_{i0}\}^2\hat{W}_{i0}, \\
&\hat{S}_{1,2}^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\{\theta^\top\hat{W}_{i0}\}\hat{W}_{i0}\hat{W}_{i0}^\top, && \hat{S}_3^\theta = \tfrac{1}{n}\textstyle\sum_{i=1}^n \hat{K}_{h,i}^\theta\hat{W}_{i0}\{(\theta-\theta_0)^\top\hat{W}_{i0}\}^2.
\end{aligned}$$

We improve the estimators $\check{\theta}$ and $\check{\beta}$ as follows. Let $\hat{\omega}_j^\theta = \hat{\omega}^\theta(\hat{W}_j)$. Starting with $(\theta, \beta) = (\check{\theta}, \check{\beta})$, calculate

$$\begin{aligned} \dot{\hat{a}}_j^\theta &= \{\hat{\omega}_j^\theta\}^{-1} \sum_{i=1}^n \hat{\omega}_{a,i}^\theta(\hat{W}_j)\{Y_i - \beta^\top \hat{U}_i\}, \\ \dot{\tilde{d}}_j^\theta &= \{\hat{\omega}_j^\theta\}^{-1} \sum_{i=1}^n \hat{\omega}_{d,i}^\theta(\hat{W}_j)\{Y_i - \beta^\top \hat{U}_i\}, \end{aligned} \tag{4.26}$$

$$\begin{aligned} \begin{pmatrix} \dot{\tilde{\beta}} \\ \dot{\tilde{\theta}} \end{pmatrix} &= \{\hat{\tilde{D}}_n^\theta\}^{-1} \sum_{j=1}^n I_n(\hat{W}_j) \begin{pmatrix} \hat{E}_1^\theta(\hat{W}_j) - \dot{\hat{a}}_j^\theta T_1^\theta(\hat{W}_j) \\ \dot{\tilde{d}}_j^\theta \hat{D}_1^\theta(\hat{W}_j) - \dot{\hat{a}}_j^\theta \dot{\tilde{d}}_j^\theta \hat{S}_1^\theta(\hat{W}_j) \end{pmatrix} \hat{\varsigma}_0^\theta \Big/ (\hat{W}_j), \\ \tilde{\theta} &:= sgn_1 \tilde{\theta}/|\tilde{\theta}|, \end{aligned} \tag{4.27}$$

where sgn_1 is the sign of first entry of $\dot{\tilde{\theta}}$ and

$$\hat{\tilde{D}}_n^\theta = \textstyle\sum_{j=1}^n I_n(\hat{W}_j) \begin{pmatrix} \hat{T}_2^\theta(\hat{W}_j) & \tilde{d}_j^\theta \hat{C}_2^\theta(\hat{W}_j) \\ \tilde{d}_j^\theta\{C_2^\theta(\hat{W}_j)^\top\} & (\tilde{d}_j^\theta)^2 \hat{S}_2^\theta(\hat{W}_j) \end{pmatrix} \Big/ \hat{\varsigma}_0^\theta(\hat{W}_j).$$

Repeat the procedure (4.26) and (4.27) with (θ, β) replaced by $(\dot{\tilde{\theta}}, \dot{\tilde{\beta}})$ until convergence. Denote the final value by $(\hat{\tilde{\theta}}, \hat{\tilde{\beta}})$ and denote the final value of $\dot{\hat{a}}_j^\theta$ by $\hat{\hat{a}}_j^\theta$.

Lemma 4.6. *Let $(\hat{\tilde{\beta}}, \hat{\tilde{\theta}})$ be the estimators based on the single-index kernel weight starting with $(\beta, \theta) = (\check{\beta}, \check{\theta})$. Under all the assumptions listed in Part b, as $b_1 \sim n^\delta$ with $1/6 < \delta < 1/4$ and that $E(\varepsilon_i | U_j, W_j, Y_j, j < i) = 0$ almost surely, we have*

$$\begin{aligned} \hat{\tilde{\theta}} - \theta_0 &= O_p\left((\rho + \varrho) b_1\right), \\ \hat{\tilde{\beta}} - \beta_0 &= O_p\left((\rho + \varrho) b_1\right). \end{aligned}$$

Proof of Lemma 4.6 Let $\hat{\tilde{\triangle}}_i(w) = Y_i - \dot{\hat{a}} - \beta_0^\top \hat{U}_i - \dot{\tilde{d}} \hat{W}_{i0}^\top \theta_0$, we have

$$\begin{pmatrix} \hat{\tilde{\beta}} \\ \hat{\tilde{\theta}} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \theta_0 \end{pmatrix} + \hat{\tilde{D}}_n^{-1}(\theta) \textstyle\sum_{j=1}^n I_n\left(\hat{W}_j\right) \sum_{i=1}^n \hat{K}_{h,i}^\theta\left(\theta^\top \hat{W}_j\right) \begin{pmatrix} \hat{U}_i \\ \hat{W}_{ij} \dot{\tilde{d}}_j \end{pmatrix} \hat{\tilde{\triangle}}_i\left(\hat{W}_j\right) \Big/ \hat{\varsigma}_0^\theta\left(\hat{W}_j\right).$$

By Lemma 4.4, we can obtain

$$\hat{\tilde{D}}_n^\theta - \tilde{D}_n^\theta = O_p\left((\rho+\varrho)b_1\right),$$

$$\hat{\varsigma}_0^\theta(\hat{W}_j) - \varsigma_0^\theta(W_j) = O_p\left((\rho+\varrho)b_1\right),$$

$$\sum_{i=1}^n \hat{K}_{h,i}^\theta(\hat{W}_j)\left(\begin{array}{c}\hat{U}_i\\ \hat{W}_{ij}\dot{\tilde{d}}_j\end{array}\right)\hat{\tilde{\triangle}}_i(\hat{W}_j) - \sum_{i=1}^n K_{h,i}^\theta(W_j)\left(\begin{array}{c}U_i\\ W_{ij}\tilde{d}_j\end{array}\right)\tilde{\triangle}_i(W_j) = O_p\left((\rho+\varrho)b_1\right),$$

where $\tilde{D}_n^\theta$, $\tilde{\triangle}_i$, $\tilde{d}_j$ and $\varsigma_0^\theta(W_j)$ are defined in Xia and Härdle (2006)

By Assumption (A2.2) and (B2.2), applying the method similar to Xia and Härdle (2006), we can obtain

$$\begin{aligned}\hat{\tilde{\theta}} - \theta_0 &= O_p\left((\rho+\varrho)b_1\right) + O_p\left(n^{-1/2}\right)\\ &= O_p\left((\rho+\varrho)b_1\right),\\ \hat{\tilde{\beta}} - \beta_0 &= O_p\left((\rho+\varrho)b_1\right) + O_p\left(n^{-1/2}\right)\\ &= O_p\left((\rho+\varrho)b_1\right).\end{aligned} \tag{4.28}$$

Proofs of Theorems

Let

$$\begin{array}{ll}\|\beta_X\|^2 = \int_T \beta_X^2(s)ds, & \|\beta_Z\|^2 = \int_T \beta_Z^2(t)dt,\\ \beta_X^{K_1}(s) = \sum_{j=1}^{K_1}\beta_X^j\phi_j(s), & \beta_Z^{K_2}(t) = \sum_{k=1}^{K_2}\beta_Z^k\psi_k(t),\\ \|\gamma\|^2 = \int_T\int_T \gamma^2(s,t)dsdt, & \gamma(s,t) = \sum_{j=1}^{K_1}\sum_{k=1}^{K_2}\gamma_{jk}\phi_j\psi_k\end{array}$$

We note that the square integrability properties imply, as $n\to\infty$,

$$R_{\beta_X,K_1} = \|\beta_X - \beta_X^{K_1}\| \to 0,$$

$$R_{\beta_Z,K_2} = \|\beta_Z - \beta_Z^{K_2}\| \to 0,$$

$$R_\gamma = \|\gamma - \gamma_{K_1,K_2}\| \to 0.$$

The weak convergence rates Ψ_{β_X}, Ψ_{β_Z} and Ψ_γ of the regression function estimators $\hat{\beta}_X$, $\hat{\beta}_Z$ and $\hat{\gamma}$ are as follows,

$$\begin{aligned}
\Psi_{\beta_X} &= \sum_{j=1}^{K_1} \frac{\pi_j^X}{n^{1/2} h_X^*} + (\rho + \varrho) b_1 + R_{\beta_X, K_1}, \\
\Psi_{\beta_Z} &= \sum_{k=1}^{K_2} \frac{\pi_j^Z}{n^{1/2} h_Z^*} + (\rho + \varrho) b_1 + R_{\beta_Z, K_2}, \\
\Psi_\gamma &= \sum_{j=1}^{K_1} \sum_{k=1}^{K_2} \left(\frac{\pi_j^X}{n^{1/2} h_X^*} + \frac{\pi_k^Z}{n^{1/2} h_Z^*} \right) + (\rho + \varrho) b_1) + R_\gamma,
\end{aligned} \tag{4.29}$$

respectively, where h_X^* and h_Z^* are the bandwidths for estimating G_X and G_Z respectively. π_j^X and π_k^Z are defined in Part b.

Proof of Theorem 1 Using Lemmas 4.1 and 4.6, it is easy to see

$$\begin{aligned}
&\left\| \hat{\beta}_X - \beta_X \right\| \\
&\leqslant \sum_{j=1}^{K_1} \left\| \hat{\hat{\beta}}_X^j \hat{\phi}_j(s) - \beta_X^j \phi_j(s) \right\| + \left\| \sum_{j=1}^{K_1} \beta_X^j \phi_j(s) \right\| \\
&\leqslant \sum_{j=1}^{K_1} \left[\left| \hat{\hat{\beta}}_X^j - \beta_X^j \right| \|\phi_j(s)\| + \left| \hat{\hat{\beta}}_X^j - \beta_X^j \right| \left\| \hat{\phi}_j(s) - \phi_j(s) \right\| + \left| \beta_X^j \right| \left\| \hat{\phi}_j(s) - \phi_j(s) \right\| \right] \\
&\quad + \left\| \sum_{j=1}^{K_1} \beta_X^j \phi_j(s) \right\| \\
&= O_p\left(\sum_{j=1}^{K_1} \frac{\pi_j^X}{n^{1/2} h_X^*} + (\rho + \varrho) b_1 + R_{\beta_X, K_1} \right)
\end{aligned}$$

Analogously, one shows $\|\hat{\beta}_Z - \beta_Z\| = O_p(\sum_{k=1}^{K_2} \frac{\pi_j^Z}{n^{1/2} h_Z^*} + (\rho + \varrho) b_1 + R_{\beta_Z, K_2})$.

It can also be observed

$$
\begin{aligned}
&\|\hat{\gamma}(s,t)-\gamma(s,t)\| \\
&\leqslant \sum_{j=1}^{K_1}\sum_{k=1}^{K_2}\left[\left|\hat{\hat{\gamma}}_{j,k}-\gamma_{j,k}\right|\|\phi_j(s)\psi_k(t)\| + \left|\hat{\hat{\gamma}}_{j,k}-\gamma_{j,k}\right|\left\|\hat{\phi}_j(s)\hat{\psi}_k(t)-\phi_j(s)\psi_k(t)\right\|\right. \\
&\quad \left. + |\gamma_{j,k}|\left\|\hat{\phi}_j(s)\hat{\psi}_k(t)-\phi_j(s)\psi_k(t)\right\|\right] + \left\|\sum_{j=1}^{K_1}\sum_{k=1}^{K_2}\gamma_{j,k}\phi_j(s)\psi_k(t)\right\| \\
&= O_p\left(\sum_{j=1}^{K_1}\frac{\pi_j^X}{n^{1/2}h_X^*} + \sum_{k=1}^{K_2}\frac{\pi_k^Z}{n^{1/2}h_Z^*} + (\rho+\varrho)b_1 + R_\gamma\right).
\end{aligned}
$$

The proof is completed.

The weak convergence rates Ψ_g of the nonparametric function estimator $\hat{g}(\cdot)$ is as follows,

$$
\Psi_g = 1/2\kappa_2 g^{''}(v) + 1/\sqrt{nb_1} + (\rho+\varrho)b_1. \tag{4.30}
$$

Proof of Theorem 2 Let $\hat{a}_j^\theta$ is the final estimator of a_j and $\hat{g}^*(v)$ is the estimator of $g(v)$ in Xia and Härdle (2006), we have

$$
\begin{aligned}
&\hat{g}(v)-\hat{g}^*(v) \\
&= \frac{\sum_{i=1}^n \hat{\hat{a}}_j^\theta K\left\{\left(v-\hat{\hat{\theta}}\hat{W}_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\hat{\theta}}\hat{W}_i\right)/b_1\right\}} - \frac{\sum_{i=1}^n \hat{a}_j^\theta K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}} \\
&= \left[\frac{\sum_{i=1}^n \hat{\hat{a}}_j^\theta K\left\{\left(v-\hat{\hat{\theta}}\hat{W}_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\hat{\theta}}\hat{W}_i\right)/b_1\right\}} - \frac{\sum_{i=1}^n \hat{\hat{a}}_j^\theta K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}}\right] \\
&\quad + \left[\frac{\sum_{i=1}^n \hat{\hat{a}}_j^\theta K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}} - \frac{\sum_{i=1}^n \hat{a}_j^\theta K\left\{\left(v-\hat{\theta}W_iW_i\right)/b_1\right\}}{\sum_{i=1}^n K\left\{\left(v-\hat{\theta}W_i\right)/b_1\right\}}\right] \\
&\equiv A_1 + A_2.
\end{aligned} \tag{4.31}
$$

By Assumption (D5) and (4.28), we obtain

$$
\begin{aligned}
A_1 &\leqslant c\left|\hat{\hat{\theta}}\hat{W}_i - \hat{\theta}W_i\right| \\
&= O_p\left((\rho+\varrho)b_1\right).
\end{aligned}
\tag{4.32}
$$

According to Lemma 4.4, we have

$$
\begin{aligned}
A_2 &\leqslant c\left|\hat{\hat{a}}_j^{\theta} - \hat{a}_j^{\theta}\right| \\
&= O_p\left((\rho+\varrho)b_1\right).
\end{aligned}
\tag{4.33}
$$

By(4.31), (4.32) and (4.33), we obtain $\hat{g}(v) - \hat{g}^*(v) = O_p((\rho+\varrho)b_1)$.

According to Theorem 1 of Xia and Härdle (2006), we can obtain

$$
\hat{g}(v) - g(v) = 1/2\kappa_2 g^{''}(v) + O_p\left(1/\sqrt{nb_1} + (\rho+\varrho)b_1\right).
$$

The proof is completed.

The weak convergence rates Φ_E of the predictions $\hat{E}[Y|X,Z]$ is as follows,

$$
\begin{aligned}
\Phi_E =\ & 1/2\kappa_2 g^{''}\left(\sum_{j=1}^{K_1}\sum_{k=1}^{K_2}\gamma_{jk}\xi_j^{*I}\zeta_k^{*I}\right) + \sum_{j=1}^{K_1}\frac{\pi_j^X}{n^{1/2}h_X^*} + \sum_{k=1}^{K_2}\frac{\pi_j^Z}{n^{1/2}h_Z^*} + K_1 N^{-1/2} + K_2 M^{-1/2} \\
& + (\rho+\varrho)b_1 + R_{\beta_X,K_1} + R_{\beta_Z,K_2} + R_\gamma + 1/\sqrt{nh}.
\end{aligned}
\tag{4.34}
$$

Chapter 5

Testing Equality of Covariance Operator for Two-sample Functional Data

5.1 Introduction

Tests of significance are essential statistical problems and are challenging particularly for functional data analysis drawing accumulated research attention, see Ledoit and Wolf (2002), Berkes et al. (2009), Benko et al. (2009), Panaretos et al. (2010), Arias-Castro et al. (2011), Horváth and Kokoszka (2012), and Horváth et al. (2013), among others. It is natural to validate whether covariance operators or matrices of two populations are equal or not before further analysis, see pp.49-53 in Ferraty (2011) and Fremdt et al. (2013), among others. Thereafter, we use the word covariance operator as was named in Ferraty and Vieu (2006). The term covariance function is also used in other monograph for functional data analysis. In this chapter, we address new methodology for significance testing of covariance operators for the stochastic process in functional data analysis. Our methodology is specifically designed for the situations where the timing of recordings is sparse and irregularly spaced, say some longitudinal data studies, refer to Zhao et al. (2004), Zhu et al. (2011), and Chen et al. (2013), but it works well also for the cases where the recordings of the curves are scheduled on a regular and dense grid.

A primary motivation for this part of this research comes from a study on comparison of emission levels of SO_2 and $PM2.5$, two well-known main air pollutant indices, detected in some southwestern industrial area of China in 2013. In industrial area, SO_2 is primarily caused by combustion procedure such as burning fuels, electric utilities, and other industrial activities. The dirty air in China, known as smog, has been blamed by World Health Organization(WHO) and the public. More policies and studies have been conducted to do against the air pollution for public health. The afore city governmental bureau invigilated and monitored the arising issue of environmental detection and protection. Therefore, by an environmental detection station in the center of the city, hourly and daily in the whole year of 2013, there are official records of emission levels of SO_2 and $PM2.5$. The hourly measurements in a day(24 hours) form natural emission level curves of the day. However, in some days, all hourly data could be observed, whereas in other days, the data recorders could only gain incompletely observed hourly data because of detecting machines out of run or meter burst by high pollutant levels. This incurred the recorded functional data containing multiple types: sometimes it is regular and dense, but sometimes it is irregular and sparse. The scatterplot of 3 selected days for SO_2 levels and 50 incompletely observed $PM2.5$ levels are displayed in Fig. 5.1. One of our interest in studying this air pollutant data is to test the equality of covariance operators of SO_2 and $PM2.5$ in working days and in non-working days or varying seasonally. More details will be stated and analyzed in the first application in this chapter.

For testing equality of covariance operators in functional data samples, although this problem is important, it is challenging and related research progress is quite limited. Benko et al. (2009) developed bootstrap procedures for testing the equality of specific functional principal components which was equivalent to testing if covariance operators were equal. Horváth et al. (2009) compared linear operators in two

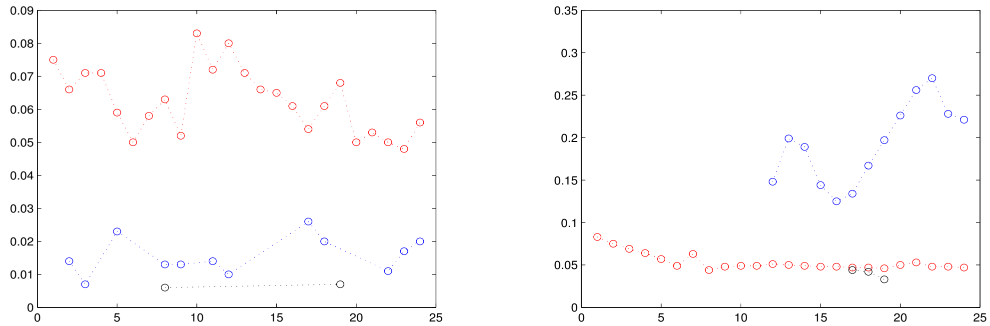

Figure 5.1: Left: SO_2 emission levels for 3 selected days; Right: $PM2.5$ emission levels for 3 selected days.

functional regression models. Panaretos et al. (2010) focused on testing the equality of the covariance operators in two samples of independent and identically distributed Gaussian functional observations. Fremdt et al. (2013) proposed a non-parametric test for the equality of the covariance structures in two functional samples. However, all aforementioned research had assumed that repeated measurements took place on the dense and regular time points for each subject. For samples with sparse and irregularly spaced observations, to the best of our knowledge, rare work could be searched. Furthermore, it is hard to decide when the observations are dense or sparse. In some functional data studies, it is possible that we have dense observations on some subjects and sparse observations on the others. It hence deserves developing unified methodologies for testing equality of two covariance operators regardless of whatever types of functional data, dense or sparse, and regular or imbalanced.

5.2 Methodology and main results

Functional data may usually be modeled as independent realizations of an underlying second-order effect stochastic process

$$Y_i(t) = \mu(t) + v_i(t) + \varepsilon_i(t),\ i = 1, \cdots, n,\ t \in \mathcal{T}$$

where $\mathcal{T} = [0, 1]$ in most literature, or any compact domain $\mathcal{T} \subset \mathbb{R}$, $\{Y_i(t)\}$ is the ith response process, $\mu(t)$ denotes the population mean function of afore stochastic process, $\{v_i(t)\}$ models random effect process or between-subject variation, and $\varepsilon_i(t)$s are i.i.d. random error process independent of $v_i(t)$, refer to Shi et al. (1996), Zhang and Chen (2007), and Horváth and Kokoszka (2012), among others.

Consider two independent samples:

$$\begin{aligned} &Y_i^{(m)}(t_{il_m}) = \mu_m(t_{il_m}) + v_i^{(m)}(t_{il_m}) + \varepsilon_{il_m}^{(m)}, \\ &m = 1, 2;\ i = 1, \cdots, n_m;\ l_m = 1, \cdots, N_i. \end{aligned} \tag{5.1}$$

where $Y_i^{(1)}(t_{il_1})$ and $Y_i^{(2)}(t_{il_2})$ are the measurements taken at time t_{il_1} and t_{il_2} from two samples with N_{n_1} and N_{n_2} the number of measurements, respectively. Without loss of generality, the $\varepsilon_{il_1}^{(1)}$s are zero-mean errors with $E(\varepsilon_{il_1}^{(1)2}) = \sigma^2$, so are $\varepsilon_{il_2}^{(2)}$s. For the subject-specific random trajectory process $v_i^{(1)}$ and $v_i^{(2)}$, denote covariance operators $\gamma_1(t, s) = \text{cov}\{v_i^{(1)}(t), v_i^{(1)}(s)\}$ and $\gamma_2(t, s) = \text{cov}\{v_i^{(2)}(t), v_i^{(2)}(s)\}$, respectively. In this section, we focus on testing if two functional samples have the same covariance operator structure, i.e.

$$H_0 : \gamma_1(t, s) = \gamma_2(t, s) \text{ vs. } H_1 : \gamma_1(t, s) \neq \gamma_2(t, s),\ t, s \in \mathcal{T}.$$

5.2.1 Estimation of covariance operator

The estimation of the covariance operator in functional data has drawn arising attention because of the importance of covariance operators in functional data analysis. Based on the functional principle component analysis, Hall and Hosseini-Nasab (2006) and Zhang and Chen (2007) considered a smooth-first-then-estimate strategy. Cai and Yuan (2010) proposed a nonparametric estimation method within a reproducing kernel Hilbert space frame. Li and Hsing (2010) estimated the covariance operator based on the local linear smoother and made it statistical inference.

Kraus (2015) presented an estimation method by looking irregular functional data as missing data, among others.

Here our estimation of covariance operator of individual and pooled sample can be conducted following the procedures below.

Step 5.1. *Estimate the eigenfunctions of the pooled samples.*

Step 1a: Obtain the initial estimations of mean functions $\mu_1(t)$ and $\mu_2(t)$. It is direct application of (A.2) and (A.3) of Yao et al. (2005). The local linear scatterplot smoother $\hat{\mu}_1^0(t) = \hat{d}_0(t)$ is obtained by optimizing

$$\arg\min_{d_0,d_1} \sum_{i=1}^{n_1} \sum_{l_1=1}^{N_i} K^{(1)} \left(\frac{t_{il_1} - t}{b_1} \right) \left[Y_i(t_{il_1}) - d_0 - d_1(t_{il_1} - t) \right]^2,$$

where $K^{(1)}(\cdot)$ is a smooth symmetric kernel density function and b_1 is a bandwidth. Analogously, one may define $\hat{\mu}_2^0(t)$ using the bandwidth b_2, the estimator of the mean function $\mu_2(t)$.

Step 1b: Obtain the initial estimators of covariance operators $\gamma_m(t,s),\ m = 1,2$. Denote $G_{1,i}(t_{il_{11}}, t_{il_{12}}) = \{Y_i(t_{il_{11}}) - \hat{\mu}_1(t_{il_{11}})\}\{Y_i(t_{il_{12}}) - \hat{\mu}_1(t_{il_{12}})\}$. The local linear surface smoother $\hat{\gamma}_1^0(t,s) = \hat{\alpha}_0(t,s)$ can be obtained by optimizing

$$\arg\min_{\alpha} \sum_{i=1}^{n_1} \sum_{1 \leqslant l_{11} \neq l_{12} \leqslant N_i} K^{(2)} \left(\frac{t_{il_{11}} - t}{h_1}, \frac{t_{il_{12}} - s}{h_1} \right) \left\{ G_{1,i}(t_{il_{11}}, t_{il_{12}}) - f(\alpha, (t,s), (t_{il_{11}}, t_{il_{12}})) \right\}^2,$$

where $K^{(2)}(\cdot,\cdot)$ is a bivariate smooth kernel density with zero mean and finite covariance, h_1 is a bandwidth, and $f(\alpha, (t,s), (t_{il_{11}}, t_{il_{12}})) = \alpha_0 + \alpha_{11}(t - t_{il_{11}}) + \alpha_{12}(s - t_{il_{12}})$ with $\alpha = (\alpha_0, \alpha_{11}, \alpha_{12})$. An essential feature is the omission of the diagonal elements $l_{11} = l_{12}$ which are contaminated with the measurement errors. Analogously, we can obtain $\hat{\gamma}_2^0(t,s)$ using the bandwidth h_2, the initial estimator of the covariance function $\gamma_2(t,s)$.

Denote $\hat{\gamma}^0(t,s) = \frac{n_2}{n_1+n_2}\hat{\gamma}_1^0(t,s) + \frac{n_1}{n_1+n_2}\hat{\gamma}_2^0(t,s)$. It is easy to see $\hat{\gamma}^0(t,s) \xrightarrow{p} \gamma(t,s) \equiv (1-\theta)\gamma_1(t,s) + \theta\gamma_2(t,s)$ where $\gamma(t,s)$ is an asymptotic pooled covariance operator of the two given samples and θ is defined at assumption 5.8 in Subsection 5.2.2. Consequently, it has functional principal components, also known as orthonormal eigenfunctions $\{\phi_k\}_{k\geqslant 1}$, as well as corresponding non-negative eigenvalues $\{\nu_k\}_{k\geqslant 1}$ with $\nu_1 \geqslant \nu_2 \geqslant \cdots$ satisfying:

$$\int_{\mathcal{T}} \gamma(t,s)\phi_k(s)ds = \nu_k\phi_k(t).$$

Step 1c: Estimate the set of orthonormal basis $\{\nu_k, \phi_k\}_{k\geqslant 1}$ of $\gamma(t,s)$. Estimation of eigenvalues and eigenfunctions $\{\nu_k, \phi_k\}_{k\geqslant 1}$ are obtained by numerical solutions $\{\hat{\nu}_k, \hat{\phi}_k\}_{k\geqslant 1}$ of suitably discretized eigenequations,

$$\int_{\mathcal{T}} \hat{\gamma}^0(t,s)\hat{\phi}_k(s)ds = \hat{\nu}_k\hat{\phi}_k(t)$$

with orthonormal constraints on $\{\hat{\phi}_k\}_{k\geqslant 1}$.

Step 5.2. *Obtain the projection estimators of covariance operators.*

Estimators of the covariance operators are projection estimators onto a space generated based on the orthonormal basis $\{\hat{\phi}_k\}_{k\geqslant 1}$. We propose the following estimators of the covariance functions

$$\hat{\gamma}_1(t,s) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \hat{\rho}_{kk'}^{(1)}\hat{\phi}_k(t)\hat{\phi}_{k'}(s);\ \hat{\gamma}_2(t,s) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \hat{\rho}_{kk'}^{(2)}\hat{\phi}_k(t)\hat{\phi}_{k'}(s),$$

where

$$\hat{\rho}_{kk'}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left[\hat{\eta}_{ik}^{(1)} - \hat{\eta}_{k}^{(1)}\right]\left[\hat{\eta}_{ik'}^{(1)} - \hat{\eta}_{k'}^{(1)}\right],\ \hat{\rho}_{kk'}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\left[\hat{\eta}_{ik}^{(2)} - \hat{\eta}_{k}^{(2)}\right]\left[\hat{\eta}_{ik'}^{(2)} - \hat{\eta}_{k'}^{(2)}\right];$$

$$\hat{\eta}_{k}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\eta}_{ik}^{(1)},\ \hat{\eta}_{ik}^{(1)} = \frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_i^{(1)}(t_{il_1})\hat{\phi}_k(t_{il_1});$$

$$\hat{\eta}_{k}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\eta}_{ik}^{(2)},\ \hat{\eta}_{ik}^{(2)} = \frac{1}{N_i}\sum_{l_2=1}^{N_i} Y_i^{(2)}(s_{il_2})\hat{\phi}_k(s_{il_2}),$$

where K is a tuning parameter. We propose a concise and practical method to select K by $K = \max\{K_1, K_2\}$, where K_1 and K_2 are minimum numbers of functional principal components that explain 95% of the total variation for sample 1 and sample 2, respectively. The Matlab package PACE is a tool to calculate K_1 and K_2 conveniently, refer to Yao et al. (2005).

Remark 5.1. *The difference between our estimating approach and that of Panaretos et al. (2010) and Fremdt et al. (2013) is typically to project the curves $Y^{(1)}(t)$ and $Y^{(2)}(t)$ on the common basis function space and obtain $\hat{\eta}_{ik}^{(1)}$ and $\hat{\eta}_{ik}^{(2)}$, and hence $\hat{\rho}_{kk'}^{(1)}$ and $\hat{\rho}_{kk'}^{(2)}$, which are the covariance of projection. The covariances $\hat{\rho}_{kk'}^{(1)}$ and $\hat{\rho}_{kk'}^{(2)}$ reflect the volatility of two functional samples on direction $\hat{\phi}_k(t)\hat{\phi}_{k'}(s)$. Advantages of our method lies in three aspects. Firstly, unlike the method proposed by Panaretos et al. (2010) and Fremdt et al. (2013), our method has a better explanation and easier to implement. Secondly, one obvious advantage of such a method is that we can always define estimators of $\hat{\eta}_{ik}^{(1)}$ and $\hat{\eta}_{ik}^{(2)}$, and hence $\hat{\gamma}_1(t,s)$ and $\hat{\gamma}_2(t,s)$, no matter functional data is sparse and irregular, dense and irregular, or dense and regular, which leads to wide applicability of the proposed test. Lastly, the distance of covariance operators of two samples can be transformed into that of the variances $\hat{\rho}_{kk'}^{(1)}$ and $\hat{\rho}_{kk'}^{(2)}$ from the expression of $\hat{\gamma}_1(t,s)$ and $\hat{\gamma}_2(t,s)$ and (5.2). This directly leads to a global statistic.*

5.2.2 Asymptotic distributions

Next we propose a testing statistic by measuring the integrated square discrepancy of the covariance operators of two functional data samples. Denote TC $= \frac{n_1 n_2}{n_1+n_2}\int\int[\hat{\gamma}_1(t,s)-\hat{\gamma}_2(t,s)]^2 dtds$. It is readily seen that

$$
\begin{aligned}
\mathrm{TC} &= \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{S}}\int_{\mathcal{T}}\left\{\sum_{k=1}^{K}\sum_{k'=1}^{K}\hat{\rho}^{(1)}_{kk'}\hat{\phi}_k(t)\hat{\phi}_{k'}(s)-\sum_{k=1}^{K}\sum_{k'=1}^{K}\hat{\rho}^{(2)}_{kk'}\hat{\phi}_k(t)\hat{\phi}_{k'}(s)\right\}^2 dtds \\
&= \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{S}}\int_{\mathcal{T}}\left[\sum_{k=1}^{K}\sum_{k'=1}^{K}\left\{\hat{\rho}^{(1)}_{kk'}-\hat{\rho}^{(2)}_{kk'}\right\}\right]^2 \phi_k^2(t)\phi_{k'}^2(s)dtds \\
&\quad + \frac{n_1 n_2}{n_1+n_2}\int_{\mathcal{S}}\int_{\mathcal{T}}\left[\sum_{k=1}^{K}\sum_{k'=1}^{K}\left\{\hat{\rho}^{(1)}_{kk'}-\hat{\rho}^{(2)}_{kk'}\right\}\right]^2\left[\hat{\phi}_k^2(t)\hat{\phi}_{k'}^2(s)-\phi_k^2(t)\phi_{k'}^2(s)\right]dtds \\
&\equiv \frac{n_1 n_2}{n_1+n_2}\left[\sum_{k=1}^{K}\sum_{k'=1}^{K}\left\{\hat{\rho}^{(1)}_{kk'}-\hat{\rho}^{(2)}_{kk'}\right\}\right]^2 + o_p(1).
\end{aligned}
\tag{5.2}
$$

The decomposition in equation (5.2) shows that the statistic TC is equivalent to the square error of two sample variances of the orthonormal basis except a subtle residual. Therefore, H_0 will be rejected if TC is large.

We demand the following assumptions in order to derive the asymptotic properties of statistic $\widehat{TC}$.

Assumption 5.1. *For the estimators of mean functions $\mu_1(s)$ and the initial estimators of covariance operators $\gamma_1(s,t)$, We require $b_1 \to 0$, $h_1 \to 0$, $nb_1^4 \to \infty$, $nh_1^6 \to \infty$, $nb_1^6 < \infty$ and $nh_1^8 < \infty$. For the estimators of mean functions $\mu_2(s)$ and the initial estimators of covariance operators $\gamma_2(t,s)$, analogous conditions are required.*

Assumption 5.2. $\sup_{t\in\mathcal{T}} E[Y^{(1)4}(t)] < \infty$ *and* $\sup_{t\in\mathcal{T}}[Y^{(2)4}(t)] < \infty$.

Assumption 5.3. *$K^{(m)}$ is absolutely integrable, that is, $\int |K^{(m)}(u)|du < \infty$, for $m = 1, 2$.*

Assumption 5.4. *Let $\delta_k^{Y^{(1)}} = \min_{1 \leqslant j \leqslant k}(\nu_j^{(1)} - \nu_{j+1}^{(1)})$ and $\delta_k^{Y^{(2)}} = \min_{1 \leqslant j \leqslant k}(\nu_j^{(2)} - \nu_{j+1}^{(2)})$ where $\nu_j^{(1)}$ and $\nu_j^{(2)}$ are the eigenvalues of covariance operators $\gamma_1(t,s)$ and $\gamma_2(t,s)$, respectively. Denote $\pi_k^{Y^{(1)}} = 1/\nu_k^{(1)} + 1/\delta_k^{Y^{(1)}}$ and $\pi_k^{Y^{(2)}} = 1/\nu_k^{(2)} + 1/\delta_k^{Y^{(2)}}$. We demand $\pi_k^{Y^{(1)}}/h_1^2 < \infty$ and $\pi_k^{Y^{(2)}}/h_2^2 < \infty$.*

Assumption 5.5. *For sample $\{Y_1^{(1)}, \cdots, Y_{n_1}^{(1)}\}$, denote the sorted time points across all subjects by $0 \leqslant t_1 \leqslant \cdots \leqslant t_{\tilde{N}} \leqslant 1$, and $\Delta^X = max\{t_{(n_2)} - t_{(n_2-1)} : n_2 = 1, \ldots, \tilde{N} + 1\}$, where $\tilde{N} = \sum_{i=1}^{n_1} N_i$. For the ith subject, suppose that the time points t_{ij} have been ordered non-decreasingly. Let $\Delta_i^{Y^{(1)}} = \max\{t_{i,j} - t_{i,j-1} : j = 1, \ldots, N_i + 1\}$, $\Delta_i^{Y^{(2)}} = \max\{\Delta_i^{Y^{(1)}} : i = 1, \ldots, n_1\}$ and $\bar{N} = \tilde{N}/n_1$. Put $N_{\max}^{Y^{(1)}} = \max\{N_i : i = 1, \ldots, n_1\}$ and $N_{\min}^{Y^{(1)}} = \min\{N_i : i = 1, \ldots, n_1\}$. Assume that $\Delta^{Y^{(1)}} = O(\min\{n_1^{-1/2} b_1^{-1}, n_1^{-1/2} h_1^{-1}\})$, and $\bar{N} \to \infty$, $N_{\max}^{Y^{(1)}} \leqslant c_2 \bar{N}$, $\Delta^{Y^{(2)}} = O(1/\bar{N})$, for some $c_2 > 0$. For sample $\{Y_1^{(2)}, \cdots, Y_{n_2}^{(2)}\}$, analogous conditions are required.*

Assumption 5.6. *$\{v_i^{(1)}(\cdot)\}_i$, $\{t_{il_1}\}_{i,l_1}$ and $\{\varepsilon_{il_1}^{(1)}\}_{i,l_1}$ are independent and identically distributed and mutually independent. Similarly, $\{v_i^{(2)}(\cdot)\}_i$, $\{s_{il_2}\}_{i,l_2}$ and $\{\varepsilon_{il_2}^{(2)}\}_{i,l_2}$ are independent and identically distributed and mutually independent.*

Assumption 5.7. *Assume $\int_0^1 \mu_m^2(t)dt < \infty$, $m = 1, 2$ and $\mu_m(t)$ may be written as $\mu_m(t) = \sum_{k=1}^{\infty} \eta_k^{(m)} \phi_k$ where $\eta_k^{(m)} = \int_0^1 \mu_m(t)\phi_k(t)dt$.*

Assumption 5.8. $\min\{n_1, n_2\} \to \infty$, $\frac{n_1}{n_1+n_2} \to \theta$ *for a fixed constant $\theta \in (0,1)$.*

Remark 5.2. *Assumptions 5.1 and 5.3 are similar to that of Yao et al. (2005) which are also regular condition for unbalance functional data analysis. Assumptions 5.2 and 5.6 are regular conditions in functional data analysis. Assumption 5.4 requires* *that the spacings between the eigenvalues are not too small. Assumption 5.5 was adopted by Müller and Yao (2008). This assumption implies the dense observation which is easier to use in our theoretical justifications. Assumption 5.8 is a regular condition in two sample test. Assumptions 5.1-5.5 and 5.8 are used to get the bound of $\hat{\phi}_j(t) - \phi_j(t)$. Assumption 5.8 is a regular condition in two sample test.*

In order to present our testing statistic, we need to use of the below asymptotic result.

Lemma 5.1. *Under assumptions 5.1-5.5 and 5.8, we have*

$$\sup_{t \in \mathcal{T}} \left| \hat{\phi}_i(t) - \phi_i(t) \right| = O_p \left(A_{(n_1,n_2)} \right).$$

where $A_{(n_1,n_2)} = \max(\frac{\pi_k^{Y^{(1)}}}{\sqrt{n_1}h_1^2}, \frac{\pi_k^{Y^{(2)}}}{\sqrt{n_2}h_2^2})$ *and* $\pi_l^{Y^{(1)}}$ *and* $\pi_l^{Y^{(2)}}$ *are stated in assumption 5.4.*

Lemma 5.2. *Under assumptions 5.1-5.8 and H_0, we have*

$$\frac{\hat{\rho}_{kk'}^{(1)} - \hat{\rho}_{kk'}^{(2)}}{\sqrt{\frac{1}{n_1}\omega_{kk'}^{(1)} + \frac{1}{n_2}\omega_{kk'}^{(2)}}} \xrightarrow{d} N(0,1), \ \min\{n_1, n_2\} \to \infty.$$

where

$$\begin{aligned}
\omega_{kk'}^{(1)} = \sigma^4 &+ \left\{ \int_0^1 \mu_1(t)\phi_k(t)dt \right\}^2 \left\{ \int_0^1 \mu_1(s)\phi_{k'}(s)ds \right\}^2 \\
&+ \left\{ \int_0^1 \mu_1(t)\phi_{k'}(t)dt \right\}^2 \int_0^1 \int_0^1 \phi_k(s_1)\gamma_1(s_1,s_2)\phi_k(s_2)ds_1 ds_2 \\
&+ \left\{ \int_0^1 \mu_1(t)\phi_k(t)dt \right\}^2 \int_0^1 \int_0^1 \phi_{k'}(s_1)\gamma_1(s_1,s_2)\phi_{k'}(s_2)ds_1 ds_2 \\
&+ \int_0^1 \int_0^1 \phi_k^2(t) \left[E\left\{ v_i^{(1)2}(t) v_i^{(1)2}(s) \right\} \right] \phi_{k'}^2(s)dtds \\
&- \left\{ \int_0^1 \int_0^1 \phi_k(t)\gamma_1(t,s)\phi_{k'}(s)dtds \right\}^2
\end{aligned}$$

and

$$
\begin{aligned}
\omega_{kk'}^{(2)} =& \sigma^4 + \left\{\int_0^1 \mu_2(t)\phi_k(t)dt\right\}^2 \left\{\int_0^1 \mu_2(s)\phi_{k'}(s)ds\right\}^2 \\
&+ \left\{\int_0^1 \mu_2(t)\phi_{k'}(t)dt\right\}^2 \int_0^1\int_0^1 \phi_k(s_1)\gamma_2(s_1,s_2)\phi_{k'}(s_2)ds_1ds_2 \\
&+ \left\{\int_0^1 \mu_2(t)\phi_k(t)dt\right\}^2 \int_0^1\int_0^1 \phi_{k'}(s_1)\gamma_2(s_1,s_2)\phi_{k'}(s_2)ds_1ds_2 \\
&+ \int_0^1\int_0^1 \phi_k^2(t)\left[E\left\{v_i^{(2)2}(t)v_i^{(2)2}(s)\right\}\right]\phi_{k'}^2(s)dtds \\
&- \left\{\int_0^1\int_0^1 \phi_k(t)\gamma_2(t,s)\phi_{k'}(s)dtds\right\}^2
\end{aligned}
$$

Based on lemma 5.1 and symmetry of $\hat{\rho}_{kk'}^{(1)} - \hat{\rho}_{kk'}^{(2)}$ for subscripts k and k', we suggest the following statistic:

$$
\widehat{\mathrm{TC}} = \frac{\frac{n_1n_2}{n_1+n_2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\frac{\left\{\hat{\rho}_{kk'}^{(1)}-\hat{\rho}_{kk'}^{(2)}\right\}^2}{\hat{\omega}_{kk'}} - \frac{K(K-1)}{2}}{\sqrt{K(K-1)}},
$$

where $\hat{\omega}_{kk'} = \frac{n_2}{n_1+n_2}\hat{\omega}_{kk'}^{(1)} + \frac{n_1}{n_1+n_2}\hat{\omega}_{kk'}^{(2)}$ and $\hat{\omega}_{kk'}^{(m)}$, $m = 1,2$ are the estimators of $\omega_{kk'}^{(m)}$, $m = 1,2$, respectively. In fact, $\omega_{kk'}^{(m)}$, $m = 1,2$ are unknown but can be substituted by their consistent estimators $\frac{1}{n_1-1}\sum_{i=1}^{n_1}\{[\hat{\eta}_{ik}^{(1)} - \hat{\eta}_k^{(1)}][\hat{\eta}_{ik'}^{(1)} - \hat{\eta}_{k'}^{(1)}] - \Phi_1\}^2$ and $\frac{1}{n_2-1}\sum_{i=1}^{n_2}\{[\hat{\eta}_{ik}^{(2)} - \hat{\eta}_k^{(2)}][\hat{\eta}_{ik'}^{(2)} - \hat{\eta}_{k'}^{(2)}] - \Phi_2\}^2$, respectively, where $\Phi_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}[\hat{\eta}_{ik}^{(1)} - \hat{\eta}_k^{(1)}][\hat{\eta}_{ik'}^{(1)} - \hat{\eta}_{k'}^{(1)}]$ and $\Phi_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}[\hat{\eta}_{ik}^{(2)} - \hat{\eta}_k^{(2)}][\hat{\eta}_{ik'}^{(2)} - \hat{\eta}_{k'}^{(2)}]$.

We are ready to present the asymptotic results of the proposed test. Theorems 5.1 and 5.2 below establish the asymptotic behaviors of the statistic $\widehat{\mathrm{TC}}$ under hypotheses H_0 and H_1, respectively. The proofs of these theorems are shown in Section 5.5.

Theorem 5.1. *Under assumptions 5.1-5.8 and H_0, we have*

$$
\widehat{\mathrm{TC}} \xrightarrow{d} N(0,1), \ \min\{n_1,n_2\} \to \infty.
$$

The null hypothesis $H_0 : \gamma_1(t,s) = \gamma_2(t,s),\ t,s \in \mathcal{T}$ *is rejected if* $|\widehat{\mathrm{TC}}| > q_\alpha$*, where* q_α *is the upper-*α *quantile of* $N(0,1)$*.*

Theorem 5.2. *Under assumptions 5.1-5.8 and* H_1*, we have*

$$\widehat{\mathrm{TC}} \xrightarrow{p} \infty,\ \min\{n_1,n_2\} \to \infty.$$

Remark 5.3. *From the expression of* $\widehat{\mathrm{TC}}$*, we can see that* $\widehat{\mathrm{TC}}$ *depends on two sample sizes and distance between* $\hat{\rho}^{(1)}_{kk'}$ *and* $\hat{\rho}^{(2)}_{kk'}$ *for* $k = 1,\cdots,K$ *and* $k' = 1,\cdots,K$ *, which reflect the difference of covariance operators* $\gamma_1(t,s)$ *and* $\gamma_2(t,s)$*. On the one hand, with the larger difference of covariance operators,* $\widehat{\mathrm{TC}}$ *will become bigger and bigger when the sample size is fixed. On the other hand,* $\widehat{\mathrm{TC}}$ *will grow with* n_1 *and* n_2 *when the difference of covariance operators are fixed. Theorems 5.1 and 5.2 thus provide clear theoretical justification of the empirical properties discussed in Section 5.3. Theorem 5.2 also shows that the behavior of the test is consistent.*

5.3 Simulation studies

The random effect function $v_i(t)$ was generated from

$$v_i(t) = A_i \sin(\pi t) + B_i \sin(2\pi t) + C_i \sin(4\pi t),$$

where $A_i = 5W_1,\ B_i = 3W_2,\ C_i = W_3$, and W_1, W_2 and W_3 were independent t-distributed random variables. All the simulation results reported were based on 1000 simulations.

Case 5.1. *Sparse design with identical mean functions.*

To illustrate the adaptivity of our test method to sparse design, we first considered the following model,

$$\begin{aligned} Y_i^{(1)}(t) &= v_i^{(1)}(t) + \varepsilon_i^{(1)}(t), \\ Y_i^{(2)}(t) &= v_i^{(2)}(t) + \varepsilon_i^{(2)}(t). \end{aligned} \tag{5.3}$$

The number of measurements was selected from $\{5, \cdots, 9\}$ with equal probability in $[0, 1]$. The measurement errors were $\varepsilon_i^{(1)}(t) \sim N(0, \sigma^2)$ for $t \in \{t_{il_1}\}_{i=1,\cdots,n_1;l_1=1,\cdots,N_i}$ and $\varepsilon_i^{(2)}(t) \sim N(0, \sigma^2)$ for $t \in \{t_{il_2}\}_{i=1,\cdots,n_2;l_2=1,\cdots,N_i}$. In order to study the empirical size and power of the test, we set $v_i^{(2)}(t) = av_i^{(1)}(t)$, $a = 0.4, 0.6, 0.8, 1.0, 1.2,$ $1.4, 1.6$, where the parameter a regulated the difference in the covariance function in two samples. The empirical size can be reached when $a = 1$ and the empirical power can be reached when $a \neq 1$. This can be viewed as a sparse design with the identical mean functions $\mu_1(t) = \mu_2(t) = 0$. We considered combinations of sample sizes $(n_1, n_2) = (50, 50)$, $(100, 100)$ and $(200, 200)$. The empirical size and power of the test are reported in Table 5.1 for the setting of $\sigma^2 = 0.5$ and $\sigma^2 = 4$, respectively.

Case 5.2. *Sparse design with distinct mean functions.*

To illustrate whether mean functions has influence for the test of covariance function, we considered the following model,

$$\begin{aligned} Y_i^{(1)}(t) &= v_i^{(1)}(t) + \varepsilon_i^{(1)}(t), \\ Y_i^{(2)}(t) &= t + v_i^{(2)}(t) + \varepsilon_i^{(2)}(t), \end{aligned} \tag{5.4}$$

where $v_i^{(1)}(t)$, $v_i^{(2)}(t)$, $\varepsilon_i^{(1)}(t)$, and $\varepsilon_i^{(2)}(t)$ followed the Case 5.1. This can be viewed as a sparse design with the different mean functions $\mu_1(t) = 0$ and $\mu_2(t) = t$. The empirical size and power of the test are also reported in Table 5.1 for the setting of $\sigma^2 = 0.5$ and $\sigma^2 = 4$, respectively.

Table 5.1: The empirical size and power of the test in Cases 5.1 and 5.2.

Case	a	$(n_1, n_2) = (50, 50)$ $\alpha = 0.01$	$\alpha = 0.05$	$(n_1, n_2) = (100, 100)$ $\alpha = 0.01$	$\alpha = 0.05$	$(n_1, n_2) = (200, 200)$ $\alpha = 0.01$	$\alpha = 0.05$
				$\sigma^2 = 0.5$			
Case 1	0.4	0.957	0.992	0.993	0.999	0.999	1.000
	0.6	0.551	0.762	0.845	0.955	0.974	0.992
	0.8	0.083	0.219	0.191	0.398	0.394	0.619
	1.0	0.010	0.046	0.011	0.047	0.014	0.058
	1.2	0.058	0.172	0.120	0.293	0.249	0.460
	1.4	0.261	0.481	0.460	0.707	0.784	0.889
	1.6	0.476	0.695	0.792	0.914	0.977	0.994
Case 2	0.4	0.952	0.987	0.998	0.999	1.000	1.000
	0.6	0.521	0.741	0.846	0.947	0.986	0.996
	0.8	0.103	0.232	0.165	0.354	0.381	0.617
	1.0	0.009	0.056	0.012	0.047	0.010	0.042
	1.2	0.055	0.150	0.112	0.278	0.249	0.475
	1.4	0.250	0.451	0.475	0.707	0.803	0.911
	1.6	0.478	0.704	0.796	0.916	0.967	0.991
				$\sigma^2 = 4$			
Case 1	0.4	0.937	0.984	0.993	0.998	0.999	1.000
	0.6	0.513	0.731	0.830	0.940	0.981	0.994
	0.8	0.070	0.209	0.182	0.377	0.389	0.622
	1.0	0.006	0.050	0.007	0.057	0.014	0.059
	1.2	0.052	0.157	0.132	0.287	0.231	0.463
	1.4	0.225	0.458	0.464	0.670	0.763	0.911
	1.6	0.492	0.706	0.775	0.917	0.969	0.993
Case 2	0.4	0.937	0.982	0.991	0.996	0.999	1.000
	0.6	0.488	0.716	0.817	0.928	0.975	0.993
	0.8	0.073	0.226	0.163	0.355	0.350	0.563
	1.0	0.015	0.061	0.008	0.050	0.010	0.051
	1.2	0.057	0.179	0.113	0.273	0.244	0.468
	1.4	0.223	0.407	0.451	0.684	0.791	0.905
	1.6	0.488	0.717	0.774	0.903	0.973	0.988

Several phenomenon can be observed from Table 5.1. First, the test based on the proposed method has correct empirical size at all levels. Second, as expected from the theory results, when a increases from 1.2 to 1.6 or decreases from 0.8 to 0.4, the power of the test becomes larger and larger. Third, from the simulations of Cases

5.1 and 5.2, we find that the power of the test increases as the sample size increases. For example, the empirical power for sample sizes $(n_1, n_2) = (200, 200)$ is better than that of $(n_1, n_2) = (100, 100)$, whereas the sample sizes are equal. Last, the empirical size and power in Case 5.1 are comparable with that in Case 5.2. This shows that mean functions have no significant influence on our testing procedure in the setting of sparse design.

Case 5.3. *Dense design with identical mean functions.*

In order to compare the proposed method, denoted by $\widehat{\mathrm{TC}}$, with that of Fremdt et al. (2013), denoted by $Fremdt$, we considered the dense design where $v_i^{(1)}(t)$ and $v_i^{(2)}(t)$ follow (5.3) in Case 5.1 except that the locations of measurements for each $v_i^{(1)}(t)$ and $v_i^{(2)}(t)$ were selected at 50 equidistant time points in $[0, 1]$. We computed the empirical size and power of the test for sample sizes $(n_1, n_2) = (50, 50), (100, 100)$ and $(200, 200)$ in Table 5.2 for the setting of $\sigma^2 = 0.5$ and $\sigma^2 = 4$, respectively.

Case 5.4. *Dense design with distinct mean functions.*

This experiment is to compare with case 5.3 and case 5.2 separately. Therefore, we used (5.4) to model the functional data but the locations of measurements for each $v_i^{(1)}(t)$ and $v_i^{(2)}(t)$ were selected at 50 equidistant time points in $[0, 1]$. The empirical size and power of the test are also reported in Table 5.2 (right side) for the setting of $\sigma^2 = 0.5$ and $\sigma^2 = 4$, respectively.

Table 5.2: Empirical sizes and power of test in Cases 5.3 and 5.4.

Sample sizes	identical mean					non-identical mean				
	a	$\widehat{\mathrm{TC}}$		*Fremdt*		a	$\widehat{\mathrm{TC}}$		*Fremdt*	
	$\sigma^2 = 0.5$					$\sigma^2 = 0.5$				
		0.01	0.05	0.01	0.05		0.01	0.05	0.01	0.05
$(n_1, n_2) =$ $(50, 50)$	0.7	0.448	0.554	0.131	0.388	0.7	0.477	0.605	0.157	0.428
	0.8	0.183	0.272	0.041	0.171	0.8	0.173	0.261	0.031	0.148
	0.9	0.046	0.087	0.013	0.069	0.9	0.054	0.097	0.004	0.042
	1.0	0.023	0.050	0.001	0.041	1.0	0.025	0.053	0.006	0.041
	1.1	0.044	0.091	0.009	0.056	1.1	0.053	0.088	0.006	0.057
	1.2	0.126	0.192	0.015	0.127	1.2	0.109	0.180	0.024	0.118
	1.3	0.255	0.362	0.052	0.215	1.3	0.278	0.376	0.065	0.231
$(n_1, n_2) =$ $(100, 100)$	0.7	0.778	0.843	0.488	0.762	0.7	0.762	0.851	0.491	0.761
	0.8	0.340	0.449	0.109	0.326	0.8	0.359	0.464	0.123	0.353
	0.9	0.078	0.132	0.015	0.088	0.9	0.076	0.134	0.014	0.094
	1.0	0.034	0.054	0.004	0.038	1.0	0.019	0.045	0.003	0.040
	1.1	0.077	0.128	0.013	0.070	1.1	0.067	0.107	0.010	0.073
	1.2	0.222	0.318	0.072	0.226	1.2	0.222	0.325	0.067	0.233
	1.3	0.750	0.836	0.487	0.745	1.3	0.474	0.572	0.199	0.466
$(n_1, n_2) =$ $(200, 200)$	0.7	0.967	0.983	0.876	0.970	0.7	0.974	0.988	0.876	0.971
	0.8	0.630	0.707	0.370	0.650	0.8	0.627	0.709	0.384	0.645
	0.9	0.124	0.193	0.042	0.142	0.9	0.133	0.218	0.033	0.154
	1.0	0.031	0.063	0.004	0.044	1.0	0.020	0.044	0.007	0.046
	1.1	0.132	0.197	0.035	0.154	1.1	0.114	0.168	0.031	0.125
	1.2	0.419	0.545	0.196	0.439	1.2	0.431	0.542	0.203	0.444
	1.3	0.768	0.845	0.552	0.774	1.3	0.789	0.866	0.536	0.793
	$\sigma^2 = 4$					$\sigma^2 = 4$				
		0.01	0.05	0.01	0.05		0.01	0.05	0.01	0.05
$(n_1, n_2) =$ $(50, 50)$	0.7	0.338	0.482	0.124	0.412	0.7	0.471	0.600	0.148	0.413
	0.8	0.103	0.192	0.032	0.147	0.8	0.170	0.244	0.032	0.152
	0.9	0.051	0.088	0.007	0.067	0.9	0.051	0.090	0.006	0.048
	1.0	0.031	0.059	0.011	0.046	1.0	0.022	0.048	0.004	0.030
	1.1	0.036	0.072	0.010	0.059	1.1	0.048	0.087	0.008	0.056
	1.2	0.090	0.160	0.018	0.114	1.2	0.115	0.183	0.025	0.109
	1.3	0.156	0.265	0.060	0.205	1.3	0.241	0.350	0.053	0.209
$(n_1, n_2) =$ $(100, 100)$	0.7	0.695	0.801	0.494	0.745	0.7	0.759	0.857	0.457	0.736
	0.8	0.242	0.358	0.112	0.349	0.8	0.360	0.471	0.117	0.354
	0.9	0.049	0.106	0.013	0.088	0.9	0.067	0.133	0.014	0.088
	1.0	0.028	0.048	0.002	0.033	1.0	0.024	0.050	0.003	0.039
	1.1	0.048	0.088	0.012	0.082	1.1	0.076	0.128	0.016	0.087
	1.2	0.172	0.264	0.005	0.224	1.2	0.221	0.315	0.056	0.214
	1.3	0.359	0.496	0.205	0.466	1.3	0.488	0.599	0.212	0.476
$(n_1, n_2) =$ $(200, 200)$	0.7	0.940	0.967	0.878	0.973	0.7	0.961	0.976	0.859	0.956
	0.8	0.539	0.647	0.365	0.622	0.8	0.614	0.707	0.347	0.633
	0.9	0.094	0.154	0.046	0.167	0.9	0.133	0.216	0.044	0.184
	1.0	0.021	0.048	0.003	0.038	1.0	0.033	0.054	0.005	0.041
	1.1	0.101	0.164	0.028	0.134	1.1	0.108	0.180	0.044	0.136
	1.2	0.338	0.445	0.195	0.452	1.2	0.439	0.543	0.220	0.454
	1.3	0.711	0.793	0.556	0.788	1.3	0.794	0.861	0.549	0.795

From Table 5.2, we can see that the tests based on $\widehat{\mathrm{TC}}$ and $Fremdt$ have correct empirical size at all levels which imply that both tests can control the type I error well. However, the test based on statistic $\widehat{\mathrm{TC}}$ has a higher power than that of $Fremdt$. Comparing Tables 5.1 with 5.2 when $a = 0.8$ and 1.2, we find that dense design performs better than sparse design.

5.4 Real data example: Environmental pollution data

We applied the proposed test method to an environmental pollution data recorded in a southwestern city of China which is an industrial zone. A primary motivation for the first part of this research comes from a study of the comparison of air pollutants SO_2 and $PM2.5$ recorded in some southwestern areas of China two years ago. In industrial area, SO_2 is primarily caused by combustion procedure such as burning fuels, electric utilities, and other industrial activities. The dirty air in China, known as smog, has been blamed by World Health Organization (WHO) and the public. More policies and studies have been conducted to do against the air pollution for public health. In one southwestern city, the emission levels of SO_2 and $PM2.5$ were measured per hour in several workstations each day during the whole year in 2013. Fig. 5.2 shows SO_2 and $PM2.5$ emission levels for 365 days in 2013 which were measured by an environmental control station close to an industrial area in the center of the city. The hourly measurements in a day (24 hours) formed natural emission level curves of the day. One of our interest in studying this air pollutant data is to test the equality of covariance operators of SO_2 and $PM2.5$ in working days and in non-working days or varying seasonally. It happened that all hourly data could be observed, or just incompletely observed in a day sometimes because of machine out of run or meter burst by high pollutant levels.

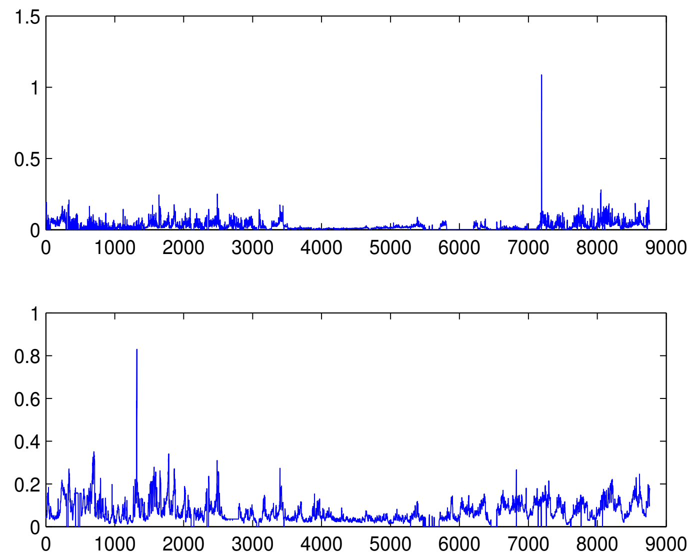

Figure 5.2: Top: SO_2 emission levels for 365 days in 2013. Bottom: $PM2.5$ emission levels for 365 days in 2013.

Example 5.1. *Testing equality of covariance operators for working day and non-working day.*

For $PM2.5$, 3 days of 250 working days and 4 days of 115 non-working days were completely unobserved. The total number of samples of working days was 247 where 38 curves had incompletely observations and 111 for non-working days where 12 curves had incompletely observations. A subsample of 20 randomly selected curves for working days and non-working days are plotted in Fig. 5.3. The numbers $A_{ij}^{(1)}$ and $A_{ij}^{(2)}$ of observation per day for working days and non-working days varied from 1 to 24, and two histograms of $A_{ij}^{(1)}$ and $A_{ij}^{(2)}$ are shown in Fig. 5.4. The initial estimated covariance operators were displayed in Fig. 5.5 for working days and non-working days. From the initial estimated covariance operators, it is seen that the volatility of $PM2.5$ levels in non-working days is higher than that of working days from 0 to 18 o'clock. However, more volatility is appeared in working days from 19 to 23 o'clock. Our interest is to test if the covariance operators of $PM2.5$ emission level curves of working and non-working days are significantly different. This motivates a two-sample covariance operators testing problem.

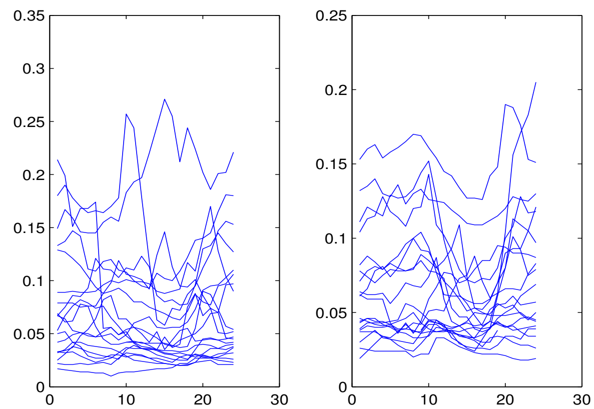

Figure 5.3: Left: A subsample of 20 randomly selected curves for working days; Right: A subsample of 20 randomly selected curves for non-working days.

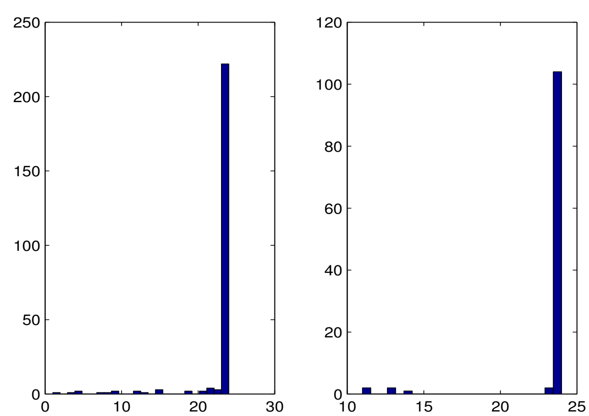

Figure 5.4: Frequency distributions of the number of observation for $PM2.5$. Left: 234 working days; Right: 106 non-working days.

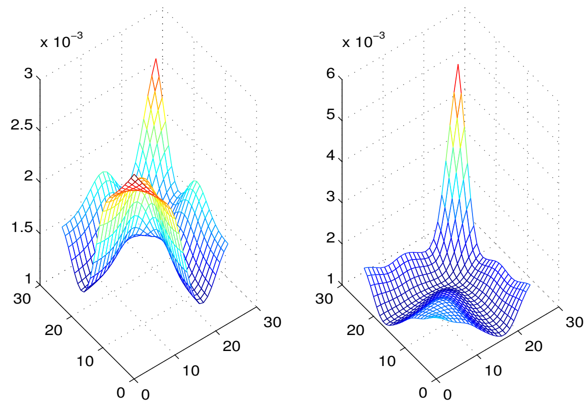

Figure 5.5: The initial estimated covariance operators. Left: working days; Right: non-working days.

According to method introduced in Section 5.2, $K = 4$ is selected. Based on the asymptotic distribution of the test statistic $\widehat{\mathrm{TC}}$, the p-value was calculated to be 0.8982. We also computed the p-value of different K in Table 5.3. All the results indicate that there is little evidence that the covariance operators are different for working and non-working days.

In reality, industrial pollution and automobile exhaust are the main sources of $PM2.5$ in the city. For the factory, production was business as usual in the non-working days. Therefore, the main factor that cause different manifestations is automobile exhaust. The number of workers in non-working days may be have more choices to stay at home or outdoor than that of working days at night. This leads to more volatility of $PM2.5$ levels in non-working days from 20 to 23 o'clock. On the contrary, there are so many more options of transport to consider for workers in working days in the daytime. Thus, it shows that the volatility of $PM2.5$ levels in working days is more larger than that of non-working days from 0 to 19 o'clock. Therefore, this leads to different modes for workdays and non-workdays. Generally speaking, we think that $PM2.5$ emission level curves of workdays and non-workdays are not different.

Table 5.3: p-values of the test for the SO_2 and $PM2.5$ data.

K	2	3	4	5
working day and non-workday				
$PM2.5$	0.3926	0.3858	0.8982	0.8145
SO_2	0.5711	0.6756	0.1721	0.2377
Spring and Summer				
$PM2.5$	0.0000	0.0000	0.0000	0.0000
SO_2	0.4724	0.7541	0.7009	0.6036
Spring and Autumn				
$PM2.5$	0.0124	0.0342	0.0524	0.1769
SO_2	0.8168	0.7888	0.9480	0.8459
Spring and Winter				
$PM2.5$	0.2319	0.6014	0.1885	0.4104
SO_2	0.0352	0.3589	0.4358	0.9459
Summer and Autumn				
$PM2.5$	0.0000	0.0000	0.0000	0.0000
SO_2	0.7714	0.4188	0.2401	0.1657
Summer and Winter				
$PM2.5$	0.0000	0.0000	0.0000	0.0000
SO_2	0.0348	0.0101	0.0021	0.0000
Autumn and Winter				
$PM2.5$	0.0000	0.0671	0.1066	0.3173
SO_2	0.0187	0.0000	0.0094	0.0133

For SO_2, we considered 234 workdays where 100 days have incomplete observations and 106 non-workdays where 47 days have incompletely observations because some days it could not be recorded. Using the method similar to that of $PM2.5$, the p-value is 0.1721 by the proposed test method. Also, the p-values of different K are listed in Table 5.3. This implies that there is little evidence that the covariance operators of the SO_2 levels are different for working and non-working days.

Example 5.2. *Testing the covariance operators for Spring, Summer, Autumn and Winter.*

In this example, we firstly consider the covariance operators testing of $PM2.5$ level for Spring, Summer, Autumn and Winter where the days are 91, 94, 90, and 90, respectively. But we only obtain the number of curves are 90, 94, 84, and 90 because some days can not be recorded. Among of curves, some are fully recorded and others have incompletely record. The curves of four seasons are plotted in Fig. 5.6, respectively. Four histograms of the number of observation per day for four seasons are shown in Fig. 5.7.

The initial estimated covariance operators were displayed in Fig. 5.8 for four seasons. For the initial estimated covariance operators, we can see that $PM2.5$ levels in Spring and Winter jumped by the biggest amount but it can be reduced to a minimum in Summer and Autumn. We are interested in whether the curves have the same covariance operators for the combinations (Spring, Summer), (Spring, Autumn), (Spring, Winter), (Summer, Autumn), (Summer, Winter), and (Autumn, Winter).

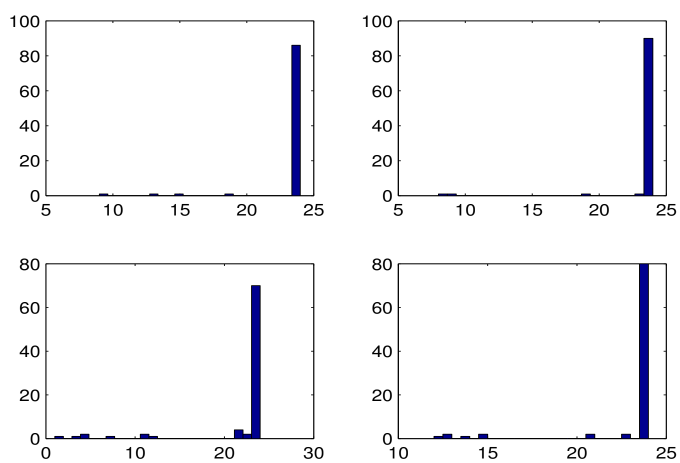

Figure 5.7: Four histograms of the number of observation per day for four seasons.

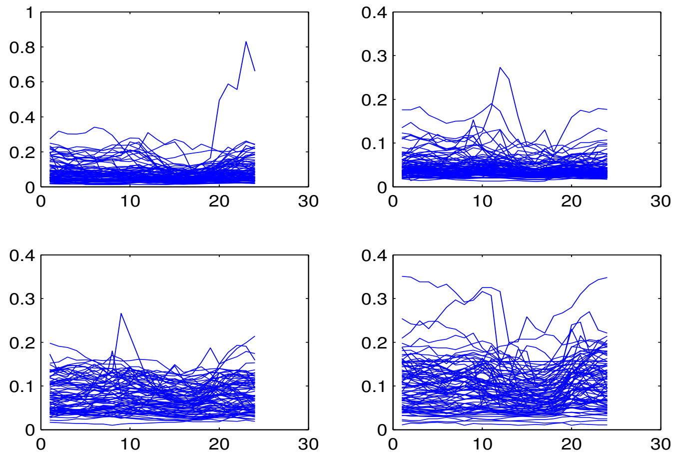

Figure 5.6: The curves of four seasons: Spring, Summer, Autumn, and Winter.

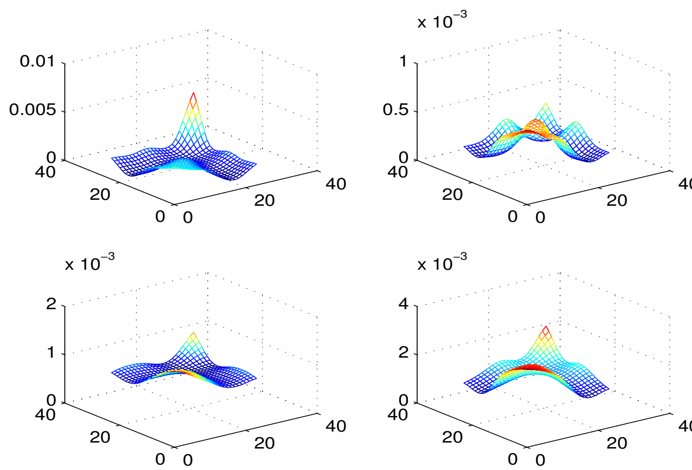

Figure 5.8: The initial estimated covariance operators for four seasons.

The p-values for different combinations were computed and collected in Table 5.3 using the proposed test method. From Table 5.3, we can see that the hypothesis $\gamma_1(t,s) = \gamma_2(t,s)$ is not reasonable except for the combination (Spring, Autumn), (Spring, Winter) and (Autumn, Winter). For the test of covariance operators of above three combinations, the p-value using the proposed test method were computed to be 0.0524, 0.1885 and 0.1066, respectively. Thus, there is little evidence that the covariance operators are different for the combination (Spring, Autumn), (Spring, Winter) and (Autumn, Winter).

By the similar method to that of $PM2.5$, we test the mean function of SO_2 for the combination (Spring, Summer), (Spring, Autumn), (Spring, Winter), (Summer, Autumn), (Summer, Winter), and (Autumn, Winter). The results are also display in Table 5.3, Different conclusions are drawn from Table 5.3. Almost of the combination can not reject the null hypothesis $H_0 : \gamma_1(t, s) = \gamma_2(t, s)$ except for the combination (Summer, Winter) and (Autumn, Winter).

5.5 Proofs of main results

Proof of lemma 5.1

$$
\begin{aligned}
&\left|\hat{\nu}_k\hat{\phi}_k(t)-\nu_k\phi_k(t)\right| \\
&=\left|\int_0^1 \hat{\gamma}_0(t,s)\hat{\phi}_k(s)ds-\int_0^1 \gamma(t,s)\phi_k(s)ds\right| \\
&=\left|\int_0^1 \left\{\frac{n_2}{n_1+n_2}\hat{\gamma}_1(t,s)+\frac{n_1}{n_1+n_2}\hat{\gamma}_2(t,s)\right\}\hat{\phi}_k(s)ds\right. \\
&\qquad\left.-\int_0^1 \left\{(1-\theta)\gamma_1(t,s)+\theta\gamma_2(t,s)\right\}\phi_k(s)ds\right| \\
&=\left|\int_0^1 \left\{\frac{n_2}{n_1+n_2}\hat{\gamma}_1(t,s)-(1-\theta)\gamma_1(t,s)\right\}\hat{\phi}_k(s)ds\right. \\
&\qquad\left.+\int_0^1 \left\{\frac{n_1}{n_1+n_2}\hat{\gamma}_2(t,s)-\theta\gamma_2(t,s)\right\}\phi_k(s)ds\right| \\
&\leqslant\left|\int_0^1 \left\{\frac{n_2}{n_1+n_2}\hat{\gamma}_1(t,s)-(1-\theta)\hat{\gamma}_1(t,s)\right\}\hat{\phi}_k(s)ds\right| \\
&\qquad+\left|\int_0^1 \left\{(1-\theta)\hat{\gamma}_1(t,s)-(1-\theta)\gamma_1(t,s)\right\}\hat{\phi}_k(s)ds\right| \\
&\qquad+\left|\int_0^1 \left\{\frac{n_1}{n_1+n_2}\hat{\gamma}_2(t,s)-\theta\hat{\gamma}_2(t,s)\right\}\phi_k(s)ds\right| \\
&\qquad+\left|\int_0^1 \left\{\hat{\gamma}_2(t,s)-\theta\gamma_2(t,s)\right\}\phi_k(s)ds\right| \\
&\equiv G_1+G_2+G_3+G_4
\end{aligned}
$$

According to assumption 5.8, we have $G_1\to 0$ and $G_3\to 0$.

For G_2, we have

$$\begin{aligned}G_2 &\leqslant (1-\theta)\int_0^1 |\hat{\gamma}_1(t,s)-\gamma_1(t,s)| \left|\hat{\phi}_k(s)\right| ds + (1-\theta)\int_0^1 |\gamma_1(t,s)| \left|\hat{\phi}_k(s)-\phi_k(s)\right| ds \\ &\leqslant (1-\theta)\sqrt{\int_0^1 \{\hat{\gamma}_1(t,s)-\gamma_1(t,s)\}^2 ds}\left\|\hat{\phi}_k(s)\right\| \\ &\quad + (1-\theta)\sqrt{\int_0^1 \{\gamma_1(t,s)\}^2 ds}\left\|\hat{\phi}_k(s)-\phi_k(s)\right\|\end{aligned}$$

According to assumptions 5.1-5.5, we can obtain $G_2 = O_p(\frac{\pi_k^{Y^{(1)}}}{\sqrt{n_1 h_1^2}})$. Analogously, $G_4 = O_p(\frac{\pi_k^{Y^{(2)}}}{\sqrt{n_2 h_2^2}})$ can be obtained.

Proof of lemma 5.2

Under assumptions 5.1-5.6, if we can prove

$$\hat{\eta}_k^{(1)} - \eta_k^{(1)} = O_p\left(n_1^{-1/2}\right), \quad \hat{\eta}_k^{(2)} - \eta_k^{(2)} = O_p\left(n_2^{-1/2}\right).$$

then lemma 5.2 can be easily proved. It can be observed

$$\begin{aligned}&\hat{\eta}_k^{(1)} - \eta_k^{(1)} \\ &= \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_{il_1}^{(1)}\phi_k(t_{il_1}) - \eta_k^{(1)}\right\} \\ &\quad + \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_{il_1}^{(1)}\hat{\phi}_k(t_{il_1}) - \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i} Y_{il_1}^{(1)}\phi_k(t_{il_1})\right\} \\ &\equiv A_1 + A_2.\end{aligned} \tag{5.5}$$

For A_1, we have

$$\begin{aligned}
A_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\left\{v_i^{(1)}(t_{il_1})+\varepsilon_{il_1}^{(1)}\right\}\phi_l(t_{il_1}) \\
&\quad+\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\mu_1(t_{il_1})\phi_l(t_{il_1})-\eta_k^{(1)}\right\} \\
&\equiv A_{11}+A_{12}.
\end{aligned} \tag{5.6}$$

It is easy to see that A_{11} is the average of independent identically distributed random variables. By the central limit theorem, we obtain

$$A_{11}=O_p\left(n_1^{-1/2}\right). \tag{5.7}$$

For A_{12}, according to assumption 5.7, we have

$$A_{12}=o\left(n_1^{-1/2}\right). \tag{5.8}$$

By (5.6), (5.7) and (5.8), we obtain

$$A_1=O_p\left(n_1^{-1/2}\right). \tag{5.9}$$

For A_2, we have

$$\begin{aligned}
A_2 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\mu_1(t_{il_1})\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\} \\
&\quad+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}v_i^{(1)}(t_{il_1})\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\} \\
&\quad+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\sum_{l_1=1}^{N_i}\varepsilon_{ik}^{(1)}\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\} \\
&\equiv A_{21}+A_{22}+A_{23}.
\end{aligned} \tag{5.10}$$

According to Cauchy-Schwarz inequality, assumption 5.2 and lemma 5.1, we have

$$\begin{aligned}A_{21} &\leqslant \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\left[\sum_{l_1=1}^{N_i}\mu_1^2(t_{il_1})\sum_{l_1=1}^{N_i}\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\}^2\right]^{1/2}\\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{\frac{1}{N_i}\sum_{l_1=1}^{N_i}\mu_1^2(t_{il_1})}\sqrt{\frac{1}{N_i}\sum_{l_1=1}^{N_i}\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\}^2}\\ &= O_p\left(n_1^{-1/2}\right).\end{aligned} \tag{5.11}$$

For A_{22}, according to Cauchy-Schwarz inequality, assumption 5.2 and lemma 5.1, we have

$$\begin{aligned}A_{22} &\leqslant \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i}\left[\sum_{l_1=1}^{N_i}v_i^{(1)2}(t_{il_1})\sum_{l_1=1}^{N_i}\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\}^2\right]^{1/2}\\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{\frac{1}{N_i}\sum_{l_1=1}^{N_i}v_i^{(1)2}(t_{il_1})}\sqrt{\frac{1}{N_i}\sum_{l_1=1}^{N_i}\left\{\hat{\phi}_k(t_{il_1})-\phi_k(t_{il_1})\right\}^2}\\ &= O_p\left(n_1^{-1/2}\right).\end{aligned} \tag{5.12}$$

Using the arguments similar to that of (5.12), it can be shown that

$$A_{23} = O_p\left(n_1^{-1/2}\right). \tag{5.13}$$

By (5.10), (5.11), (5.12), and (5.13), we have

$$A_2 = O_p\left(n_1^{-1/2}\right). \tag{5.14}$$

By (5.5), (5.9) and (5.14), we obtain $\hat{\eta}_k^{(1)} - \eta_k^{(1)} = O_p(n_1^{-1/2})$. Similarly, we can prove $\hat{\eta}_k^{(2)} - \eta_k^{(2)} = O_p(n_2^{-1/2})$.

Now we proof lemma 5.2. It can be observed that

$$
\begin{aligned}
\hat{\rho}_{kk'}^{(1)} &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\hat{\eta}_{ik}^{(1)}-\hat{\eta}_{k}^{(1)}\right\}\left\{\hat{\eta}_{ik'}^{(1)}-\hat{\eta}_{k'}^{(1)}\right\}\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\eta}_{ik}^{(1)}\hat{\eta}_{ik'}^{(1)}-\hat{\eta}_{k}^{(1)}\hat{\eta}_{k'}^{(1)}\\
&\equiv B_1 - B_2.
\end{aligned}
$$

B_1 can be decomposed as

$$
\begin{aligned}
B_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{1}{N_i}\sum_{l_1=1}^{N_i}Y_i^{(1)}(t_{il_1})\hat{\phi}_k(t_{il_1})\right\}\left\{\frac{1}{N_i}\sum_{l_1=1}^{N_i}Y_i^{(1)}(t_{il_1})\hat{\phi}_{k'}(t_{il_1})\right\}\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}Y_i^{(1)}(t_{il_1})^2\hat{\phi}_k(t_{il_1})\hat{\phi}_{k'}(t_{il_1})+\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}Y_i^{(1)}(t_{il_{11}})Y_i^{(1)}(t_{il_{12}})\hat{\phi}_k(t_{il_{11}})\hat{\phi}_{k'}(t_{il_{12}})\right\}\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left[\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\mu_1(t_{il_1})+v_i^{(1)}(t_{il_1})+\varepsilon_{il_1}^{(1)}\right\}^2\hat{\phi}_k(t_{il_1})\hat{\phi}_{k'}(t_{il_1})\right.\\
&\quad\left.+\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\left\{\mu_1(t_{il_{11}})+v_i^{(1)}(t_{il_{11}})+\varepsilon_{il_{11}}^{(1)}\right\}\left\{\mu_1(t_{il_{12}})+v_i^{(1)}(t_{il_{12}})+\varepsilon_{il_{12}}^{(1)}\right\}\hat{\phi}_k(t_{il_{11}})\hat{\phi}_{k'}(t_{il_{12}})\right]\\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\mu_1(t_{il_1})+v_i^{(1)}(t_{il_1})+\varepsilon_{il_1}^{(1)}\right\}^2\phi_k(t_{il_1})\phi_{k'}(t_{il_1})\\
&\quad+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\left\{\mu_1(t_{il_{11}})+v_i^{(1)}(t_{il_{11}})+\varepsilon_{il_{11}}^{(1)}\right\}\left\{\mu_1(t_{il_{12}})+v_i^{(1)}(t_{il_{12}})+\varepsilon_{il_{12}}^{(1)}\right\}\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}})\\
&\quad+\frac{1}{n_1}\sum_{i=1}^{n}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\mu_1(t_{il_1})+v_i^{(1)}(t_{il_1})+\varepsilon_{il_1}^{(1)}\right\}^2\left\{\hat{\phi}_k(t_{il_1})\hat{\phi}_{k'}(t_{il_1})-\phi_k(t_{il_1})\phi_{k'}(t_{il_1})\right\}\\
&\quad+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\left\{\mu_1(t_{il_{11}})+v_i^{(1)}(t_{il_{11}})+\varepsilon_{il_{11}}^{(1)}\right\}\left\{\mu_1(t_{il_{12}})+v_i^{(1)}(t_{il_{12}})+\varepsilon_{il_{12}}^{(1)}\right\}\\
&\qquad\times\left\{\hat{\phi}_k(t_{il_{11}})\hat{\phi}_{k'}(t_{il_{12}})-\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}})\right\}\\
&\equiv B_{11}+B_{12}+B_{13}+B_{14}.
\end{aligned}
$$

For B_{11}, we have

$$\begin{aligned}
B_{11} = & \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\{\mu_1(t_{il_1})\}^2\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
& + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{v_i^{(1)}(t_{il_1})\right\}^2\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
& + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\varepsilon_{il_1}^{(1)}\right\}^2\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
& + \frac{2}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\mu_1(t_{il_1})v_i^{(1)}(t_{il_1})\right\}\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
& + \frac{2}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{\mu_1(t_{il_1})\varepsilon_{il_1}^{(1)}\right\}\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
& + \frac{2}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_1=1}^{N_i}\left\{v_i^{(1)}(t_{il_1})\varepsilon_{il_1}^{(1)}\right\}\phi_k(t_{il_1})\phi_{k'}(t_{il_1}) \\
\equiv & B_{111} + B_{112} + B_{113} + B_{114} + B_{115} + B_{116}.
\end{aligned}$$

For B_{12}, we have

$$\begin{aligned}
B_{12} =& \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\mu_1(t_{il_{11}})\mu_1(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\mu_1(t_{il_{11}})v_i^{(1)}(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\mu_1(t_{il_{11}})\varepsilon_{il_{12}}^{(1)}\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}v_i^{(1)}(t_{il_{11}})\mu_1(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}v_i^{(1)}(t_{il_{11}})v_i^{(1)}(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}v_i^{(1)}(t_{il_{11}})\varepsilon_{il_{12}}^{(1)}\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\varepsilon_{il_{11}}^{(1)}\mu_1(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\varepsilon_{il_{11}}^{(1)}v_i^{(1)}(t_{il_{12}})\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
&+\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{N_i^2}\sum_{l_{11}\neq l_{12}}\varepsilon_{il_{11}}^{(1)}\varepsilon_{il_{12}}^{(1)}\phi_k(t_{il_{11}})\phi_{k'}(t_{il_{12}}) \\
\equiv& B_{121}+B_{122}+B_{123}+B_{124}+B_{125}+B_{126}+B_{127}+B_{128}+B_{129}.
\end{aligned}$$

For B_{125}, we have

$$B_{125}\xrightarrow{d} N\left(\iota_1, \pi_{1k}^{(1)}\right)$$

where $\iota_1=\int_0^1\int_0^1\phi_k(t)\gamma_1(t,s)\phi_{k'}(s)dtds$ and $\pi_{1k}^{(1)}=\int_0^1\int_0^1\phi_k^2(t)[E\{v_i^{(1)2}(t)v_i^{(2)2}(s)\}]\phi_{k'}^2(s)dtds-\{\int_0^1\int_0^1\phi_k(t)\gamma_1(t,s)\phi_{k'}(s)dtds\}^2$.

By assumption 5.7, we may obtain

$$B_{121} - \hat{\eta}_k^{(1)} \hat{\eta}_{k'}^{(1)} = o_p \left(n_1^{-1/2} \right).$$

By assumptions 5.2 and 5.5, we have

$$B_{11g} = o_p \left(n_1^{-1/2} \right), \ g = 1, 2.$$

Analogously, we have $B_{11g} = o_p(n_1^{-1/2}), \ g = 3, 4, 5, 6$. Because $B_{122} + B_{123} + B_{124} + B_{126} + B_{127} + B_{128} + B_{129}$ are the average of independent identically distributed random variables with mean zeros and variance $\pi_{1k}^{(2)}$, where

$$\begin{aligned}
\pi_{1k}^{(2)} = & \left\{ \int_0^1 \mu_1(t)\phi_k(t)dt \right\}^2 \left[\int_0^1 \mu_1(s)\phi_{k'}(s)ds \right]^2 \\
& + \left\{ \int_0^1 \mu_1(t)\phi_{k'}(t)dt \right\}^2 \int_0^1 \int_0^1 \phi_k(s_1)\gamma_1(s_1, s_2)\phi_k(s_2)ds_1 ds_2 \\
& + \left\{ \int_0^1 \mu_1(t)\phi_k(t)dt \right\}^2 \int_0^1 \int_0^1 \phi_{k'}(s_1)\gamma_1(s_1, s_2)\phi_{k'}(s_2)ds_1 ds_2 + \sigma^4.
\end{aligned}$$

So, we obtain

$$B_{122} + B_{123} + B_{124} + B_{126} + B_{127} + B_{128} + B_{129} \xrightarrow{d} N \left(0, \pi_{1k}^{(2)} \right).$$

According to Cauchy-Schwarz inequality and lemma 5.1, we have

$$B_{1g} = o_p \left(n_1^{-1/2} \right), \ g = 3, 4.$$

By the central limit theorem, we obtain

$$\sqrt{n_1} \left(\hat{\rho}_{kk'}^{(1)} - \rho_{kk'}^{(1)} \right) \xrightarrow{d} N \left(\iota_1, \omega_{kk'}^{(1)} \right),$$

where $\iota_1 = \int_0^1 \int_0^1 \phi_k(t)\gamma_1(t, s)\phi_{k'}(s)dtds$ and $\omega_{kk'}^{(1)} = \pi_{1k}^{(1)} + \pi_{1k}^{(2)}$.

Similarly, we can prove $\sqrt{n_2}(\hat{\rho}^{(2)}_{kk'} - \rho^{(2)}_{kk'}) \xrightarrow{d} N(\iota_2, \omega^{(2)}_{kk'})$, where $\iota_2 = \int_0^1\int_0^1 \phi_k(t)\gamma_2(t,s)\phi_{k'}(s)dtds$ and $\omega^{(2)}_{kk'}$ is defined in Section 5.2.

Under H_0, we obtain the conclusion of lemma 5.2. The proof is then completed.

Proof of theorem 5.1 According to lemma 5.1, lemma 5.2 and Slutsky theory, we can easily obtain the the conclusion of theorem 5.1. The proof is then completed.

Proof of theorem 5.2 According to lemma 5.2, we have

$$\hat{\rho}^{(1)}_{kk'} \xrightarrow{p} \int_0^1\int_0^1 \phi_k(t)\gamma_1(t,s)\phi_{k'}(s)dtds,\ \hat{\rho}^{(2)}_{kk'} \xrightarrow{p} \int_0^1\int_0^1 \phi_k(t)\gamma_2(t,s)\phi_{k'}(s)dtds.$$

Under H_1, we obtain

$$\widehat{TC} \xrightarrow{p} \frac{\frac{n_1 n_2}{n_1+n_2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\frac{\left[\int_0^1\int_0^1 \phi_k(t)\{\gamma_1(t,s)-\gamma_2(t,s)\}\phi_{k'}(s)dtds\right]^2}{\frac{1}{n_1}\omega^{(1)}_{kk'}+\frac{1}{n_2}\omega^{(2)}_{kk'}} - \frac{K(K-1)}{2}}{\sqrt{K(K-1)}} \to \infty,$$

the proof is then completed.

Chapter 6

Testing Equality of Covariance Matrix for High-dimensional Data

6.1 Introduction

Initiating from functional data analysis, we gain much insight in presenting a novel two-sample testing procedure on high-dimensional covariance matrices under the non-normality assumption and "*large* p, *small* n" paradigm.

Testing the equality of two covariance matrices Σ_1 and Σ_2 is an significant problem in multivariate analysis. Many statistical procedures including the classical Fisher's linear discriminant analysis depend on the assumption of equal covariance matrices has been studied, see Sugiura and Nagao (1968), Gupta and Giri (1973), Perlman (1980), Gupta and Tang (1984), O'Brien (1992), and Anderson (2003). In particular, the likelihood ratio test (LRT) is commonly used and enjoys certain optimality under regular conditions. However, the abovementioned work are based on the low-dimensional data.

High-dimensional data are increasingly encountered in many statistical applications with the most prominently in biological and financial studies. A common feature of high-dimensional data is that the data dimension is much larger than the sample size, namely the "large p, small n". Tests of significance are challenging for high-dimensional data analysis arisen accumulated research interest. Chen and Qin (2010) proposed a two-sample test for the means without requiring explicit conditions in the relationship between the data dimension p and sample size n under the data structure of i.i.d. within each sample. Zhong and Chen (2011) introduced simultaneous test for coefficients in high-dimensional linear regression models with factorial designs. Qiu and Chen (2012) introduced a test for bandedness of high-dimensional covariance matrices and bandwidth estimation without assuming a specific parametric distribution. Zhong et al. (2013) considered two alternative tests to the Higher Criticism test of Donoho and Jin (2004) for high-dimensional means under the sparsity of the nonzero means for sub-Gaussian distributed data with unknown column-wise dependence. Under dependence assumption, Cai et al. (2014) developed a test for testing the equality of two mean vectors based on a linear transformation of the data by the precision matrix which incorporates the correlations between the variables. Wang et al. (2015b) concerned with testing the population mean vector of nonnormal high-dimensional multivariate data and proposed a test statistic based on the spatial sign function of the observed data.

The conventional testing procedures such as the LRT for covariance matrices either perform poorly or are not even well defined under such high-dimensional data setting. Several tests for the equality of two large covariance matrices have been proposed, Ledoit and Wolf (2002) showed the locally best invariant test based on John's U statistic to be (n, p)-consistent when $p/n \to 0 < \infty$, where c is a constant known as the concentration. Srivastava (2005) proposed a test based on the first and second arithmetic means of the eigenvalues of the sample covariance but only requires the more general condition $n = O(p^\delta),\ 0 < \delta \leqslant 1$. Schott (2005) introduced a simple statistic for testing the complete independence of random variable under a multivariate normal distribution and compared the finite sample size performance with the Likelihood ratio test. Schott (2006) proposed a testing procedure for the test that the smallest eigenvalues of a covariance matrix are equal based on Ledoit and Wolf (2002). Schott (2007) introduced a test for the equality of several covariance matrices based on the Frobenius norm of the deviation of any two covariance matrices. Srivastava and Yanagihara (2010) constructed a test that relied on a measure of distance by $\mathrm{tr}(\Sigma_1^2)/(\mathrm{tr}(\Sigma_1))^2 - \mathrm{tr}(\Sigma_2^2)/(\mathrm{tr}(\Sigma_2))^2$. Both of these two tests are designed for the multivariate normal populations. Without the strong assumption of Gaussian distribution of two vector samples, Chen et al. (2010) resented nonparametric testing statistics for sphericity and identity of the covariance matrix based on estimators for traces of covariance matrix and its square when p may be a larger order of n. Fisher et al. (2010) developed a new test procedure of the covariance matrix based on Cauchy-Schwarz inequality utilizing the ratio of the second and fourth arithmetic means of the sample eigenvalues. Fisher (2012) explored the problem of testing the covariance matrix is an identity matrix when the dimensionality is equal to the sample size or larger. Li and Chen (2012) constructed a testing statistic based on an unbiased estimator of the Frobenius norm of the difference of two covariance matrices allowing the dimension to be much larger than the sample sizes, whereas their empirical size and power do not perform well when n is comparatively smaller than p. Cai and Ma (2013) proposed a covariance matrix test based on U-statistics in the high-dimensional setting under the data structure of i.i.d. within each sample. Cai et al. (2013) developed a whole variance-covariance matrices test based on the maximum of the standardized differences between the entries of the two sample covariance matrices Σ_1 and $\Sigma_@$ under the data structure of independence between the two samples and i.i.d. within each sample. Also, they considered the support recovery of difference between two covariance matrices under the null hypothesis is rejected as well as testing them row by row. Li and Qin (2014) proposed tests for an identity matrix and for the equality of two covariance matrices based on empirical spectral distributions (SD) when the data dimension and the sample size are both large.

The novelty lies in the fact that high-dimensional data could be converted into a stochastic process by stringing via existed methods and software packages, refer to Chen et al. (2011b).

Stimulation for this part of this research comes from the close relationship between functional data and high-dimensional data. Chen et al. (2011b) constructed stringing as a tool to reorder the components of high-dimensional vector data by multi-dimensional scaling (MDS), thus transforming the high-dimensional vectors into functional data. MDS projects data into a low-dimensional target space, where the configuration in the target space aims to reproduce the proximity relations in the original space, by minimizing a cost function. as well as transformed Pearson correlation as proximity measures in the original high-dimensional predictor space (Cox and Cox, 2001). The configuration obtained by MDS projection into one dimension provides an ordering of the predictors and assigns a location to each predictor, aligning the predictors within a one-dimensional interval like pearls on a string. Predictors with high proximity will tend to be positioned closely together after MDS projection, enabling the construction of smooth trajectories in function space. Once the data have been converted into a smooth stochastic process by stringing, functional principal component analysis (FPCA) can be used to summarize and further analyze the high-dimensional data. Its implementation is readily conducted by the option *PACE-Stringing* from PACE 2.15 package in Matlab, see `http://anson.ucdavis.edu/~mueller/data/pace.html`. This motivates us to take advantage of methodologies from functional data analysis by mapping high-dimensional predict vectors into infinite-dimensional smooth random functions (stochastic process).

6.2 Methodology and main results

Now we come to develop testing procedure on equality of covariance matrices from two high-dimensional data samples. Let $X_1^{(1)}, \cdots, X_{n_1}^{(1)}$ and $X_1^{(2)}, \cdots, X_{n_2}^{(2)}$ be independent and identically distributed samples drawn from two p-dimensional random vectors with variance-covariance matrices $\Sigma_1^{(0)}$ and $\Sigma_2^{(0)}$, respectively, where $p/n_1 \to \infty$ and $p/n_2 \to \infty$ within a large p, small n setting. Our objective is to test

$$H_0^{®} : \Sigma_1^{(0)} = \Sigma_2^{(0)} \text{ vs. } H_1^{®} : \ \Sigma_1^{(0)} \neq \Sigma_2^{(0)}. \tag{6.1}$$

Similar to Section 5.2, we propose a three-step algorithm below.

Step 6.1. *String high-dimensional data into functional data type.*

In stringing, every high-dimensional vector is thought of as being generated by a hidden smooth stochastic process $\{Z(t), t \in [0,1]\}$, where each element of a grid of support points $t_i \in [0,1]$ indices one possible predictor, s_i being the position of the corresponding predictor and $Z(t_i)$ its value. The distance between predictor positions which can be derived from proximities such as empirical Pearson correlation is interpreted as a measure of the relatedness of the predictors. Once a distance matrix has been determined, the predictors are stringed into the real line by minimizing the stress function, and the detail can be seen in Chen et al. (2011b). So, each p-dimensional random vector is converted into a random function where the recordings of every random function are scheduled on a regular and dense grid on interval $[0,1]$. Denote $Y_1^{(1)}, \cdots, Y_{n_1}^{(1)}$ and $Y_1^{(2)}, \cdots, Y_{n_2}^{(2)}$ to be the reordered random vectors with covariance matrices Σ_1 and Σ_2, respectively. Then test (6.1) is equivalent to test

$$H_0^{®} : \Sigma_1 = \Sigma_2 \text{ vs. } H_1^{®} : \ \Sigma_1 \neq \Sigma_2.$$

Step 6.2. *Obtain a pooling covariance matrix and spectrum decomposition.*

Let $\widehat{\Sigma}_1^{(0)} = \frac{1}{n_1}\sum_{i=1}^{n_1}(Y_i^{(1)} - \overline{Y}_{n_1}^{(1)})^\top(Y_i^{(1)} - \overline{Y}_{n_1}^{(1)})$ and $\widehat{\Sigma}_2^{(0)} = \frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i^{(2)} - \overline{Y}_{n_2}^{(2)})^\top(Y_i^{(2)} - \overline{Y}_{n_2}^{(2)})$ be the sample covariance matrices of two sample where $\overline{Y}_{n_1}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i^{(1)}$ and $\overline{Y}_{n_2}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i^{(2)}$, respectively. Denote $\widehat{\Sigma}^{(0)} \equiv \frac{n_2}{n_1+n_2}\widehat{\Sigma}_1^{(0)} + \frac{n_1}{n_1+n_2}\widehat{\Sigma}_2^{(0)}$. It is readily seen that $\widehat{\Sigma}^{(0)} \xrightarrow{p} \Sigma \equiv (1-\alpha)\Sigma_1 + \alpha\Sigma_2$, where α is defined at assumption 6.5 and Σ is an asymptotic pooled covariance matrices. Consequently, it has orthonormal eigenvector $\{e_k\}_{k\geqslant 1}$ and corresponding decreasing sequence of non-negative eigenvalues $\{\lambda_k\}_{k\geqslant 1}$ such that,

$$\Sigma = \lambda_k e_k^\top e_k.$$

Estimation of eigenvalues and eigenvectors $\{(\lambda_k, e_k)\}_{k\geqslant 1}$ is obtained by eigenequations,

$$\widehat{\Sigma}^{(0)}\hat{e}_k = \hat{\lambda}_k\hat{e}_k,$$

with orthonormal constraints on $\{\hat{e}_k\}_{k\geqslant 1}$.

Step 6.3. *Obtain the projection estimators of covariance matrices. We propose the following estimators of covariance matrices.*

$$\widehat{\Sigma}_1 = \sum_{k=1}^{K^{®}}\sum_{k'=1}^{K^{®}} \hat{\vartheta}_{kk'}^{(1)}\hat{e}_k^\top\hat{e}_{k'}, \quad \widehat{\Sigma}_2 = \sum_{k=1}^{K^{®}}\sum_{k'=1}^{K^{®}} \hat{\vartheta}_{kk'}^{(2)}\hat{e}_k^\top\hat{e}_{k'},$$

where

$$\hat{\vartheta}_{kk'}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\hat{\xi}_{ik}^{(1)} - \hat{\xi}_k^{(1)}\right\}\left\{\hat{\xi}_{ik'}^{(1)} - \hat{\xi}_{k'}^{(1)}\right\}, \ \hat{\xi}_k^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\xi}_{ik}^{(1)}, \ \hat{\xi}_{ik}^{(1)} = \frac{1}{p}Y_i^{(1)\top}\hat{e}_k;$$

$$\hat{\vartheta}_{kk'}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\left\{\hat{\xi}_{ik}^{(2)} - \hat{\xi}_k^{(2)}\right\}\left\{\hat{\xi}_{ik'}^{(2)} - \hat{\xi}_{k'}^{(2)}\right\}, \ \hat{\xi}_k^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\xi}_{ik}^{(2)}, \ \hat{\xi}_{ik}^{(2)} = \frac{1}{p}Y_i^{(2)\top}\hat{e}_k.$$

Here the tuning parameter $K^{®}$ can be selected by $K^{®} = \max\{K_1^{®}, K_2^{®}\}$, with $K_m^{®}$ being the minimum number of the eigenvalues of $\hat{\Sigma}_m$, $m = 1, 2$, which explains 95% of the total variation for the transformed random function sequences $Y_i^{(1)}, i = 1, \cdots, n_1$ and $Y_i^{(2)}, i = 1, \cdots, n_2$. However, it is noted that it is time consuming when p is too large in our simulation. To overcome this problem, we adopt a practical way to select $K^{®}$ in the simulation: we conduct a pilot 50 trials to select $\{K_i^{®}\}_{i=1}^{50}$, the median of which is employed as our $K^{®}$ in all simulated trials.

Remark 6.1. *In fact, it is easy to extend the above method to the two sample test of high-dimensional vector with partial missing and partial completely observation. We can rearrange the high-dimensional data using the completely observed subsample. So, random vectors are converted into partial densely and partial sparsely random functions. According to the remark 5.1, the proposed method can also be applied for this kind of data.*

We demand the following assumptions in order to derive the asymptotic properties of statistic $\widehat{TC}^{®}$.

Assumption 6.1. $Y_i^{(1)} = \mu_1 + w_i^{(1)}$ *and* $Y_i^{(2)} = \mu_2 + w_i^{(2)}$ *where* $E(w_i^{(1)}) = E(w_i^{(2)}) = 0$, $cov(w_i^{(1)}, w_i^{(1)}) = \Sigma_1$ *and* $cov(w_i^{(2)}, w_i^{(2)}) = \Sigma_2$.

Assumption 6.2. *The eigenvalues of covariance matrices* Σ *such that* $\lambda_1 > \lambda_2 > \cdots$.

Assumption 6.3. μ_m, $m = 1, 2$ *may be written as* $\mu_m = \sum_{l=1}^{\infty} \xi_l^{(m)} e_l$, *where* $\xi_l^{(m)} = \frac{1}{p}\mu_m e_l^{\top}$.

Assumption 6.4. $\min\{n_1, n_2\} \to \infty$, $\frac{n_1}{n_1+n_2} \to \alpha$ *for a fixed constant* $\alpha \in (0, 1)$.

Assumption 6.5. *Assume the condition of Chen et al. (2011b) satisfies.*

Remark 6.2. *Assumption 6.1 is the slightly weaker than that of Li and Chen (2012) where* $w_i^{(1)}$ *and* $w_i^{(2)}$ *are divided into the product of two factor. Assumption 6.2 is* *used to guarantee the consistency of the estimators for eigenvector. Assumption 6.3 is used to prove the asymptotic normality of $\hat{\vartheta}^{(1)}_{ll'}$. Assumption 6.4 is a regular condition in two sample test. Assumption 6.5 is useful to guarantee high dimensional data which can be converted into a random function.*

In order to present our testing statistic, we need to use of the below asymptotic result.

Lemma 6.1. *Under assumptions 6.1-6.5 and $H_0^{®}$, we have*

$$\frac{\hat{\vartheta}^{(1)}_{kk'} - \hat{\vartheta}^{(2)}_{kk'}}{\sqrt{\frac{1}{n_1}\varpi^{(1)}_{kk'} + \frac{1}{n_2}\varpi^{(2)}_{kk'}}} \xrightarrow{d} N(0,1),$$

where

$$\varpi^{(1)}_{kk'} = \left\{u_1^\top e_{k'}\right\}^2 \left\{e_k^\top \Sigma_1 e_k\right\} + \left\{\mu_1^\top e_k\right\}^2 \left\{e_{k'}^\top \Sigma_1 e_{k'}\right\} + E\left[\left\{e_k^\top w_i^{(1)}\right\}^2 \left\{e_{k'}^\top w_i^{(1)}\right\}^2\right] - \left\{e_k^\top \Sigma_1 e_{k'}\right\}^2,$$

$$\varpi^{(2)}_{kk'} = \left\{\mu_2^\top e_{k'}\right\}^2 \left\{e_k^\top \Sigma_2 e_k\right\} + \left\{\mu_2^\top e_k\right\}^2 \left\{e_{k'}^\top \Sigma_2 e_{k'}\right\} + E\left[\left\{e_k^\top w_i^{(2)}\right\}^2 \left\{e_{k'}^\top w_i^{(2)}\right\}^2\right] - \left\{e_k^\top \Sigma_2 e_{k'}\right\}^2,$$

where $w_i^{(1)}$ and $w_i^{(2)}$ are stated in assumption 6.1.

Based on lemma 6.1, we then propose the following statistic:

$$\widehat{TC}^{®} = \frac{\frac{n_1 n_2}{n_1+n_2}\sum_{k=1}^{K^{®}}\sum_{k'=1}^{K^{®}} \frac{\left\{\hat{\vartheta}^{(1)}_{kk'} - \hat{\vartheta}^{(2)}_{kk'}\right\}^2}{\hat{\varpi}_{kk'}} - \frac{K^{®}(K^{®}-1)}{2}}{\sqrt{K^{®}(K^{®}-1)}},$$

where $\hat{\varpi}_{kk'} = \frac{n_2}{n_1+n_2}\hat{\varpi}^{(1)}_{kk'} + \frac{n_1}{n_1+n_2}\hat{\varpi}^{(2)}_{kk'}$ with $\hat{\varpi}^{(m)}_{kk'}$, $m = 1,2$ being the consistent estimators of $\varpi^{(m)}_{kk'}$, $m = 1,2$, respectively. In fact, $\varpi^{(m)}_{kk'}$, $m = 1,2$ are unknown and can be substituted by their consistent estimators $\frac{1}{n_1}\sum_{i=1}^{n_1}[\{\hat{\xi}^{(1)}_{ik} - \hat{\xi}^{(1)}_{k}\}\{\hat{\xi}^{(1)}_{ik'} - \hat{\xi}^{(1)}_{k'}\} - \Psi_1]^2$ and $\frac{1}{n_2}\sum_{i=1}^{n_2}[\{\hat{\xi}^{(2)}_{ik} - \hat{\xi}^{(2)}_{k}\}\{\hat{\xi}^{(2)}_{ik'} - \hat{\xi}^{(2)}_{k'}\} - \Psi_2]^2$, respectively, where $\Psi_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\{\hat{\xi}^{(1)}_{ik} - \hat{\xi}^{(1)}_{k}\}\{\hat{\xi}^{(1)}_{ik'} - \hat{\xi}^{(1)}_{k'}\}$ and $\Psi_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\{\hat{\xi}^{(2)}_{ik} - \hat{\xi}^{(2)}_{k}\}\{\hat{\xi}^{(2)}_{ik'} - \hat{\xi}^{(2)}_{k'}\}$.

We are ready to present the asymptotic results of the proposed test. Theorems 6.1 and 6.2 below establish the asymptotic behaviors of the statistic $\widehat{TC}^{®}$ under hypotheses $H_0^{®}$ and $H_1^{®}$, respectively. The proofs of these theorems are provided in Section 6.5.

Theorem 6.1. *Under assumptions 6.1-6.5 and $H_0^{®}$, we have*

$$\widehat{TC}^{®} \xrightarrow{d} N(0,1),\ \min\{n_1, n_2\} \to \infty.$$

Theorem 6.2. *Under assumptions 6.1-6.5 and $H_1^{®}$, we have*

$$\widehat{TC}^{®} \xrightarrow{p} \infty,\ \min\{n_1, n_2\} \to \infty.$$

6.3 Simulation studies

To compare the proposed method, denoted by $\widehat{TC}^{®}$, with Li and Chen (2012), denoted by $Li - Chen$, we carry out simulations for scenarios where p is much larger than n. We choose a set of data with p ranging from 32 to 700 and n ranging from 20 to 100, respectively. We consider a moving average model

$$Y_{il_1}^{(1)} = Z_{il_1}^{(1)} + 2Z_{i,l_1+1}^{(1)}, \tag{6.2}$$

as the null model of both populations for size evaluation. To assess the power performance, the first population is generated according to (6.2), while the second is from

$$Y_{il_2}^{(2)} = Z_{il_2}^{(2)} + 2Z_{i,l_2+1}^{(2)} + Z_{i,l_2+2}^{(2)}, \tag{6.3}$$

where $\{Z_{il_1}^{(1)}\}_{i=1,\cdots,n_1;k=1,\cdots,p}$ and $\{Z_{il_2}^{(2)}\}_{i=1,\cdots,n_2;q=1,\cdots,p+2}$ are $i.i.d$ sequences. Three combinations of distributions are experimented in models (6.2) and (6.3), respectively. They are: (i) both sequences are standard normal; (ii) centralized Gamma(4, 0.5) for Sample 1 and centralized Gamma(0.5,$\sqrt{2}$) for Sample 2; (iii) standard normal for Sample 1 and centralized Gamma(0.5,$\sqrt{2}$) for Sample 2. The last two combinations are designed to assess the performance under nonnormality. All the simulation results reported are based on 1000 simulations with the nominal significance level to be 5%. The empirical size and power of the test are reported in Table 6.1.

From Table 6.1, we can see when $\widehat{TC}^{®}$ and $Li-Chen$ have controllable empirical size. It is noticeable that $\widehat{TC}^{®}$ has higher power than that of $Li-Chen$. A significant improvement can be obtained particularly when n_1 is relatively smaller, say $n_1 = 20$, and p is large. This indicates that the proposed test is especially adaptive to the problem of "large p, small n". There are two reasons that might explain why our method performs better. On one hand, stringing method itself works better when dimension p is larger. On the other hand, observation on interval $[0, 1]$ means that it becomes denser when dimension p is increasing. This result is consistent with that in Section 5.3, i.e. dense design performs better than sparse design.

6.4 Real data example: Mitochondrial calcium concentration data

Mitochondrial Calcium Concentration Data has been studied by some authors. For example, Ruiz-Meana et al. (2003) investigated whether cariporide could inhibit mitochondrial Na+/H+ exchanger during ischemia, delaying H+ gradient dissipation and ATP exhaustion. Gregory et al. (2015) analyzed Mitochondrial Calcium Concentration Data to illustrate the test of two population mean vectors in the "*large* p, *small* n" setting. The mitochondrial calcium was measured in two groups (control and treatment). To a treatment group, a dose of cariporide was administered, which is believed to inhibit cell death due to oxidative stress. The investiga-

Table 6.1: Empirical size and power of test based on statistics $\widehat{TC}^{®}$ and $Li-Chen$.

	n_1,n_2	Test Statistics	64	128	256	512	700
			Size				
Sample 1∼	20	$\widehat{TC}^{®}$	0.026	0.031	0.015	0.022	0.031
		$Li-Chen$	0.054	0.051	0.048	0.051	0.038
$N(0,1)$	50	$\widehat{TC}^{®}$	0.037	0.029	0.017	0.022	0.032
		$Li-Chen$	0.060	0.033	0.043	0.054	0.049
Sample 2∼	80	$\widehat{TC}^{®}$	0.035	0.029	0.031	0.023	0.024
		$Li-Chen$	0.060	0.047	0.048	0.052	0.053
$N(0,1)$	100	$\widehat{TC}^{®}$	0.029	0.034	0.028	0.023	0.023
		$Li-Chen$	0.049	0.052	0.046	0.049	0.048
			Power				
	20	$\widehat{TC}^{®}$	0.352	0.470	0.683	0.863	0.933
		$Li-Chen$	0.256	0.267	0.277	0.282	0.291
	50	$\widehat{TC}^{®}$	0.927	0.988	0.999	1.000	1.000
		$Li-Chen$	0.821	0.830	0.837	0.832	0.849
	80	$\widehat{TC}^{®}$	0.996	1.000	1.000	1.000	1.000
		$Li-Chen$	0.992	0.991	0.998	0.999	0.998
	100	$\widehat{TC}^{®}$	1.000	1.000	1.000	1.000	1.000
		$Li-Chen$	1.000	0.999	1.000	1.000	1.000
			Size				
Sample 1∼	20	$\widehat{TC}^{®}$	0.045	0.025	0.026	0.019	0.020
		$Li-Chen$	0.117	0.069	0.063	0.051	0.040
$Gamma(4,0.5)$	50	$\widehat{TC}^{®}$	0.033	0.031	0.021	0.029	0.020
		$Li-Chen$	0.110	0.094	0.052	0.053	0.051
Sample 2∼	80	$\widehat{TC}^{®}$	0.041	0.036	0.027	0.038	0.033
		$Li-Chen$	0.111	0.093	0.067	0.064	0.044
$Gamma(0.5,\sqrt{2})$	100	$\widehat{TC}^{®}$	0.050	0.050	0.042	0.055	0.043
		$Li-Chen$	0.120	0.084	0.056	0.058	0.053
			Power				
	20	$\widehat{TC}^{®}$	0.183	0.311	0.464	0.628	0.697
		$Li-Chen$	0.282	0.290	0.309	0.265	0.277
	50	$\widehat{TC}^{®}$	0.770	0.914	0.981	0.996	0.997
		$Li-Chen$	0.665	0.693	0.750	0.801	0.828
	80	$\widehat{TC}^{®}$	0.974	0.996	1.000	1.000	1.000
		$Li-Chen$	0.886	0.942	0.968	0.991	0.986
	100	$\widehat{TC}^{®}$	0.995	1.000	1.000	1.000	1.000
		$Li-Chen$	0.945	0.986	0.995	0.998	1.000
			Size				
Sample 1∼	20	$\widehat{TC}^{®}$	0.040	0.025	0.021	0.019	0.023
		$Li-Chen$	0.099	0.076	0.059	0.070	0.050
$N(0,1)$	50	$\widehat{TC}^{®}$	0.029	0.021	0.038	0.026	0.017
		$Li-Chen$	0.111	0.069	0.068	0.057	0.053
Sample 2∼	80	$\widehat{TC}^{®}$	0.046	0.052	0.035	0.029	0.044
		$Li-Chen$	0.099	0.091	0.065	0.064	0.060
$Gamma(0.5,\sqrt{2})$	100	$\widehat{TC}^{®}$	0.047	0.042	0.049	0.041	0.044
		$Li-Chen$	0.122	0.085	0.069	0.056	0.047
			Power				
	20	$\widehat{TC}^{®}$	0.213	0.305	0.467	0.635	0.727
		$Li-Chen$	0.296	0.278	0.297	0.276	0.295
	50	$\widehat{TC}^{®}$	0.794	0.924	0.982	0.997	0.999
		$Li-Chen$	0.659	0.724	0.766	0.824	0.823
	80	$\widehat{TC}^{®}$	0.979	1.000	1.000	1.000	1.000
		$Li-Chen$	0.890	0.950	0.977	0.989	0.992
	100	$\widehat{TC}^{®}$	1.000	1.000	1.000	1.000	1.000
		$Li-Chen$	0.958	0.982	0.996	0.999	1.000

Table 6.2: The p-values of the test for the intact cells and permeabilized cells data.

K	2	3	4	5
intact cells data	0.0011	0.0041	0.0000	0.0000
permeabilized cells data	0.9645	0.8103	0.1922	0.1440

tors measured the mitochondrial concentration of Ca2+ every ten seconds during the hour. In fact, due to technical reasons, the original experiment was performed twice, using both the *intact*, original cells and *permeabilized* cells (a condition related to the mitochondrial membrane), see Ruiz-Meana et al. (2003). The data have been made available by Febrero-Bande and Oviedo de la Fuente (2012) in the R package. The first 180 seconds of the data are removed, given the erratic behavior of the curves, leaving $p = 342$ time points. The tests were applied to both the intact and permeabilized data to test for equality between the true treatment and control covariance matrices.

For the intact cells data, the sample sizes are 89 where the sample sizes of control group and treatment group are 44 and 45, respectively. Let Σ_1 and Σ_2 be, respectively, the covariance matrices of control group and treatment group. We apply the test procedure in Section 6.2 to check the hypotheses $H_0 : \Sigma_1 = \Sigma_2$ vs $H_1 : \Sigma_1 \neq \Sigma_2$. $K = 2$ is selected by our method. Based on the asymptotic distribution of the test statistic $\widehat{TC}^{®}$, the p-value is calculated to be 0.0011. At the standard significant level $\alpha = 0.05$, the null hypothesis $H_0 : \Sigma_1 = \Sigma_2$ is thus rejected. For more illustration, we also compute the p-values under different K values in Table 6.2. Such results show that it is not reasonable to assume $\Sigma_1 = \Sigma_2$ in applying a classifier to this data set.

The equality testing of covariance matrices is also conducted for permeabilized cells data, where the sample sizes of control group and treatment group are 45. The p-values under different Ks are also included in Table 6.2. We can see that there is little evidence that the covariance structures for the permeabilized cells data are different for the control group and the treatment group.

6.5 Proofs of main results

Proof of lemma 6.1 Under assumptions 6.1-6.3, if we can prove

$$\hat{\xi}_k^{(1)} - \xi_k^{(1)} = O_p(n_1^{-1/2}),$$

$$\hat{\xi}_k^{(2)} - \xi_k^{(2)} = O_p(n_2^{-1/2}),$$

then lemma 6.1 can be easily proved. It can be observed

$$\begin{aligned}
&\hat{\xi}_k^{(1)} - \xi_k^{(1)} \\
&= \left\{ \frac{1}{n_1 p} \sum_{i=1}^{n_1} Y_i^{(1)} e_k^\top - \xi_k^{(1)} \right\} + \left\{ \frac{1}{n_1 p} \sum_{i=1}^{n_1} Y_i^{(1)} \hat{e}_k^\top - \frac{1}{n_1 p} \sum_{i=1}^{n_1} Y_i^{(1)} e_k^\top \right\} \qquad (6.4) \\
&\equiv C_1 + C_2.
\end{aligned}$$

For C_1, we have

$$C_1 = \tfrac{1}{n_1} \sum_{i=1}^{n_1} \tfrac{1}{p} Y_i^{(1)\top} e_k - \tfrac{1}{p} \mu_1^\top e_k. \qquad (6.5)$$

It is easy to see that C_1 is the average of independent and identically distributed random variables. By the central limit theorem, we obtain

$$C_1 = O_p\left(n_1^{-1/2}\right). \qquad (6.6)$$

For C_2, we have

$$C_2 = \frac{1}{n_1 p} \sum_{i=1}^{n_1} Y_i^{(1)\top} \left\{ \hat{e}_k - e_k \right\}. \qquad (6.7)$$

It is easy to see that $\widehat{\Sigma}_1^{(0)} - \Sigma_1 = O_p(n_1^{-1/2})$ and $\widehat{\Sigma}_2^{(0)} - \Sigma_2 = O_p(n_1^{-1/2})$, so we have $\widehat{\Sigma}^{(0)} - \Sigma = O_p(n_1^{-1/2})$ and $\hat{e}_k^\top - e_k^\top = O_p(n_1^{-1/2})$. According to Cauchy-Schwarz inequality, we have

$$C_2 \leqslant \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{p} \left\{ \left\| Y_i^{(1)} \right\|^2 \left\| \hat{e}_k^\top - e_k^\top \right\|^2 \right\}^{1/2} = O_p\left(n_1^{-1/2}\right). \qquad (6.8)$$

By (6.4), (6.6) and (6.8), we obtain $\hat{\xi}_k^{(1)} - \xi_k^{(1)} = O_p(n_1^{-1/2})$. Similarly, we can prove $\hat{\xi}_k^{(2)} - \xi_k^{(2)} = O_p(n_2^{-1/2})$.

Now we proof lemma 6.1. It can be observed

$$\begin{aligned}
\hat{\vartheta}_{kk'}^{(1)} &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\hat{\xi}_{ik}^{(1)} - \hat{\xi}_k^{(1)}\right\}\left\{\hat{\xi}_{ik'}^{(1)} - \hat{\xi}_{k'}^{(1)}\right\} \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\xi}_{ik}^{(1)}\hat{\xi}_{ik'}^{(1)} - \hat{\xi}_k^{(1)}\hat{\xi}_{k'}^{(1)} \\
&\equiv D_1 - D_2.
\end{aligned}$$

D_1 can be decomposed as

$$\begin{aligned}
D_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{p}Y_i^{(1)\top}\hat{e}_k\frac{1}{p}Y_i^{(1)\top}\hat{e}_{k'} \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{1}{p}Y_i^{(1)\top}e_k\right\}\left\{\frac{1}{p}Y_i^{(1)\top}e_{k'}\right\} + \frac{1}{p^2}e_k^\top\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}Y_i^{(1)\top}Y_i^{(1)}\right\}(\hat{e}_{k'} - e_{k'}) \\
&+ \frac{1}{p^2}(\hat{e}_k - e_k)^\top\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}Y_i^{(1)\top}Y_i^{(1)}\right\}e_{k'} + \frac{1}{p^2}(\hat{e}_k - e_k)^\top\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}Y_i^{(1)}Y_i^{(1)\top}\right\}(\hat{e}_{k'} - e_{k'}) \\
&\equiv D_{11} + D_{12} + D_{13} + D_{14}.
\end{aligned} \tag{6.9}$$

It is easy to show that D_{12}, D_{13} and D_{14} equal to $o_p(n^{-1/2})$. For D_{11}, we have

$$\begin{aligned}
D_{11} &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left[\frac{1}{p}\left\{w_i^{(1)} + \mu_1\right\}^\top e_k\right]\left[\frac{1}{p}\left\{w_i^{(1)} + \mu_1\right\}^\top e_{k'}\right] \\
&= \frac{1}{n_1p^2}\sum_{i=1}^{n_1}\left\{w_i^{(1)\top}e_k\right\}\left\{w_i^{(1)\top}e_{k'}\right\} + \frac{1}{n_1p^2}\sum_{i=1}^{n_1}\left(\mu_1^\top e_k\right)\left(\mu_1^\top e_{k'}\right) \\
&+ \frac{1}{n_1p^2}\sum_{i=1}^{n_1}e_k^\top\left\{w_i^{(1)}\mu_1^\top\right\}e_k + \frac{1}{n_1p^2}\sum_{i=1}^{n_1}e_k^\top\left\{\mu_1 w_i^{(1)\top}\right\}e_k \\
&\equiv D_{111} + D_{112} + D_{113} + D_{114}.
\end{aligned} \tag{6.10}$$

According to assumption 6.4, we have

$$D_{111} \xrightarrow{d} N\left(e_k^\top \Sigma_1 e_{k'}, E\left[\left\{e_k^\top w_i^{(1)}\right\}^2 \left\{e_{k'}^\top w_i^{(1)}\right\}^2\right] - \left(e_k^\top \Sigma_1 e_{k'}\right)^2\right).$$

By assumption 6.3, we can obtain

$$D_{112} - D_2 = o_p\left(n_1^{-1/2}\right).$$

Because $D_{113}+D_{114}$ is the average of independent and identically distributed random variables with mean 0 and variance $(\mu_1^\top e_{k'})^2(e_k^\top \Sigma_1 e_k) + (\mu_1^\top e_k)^2(e_{k'}^\top \Sigma_1 e_{k'})$. By the central limit theorem, we can obtain

$$\sqrt{n_1}\left\{\hat{\vartheta}_{kk'}^{(1)} - \vartheta_{kk'}^{(1)}\right\} \xrightarrow{d} N\left(e_k^\top \Sigma_1 e_{k'}, \varpi_{kk'}^{(1)}\right),$$

where $\varpi_{kk'}^{(1)} = (\mu_1^\top e_{k'})^2(e_k^\top \Sigma_1 e_k)+(\mu_1^\top e_k)^2(e_{k'}^\top \Sigma_1 e_{k'})+E\{(e_k^\top w_i^{(1)})^2(e_{k'}^\top w_i^{(1)})^2\}-(e_k^\top \Sigma_1 e_{k'})^2$.

Similarly, we can prove $\sqrt{n_2}\{\hat{\vartheta}_{kk'}^{(2)} - \vartheta_{kk'}^{(2)}\} \xrightarrow{d} N(e_k^\top \Sigma_2 e_{k'}, \varpi_{kk'}^{(2)})$, where $\varpi_{kk'}^{(2)} = (\mu_2^\top e_{k'})^2(e_k^\top \Sigma_2 e_k) + (\mu_2^\top e_k)^2(e_{k'}^\top \Sigma_2 e_{k'}) + E[\{e_k^\top w_i^{(2)}\}^2\{e_{k'}^\top w_i^{(2)}\}^2] - (e_k^\top \Sigma_2 e_{k'})^2$. Under $H_0^{®}$, we obtain the conclusion of lemma 6.1. The proof is then completed.

Proof of theorem 6.1 According to lemma 6.1 and the Slutsky theorem, we can easily obtain the the conclusion of theorem 6.1. The proof is then completed.

Proof of theorem 6.2 According to lemma 6.1, we have

$$\hat{\vartheta}_{kk'}^{(1)} \xrightarrow{p} e_k^\top \Sigma_1 e_{k'},\ \hat{\vartheta}_{kk'}^{(2)} \xrightarrow{p} e_k^\top \Sigma_2 e_{k'}.$$

Under $H_1^{®}$, we obtain

$$\widehat{TC}^{®} \xrightarrow{p} \frac{\frac{n_1 n_2}{n_1+n_2}\sum_{k=1}^{K^{®}}\sum_{k'=1}^{K^{®}} \frac{\left\{e_k^\top(\Sigma_1-\Sigma_2)e_{k'}\right\}^2}{\frac{1}{n_1}\varpi_{kk'}^{(1)}+\frac{1}{n_2}\varpi_{kk'}^{(2)}} - \frac{K^{®}(K^{®}-1)}{2}}{\sqrt{K^{®}\left(K^{®}-1\right)}} \to \infty,$$

the proof is then completed.

Chapter 7

Conclusions & Future Work

The title name contains the term longitudinal has three reasons: First, in Chapter 2, data are collected by repeated measurements which can be regarded as longitudinal data. For the random curve samples, we considered the two-sample mean curve test and covariance function test in sparse and irregular data. It also can be seen as longitudinal study. In addition, in real data analysis, precipitation and CD4 study are also longitudinal studies. Therefore, When functional data are observed at irregular time points, perhaps just a few time points per subject, they are usually referred as longitudinal data since they often arise from longitudinal studies.

The thesis has motivated further research in hands. Here we mainly report two ongoing pieces of work.

The first one is to extend two-sample test of equality of covariance functions to multiple sample scenarios. Suppose we have G independent samples

$$Y_{g,i,k} = \mu(t_{g,i,k}) + v_{g,i}(t_{g,i,k}) + \epsilon_{g,i,k},$$

where $g = 1, \cdots, G$, $i = 1, \cdots, n_g$, and $k = 1, \cdots, n_{g,i}$. The hypotheses are

$$H_0 : \gamma_1(s,t) = \cdots = \gamma_g(s,t) \quad \text{vs.} \quad H_1 : \exists j \neq k \text{ such that } \gamma_i(s,t) \neq \gamma_j(s,t).$$

Let n be the whole sample size, i.e. $n = \sum_{g=1}^{G} n_G$. When $G = 2$, it is easy to see that

$$TC \equiv \frac{n_1 n_2}{n_1 + n_2} ||\hat{\gamma}_1 - \hat{\gamma}_2||_2^2 = \frac{1}{G} \sum_{g=1}^{G} n_g ||\hat{\gamma}_0 - \hat{\gamma}_g||_2^2,$$

where $\hat{\gamma}_0 = \frac{1}{n} \sum\limits_{g=1}^{G} n_g \hat{\gamma}_g$ and $||f - g||_2^2 = \iint \{f(s,t) - g(s,t)\}^2 dsdt$. Therefore, a natural extension to G populations framework of the proposed test statistic is to consider

$$TC_G \equiv \frac{1}{G} \sum_{g=1}^{G} n_g ||\hat{\gamma}_0 - \hat{\gamma}_g||_2^2.$$

In this new setting, one should expect

$$TC_G = \frac{1}{G} \sum_{g=1}^{G} n_g \left\{ \sum_{l=1}^{K} \left[\frac{1}{n} \sum_{g'=1, g' \neq g}^{G} n_{g'} \hat{\rho}_l^{(g)} \right]^2 \right\} + o_p(1),$$

where $\hat{\rho}_l^{(g)}$'s are similarly defined by involving $\{Y_{g',i,k}; i = 1, \cdots, n_{g'}, k = 1, \cdots, n_{g',i}\}$.

The next one is to consider the partial derivative of hypersurface functional data. In practice, one may be more interested in the rate of change of function instead of the function values themselves. Here by term "rate of change" of a function we mean the derivatives of a function, and in multivariate case the partial derivatives. Therefore, developing new methods for estimating the partial derivatives of functional data will also be an interesting topic. In addition, appropriate mean regression that can predict multivariate functional response deserves further investigation.

Bibliography

Ait-Saïdi, A., Ferraty, F., Kassa, R. and Vieu, P. (2008) Cross-validated estimations in the single-functional index model. *Statistics*, **42**, 475–494.

Anderson, T. (2003) *An Introduction to Multivariate Statistical Analysis (3rd ed.)*. New York: Wiley-Interscience.

Arias-Castro, E., Candès, E. and Plan, Y. (2011) Global testing under sparse alternatives: anova, multiple comparisons and the higher criticism. *Ann. Statist.*, **39(5)**, 2533–2556.

Ash, R. and Gardner, M. (1975) *Topics in Stochastic Processes*. Academic Press.

Aston, J. and Kirch, C. (2012) Detecting and estimating changes in dependent functional data. *J. Multi. Anal.*, **109**, 204–220.

Aston, J., Pigoli, D. and Tavakoli, S. (2017) Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.*, **45(4)**, 1431–1461.

Benko, M., Härdle, W. and Kneip, A. (2009) Common functional principal components. *Ann. Statist.*, **37(1)**, 1–34.

Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009) Detecting changes in the mean of functional observations. *J. R. Statist. Soc.* B, **71(5)**, 927–946.

Billingsley, P. (1995) *Probability and Measure (3rd edition)*. John Wiley & Sons Inc., New York.

Bosq, D. (2000) *Linear Processes in Function Spaces*. Springer.

Cai, T. and Hall, P. (2006) Prediction in functional linear regression. *Ann. Statist.*, **34**, 2159–2179.

Cai, T., Liu, W. and Xia, Y. (2013) Two-sample covariance matrix testing and support recoverary in high-dimensional and sparse settings. *J. Am. Statist. Ass.*, **108(501)**, 265–277.

Cai, T., Liu, W. and Xia, Y. (2014) Two-sample test of high dimensional means under dependence. *J. R. Statist. Soc.* **B**, **76(2)**, 349–372.

Cai, T. and Ma, Z. (2013) Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, **19(5B)**, 2359–2388.

Cai, T. and Yuan, M. (2010) Nonparametric covariance function estimation for functional and longitudinal data. *Tech. Rep.*

Cardot, H., Ferrarty, F. and Sarda, P. (2003) Spline estimators for the functional linear model. *Statist. Sin.*, **13**, 571–591.

Castro, P., Lawton, W. and Sylvestre, E. (1986) Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28(4)**, 329–337.

Chen, D., Hall, P. and Müller, H.-G. (2011a) Single and multiple index functional regression models with nonparametric link. *Ann. Statist.*, **39**, 1720–1747.

Chen, K., Chen, K., Müller, H.-G. and Wang, J.-L. (2011b) Stringing high-dimensional data for functional analysis. *J. Am. Statist. Assoc.*, **106(493)**, 275–284.

Chen, K., Delicado, P. and Müller, H.-G. (2017) Modeling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Statist. Soc.* B, **79(1)**, 177–196.

Chen, L., Dou, W. and Qiao, Z. (2013) Ensemble subsampling for imbalanced multivariate two-sample tests. *J. Am. Statist. Assoc.*, **108(504)**, 1308–1323.

Chen, S. and Qin, Y.-L. (2010) A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, **38(2)**, 808–835.

Chen, S., Zhang, L. and Zhong, P. (2010) Tests for high-dimensional covariance matrices. *J. Am. Statist. Ass.*, **105(490)**, 810–819.

Chiou, J.-M., Müller, H.-G. and Wang, J.-L. (2003) Functional quasi-likelihood regression models with smooth random effects. *J. R. Statist. Soc.* B, **65(2)**, 405–423.

Cox, T. and Cox, M. (2001) *Multidimensional Scaling.* London: Chapman & Hall.

Cuevas, A., Febrero, M. and Fraiman, R. (2004) An anova test for functional data. *Computnl Statist. Data Anal.*, **47**, 111–122.

Donoho, D. and Jin, J. (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, **32**, 962–994.

Estévez-Pérez, G. and Vilar, J. (2013) Functional ANOVA starting from discrete data: an application to air quality data. *Environ. Ecol. Statist.*, **20**, 495–517.

Fan, J. and Lin, S.-K. (1998) Test of significance when data are curves. *J. Am. Statist. Assoc.*, **98(443)**, 1007–1021.

Febrero, M., Galeano, P. and González-Manteiga, W. (2008) Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics*, **19(4)**, 331–345.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012) Statistical computing in functional data analysis: the R package fda. usc. *J. Stat. Softw.*, **51(4)**, 1–28.

Ferraty, F. (2011) *Recent Advances in Functional Data Analysis and Related Topics.* Springer.

Ferraty, F., Keilegom, I. and Vieu, P. (2012) Regression when both response and predictor are functions. *J. Multi. Anal.*, **109**, 10–28.

Ferraty, F. and Romain, Y. (2011) *The Oxford Handbook of Functional Data Analysis.* Oxford University Press.

Ferraty, F. and Vieu, P. (2002) The functional nonparametric model and application to spectrometric data. *Computnl. Statist.*, **17**, 545–564.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis.* Springer.

Ferraty, F. and Vieu, P. (2009) Additive prediction and boosting for functional data. *Computnl Statist. Data Anal.*, **53**, 1400–1413.

Fischl, M., Ribaudo, H., Collier, A., Erice, A., Giuliano, M., Dehlinger, M., Eron, J., Saag, M., Hammer, S. and Vella, S. (2003) A randomized trial of 2 different 4-drug antiretroviral regimens versus a 3-drug regimen, in advanced human immunodeficiency virus disease. *J. Infect. Dis.*, **188(5)**, 625–634.

Fisher, T., Sun, X. and Gallagher, C. (2010) A new test for sphericity of the covariance matrix for high dimensional data. *J. Multi. Anal.*, **101(10)**, 2554–2570.

Fisher, T. J. (2012) On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *J. Statist. Planng Inf.*, **142(1)**, 312–326.

Fremdt, S., Horváth, L., Kokoszka, P. and Steinebach, J. (2014) Functional data analysis with increasing number of projections. *J. Multi. Anal.*, **124**, 313–332.

Fremdt, S., Steinebach, J., Horváth, L. and Kokoszka, P. (2013) Testing the equality of covariance operators in functional samples. *Scand. J. Statist.*, **40(1)**, 138–152.

Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *J. Am. Statist. Assoc.*, **90(432)**, 1179–1188.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L. and Prader, A. (1984) Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12(1)**, 210–229.

Ghiglietti, A., Leva, F. and Paganoni, A. (2017) Statistical inference for stochastic processes: two-sample hypothesis tests. *J. Statist. Planng Inf.*, **180**, 49–68.

Górecki, T. and Smaga, Ł. (2015) A comparison of tests for the one-way ANOVA problem for functional data. *Comput. Stat.*, **30(4)**, 987–1010.

Gregory, K., Carroll, R., Baladandayuthapani, V. and Lahiri, S. (2015) A two-sample test for equality of means in high dimension. *J. Am. Statist. Assoc.*, **110(510)**, 837–849.

Grenander, U. (1950) Stochastic processes and statistical inference. *Ark. Mat.*, **1(3)**, 195–277.

Gromenko, O., Kokoszka, P. and Reimherr, M. (2017) Detection of change in the spatiotemporal mean function. *J. R. Statist. Soc.* B, **79(1)**, 29–50.

Gupta, A. and Giri, N. (1973) Properties of tests concerning covariance matrices of normal distributions. *Ann. Statist.*, **6**, 1222–1224.

Gupta, A. and Tang, J. (1984) Distribution of likelihood ratio statistic for testing equality of covariance matrices of multivariate Gaussian models. *Biometrika*, **71**, 555–559.

Hall, P. and Horowitz, J. (2007) Methodology and convergence rates for functional linear regression. *Ann. Statist.*, **35**, 70–91.

Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis. *J. R. Statist. Soc.* B, **68(1)**, 109–126.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34(3)**, 1493–1517.

He, G., Müller, H.-G., Wang, J.-L. and Yang, W. (2010) Functional linear regression via canonical analysis. *Bernoulli*, **16(3)**, 705–729.

Horváth, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications.* Springer series in statistics, Springer, New York, NY.

Horváth, L., Kokoszka, P. and Reeder, R. (2013) Estimation of the mean of functional time series and a two-sample problem. *J. R. Statist. Soc.* B, **75(1)**, 103–122.

Horváth, L., Kokoszka, P. and Reimherr, M. (2009) Two sample inference in functional linear models. *Can. J. Statist.*, **37(4)**, 571–591.

Horváth, L., Kokoszka, P. and Rice, G. (2014) Testing stationarity of functional time series. *J. Econometrics*, **75(1)**, 103–122.

Horváth, L. and Rice, G. (2015a) An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.*, **28**, 505–548.

Horváth, L. and Rice, G. (2015b) Testing equality of means when the observations are from functional time series. *J. Time Ser. Anal.*, **36**, 84–108.

Jiang, C.-R. and Wang, J.-L. (2011) Functional single index models for longitudinal data. *Ann. Statist.*, **39(1)**, 362–388.

Kokoszka, P. and Reimherr, M. (2017) *Introduction to Functional Data Analysis.* Chapman and Hall/CRC.

Kraus, D. (2015) Components and completion of partially observed functional data. *J. R. Statist. Soc.* B, **77(4)**, 777–801.

Ledoit, O. and Wolf, M. (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, **30(4)**, 1081–1102.

Lee, E. and Park, B. (2012) Sparse estimation in functional linear regression. *J. Multi. Anal.*, **105(1)**, 1–17.

Li, J. and Chen, S. (2012) Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, **40(2)**, 908–940.

Li, W. and Qin, Y. (2014) Hypothesis testing for high-dimensional covariance matrices. *J. Multi. Anal.*, **128**, 108–119.

Li, Y. and Guan, Y. (2014) Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *J. Am. Statist. Assoc.*, **109(507)**, 1205–1215.

Li, Y. and Hsing, T. (2007) On rates of convergence in functional linear regression. *J. Multiv. Anal.*, **98**, 1782–1804.

Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, **38(6)**, 3321–3351.

Li, Y., Wang, N. and Carroll, R. (2010) Generalized functional linear models with semiparametric single-index interactions. *J. Am. Statist. Assoc.*, **105**, 621–633.

Lindquist, M. (2008) The statistical analysis of fMRI data. *Statist. Sci.*, **23(4)**, 439–464.

Ma, S. (2016) Estimation and inference in functional single-index models. *Ann. Inst. Stat. Math.*, **68**, 181–208.

Ma, S., Yang, L. and Carroll, R. (2012) A simultaneous confidence band for sparse longitudinal regression. *Statist. Sin.*, **22**, 95–122.

Müller, H.-G. (2014) Functional data analysis. *International Encyclopedia of Statistical Science*, 554–555.

Müller, H.-G. (2016) Peter Hall, functional data analysis and random objects. *Ann. Statist.*, **44**, 1867–1887.

Müller, H.-G. and Yao, F. (2008) Functional additive models. *J. Am. Statist. Assoc.*, **103(484)**, 1534–1544.

O'Brien, P. (1992) Robust procedures for testing equality of covariance matrices. *Biometrics*, **48**, 819–827.

Oeppen, J. and Vaupel, J. (2002) Broken limits to life expectancy. *Science*, **296**, 1029–1031.

Panaretos, V., Kraus, D. and Maddocks, J. (2010) Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Am. Statist. Ass.*, **105(490)**, 670–682.

Park, S. and Staicu, A. (2015) Longitudinal functional data analysis. *Stat*, **4**, 212–226.

Perlman, M. (1980) Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *Ann. Statist.t*, **8**, 247–263.

Pomann, G., Staicu, A. and Ghosh, S. (2016) A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *J. R. Statist. Soc.* C, **65(3)**, 395–414.

Pryor, S. (2013) *Climate Change in the Midwest: Impacts, Risks, Vulnerability, and Adaptation.* Indiana University Press.

Qiu, Y. and Chen, S. (2012) Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.*, **40(3)**, 1285–1314.

Ramsay, J. (1982) When the data are functions. *Psychometrika*, **47(4)**, 379–396.

Ramsay, J. and Dalzell, C. (1991) Some tools for functional data analysis. *J. R. Statist. Soc.* B, **53(3)**, 539–572.

Ramsay, J., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and MATLAB.* New York: Springer.

Ramsay, J. and Silverman, B. (2002) *Applied Functional Data Analysis*. New York: Springer.

Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis*. New York: Springer.

Rao, C. (1958) Some statistical methods for comparison of growth curves. *Biometrics*, **14(1)**, 1–17.

Rice, J. and Silverman, B. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B, **53**, 233–243.

Ruiz-Meana, M., Garcia-Dorado, D., Pina, P., Inserte, J., Agullo, L. and Soler-soler, J. (2003) Cariporide preserves mitochondrial proton gradient and delays ATP depletion in cardiomyocytes during ischemic conditions. *Am. J. Physiol. Heart Circ. Physiol.*, **285(3)**, H999–H1006.

Schott, J. (2005) Testing for complete independence in high dimensions. *Biometrika*, **92(4)**, 951–956.

Schott, J. (2006) A high-dimnsional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multi. Anal.*, **97(4)**, 827–843.

Schott, J. (2007) A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Stat. Data Anal.*, **51(12)**, 6535–6542.

Shi, J. and Choi, T. (2011) *Gaussian Process Regression Analysis for Functional Data*. Chapman and Hall/CRC.

Shi, M., Weiss, R. and Taylor, J. (1996) An analysis of Paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *J. R. Statist. Soc.* C, **45(2)**, 151–163.

Srivastava, M. (2005) Some tests concerning the covariance matrix in high dimensioanl data. *J. Janpan Statist. Soc.*, **35(2)**, 251–272.

Srivastava, M. and Yanagihara, H. (2010) Testing the equality of serveral covariance matrices with fewer observations than the dimension. *J. Multi. Anal.*, **101(6)**, 1319–1329.

Staicu, A.-M., Lahiri, S. and Carroll, R. (2015) Significance tests for functional data with complex dependence structure. *J. Statist. Planng Inf.*, **156**, 1–13.

Staicu, A.-M., Li, Y., Crainiceanu, C. and Ruppert, D. (2014) Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scand. J. Statist.*, **41**, 932–949.

Sugiura, N. and Nagao, H. (1968) Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Statist.*, **39**, 1682–1692.

Torgovitski, L. (2015) Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components. *Preprint ArXiv: 1509.07409*.

Usset, J., Staicu, A.-M. and Maity, A. (2016) Interaction models for functional regression. *Computnl Statist. Data Anal.*, **94**, 317–329.

Vaupel, J., Carey, J., Christensen, K., Johnson, T., Yashin, A., Holm, N., Iachine, I., Kannisto, V., Khazaeli, A., Liedo, P., Longo, V., Zeng, Y., Manton, K. and Curtsinger, J. (1998) Biodemographic trajectories of longevity. *Science*, **280(5365)**, 855–860.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2015a) Review of functional data analysis. *Ann. Rev. Statist.*, **104**, 1–41.

Wang, L., Peng, B. and Li, R. (2015b) A high-dimensional nonparametric multivariate test for mean vector. *J. Am. Statist. Assoc.*, **110(512)**, 1658–1669.

Wilmoth, J., Andreev, K., Jdanov, D. and Glei, D. (2007) Methods protocol for the human mortality database, version 5. *Technical Report*.

Xia, Y. and Härdle (2006) Semi-parametric estimation of partially linear single-index models. *J. Multiv. Anal.*, **97(5)**, 1162–1184.

Yang, J. and Zhang, T. (2017) Unified two-sample testing of mean functions for sparse or dense functional data. *Comput. Stat. Data Anal.*, **Submitted**.

Yang, J., Zhang, T., Liu, C. and Ferraty, F. (2017) Testing equality of covariance operators/matrices for two-sample functional/high-dimensional data. *Journal of the Royal Statistical Society, Series B*.

Yao, F. (2007) Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multi. Anal.*, **98(1)**, 40–56.

Yao, F. and Müller, H.-G. (2010) Functional quadratic regression. *Biometrika*, **97**, 49–64.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.*, **100**, 577–590.

Yuan, M. and Cai, T. (2010) A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.*, **38(6)**, 3412–3444.

Zhang, J.-T. (2013) *Analysis of Variance for Functional Data*. CRC press.

Zhang, J.-T. and Chen, J. (2007) Statistical inferences for functional data. *Ann. Statist.*, **35(3)**, 1052–1079.

Zhang, J.-T. and Liang, X. (2014) One-way ANOVA for functional data via globalizing the pointwise F-test. *Scand. J. Statist.*, **41(1)**, 51–71.

Zhang, J.-T., Liang, X. and Xiao, S. (2010) On the two-sample Behrens-Fisher problem for functional data. *J. Stat. Theory Pract.*, **4**, 571–587.

Zhang, T., Liu, C. and Yang, J. (2017a) Interaction models with nonlinear link for functional regression. *Can. J. Statist.*, **Submitted**.

Zhang, T., Yang, J., Liu, C., Yuen, K. and Liu, A. (2017b) Testing for equality of mean surfaces for two-sample bivariate functional data. *J. R. Statist. Soc.* **C**, **Submitted**.

Zhang, X., Shao, X., Hayhoe, K. and Wuebbles, D. (2011) Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Statist.*, **5**, 1765–1796.

Zhao, X., Marron, J. and Wells, M. (2004) The functional data analysis view of longitudinal data. *Statist. Sin.*, **14(3)**, 789–808.

Zhong, P.-S. and Chen, S. (2011) Tests for high-dimensional regression coefficients with factorial designs. *J. Am. Statist. Assoc.*, **106(493)**, 260–274.

Zhong, P.-S., Chen, S. and Xu, M. (2013) Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.*, **41(6)**, 2820–2851.

Zhu, B., Taylor, J. and Song, P.-K. (2011) Semiparametric stochastic modelling of the rate function in longitudinal studies. *J. Am. Statist. Assoc.*, **106(496)**, 1485–1495.